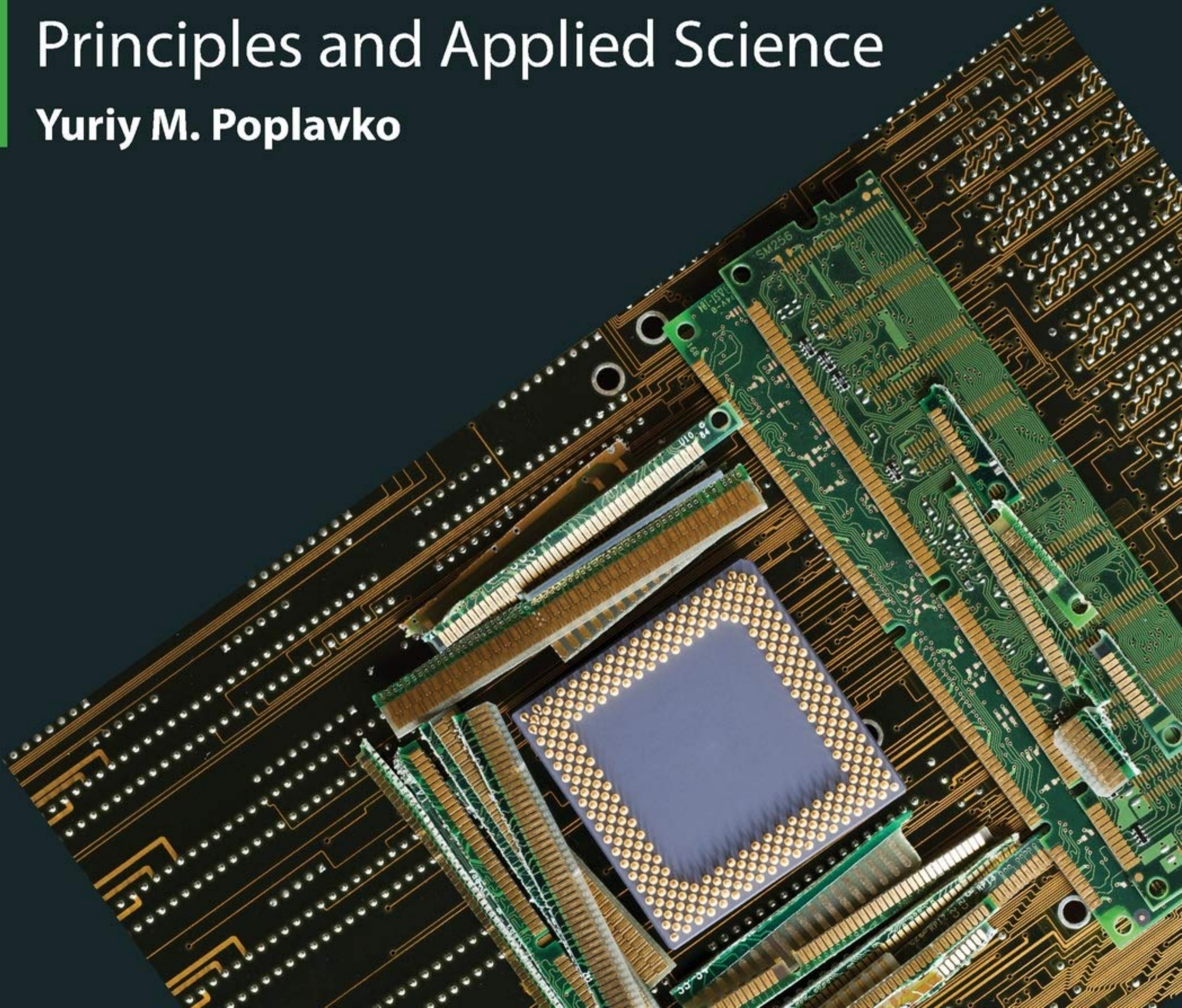




ELECTRONIC MATERIALS

Principles and Applied Science

Yuriy M. Poplavko



Electronic Materials

Electronic Materials

Principles and Applied Science

Yuriy M. Poplavko



Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States

© 2019 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-0-12-815780-0

For information on all Elsevier publications
visit our website at <https://www.elsevier.com/books-and-journals>



Publisher: Matthew Deans

Acquisition Editor: Kayla Dos Santos

Editorial Project Manager: Katie Chan

Production Project Manager: Surya Narayanan Jayachandran

Cover Designer: Victoria Pearson

Typeset by SPi Global, India

Preface

The author of the book is Dr. Yuriy Poplavko, a professor at the Microelectronics Department of the Kiev Polytechnic Institute, founded 120 years ago by Dmitry Mendeleev and now called the National Technical University of Ukraine named after Igor Sikorsky. The electronic material sciences is studied and taught at the department for many years, and Dr. Yuriy Poplavko published in this area several books in Russian and Ukrainian; this book extends this experience.

Electronic materials science is part of “Solid-state physics” but accommodated to the electronic engineering. In general, modern materials sciences cover a very wide range of issues, but in this book mainly the *electrical properties* of metals, semiconductors, dielectrics, and magnets are described; in particular, those that are of importance for specialists in electronics.

Electronics is a science and engineering discipline that concerns study and application of electrical phenomena inherent in substances (mainly, solids). Based on these studies, electronic devices are created, as well as the art of electronic circuits and systems construction is developed. It is also possible to define electronics as a science of electrons interaction with electromagnetic fields or as a science and methodology of creating electronic materials, instruments, and devices. Theoretical problems of electronics concern with the study of electrons interaction with macroscopic fields inside the workspace of an electronic device as well as with the study of interaction of electrons with microscopic fields of ions, atoms, molecules, or crystal lattice. Practical problems of electronics boil down to design of electronic devices that perform various functions, such as conversion and transmission of information, control, computing as well as the energy supplying.

Thus, electronics materials sciences are important for specialists in electronics. It should be also noted that at present not only educational, but also monographic literature, cannot keep pace with rapid development of materials science, in other words, of an applied solid-state physics. This concerns many areas of knowledge and technology—from the preparation of materials to the electronic devices. In this context, nanophysics and nanotechnology are particularly fast-growing areas, and it is clear that they are never mentioned in the books on solid-state physics issued 20–30 years ago.

This book presents considerably simplified mathematical treatment of theories, while emphasis is made on the basic concepts of physical phenomena in electronic materials and their simple explanation. Most chapters are devoted to the advanced scientific and technological problems of electronic materials; in addition, some new insights of theoretical facts relevant for technical devices are presented. This approach of presentation is due to the contemporary tendency for mutual penetration and synthesis from different fields that at first glance belong to different areas of science.

First Vice-Rector of National Technical University of Ukraine
“Igor Sikorsky Kiev Polytechnic Institute”

Full Member of the Academy of Sciences of Ukraine Yuriy Yakimenko

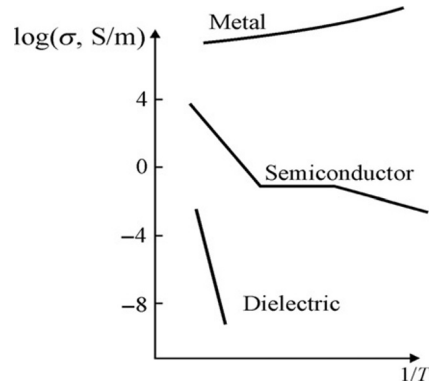
Acknowledgments

My sincerest thanks to my colleagues from Igor Sikorsky Kiev Polytechnic Institute for their continued interest in this work, and for a number of their valuable comments. I am also thankful to many students who pointed to hard-to-understand places and asked me for clarification.

I want to give special thanks to Prof., Dr. Yuriy I. Yakimenko for his continued interest in this work as well as to Dr. Victor A. Kazmirenko and Dr. Yuriy V. Didenko for great help in the design of the book and its preparation for publication. The responsibility for any remaining errors or shortcomings is, of course, mine.

Finally, I am deeply grateful to the staff of the Elsevier Publishing House: Kaila Dos Santos, Katie Chan, and Surya Narayanan Jayachandran for their patience in eliminating the shortcomings of the manuscript.

Introduction



Electronics primarily use crystals, polycrystals (ceramics), glasses, composites, amorphous substances, liquid crystals, and substances produced by the compaction of nanocrystalline structures.

Crystals are characterized by well-ordered (near-perfect) internal structure that can be described by three-dimensional (3D) spatial periodic pattern. Distinctive property of crystals is their translational symmetry, that is to say, the elementary cell composed of a few atoms can be “infinitely” translated in all directions, creating a regular crystal lattice. From outside, crystals are usually separated by the faces, that is, smooth flat surfaces that converge at *strictly defined angles*.

Although a crystal may not be shaped like a polyhedron, it will still manifest such characteristics that will allow distinguishing strictly ordered crystalline state from any disordered glassy or amorphous state. Crystals are characterized by a certain symmetry of their physical properties that *correspond to symmetry of internal structure*. This symmetry determines many physical characteristics of crystal, especially the anisotropy of its electrical, thermal, mechanical, and magnetic parameters. [Fig. I.1](#) shows photos of natural quartz (SiO_2) and KH_2PO_4 crystals artificially grown from aqueous solution, and these crystals are widely used in electronics.

Polycrystals consist of a large number of very small crystals (crystallites). Although the macroscopic structure of polycrystalline sample seems disordered, its microscopic component parts (i.e., crystallites or blocks) are high-grade crystals with perfect microscopic structure, and, therefore, the polycrystal practically has the same properties as single monolithic crystals ([Fig. I.2](#)).

The *glass-like* and *amorphous-state* solids are distinguished by the absence of any distant (translational) symmetry. Distribution of atoms in these bodies is characterized not by a long-range ordering (as in crystals) but by the neighbor ordering. In the range of several nearby atoms, the structure of glass appears as being ordered, thus enabling to determine specific coordination number for neighbors.

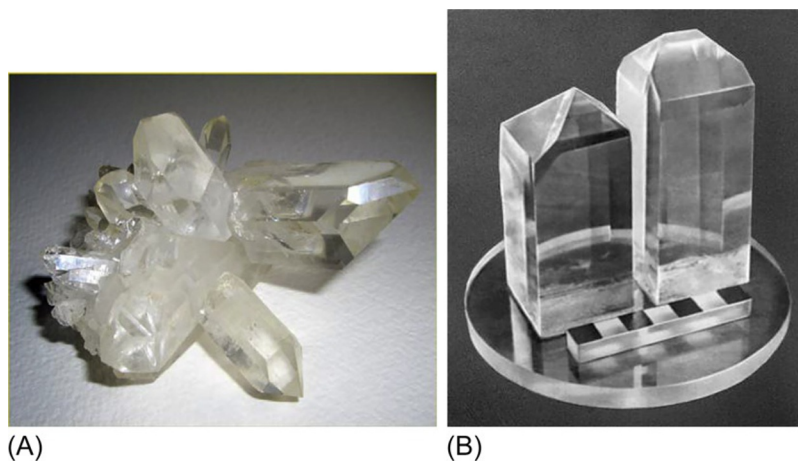


FIG. 1.1
Crystals of quartz (A) and ammonium dihydrophosphate (B).

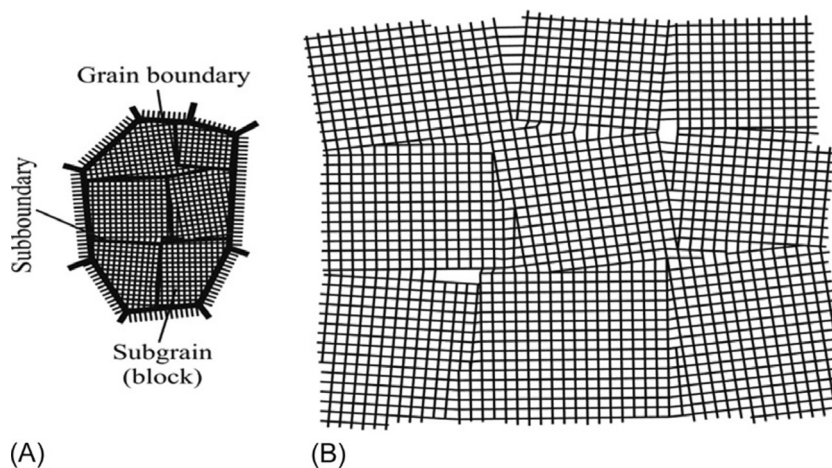


FIG. 1.2
Scheme of crystallites (grains) with its borders (A) and block (mosaic) structure inside the crystallites (B).

However, the crosscorrelation arrangement of remote atoms in a glass is violated. Nevertheless, the ordering of glassy state is higher than that in the amorphous state, which means that the coordination number is more definite in glass than in the amorphous state of a solid.

Other *ordered solids* also may have great importance for practical use in electronics. This applies primarily to two-dimensional (2D) systems (like films). In the 2D

systems, a strictly ordered structure is seen only in plane. If such planar structure is regularly repeating in the semiconductor chip (creating superstructure), its electronic properties can be characterized by so-called quantum wells that is a typical characteristic of 2D nanostructures.

Fig. I.3 shows well-known 2D system—graphene, which has great prospects in electronics.

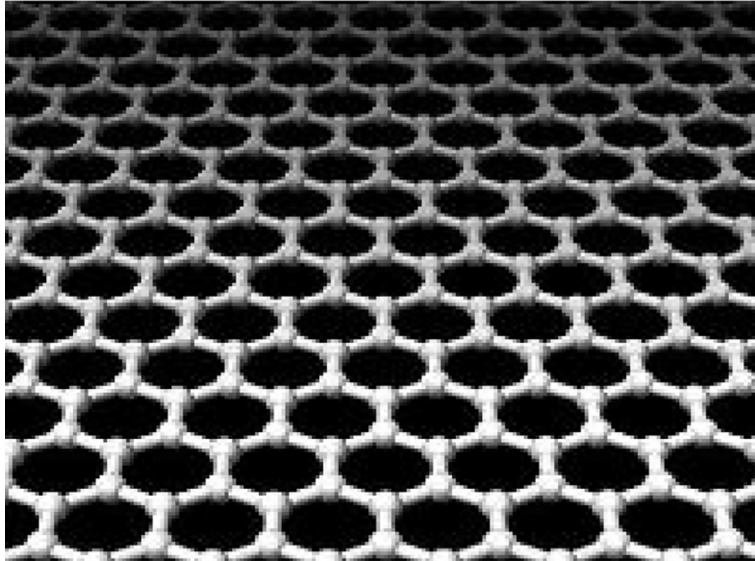


FIG. I.3

Crystal structure of graphene—2D hexagonal lattice.

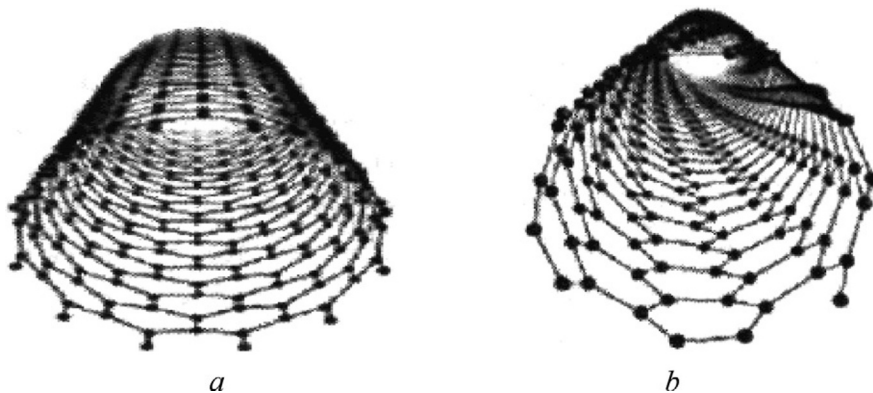


FIG. I.4

Models of nanotubes.

Accordingly, the *linear* (wire-like) systems belong to 1D nanostructures, where translation-like ordering is observed only along one direction (“quantum threads”). The porous silicon can be, in particular, attributed to these systems. However, another well-studied quasi-1D structure is shown in Fig. 1.4, carbon nanotubes.

There are also systems whose dimensions along all three orthogonal directions are commensurate with the distance between atoms. Such zero-dimension (0D) systems can be considered as “quantum dots,” in which only $10\text{--}10^3$ atoms are the ordered system. Fig. 1.5 demonstrates germanium quantum dot grown on silicon substrate. On the area of one square micron, more than 1,000 of these quantum dots can be accommodated.

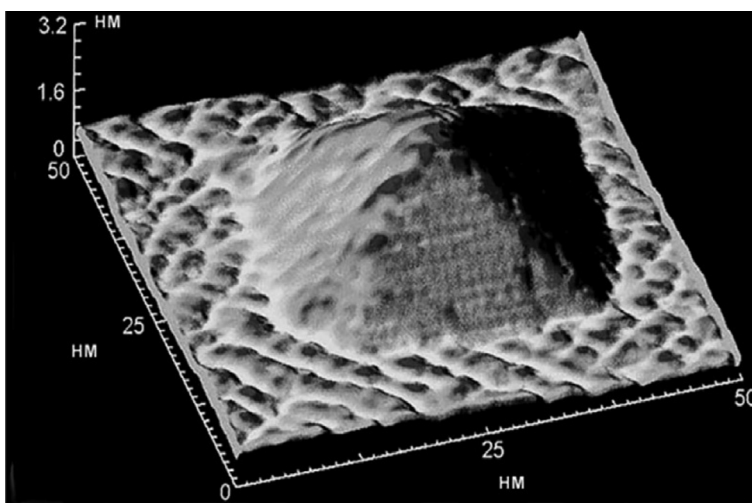


FIG. 1.5

Three-dimensional images of quantum dot image using a scanning electronic microscope.

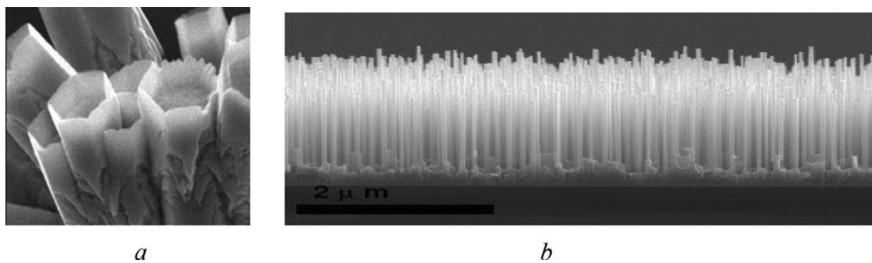


FIG. 1.6

Zinc oxide nanostructures used in sensors: (A) hexagonal nanocrystals; (B) thin tubular nanocrystals.

Nanostructures are characterized by a huge variety of shapes. Fig. I.6 gives examples of ZnO nanostructures that can be used in various sensors (humidity, gas, and even smell sensors).

“Scalp-like” finest ZnO nanocrystals shown in Fig. I.6B actually have an arrangement that is thousands of times denser than human hair.

I.1 CONNECTION BETWEEN ELECTRONIC MATERIALS PHYSICS AND GENERAL PHYSICS

Technical university course of electronic materials science, also known as *condensed matter physics* or *solid-state physics*, usually is studied in the final part of a series of physics courses so that it can be knowledge gained by students in previous courses of physics. The mapping between solid-state physics, classical mechanics, quantum mechanics, and relativistic mechanics is illustrated in Fig. I.7. Solids-state physics is located between classical and quantum mechanics.

Fig. I.7 shows that classical physics is a field of research of low velocities v (as compared with the speed of light c) and a certain ratio of Planck constant \hbar to the action parameter S . Planck constant has a physical dimension of “momentum” that is the product of “energy \times time = impulse \times length.” The action parameter S characterizes movement of particle when its way is multiplied by its impulse. At that, both axes used in Fig. I.7 are dimensionless, because the dimension of S is the same as the dimension of Planck’s constant \hbar .

Condensed matter physics (which is the generalization of solid-state physics) is based not only on classical mechanics, but also on principal methods and notions of quantum mechanics (see Fig. I.7). Einstein’s relativistic mechanics is not used in the solid-state physics, but some of laws of relativistic quantum mechanics are important

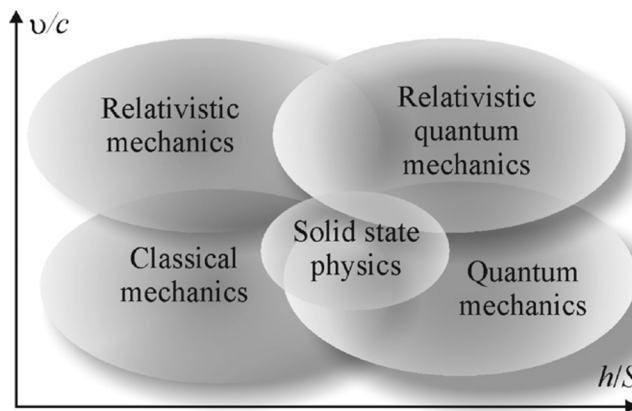


FIG. I.7

Connection between different areas of physics.

for understanding the nature of magnetism in atoms, thin films, and nanosized structures. Solid-state physics explores the nature of solid body formation and its structure (atomic and energetic), as well as its basic physical properties—electromagnetic, optical, thermal, mechanical, and others—that contribute to widespread use of solids in electronics, instrumentation, electrical, and mechanical engineering.

Electronic materials science examines rather complex spatial lattices composed of microscopic particles—atoms, ions, or molecules. The forces acting between particles have a predominantly electrostatic origin. Although atom as a particle is neutral, its electrical charges are not localized at a single point; they are slightly spaced. Therefore, during the formation of a solid body, in which atoms are located close to each other, the opposite charges attract each other, while charges with equal sign push off. Thus, the forces acting between atoms have both attraction and repulsion in them. The influence of one atom on the motion of electrons in another atom is such that the resultant force is always the force of attraction.

The mutual attraction of atoms (or ions, or molecules) that acts at a long distance is actually the cause of formation and existence of solids. However, the attraction dominates only until atoms come near each other so closely that they almost collide. Then the repulsion begins to dominate, as those forces are short-range ones. At a sufficiently small distance, the repulsive force becomes equal to the force of attraction; then a few atoms or ions form the molecule, whereas a whole set of atoms constitutes the solid body.

For materials science, it is important to establish the nature of repulsive forces. As electron is 100,000 times smaller than atom (so as atomic nucleus), from the point of view of classical mechanics the atom looks like “emptiness,” as the space occupied by electrons and nucleus is so small. However, solid-state physics (especially in its important section—crystallography) makes quite a reasonable assumption that atom (or ion) behaves as slightly deformable but solid ball. For this reason, a widely used concept in materials science and solid-state physics is the atomic or ionic *radius* (data for atoms and ions of various elements can be seen in reference tables). One can assume that crystal can be represented in a form of regular lattice, composed of solid balls (ions, atoms, or molecules). High “hardness” of apparently empty atom balls is explained by quantum mechanics; namely, when atoms approach each other, the possible space for bound electrons rapidly decreases, because uncertainty of respective coordinates diminishes. According to Heisenberg’s uncertainty principle, the position x and impulse p of a particle cannot be established simultaneously with arbitrarily high precision: $\Delta x \cdot \Delta p > \hbar/2$. The shorter distances between atoms lead to greater uncertainty in impulse and, thus, to an increase in the magnitude of impulse. As a result, the kinetic energy of electrons increases as well as their total energy, which finally results in repulsion. Energetically favorable position for atoms imposes certain distances between them. Therefore, repulsive force, which provides balance in the solid structure, has quantum nature.

Basic materials of electronics are semiconductors, metals, insulators, magnetics, and nanomaterials. Obviously, most important for electronics are their electrical characteristics, and, among them, electrical conductivity (σ). Besides, this parameter

is very convenient for the *classification* of solids. The unit for measurement of conductivity is [S/m] (Siemens per meter). It determines current density j [A/m²] in a given material, arising under applied electrical field E [V/m] according to Ohm's law: $j = \sigma E$. It is important to note that temperature dependence of conductivity $\sigma(T)$ is quite different for various types of electronics materials (Fig. I.8).

In dielectrics and metals, the $\sigma(T)$ dependences are opposite. While in dielectrics, with temperature rise, conductivity increases exponentially (because atomic thermal motion in a matter generates new charge carriers), in metals conductivity decreases approximately as T^{-1} owing to charge carrier scattering on the lattice thermal vibrations. Therefore, at low temperatures, conductivity in metals becomes very large, reaching infinity when at zero temperature (in case of superconductivity). In dielectrics, in contrast, low-temperature value of σ becomes close to zero, because charge carriers are not generated in the absence of thermal motion (and any radiation effects). Even at room temperature ($T \approx 300$ K), dielectrics have very low conductivity ($\sigma < 10^{-10}$ S/m), and for this reason they are often referred to as "electrical insulators."

With regard to $\sigma(T)$ dependence, metals and dielectrics are qualitatively different (see Fig. I.8). Yet in semiconductors the $\sigma(T)$ dependence looks like in dielectrics; the difference lies in much higher conductivity of semiconductor. If temperature becomes higher, this difference becomes less noticeable (in dielectrics and semiconductors, charge carriers appear through the temperature-driven activation process).

Various properties of solid materials are related to the nature of their chemical bonds and to energy spectrum of electrons. The study of atomic structure and electronic energy spectra of crystalline and noncrystalline solids is a fundamental problem of solid-state physics, because it facilitates a conscious search for materials with predetermined properties.

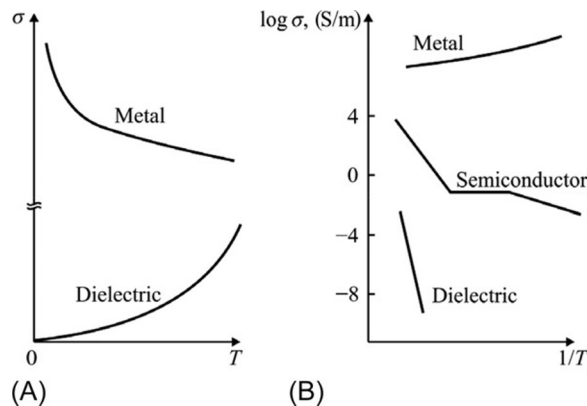


FIG. I.8

Temperature dependence of conductivity in solids: (A) normal scale, (B) logarithmic scale for σ and inversed temperature.

Distinguishing characteristic of solids is their structural ordering that is dependent on positions of neighboring atoms. The correlation of their position may manifest itself exclusively in the short-range ordering of atoms. In the amorphous solids, a short-range ordering is significantly limited; sometimes, their atomic regulation may be restricted to the microcrystallites that are disordered with respect to each other. However, most solids have the long-range ordering, that is, undistorted crystal lattice that extends to relatively large areas. A wide variety of possible geometric structures and spatial relationships in lattices leads to a large number of different phenomena in solids.

Any crystal may manifest some deviation from ideal geometric structure. In addition, any physical body has a finite size; thus a crystal is always limited, either by external surfaces or by internal borders between crystallites. This assertion is trivial, but it is essential for many physical phenomena. It is not possible to entirely neglect the disturbances in body of a real crystal, caused both by the inclusions of foreign atoms in main lattice and by the impurities and violations of local lattice periodicity.

The thermal motion of atoms in crystal lattices also leads to a deviation from strict periodicity. Indeed, periodic lattice reflects not real positions of atoms, but their imaginary equilibrium positions, where they would find themselves at temperature of absolute zero; only this case corresponds to the ground state of a crystal. Each deviation from the ground state means a disturbance. However, at normal temperatures, these deviations are so small that ordered lattice dominantly determines properties of a crystal.

The properties of electronics materials are, at first approximation, reduced to two complex problems:

- Determining the ground state of solids and reasons for its stability (i.e., clarifying the nature of forces that hold atoms in lattices).
- Describing behavior of solids under external influences (i.e., justifying and predicting various physical properties of solids such as electrical, thermal, mechanical, etc.).

The first set of complex problems are characterized by terms such as crystal structure, nature of chemical bonds, cohesion forces, and energy of bonding. However, it is only at first glance the solution of these problems is independent of behavior of solids under the external influences. In fact, solution to the first problem can be obtained only through a settlement of the second problem, because each experiment implies the disturbance of ground state. Any conclusions about properties of solids in the ground state can be made after investigation on how it is affected by the applied electrical field, temperature, exposure to light, etc.

Several external important impacts will be discussed in this book:

- *Electrical field.* First of all, the charge transfer should be studied, that is, electrical current. This investigation would enable phenomenological classification of solids as metals, semiconductors, and dielectrics. The mechanism of electrical charge transfer in an electrical field also makes it possible to determine

mechanisms of conductivity (electronic or ionic). Second, the investigation of solids in the electrical field enables to study the mechanism of electrical *charge separation*, that is, electrical polarization, whose nature may be electronic, ionic, and dipole.

- *Magnetic field*. Different responses of solids to the impact of magnetic field depend on their chemical composition and structure. They give rise to such phenomena as diamagnetism, paramagnetism, ferromagnetism, and antiferromagnetism (and their combinations—ferrimagnetism). The widely used investigation method is the magnetic field application during charge transfer phenomena study in the electrical field. These additional parameters give rise to a number of emerging effects and allow to obtain significant information about main properties of solids.
- *Temperature gradient*. It determines the direction of energy flow from hotter to colder areas of solids. It is noteworthy that simultaneously with heat transfer the electrical charge transfer is also possible. The energy and charge transfer can be described by various mechanisms.
- *Illumination by light*. Absorption, reflection, and scattering of light provide important information about interaction of electromagnetic waves with various solids.
- *Irradiation by electrons*, positrons, neutrons, and other corpuscular particles serve as probe to study various properties of solids.
- *Dosed implantation of additives* into the crystal lattice enables to get important information about crystal properties. This might be done through foreign atom inclusion into crystal structure, formation of vacancies, and atomic substitutions in a lattice, etc.

1.2 SOME COMMENTS ON THEORETICAL APPROACHES

Theoretical description of all the listed phenomena by using any universal model seems impossible. Therefore it is necessary to apply approximations. Therefore in specific cases of investigation, the simplified models should be used, which are suitable for a given problem. The purpose of various theories of solids should be eventual reduction of various theoretical facets of phenomena into unified concepts.

One of such concepts is based on the idea of *quasiparticles*, which are elementary excitations in the solids. In fact, even in case of very weak impact on solid the introduced energy cannot be delivered exclusively to one element of crystal independently of all others. The point is that the strong interaction exists between all particles of a crystal (atoms, ions, and electrons), and, therefore, the energy, even being applied to a single particle, should be rapidly distributed between other particles. This process can be modeled by emission and absorption of quanta of energy attributed to imaginary particles. The concept of quasiparticles, which corresponds to various forms of excitation in solids, can be considered as the basic principle for solid-state physics.

Thus the object of study is a solid in the *excited state*. Excitation energy can be *thermal energy* that is imposed from outside or appear as external violation of crystal lattice. The energy of thermal excitation is transferred to different subsystems of a solid. For example, it can be transferred to valence electron or to ion, as well as it can appear as kinetic energy of ionic lattice vibrations or as the energy of coupled spins in a ferromagnetic.

The excited state that lies close to the energy of ground state may be characterized by relatively small number of the *independent oscillators*. This method is used in the lattice dynamics of solids to describe small oscillations of a lattice around its equilibrium positions. Thus complicated collective vibration state is decomposed into the normal independent vibrations. These vibrations can be quantized, and the corresponding quanta are the *phonons*, the example of elementary disturbances. In some sense, they resemble elementary excitation quanta of electromagnetic field—photons.

The second example of formal description, simplifying multiparticle system with strong collective interaction, is as follows. An idea is introduced that the motion of charged particles can be described as a “gas” of charged particles. Obviously, each particle must push off other similar particles from its surroundings. Formally, this case can be described by the assumption that there is no interaction between particles, because any observed particle is accompanied by the “cloud” of opposite sign charges that partly compensate the charge of given particle. Thus, the interaction, which means other particles’ influence on a movement of given particle, can be replaced by the inertial charged cloud that the given particle should “entail” during its motion. In such a way, the system of interacting particles is replaced by the system of noninteracting particles; however, dynamic properties of new quasiparticles are changed.

In solid-state physics, there are many examples of elementary excitation. Along with phonons, which are quanta of lattice vibrations, there are collective excitations of electrons in metals, called *plasmons*. In the same way, magnetic spin system can be described by spin waves with corresponding quanta—the *magnons*. In addition, in dielectrics the elementary electronic excitations are the *polarons*, while in the semiconductors they might be the *excitons*. The nature of various quasiparticles might be different. When describing electrons, it is necessary to consider that during their movement through a crystal they are exposed to different interactions. Therefore moving electrons can be described as different quasiparticles, depending on the nature of interaction. Electrons can behave as free electrons, or polarons (escaped electrons), or excitons (bound by local interactions), or Cooper pairs, etc.

The violations of crystal lattice, such as localized impurity atoms (or vacancies) in the crystal lattice, can also be modeled by elementary excitations. These elementary excitations, at first approximation, can be considered as noninteracting. In a more refined approach, their interactions also must be taken into consideration. However, in this case, a concept of “excitation” is also applicable, and it can be taken into consideration by methods of the *perturbation theory*.

Even when it is possible to neglect interaction of elementary excitations of the same type, the question of interaction between different types of excitations remains an important problem. With this approach, a significant variety of phenomena in solids can be investigated. The process of establishing equilibrium in solids also requires considering the interaction of quasiparticles, that is, energy exchange between different systems of elementary excitations.

The concept of elementary excitations should be applied only in case of weak deviations from a ground state. If the number of collective excitations and quasiparticles is quite large and the relationship between them is too strong, then the described theoretical model becomes very complicated due to a large number of details, and the concept of quasiparticles ceases to be effective.

Structure of electronic materials

CONTENTS

1.1 Atomic Bonding in Metals, Semiconductors, and Dielectrics	1
1.2 Symmetry of Crystals	12
1.3 Basic Structures of Crystals Used in Electronics	26
1.4 Lattice Defects in Crystals	37
1.5 Structure and Symmetry of Quasicrystals and Nanomaterials	47
1.6 Structures of Composites and Metamaterials	56
1.7 Summary	62
References	68

Before describing the most diverse properties of materials used in electronics, it is necessary to consider the features of their structures, in which these properties are realized.

The formation of crystal, amorphous, and other substances from atoms is accompanied by a decrease of the energy in a system as compared to unconnected atoms. The minimum energy in solids corresponds to a regular arrangement of atoms that agrees with the specific distribution of electronic density between them. In accordance with the electronic theory of valence, interatomic bonds are formed due to the redistribution of electrons in their valence orbitals, resulting in a stable electronic configuration of noble gas (octet) due to formation of ions or of shared electron pairs between atoms.

1.1 ATOMIC BONDING IN METALS, SEMICONDUCTORS, AND DIELECTRICS

Any connections of atoms, molecules, or ions are conditioned by electrical and magnetic interactions. At longer distances, electrical forces of *attraction* dominate between particles whereas, at short distances, *repulsion* of particles increases sharply. The balance between such long-range attraction and short-range repulsion is the cause of the basic properties of substances. The atomic connection is attributable to the restructuring of atomic electronic shells, thus creating *chemical bonds*. In other words, chemical bonds are the phenomenon of atomic interaction by means of overlap of their electronic clouds, and this is accompanied by a decrease of the total energy of a system.

Chemical bonding is characterized by both *energy* and *length*. A measure of bond strength is the energy, expended in case of bond destruction, or the energy gained during compound formation from individual atoms. Consequently, the energy of chemical bonds equals the work that must be expended to separate particles that are constrained, or to alienate them from each other on the infinite distance [1].

During the formation of chemical bonds, exactly those electrons that belong to the *valence shells* play a major role because their contribution to solid body formation is much greater than that of the inner electrons of atoms. However, division into ionic residues and valence electrons is a matter of convention. For example, in metals it is sufficient to consider that valence electrons are transformed into conduction electrons whereas all other electrons belong to ionic residues.

In the *atoms* of a metal, their outer electronic orbits are filled with a relatively small number of electrons that have *low ionization energy*. When these atoms come together (i.e., when crystal is formed from atoms), the orbits of valence electrons strongly overlap. As a result, valence electrons in metals become uniformly distributed in a space between cations, and these electrons have a *common wave function*. Therefore valence electrons in most metals are fully collectivized, and thus such crystals constitute a lattice of positively charged ions crowded by “electronic gas.” Unlike, for example, covalent bonds, the complete delocalization of electrons is a distinctive feature of metallic bonds.

It is in this way that the spatial distribution of valence electrons lies at the heart of the classification of solids (dielectrics, semiconductors, and metals). The division of crystals into different classes suggests that solids consist of:

- *ionic residues*, that is, nuclei themselves and those electrons that are so strongly associated with their nuclei that the residues formed cannot significantly change their configuration as compared with the atom;
- *valence electrons*, that is, electrons, the distribution of which, in solids, may differ significantly from the configuration existing in isolated atoms.

The spatial distribution of electronic orbitals of certain atoms has a strong influence on the bond strength and their direction. Fig. 1.1 schematically shows how major electronic orbitals for *s*-, *p*-, and *d*-states of electrons in the atoms might look. Only the *s*-orbital is characterized by spherical symmetry. In contrast, the *p*-orbital has a very specific form, and this is especially true for the *d*-orbitals: their forms are considered to contribute to the specific properties of transition metals. Rare earth metals have *f*-electrons, and they may play a dual role: as residue electrons of “atomic core” and as “free” electrons (because of their complexity, *f*-orbitals are not shown in Fig. 1.1).

Thus during chemical bond formation, valence electrons play a dominant role because, at crystal formation, their contribution is much greater than that of electrons, which form atomic internal orbitals in the residues.

A classification of the possible bonds of particles in crystals is shown in Fig. 1.2. This division is rather conditional, because it corresponds to simplified models. In many cases, the actual bonding is more complicated and often presents as an intermediate case between simple models.

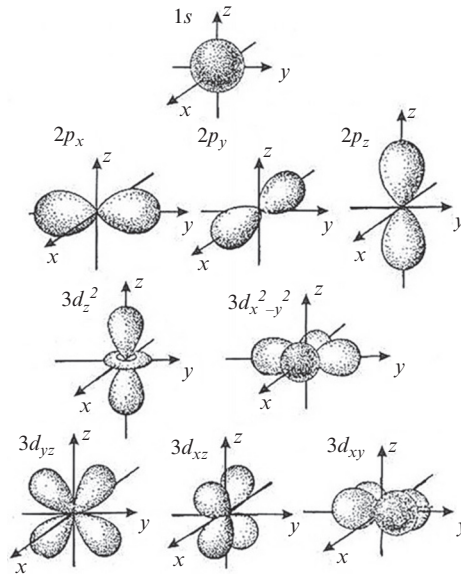


FIG. 1.1

Forms of s -, p -, and d -orbitals: angular dependence of square wave functions.

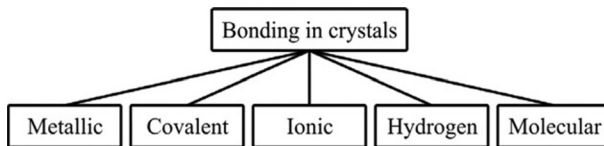


FIG. 1.2

Various models of atomic bonds in crystals [2].

Molecular and metallic bonds are shown at the opposite sides of the scheme, because they are absolute opposites. In molecular crystals, electrons usually are *completely locked* in their molecules (or atoms; Fig. 1.3A). The nuclei are surrounded by spaces (shown as black balls), where the density of the electronic cloud reaches significant values.

The simplest examples of molecular bonds are atomic crystals of solid inert gases: neon, argon, krypton, and xenon. These have completely filled electronic shells, and such a stable electronic configuration undergoes only minor changes during the formation of solids. Therefore, the inert gas crystal is an example of a rigid body with strong electron bonding exclusively *inside* atoms, whereas the electron density between atoms is rather small, because all electrons are well localized near their own nuclei.

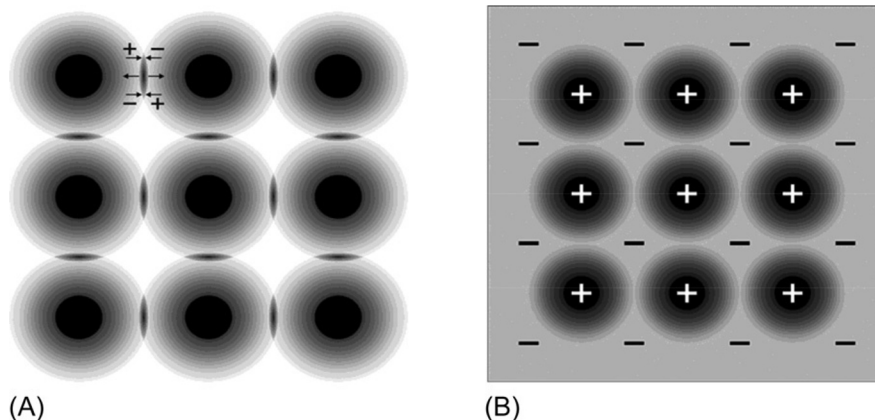


FIG. 1.3

Two-dimensional image of electrical charge distribution: (A) molecular crystal, in which quadrupole electronic fluctuation (+ - ... - +) results in the attraction of atoms, whereas partial overlapping of electronic shells leads to repulsion (←...→), thereby balancing this attraction; (B) metal crystal, black circles represent positively charged atomic residues, immersed in electronic gas.

The metal bond. As already noted, in metal atoms their outer electronic orbitals contain a rather small amount of electrons that have low ionization energy. When such atoms come closer (i.e., when a metal crystal or alloy is formed), orbitals of valence electrons largely overlap each other. As a result, these electrons become distributed almost uniformly in the space between ions (Fig. 1.3B). Indeed, X-ray studies have practically indicated a uniform electronic density in the lattice of metals.

Therefore valence electrons in metals are a joint collective in the crystal as a whole, and metal represents the lattice of positively charged ions wherein the “electronic gas” exists. This is a reason for the *delocalization* of metal bonds; moreover, metal bonds are *unsaturated* and *nondirectional*. Metals are, among crystals, characterized by the highest *coordination number* (CN) of ions (usually in metals, this number is 12; it is the number of the nearest neighbors to a given particle). For comparison, it should be noted that in ionic crystals this number is often 6 or 8. Similarly, the CN in the covalent crystals is even smaller—it is 4 for semiconductors with a diamond structure.

The bonding in **dielectrics and semiconductors** differs significantly from metal bonds. Fig. 1.4 schematically shows the *energy dependence* on the distance between atoms for bonds in basic types of dielectrics and semiconductors (metal bonds are not shown). Between particles (atoms, molecules, or ions) when creating a semiconductor or dielectric material, at relatively large distances, the forces of attraction dominate: the corresponding energy is *negative* and characterized by the curve 1. At short distances, the force of repulsion becomes much more powerful; its energy is *positive*

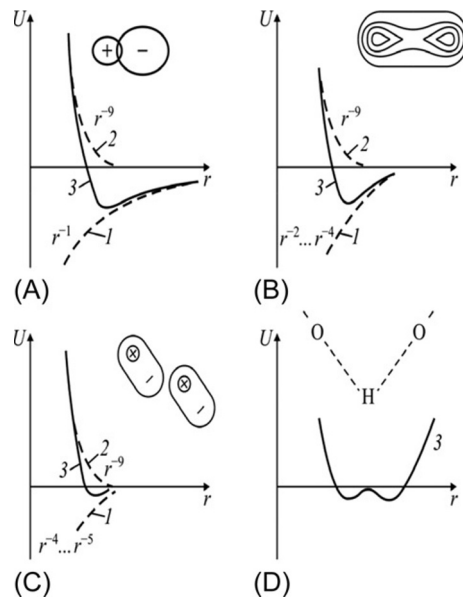


FIG. 1.4

Dependence of attraction energy (1), repulsion energy (2), and total energy (3) on the distance between particles r : (A) ionic bond, (B) covalent bond; (C) molecular (quadrupole) bond; and (D) hydrogen bond.

and characterized in Fig. 1.4 by curve 2. The total potential energy of interaction between particles is shown by curve 3 that the minimum energy corresponds to a stable distance between the interacting particles (this is *parameter of lattice*).

The strong repulsion between approaching atoms or ions can be modeled by drastic energy dependence: $U_{\text{rep}} \sim r^{-8} \dots r^{-12}$; this dependence characterizes the mutual *impenetrability* of electronic orbitals: electronic shells of neighbor atoms or ions can penetrate each other only very slightly. This is the reason that atoms, ions, or molecules can be presented by the “hard spheres” of certain radii, the size of which remains practically unchanged [3].

The attraction forces that tie atoms, ions, and molecules together in solids are of an electrical nature. It should be noted that crystals are classified just by the *nature of attraction forces*. As shown in Fig. 1.4, the main types of chemical bonds in dielectrics and semiconductors are the *covalent*, *ionic*, *molecular*, and *hydrogen* bonds. The metal bond (not shown in Fig. 1.4) can be considered a limiting case of the covalent bond.

The ionic bond. Ionic crystals (such as sodium chloride, Na^+Cl^-) are chemical compounds formed from metal and nonmetallic elements. The energy of ionic attraction varies with distance rather slowly; therefore ionic bonds are the most *long*

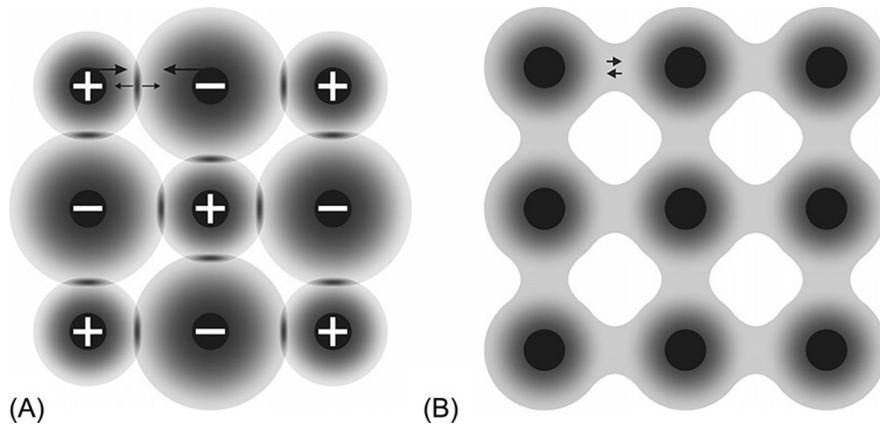


FIG. 1.5

Two-dimensional image of electronic charge distribution in: (A) ionic crystal, where ion attraction is balanced by partial overlapping of electronic shells; (B) covalent crystal, *black diffused circles* represent atomic residues surrounded by regions, where electronic density reaches significant values.

ranging in comparison with others. Similar to atomic or molecular crystals (shown in Fig. 1.3A), ionic crystals can be characterized by such a distribution of electronic charge that is almost completely *localized near ions*. In the simplest model of the ionic crystal (Fig. 1.5A), ions are “nearly impenetrable charged balls.” This approximation is rather suitable for ions that have completely filled electronic shells.

Typically, cations and anions acquire electronic configuration of the inert gas, and therefore the charge distribution in them has an almost *spherical symmetry*. Ions with opposite charges attract each other due to long-range Coulomb forces; therefore the energy of their attraction varies with distance very slowly: $U_{\text{att}} \sim r^{-1}$ (Fig. 1.4A). At the same time, the repulsive energy of ions is inversely proportional to the interatomic distance: $U_{\text{rep}} \sim r^{-8} \dots r^{-12}$ (depending on the properties of the given crystal). Therefore the ionic crystal can be considered a molecular crystal in which the lattice is built, not from atoms, but from the ions (e.g., ions Na^+ and Cl^- in the rock salt). Thus charge distribution in the ion, located in a solid body, is only slightly different as though it were an isolated ion. It is important that particles, which form ionic crystals are not neutral atoms: between ions, *large electrostatic forces* exist that play a major role and determine the main properties of ionic crystals (that differ significantly from the properties of molecular crystals).

Thus in the simplest model of an ionic crystal, all ions are presented as “interacting nearly impenetrable charged spheres”; this approximation is sufficient for ions with entirely filled electronic shells. Whereas in the atomic (or molecular) crystal, all electrons remain *locked* in their native atoms, in the ionic crystal, valence electrons are *moved* from cations to anions.

Therefore the ionic bond occurs between particles of two types, one of which easily loses electrons, forming positively charged ions (cations) and other atoms that

readily get electrons then form, respectively, negatively charged ions (anions). Most of the electropositive cations belong to groups I and II of the periodic table, whereas most anions belong to groups VI and VII.

As a rule, ions in crystals are packed tightly, as each of them is surrounded by the *largest number* of oppositely charged ions. Stabilization of the ionic solid structure takes place at CNs 6, 8, and, occasionally, even 12. It should be noted that ionic radii vary noticeably with the value of the CN. Ionic bonds, unlike metal bonds, are *saturated*, but, as in metals, they are *not directed* [4].

The covalent bond in crystals is typical for *semiconductors*. The dependence of binding energy on interatomic distance is shown in Fig. 1.5B; attractive forces in case of covalent bonds are *not so long ranging* as in the case of the ionic bond: the attraction energy changes with distance as $r^{-2} \dots r^{-4}$.

In principle, the nature of the covalent bond is very close to that of the metal bond; however, in covalent crystals, valence electrons are shared only between the *nearest neighboring atoms* whereas, in metals, valence electrons are shared within the crystal lattice. Usually, a covalent bond (i.e., *homeopolar* bond) is formed with a pair of valence electrons that have opposite spin directions. During covalent chemical bond formation, the reduction of total energy is achieved by the quantum effect of *exchange interaction*. The simplest example of a covalent bond is the hydrogen molecule H_2 , wherein both electrons belong simultaneously to both atoms.

The *diamond* might be a classic example of a covalent crystal (Fig. 1.5B), where carbon atoms are located in a rather roomy configuration: their CN is only 4. Therefore, diamond (as with semiconductors of similar structure—germanium and silicon) is characterized by a comparatively high-density electronic cloud in the atomic interstitials: electrons are concentrated mainly near the lines connecting each carbon *atom* with its four nearest neighbors. Although diamond is dielectric, the high charge density in areas between atoms is a characteristic feature of semiconductors.

Covalent bonds, unlike metal bonds, are *strongly directed*; moreover, they are *saturated*. The saturation of a covalent bond is the ability of atoms to form a *limited number* of covalent bonds. The number of bonds formed by the atom is determined by its outer electronic orbital. The directivity of covalent bonds is caused by their peculiar electronic structure and geometrical shape of electronic orbitals (the angles between two bonds are the *valence angles*).

Sometimes, covalent bonding might have pronounced *polarity* and increased *polarizability* that determines many chemical and physical properties of correspondent compounds. The polarity of a covalent bond is due to the uneven distribution of electronic density accruable to a difference in the electronegativity of atoms (therefore covalent bonds are divided into nonpolar and polar bonds). The polarizability of bonds can be expressed by the nonsymmetric spontaneous displacement of binding electrons [5].

The following polar connections are distinguished [4]:

- The *nonpolar* (simple) covalent bond arises from the fact that each atom provides one of its unpaired electrons, but the formal charge of atoms remains unchanged, because atoms that form the bond equally have a socialized electron pair.

- The *polar* covalent bond, wherein atoms are different, the degree of overlap of the socialized pair of electrons is determined by the difference in the electronegativity of atoms. The atom with a greater electronegativity more strongly attracts electrons; therefore its real charge becomes more negative. With the electronegative charge an atom acquires, there is additional positive charge of the same magnitude.
- The *donor-acceptor* bond arises when both connecting electrons are provided by one of the atoms (called the *donor*) whereas the second atom, involved in the formation of a bond, is the *acceptor*. When creating this pair, formal charge of the donor is increased by one and the formal charge of the acceptor is reduced by one. The electron pair of one atom (donor) goes into a common use, whereas another atom (acceptor) provides its free orbital, because the donor atoms usually serve atoms that have more than four valence electrons.
- The σ -bond and π -bond are approximate descriptions of some types of covalent bonds in different compounds. Therefore the σ -bond is characterized by the maximum electronic cloud density along the axis joining the nuclei of atoms. The formation of the π -bond is characterized by the lateral overlap of electronic clouds “above” and “below” the plane of the σ -bond.

Unlike metallic coupling, the emergence of a covalent bond is accompanied by such redistribution of electronic density that its maximum localizes *between* the interacting atoms. As in metals, in case of a covalent bond, the *collectivization* of the outer valence electrons is seen, but the nature of electronic allocation is different from that in metals. In the ground state of covalent crystals, that is, at $T = 0$ K, there are no partially filled electronic energy bands.

In other words, the covalent crystal cannot be described by uniform distribution of electronic density between atoms, as is typical for simple metals. Conversely, in covalent crystals, the electronic density is increased along the “best destinations,” leading to *chemical bonds*. The stronger the covalent bond, the greater the overlap of electronic clouds of interacting atoms. If this bond is formed between similar atoms, the covalent bond is the *homeopolar* and, when atoms are different, it is the *heteropolar*.

In cases where two interacting atoms share one electron pair, a *single* connection is formed; when there are two electron pairs, the *double* bond is created, and when there are three electron pairs, a *triple* bond is created. The distance between bound nuclei is defined as the *length* of the covalent bond. Bond length decreases when the order of the bond increases. For example, the length of a “carbon-to-carbon” bond depends on multiplicity: for a C—C bond, its length is 1.54×10^{-1} nm; in case of a C=C bond, the length is 1.34×10^{-1} nm, whereas for C \equiv C, it is only 1.20×10^{-1} nm [3]. With an increase of the bond order, its energy increases.

The *directivity* of covalent bonds characterizes the features of electronic density distribution in atoms. For instance, in germanium and silicon crystals (that have a diamond structure), each atom is located in the center of a tetrahedron, formed by

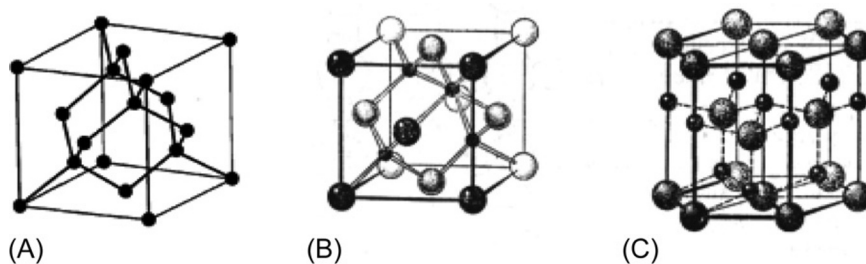


FIG. 1.6

Structure of main semiconductors: diamond (A), sphalerite (B), and wurtzite (C) [6].

four atoms, their closest neighbors (Fig. 1.6A). In this case, tetrahedral bonds are formed when each atom has only four nearest neighboring atoms.

Most covalent bonds are created by two valence (hybridized) electrons—one from each interacting atom. In case of such a connection, the electrons are localized in the space between the two atoms; thus the spins of these electrons are antiparallel. As shown in Fig. 1.5B, the plane scheme can give only an approximate representation of the actual location of atoms. In fact, the relative position of these atoms in real crystals can be quite complex, as shown in Fig. 1.6B and C. The structure of the mineral *sphalerite* (zinc sulfide, ZnS) is typical of $A^{III}B^V$ semiconductors, such as gallium arsenide. The *wurtzite* structure (calcium selenide, CaSe) is typical of $A^{II}B^{VI}$ semiconductors.

Simplified schemes of electronic density distribution in covalent and ionic crystals are shown in Fig. 1.7A and B. However, sphalerite and wurtzite belong to polar crystals that have a hybrid bonding.

The hybrid ionic-covalent bond. As with the model of a “purely covalent” structure, a model of a “purely ionic” crystal is idealized. In real crystals (especially, in some $A^{III}B^V$ and $A^{II}B^{IV}$ types of semiconductors and in active dielectrics), the intermediate case between ionic and covalent bonds exists. In the covalent silicon crystal (Fig. 1.7A), electrons are equally distributed around atoms; therefore, the electronic density between atoms is rather large. In an ionic crystal, the attraction of cation and anion is compensated by the repulsion of partially overlapping electronic shells.

The concept of intermediate type bonds agrees with the theory of ion deformation by their polarization. This may occur, for example, by the distortion of an anion’s electronic orbital, mainly by the different electronegativities of adjacent ions. Therefore, the electronic density between ionic residues increases, that is, the mixed covalent-ionic bond with a greater degree of charge separation becomes the *polar bond*. The exact presence of such bonds determines the noncentrosymmetric structure of some crystals. The hybrid ionic-covalent bonding is the main cause of pyroelectric, ferroelectric, and piezoelectric properties. Most such *active (functional) dielectrics* belong to crystals or to other ordered polar systems (liquid crystals, electrets, polar polymers, etc.). Thus the physical hypothesis, relevant to the nature of the

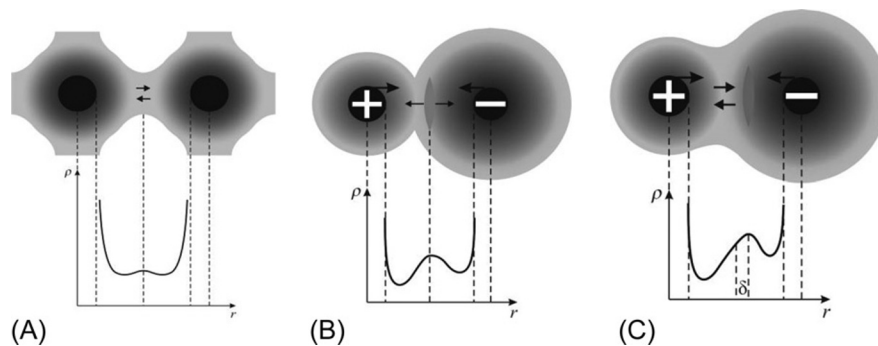


FIG. 1.7

Simplified scheme of transition from covalent and ionic bonds to mixed polar bond: (A) in covalent bond, electronic density (ρ) distribution is quite symmetric (*arrows* symbolize opposite orientation of spins in connecting electron pair); (B) in ionic bond, cation and anion are attracted (*big arrows*), whereas a small overlap of electronic shells ensures repulsion (*small arrows*); electronic density distribution is almost symmetric, (C) asymmetric mixed bond that leads to polar properties of crystals, wherein both attraction and repulsion are seen: a covalent bond is formed by electron pair with opposite spins; electronic density distribution is asymmetric, and can be characterized by displacement δ .

internal polarity (which is not caused by an external electrical field), deserves particular attention. This *hidden* (or latent) polarity manifests itself in polar crystals as the ability to provide electrical (vectorial) response to any nonelectrical scalar, vector, or more complicated tensor impacts [5].

The tendency of **polar crystals** to generate an electrical response on nonelectrical impact leads to their generation of electrical potential under uniform heating of crystals (pyroelectricity) or under uniform deformation (piezoelectricity). These are mostly crystals with *hybrid ionic-covalent bonding*. Exactly this peculiarity causes a reduction in crystal symmetry; therefore polar crystals always belong to noncentrosymmetric classes of symmetry.¹

It is obvious that the primary cause of the peculiarities of polar crystals is the asymmetry of electronic density distribution along atomic bonds. The fundamental reason for this asymmetry is a distinction in the *electronegativity* of atoms (a physical property that describes the tendency of an atom to attract electrons). Electronegativity depends on atomic number, as well as on the size and structure of outward (valence) electronic orbitals [1]. The higher the atomic electronegativity, the stronger the aptitude of atoms to attract electrons toward themselves.

¹*Comments.* In contrast, crystals with exclusively ionic bonds as well as crystals with exclusively covalent bonds are nonpolar. Usually, they belong to the centrosymmetric classes of crystals: in typically ionic crystals, a *central symmetry* exists, and there are no special orientations in atomic connections. In the same way, simple covalent crystals also belong in centrosymmetric structures: each atom provides for bond one unpaired electron; thus four socialized electron pairs are located symmetrically.

The difference of atoms by electronegativity might be very substantial. Therefore atoms with higher electronegativity strongly attract conjunctive electrons, and its true charge becomes more negative. Conversely, the atom with lower electronegativity acquires an increased positive charge. Together, these atoms create a polar connection and, correspondingly, the *noncentrosymmetric structure*. Simultaneously, such connections do not lead to the appearance of internal fields, but can provide a peculiar response to external impact that is quite different in various noncentrosymmetric crystals.

For example, in case of directional *mechanical* influence onto a polar crystal, an *electrical* response arises (piezoelectric effect). The point here is that the elastic displacement of atoms compresses (or stretches) their asymmetric connections, and thereby induces electrical charges on the crystal surface (piezoelectric polarization). In contrast, if atomic connections in crystal are centrosymmetric, no electrical response is possible to any *uniform* mechanical impact (however, the *inhomogeneous* thermal or mechanical impact makes atomic bonds asymmetric, which results in the appearance of an electrical response in any crystal).

The fact is that, in many crystals (e.g., in various semiconductor compounds), the type of bonding has an intermediate character between covalent and ionic. It is noteworthy that under conditions of *very high pressure*, any material with ionic or covalent bonding would acquire the property of a *metal bond*, and the material would turn into a metal. Thus very high pressure leads to a forced convergence of atoms with great overlap of their outer electron shells. (It should be noted that, in some rare cases, even at normal pressure, a phase transition of “dielectric-metal” is possible; this transition might be stimulated by temperature change or by an external electrical or magnetic field) [5].

The energy of ionic, covalent, and metallic chemical bonding is characterized by similar orders of magnitude. In this respect, they are much inferior to molecular bonds.

Molecular bonds (van der Waals bonds) always exist, but only when much stronger valence bonds are absent do these molecular bonds become the main type of chemical connection, primarily, in molecular crystals. Forces of attraction in this case are relatively small, being *short range*: the energy of intermolecular attraction varies with distance as $U_{\text{att}} \sim r^{-4} \dots r^{-6}$ (Fig. 1.4C). It is evident that this kind of attraction is weak in comparison with ionic and covalent forces; therefore van der Waals bonds are *additive* and *nonsaturated*.

In case of nonpolar molecules, the forces of attraction are due to the *accidental deformations* of electronic shells. Quantum fluctuations of electronic density in molecules always exist; thereby, virtual electrical dipoles (or quadrupoles) lead to molecular attraction (in Fig. 1.4C, van der Waals bonding is shown schematically and only as dipole-to-dipole interaction). The electronic polarizability of orbitals determines optical dispersion in materials; therefore forces of attraction of this type, sometimes, are called *dispersion* forces.

In case of *polar* molecules, the *orientation interaction* also contributes to the usual molecular interaction. The influence of a molecule’s intrinsic (permanent) dipole onto the induced dipole of another molecule is the *inductive interaction*.

In general, in case of van der Waals bonding, the main contribution is provided by the dispersive forces; however, when molecules have large dipole movements, the contribution of the orientation effect might be significant. As a rule, the inductive interaction is negligible [4].

The **hydrogen bond** appears between hydrogen atoms and the electronegative atoms P, O, N, Cl, and S belonging to other molecule. The nature of this bond lies in the redistribution of electronic density between atoms, conditioned by the hydrogen ion H^+ (proton; Fig. 1.4D). Crystals with hydrogen bonds (dielectrics and semiconductors) show properties similar to molecular crystals, but there is a reason to allocate them to a special class. Hydrogen is unique in the following respects:

- the residue of hydrogen ion is a “bare” proton measuring approximately 10^{-13} cm (i.e., 10^5 times smaller than any other ion);
- hydrogen needs only one electron to constitute a stable helium type; the electronic shell (unique among other stable configurations having only two electrons in the outer shell);
- the ionization potential (energy required to remove an electron from an atom) in hydrogen is high: 13.6 eV (in alkali-halide metals, it is ~ 4 eV).

Because of these properties, during crystal structure formation, the effect of hydrogen may differ significantly from the influence of other elements. Due to the high ionization potential of the hydrogen atom, it is difficult to completely remove its lone electron. Therefore the formation of ionic crystals with hydrogen occurs differently than, for example, in the case of alkali-halide metal crystals [2]. The hydrogen atom may not behave in a crystal as a typical covalent atom: when the H atom loses its electron, it can create only a *single* covalent bond, shared with another atom.

Because the size of the proton is approximately 10^{-13} cm, it is localized in the surface of large negative ions; therefore such a structure arises, which cannot be formed with any other positive ions. The energy of the hydrogen bond is less by order of magnitude than the energy of the covalent bond, but it is greater than the energy of van der Waals interactions in the order of magnitude. Although hydrogen bonds are not very strong, they play an important role in the properties of correspondent crystals.

The hydrogen bond is *directional*; molecules that form the hydrogen bond tend to have a *dipole moment* that indicates the polar nature of this bonding. In some crystals, the hydrogen bond leads to their piezoelectric, pyroelectric, and ferroelectric properties. Furthermore, it should be noted that molecular and hydrogen bonds are very important in various structures of *liquid crystals*.

1.2 SYMMETRY OF CRYSTALS

In many solids, structural symmetry plays a crucial role for the explanation of properties. Different materials are most frequently used in electronic special effects in crystals, polycrystalline materials, and polymeric films due to the peculiarities of their macro- and microsymmetries.

A crystal is a body that, due to its intrinsic properties, is limited by flat surfaces—crystal faces. A more complete definition of crystal should characterize such peculiar intrinsic properties that distinguish crystallized substance from amorphous materials and can explain the multifaceted shape of a crystal. The relationship between the outward geometry and the internal structure of crystal and its physical properties is settled by *crystallography*. It studies the physical properties of crystals through a specific method—the *symmetry* that connects physical properties of crystals with their structure. The physics of crystals formulates certain principles that establish a community of crystal symmetry and their physical phenomena; these major principles were advanced by Neumann and Curie [7].

The manifestation of symmetry in geometric forms is the ability of the shape to regularly repeat its parts. In other words, the reason for a geometrically correct external crystal shape is the regularity of the internal structure that lies in the spatial lattice of a crystal. This spatial lattice is the abstraction that allows the description of proper and regular alternation of atoms, molecules, or ions, and results in the macroscopic shape of the crystal. This lattice is infinite, and it is constructed by the translation of the unit cell of the crystal along crystallographic coordinates by endless repetition in a space with identical structural units. As a simple example, Fig. 1.8 shows various *two-dimensional* (2D) translations of unit cells on the surface. All imaging pairs of vectors \mathbf{a}_i and \mathbf{b}_i are *lattice translation* vectors, but they are not primitive vectors.

The metric of the unit cell of crystal is determined by the ideal distance between the nearest atoms or ions of a similar kind. In most simple crystals, for example, in the majority of metals the structural unit consists of only a single atom. In dielectric crystals, the unit cell may comprise a plurality of atoms, ions, or molecules. The crystal lattice may be built as a result of unit cell translational transformation, in other words, by various *point symmetry operations* [6].

The symmetry elements and operations. To describe the symmetry of a crystal's physical properties as well as to determine the symmetry of geometric forms, a quite ordinary idea is to consider only single space elements (the *unit cell* of crystal). In the theory of symmetry, the object of study is the figure, that is, a certain set of spatial points.

The imaginary geometrical object, over which symmetry operations are performed, is the *symmetry element* of the finite figure. As symmetry elements, the *planes*, *axis*, and *center* of symmetry (center of inversion) may be used.

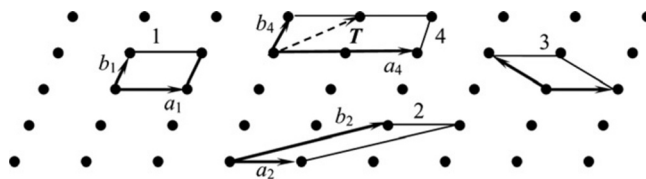


FIG. 1.8

Unit cells in a two-dimensional lattice.

The symmetry *operation* involves combining the point (or part of a figure) with another point (or part of the figure). Both combined parts of figure are symmetric. The *point symmetry* operation should be left in place at least on one point of the figure. It is the intersection point of all elements of symmetry. If symmetry operations are applied to three-dimensional (3D) figures, the twists and turns as well as the inverted turns and the reflection in a plane of symmetry are selected.

The symmetry *elements* are distinguished as first and second types. The former includes the symmetry plane, rotary axis of symmetry, and the center of inversion (center of symmetry). Complicated symmetry elements, such as inversion axis and rotary-reflection axis, belong to the second type of symmetry elements.

The *symmetry plane* is a mirror-reflecting plane that provides a combination of symmetrically equal points; when recording symmetry elements of a particular class of a crystal, the plane of symmetry can be referred to as *P*. For example, the mirror plane, being a plane in the cube diagonal, divides the cube into two equal mirror-mating parts. In the international system, the mirror plane is represented by the letter *m*. It perpendicularly bisects all segments, connecting balanced (symmetrically equal) points.

The *rotation symmetry axis* of *n*-order is denoted as L_n . When a figure turns by a specific angle of $\alpha = 360^\circ/n$ (called elementary angle), a superposition of symmetric points (equal compatible) can be realized. The rotary axes are denoted as 1, 2, 3, 4, 5, 6, 7, ..., ∞ , where the numbers indicate the order of axis.

For example, Fig. 1.9 shows a set of elements of the symmetry of a cube, which has a center of symmetry (in the cube geometric center), three axes 4 (fourth order), four axes 3 (third order), six axes 2 (second order), three planes of symmetry parallel to the faces of a cube, and six diagonal planes of symmetry. Due to the large amount of symmetry elements, the crystals of cubic symmetry are called *highly symmetric*. Other classes of crystals have much smaller number of symmetry elements.

If an *arbitrary figure*, and not a crystal, is considered, there could be any order of the rotary axis. For example, the sphere has an infinite number of rotational axes, including axes of infinite order. The cylinder has a single axis of infinite order and an infinite number of axes of order 2 (Fig. 1.10).

In certain cases, combining of a figure with its initial position must be made not only by elementary rotation angle, but also by the auxiliary reflection plane, perpendicular to the axis about which the figure rotates. The complex axis (or axis of complex symmetry) is the *mirror-rotary axis* L_{ni} . Operations that function by mirror-rotary axes can be implemented with the help of an *inversion axis* (denoted also as L_{ni}).

The order of the rotary axis of the crystal and mirror rotary axes is strictly limited. These axes can be only of the first, second, third, fourth, and sixth orders. If there are several symmetric axes, the axis with an order higher than 2 is the *principal*.

Both ends of a rotary symmetry axis might be different, in this case, it is the *polar axis*; for instance, in Fig. 1.11A, the polar axis 4 extends through a tetragonal pyramid. Polar axes are typical of a certain type of crystals (non centrosymmetric classes). As shown in Fig. 1.11, the *b* plane of symmetry *m* is *perpendicular* to the axis 4;

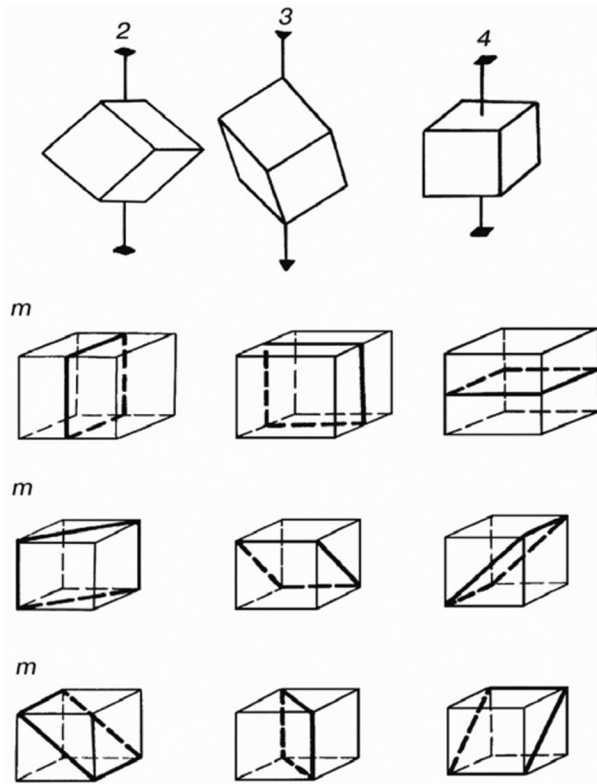


FIG. 1.9

Elements of symmetry of cube: axes of symmetry are numbered whereas planes of symmetry are denoted by the letter *m*.

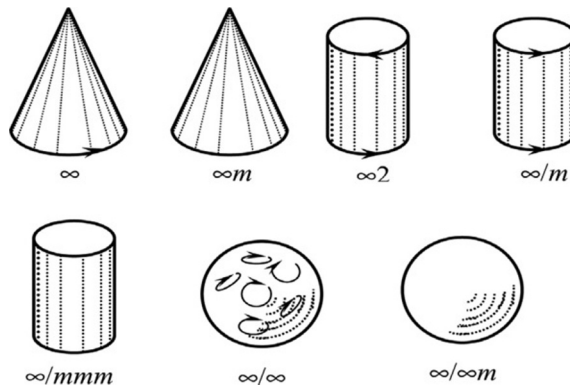


FIG. 1.10

Geometric figures representing the limiting symmetry group [7].

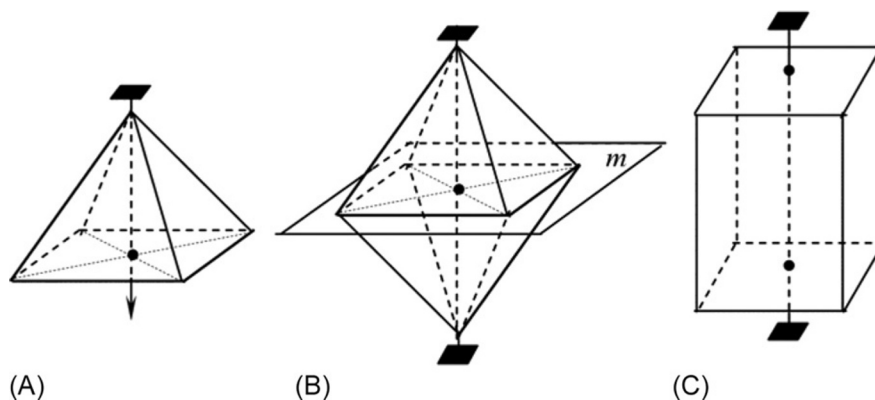


FIG. 1.11

Polar and bipolar rotary axes of 4th order: (A) tetragonal pyramid; (B) tetragonal bipyramid; and (C) tetragonal prism.

in this case, the symmetry of a figure is referred to as $4/m$. If the axis lies *in the plane* of symmetry, delimiters are not needed: $4m$. To refer to the symmetry of various crystals, it is possible to use designations: m , $2m$, $3m$, $4m$, and $6m$. Markings of the first-order axis of symmetry are not used: that is, “1” near the sign “ m ” is not needed as an axis of symmetry of first order is *always present* (when one turns the figure on 360° , any figure will coincide).

Besides the usual symmetry axes, the *inversion axes* exist. Such an axis of order n (axis L_{ni}) combines the joint action of a rotary axis and inversion center. The center of symmetry (the *inversion center*) is a singular point inside a shape (or inside the unit cell) that is characterized by the fact that any straight line drawn through the center of symmetry (denoted by symbol C) meets the same (respectively) point figures on the opposite side of the center at equal distances. A symmetric transformation in the center of symmetry is the mirror image point (Fig. 1.11B). At this point, as in a photographic lens, the image is inverted.

Sometimes, two symmetrically equal figures cannot be superposed other than by reflection. For example, in Fig. 1.12, two molecules of an organic compound are displayed (these do not have rotational axes of symmetry). The figures that can be superposed with each other only by the mirroring are the *enantiomorphous* ones. The phenomenon of enantiomorphism in crystals is expressed by the formation of left and right forms (e.g., in quartz crystal), which reflects “enantiomorphism” in its physical properties. For example, in the left form of crystals, the rotation of the polarized light plane is clockwise, whereas in right form, it is counterclockwise. This phenomenon is important from the viewpoint of the practical use of such crystals.

The concept of the “symmetry element” is broader than the concept of “symmetry operation.” The symmetry element includes all degrees of operation. For example, the axis of symmetry 4 (otherwise denoted L_4) implies a set of operations, including

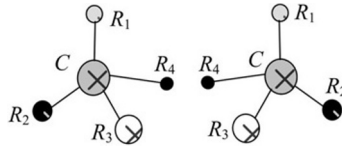


FIG. 1.12

Enantiomorphism in certain molecules [6].

$4^0 = 1$, $4^1 = 4$, $4^2 = 2$, and $4^3 = 4^{-1}$. The first operation is the operation of identification, the second is the turn on 90° , the third is the turn on 180° (turns on 180° in opposite directions are equivalent), and operation 4^3 is the turn on 270° in a certain direction, equal to the rotation in the opposite direction 90° (4^{-1}).

The **classes of symmetry** are characterized by a set of crystal symmetry elements that describes a possible symmetric transformation. For each crystal, the *unit cell* can be chosen from which the whole crystal lattice can be built via translations (*translations* are such displacements that can multiply the unit cell to create a crystal).

As a simple example, Fig. 1.13 illustrates some 2D cells (2D lattice). On a plane, each unit cell is defined by two axes (the *basis vectors*), from which the basic *elementary parallelogram* can be constructed. Such parallelograms must fill the entire plane of the 2D crystal *with no gaps*. It is important to note that, in the 2D crystal, only *five* different types of lattices are possible with a different set of symmetry elements (such elementary lattices are the *Bravais lattices*).

In the 3D space, the unit cell of a crystal lattice is the *parallelepiped*, built on three basic vectors (Fig. 1.14). The points of intersection of base vectors, composing

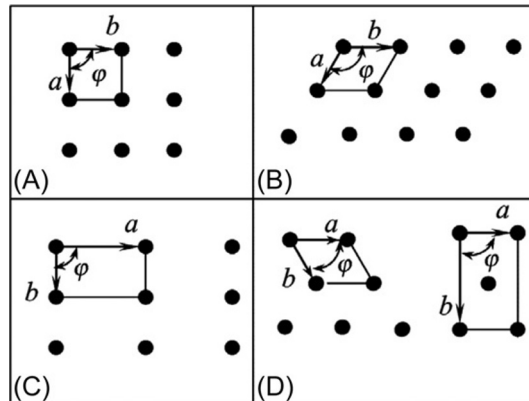


FIG. 1.13

Basic two-dimensional lattices: (A) square with $|\mathbf{a}| = |\mathbf{b}|$, $\varphi = 90^\circ$; (B) hexagonal with $|\mathbf{a}| = |\mathbf{b}|$, $\varphi = 120^\circ$; (C) rectangular, $|\mathbf{a}| \neq |\mathbf{b}|$, $\varphi = 90^\circ$; and (D) centric rectangular (axes are shown as for primitive, so for rectangular unit cells $|\mathbf{a}| \neq |\mathbf{b}|$ and $\varphi = 90^\circ$).

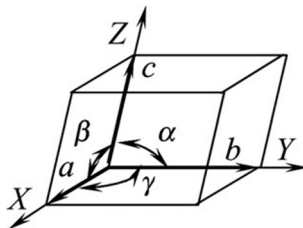


FIG. 1.14

Choice of unit cells for different classes of crystals: (A, B, and C) elementary translations on X, Y, and Z directions (called crystallographic coordinate system); α —angle opposite to X-axis; β —angle opposite Y axis; and γ —angle opposite Z-axis.

the spatial lattice, are the *junctions*. A junction may be located, as in material particles, in the center of gravity of the particle (or group of particles). As in the 2D (plane) lattice grid, the volumetric 3D primitive unit cell of a crystal does not depend on its shape and it has a constant size for a given lattice.

The spatial crystal lattice is based on parallel transport of the unit cells that are touching each other by whole faces, filling the entire space without gaps. Thus the choice of elementary translations is not unique; therefore the *shortest* of these are usually selected that correspond to basis vectors a , b , and c of a lattice. This choice is always carried out by such a way when unit cell would have the *maximal number* of symmetric elements, and thus can represent the point group of symmetry of the entire lattice. The symmetry of the crystal structure limits the choice of unit cells that can describe it. The choice of the base, and therefore the lattice itself, must comply with the symmetry of crystal structure.


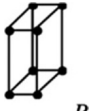
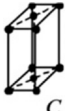
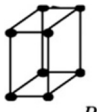
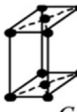



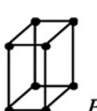

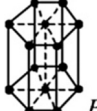

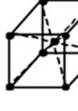

All variety of crystals in 3D structures can be described using only 14 types of lattices (Bravais lattices). They differ by the choice of unit cells and are classified by crystal *syngony*. Therefore three directions, outgoing from a single point of the selected parallelepiped, should be taken as the *coordinate axes* of a crystal, and thus define the crystallographic axes X, Y, and Z (Fig. 1.14).

The rest of the unit cell parameters are the angles between axes: α —between axes Y and Z; β —between axes Z and X; and γ —between axes X and Y. Primitive Bravais cells are the main cells that allow crystal classification by the crystallographic syngonies. Any crystalline structure can be presented with one of 14 Bravais cells listed in Table 1.1.

Any linear periodic structure can be obtained by elementary translation. To choose a cell, three guided conditions should be used:

- the symmetry of the unit cell must correspond to the highest symmetry of the crystal;
- the unit cell should have largest possible number of identical angles, or corners and edges; and
- the unit cell should have the minimal volume.

Table 1.1 Fourteen Bravais Cells

Crystal System (Lattice Basis)	Lattice type			
	Primitive	Base-Centered	Body-Centered	Face-Centered
Triclinic $a \neq b \neq c$; $\alpha \neq \beta \neq \gamma \neq 90^\circ$	 <i>P</i>			
Monoclinic $a \neq b \neq c$; $\alpha = \gamma = 90^\circ$; $\beta \neq 90^\circ$	 <i>P</i>	 <i>C</i>		
Rhombic (orthorhombic) $a \neq b \neq c$; $\alpha = \beta = \gamma = 90^\circ$	 <i>P</i>	 <i>C</i>	 <i>I</i>	 <i>F</i>
Trigonal (rhombohedral) $a = b = c$; $\alpha = \beta = \gamma \neq 90^\circ$	 <i>R</i>			
Tetragonal $a = b \neq c$; $\alpha = \beta = \gamma = 90^\circ$	 <i>P</i>		 <i>I</i>	
Hexagonal $a = b \neq c$; $\alpha = 120^\circ$; $\beta = \gamma = 90^\circ$	 <i>P</i>			
Cubic $a = b = c$; $\alpha = \beta = \gamma = 90^\circ$	 <i>P</i>		 <i>I</i>	 <i>F</i>

Next, different types of Bravais lattices are usually distinguished as primitive, volume-centered, border-centered, base-centered, and rhombohedral. Division along the crystallographic syngonies determines a choice of the coordinate system and the triples of basis vectors \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 , or, in other words, the metric (γ , β , and α and a , b , or c).

Table 1.2 Distribution of Syngonies of Crystallographic Point Groups

	Syngonies	Classes of Symmetry
1	Triclinic	$1, \bar{1}$
2	Monoclinic	$m, 2/m$
3	Orthorhombic	$mm2, 222, mmm$
4	Trigonal	$3, \bar{3}, 3m, 32, \bar{3}m$
5	Tetragonal	$4, 4/m, 4mm, 422, 4/mmm, \bar{4}, \bar{4}2m$
6	Hexagonal	$6, 6/m, 6mm, 622, 6/mmm, \bar{6}, \bar{6}2m$
7	Cubic	$23, m\bar{3}, \bar{4}3m, 432, m3m$

According to the syngonies, there are *seven types* of crystal structures: *triclinic*, *monoclinic*, *hexagonal*, *rhombohedral* (trigonal), *orthorhombic*, *tetragonal*, and *cubic*. There are 32 classes of symmetry possible in structures with 3D point symmetry, and their distribution in crystal syngonies are shown in [Table 1.2 \[7\]](#).

The items used in this set of symmetry elements are not a random choice but are strictly legitimate mathematical groups. The aggregates of these symmetry elements are introduced according to certain rules. Symbolic images of 10 polar crystal classes are shown further in this work; it is only within these classes that pyroelectric and ferroelectric properties can possibly exist.

The principal peculiarity of the *lowest category of symmetry* is the lack of such symmetry axes that they are greater than the second order. The lowest categories of symmetry are classified as triclinic, monoclinic, and orthorhombic syngonies. The triclinic lattice is the only one that has no elements of rotational symmetry or mirror planes. The lowest-category structures are rarely encountered in nature, and they have lattices based on a cell with three unequal edges and three unequal angles.

In crystals belonging to the *middle category of symmetry*, a symmetry axis obviously exists, which is beyond the second order (major axis). In addition to the major axis, the structure might have a plane m and the center of symmetry C . The middle categories of symmetry are the trigonal, tetragonal, and hexagonal crystal systems.

The crystals of *highest category of symmetry* are crystals of cubic symmetry (cubic system); their main distinguishing feature is the presence of four threefold axes ($4L_3$).

Among the many thousands of natural and artificially synthesized crystal structures studied to date, more than half account for the highest category of symmetries. Almost all metals and their alloys crystallize in a cubic system with class $m3m$ or in a hexagonal class $6/mmm$. The semiconductors, such as germanium and silicon, are also classified as $m3m$ class; however, the vast majority of semiconductors, including gallium arsenide (GaAs), is related to the point group of cubic symmetry $43m$ (sphalerite-type structure), as well as to the point group of the hexagonal system $6m2$ (wurtzite structure).

Almost no substances crystallize in the groups 4, 3, 6, and 432.

Connection between symmetry and physical properties. The characteristic features of crystals are their symmetry and anisotropy. The properties of some anisotropic crystal, unlike isotropic crystal, exhibit very high sensitivity to external influences. Influences such as heat, mechanical stress, and electrical and magnetic fields (or adding foreign atoms into crystal) can change conditions of the dynamic equilibrium of the crystal's constituent particles, that is, they can change the symmetry of the crystal and, thus, change its properties. The ability to manage properties using the abovementioned external influences allows the creation of crystal-based converters of various types of energy.

The connection between the physical properties of crystals and their symmetry was formulated by Von Neumann:

The symmetry of physical properties of a crystal is not lower than the symmetry of its structure.

This means that the structure of crystal, in any case, comprises all elements of the properties' symmetry (but may also have other symmetric elements).

The application of Neumann principle in specific experimental situations was detailed by Pierre Curie. In accordance with the Curie principle, a crystal, being under the impact of external influence, demonstrates such symmetry elements as are common to a crystal in the absence of impact, and the symmetry of impact itself in the absence of a crystal. Thus in the system "crystal-impact," only the common element of symmetry remains. A geometric illustration of the Curie principle may be the superposition of two symmetric figures: this gives a figure possessing only those symmetry elements that are common to both figures at a predetermined point in their mutual arrangement.

Thus the concept of symmetry is expanding. Symmetry is regarded as a state of space that is characteristic of the environment in which a phenomenon occurs. For example, the following factors should be taken into account during crystal growth:

- the status and structure of the environment (e.g., solution or melt);
- the movement of seed during crystal growth with respect to the environment in which the crystal is forming; and
- the impact of other physical factors on growing crystals.

The shape of the grown crystal retains only those elements of its own symmetry that coincide with the symmetry of the medium; therefore some of the crystal's symmetry elements disappear externally. Only those elements of proper symmetry should be accepted that coincide with the symmetry of the environment.

Thus the crystal responds to changes in the conditions of crystallization. Therefore different natural shapes of a crystal correspond to crystallization conditions. Under essentially different physical and chemical conditions, the minerals of the same composition acquire different structures (phenomenon of *polymorphism*). For example, the forms of pure carbon are polymorphic: cubic diamond, hexagonal diamond, multilayer graphite, globular fullerene, quasi-1D carbon, flat graphene, cylindrical nanotubes, etc.

Using the known point symmetry of a crystal, the Curie principle allows to predict what physical effects (typical for a given symmetry) can occur. However, the symmetry conditions (that follow from fundamental laws) are considered as not necessary, but rather as sufficient, because of their “abstract” nature for implementation to a physical phenomenon.

Direct and inverse piezoelectric effects can occur only in 20 of the 32 possible classes of crystals, each of which is characterized by its symmetry group. These groups comprise a set of symmetry elements: axes of symmetry and planes of symmetry. Crystals with the center of symmetry cannot be piezoelectric. There are 11 such classes of crystals out of 32 (however, there is another class 432 that refers to the noncentrosymmetric classes, but piezoelectricity is not observed in it in spite of the center of symmetry being absent).

The 10 polar groups of crystals that exhibit a *pyroelectric* effect are called the “pyroelectric group” (Fig. 1.15). A common feature of this group is that they lack certain elements of symmetry: the center of symmetry, transverse plane of symmetry, and any axes of symmetry of an infinite number, perpendicular or oblique with respect to the current axis.

Physical and crystallographic installations of crystals. To investigate the relationship between crystal properties and crystal symmetry, it is necessary to consider the orientation of a plate sample cut from a single crystal (as well as the orientation of crystalline rods or disks) as to the crystal elements of symmetry. This operation is performed using *crystal installation* that is determined by the choice of the coordinate system with respect to the symmetry elements of the crystal [8].

As a rule, two settings of crystals are used: the *crystallographic* installation (used during electronic spectra study in semiconductors and metals) and the *physical* installation (used in crystal physics including material sciences).

In the crystallographic installation, the coordinate axes should be chosen parallel to the directions of space lattice translations. In this case, the crystallographic axes of a coordinate system may not be mutually perpendicular to each other (in crystals belonging to the triclinic, monoclinic, rhombohedral, or hexagonal systems).

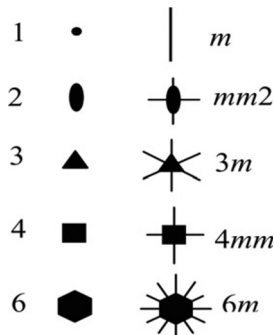


FIG. 1.15

Ten polar crystallographic symmetry groups [6].

Together with the crystallographic coordinate system that is not orthogonal for all classes of crystals, the *orthogonal* coordinate system, with axes denoted by either X , Y , and Z , or 1, 2, and 3, is selected to describe the physical properties of crystals. In this case, the symmetry axes or normals to planes of symmetry are chosen as coordinate axes.

For example, in monoclinic syngony, the Y -axis is oriented along a single axis of the second order, or along a direction perpendicular to the single plane of symmetry. The remaining two axes X and Z can be chosen arbitrarily, usually by a “binding” to the most advanced face or edge of a crystal. In the orthorhombic system, the crystallographic axis must be directed along the axes of second order, or perpendicular to the plane of symmetry m . In the class $mm2$, the symmetry axis is defined as axis Z ; for a tetragonal crystal, the Z -axis is the axis of the fourth order.

In all classes of point symmetry, the X - and Y -axes (except for 4 , $\bar{4}$, and $4/m$, where they are chosen randomly) are oriented along twofold axes or perpendicular to the plane of symmetry m . In the hexagonal system, the Z -axis is oriented along the axis of symmetry of the highest order. In classes $3m$ and $\bar{6}m2$, the X -axis must be directed perpendicular to the plane of symmetry. In cubic crystals, the axis 2 is selected as Z -axis (for classes 23 and $m\bar{3}$), or 4-axis and $\bar{4}$ -axis (for other classes). The X - and Y -axes are oriented along the axes of symmetry. Importantly, in all cases the X -axis and the Y -axis are selected in such a way as to form the *right-hand* coordinate system.

In case of any spatial lattice symmetry, the size of the unit cell (\mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3) is selected as a scale (individual segments). Coordinates of any point of crystal are uniquely determined by the direction of the symbol.

The *crystallographic direction* is a direction of line that runs at least two lattice points. One of these points can be taken as the origin: $[000]$. The crystallographic direction \mathbf{r} is completely determined by aligning on it the lattice point closest to the origin, and it is denoted as $[mnp]$, where m , n , and p are the *Miller indices*. The vector \mathbf{R} that coincides with the given direction can be written as:

$$\mathbf{R} = m\mathbf{a}_1 + n\mathbf{a}_2 + p\mathbf{a}_3.$$

Irrespective of the angle between the coordinate axes, the crystallographic axes must follow Miller indices: the X -axis is $[100]$, the axis Y is $[010]$, while the Z -axis is $[001]$. The indices of axes are written in square brackets. The position (orientation) of each face of a crystal can be described by using the ratio of unit segments a , b , and c to segments A , B , and C that cut off axes by a given face (Fig. 1.16). The set of relations a/A , b/B , and c/C can always be expressed as the ratio of integers $a/A : b/B : c/C = h : k : l$. These three numbers h , k , and l determine the position of each edge of the crystal, and they are commonly called *Miller indices of edge*, written in parentheses as (hkl) .

In this description, the crystal face is displayed by the position of unit normal \mathbf{n} to it, while a set of Miller indices is the component of the normal vector \mathbf{N} to a given face relative to the basis of the reciprocal lattice of the crystal: \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_3 , which is also called the *reciprocal basis*; that is,

$$\mathbf{N} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3,$$

because $n_1 : n_2 : n_3 = A : B : C = 1/a : 1/b : 1/c$.

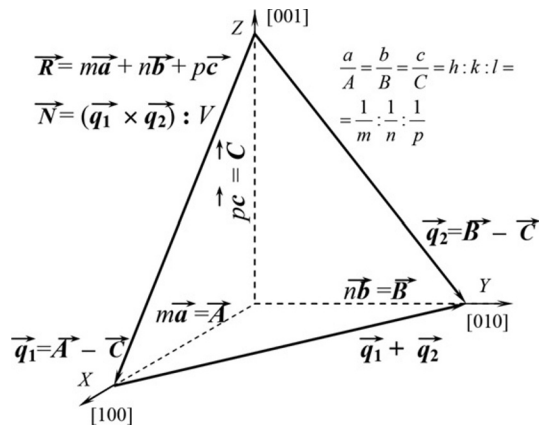


FIG. 1.16

Explanation of symbols: the position of the plane is uniquely determined by integer intercepts on the crystallographic axes of the coordinates.

Fig. 1.17 shows the main crystallographic directions with an example of a cubic lattice.

The reciprocal lattice. This notion is introduced by Gibbs and represents the Fourier transformation of the Bravais lattice that is also known as the *direct lattice* (that exists in real space). The reciprocal lattice exists in the reciprocal space, also known as the *impulse (momentum) space*.

If one makes a Fourier transformation with the reciprocal lattice, the original direct lattice will be found again, because the two lattices are Fourier transformations of each other [3]. The concept of a reciprocal lattice is used to solve many problems related to *wave processes* in crystals, for example, in X-ray experimental study or when using electronographic and neutronographic methods of crystal investigation. Moreover, the reciprocal lattice is widely used in the physics of semiconductors and

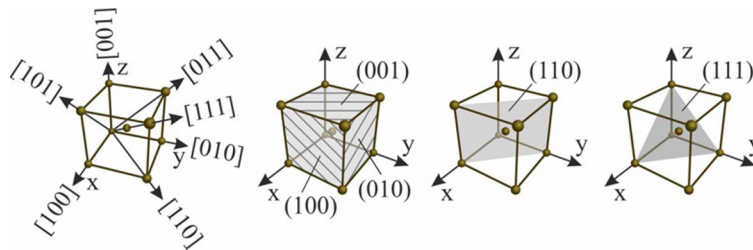


FIG. 1.17

Marking of main crystallographic directions and planes.

metals to describe the motion of electrons in the periodic structure. In that case, the concept of the *Brillouin zone* is introduced.

If a normal direct lattice is based on the translation vectors $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$, the axes of the reciprocal lattice $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ are defined as the vector products:

$$\mathbf{b}_1 = [\mathbf{a}_2 \times \mathbf{a}_3], \mathbf{b}_2 = [\mathbf{a}_3 \times \mathbf{a}_1], \mathbf{b}_3 = [\mathbf{a}_1 \times \mathbf{a}_2].$$

Furthermore, it is possible to set a reciprocal lattice by the scalar products:

$$(\mathbf{b}_1 \cdot \mathbf{a}_1) = (\mathbf{b}_2 \cdot \mathbf{a}_2) = (\mathbf{b}_3 \cdot \mathbf{a}_3) = 1,$$

and

$$(\mathbf{b}_1 \cdot \mathbf{a}_2) = (\mathbf{b}_1 \cdot \mathbf{a}_3) = (\mathbf{b}_2 \cdot \mathbf{a}_3) = (\mathbf{b}_2 \cdot \mathbf{a}_1) = (\mathbf{b}_3 \cdot \mathbf{a}_2) = (\mathbf{b}_3 \cdot \mathbf{a}_1) = 0,$$

inasmuch as

$$\mathbf{b}_1 \perp \mathbf{a}_2, \mathbf{b}_1 \perp \mathbf{a}_3, \mathbf{b}_2 \perp \mathbf{a}_3, \mathbf{b}_2 \perp \mathbf{a}_1, \mathbf{b}_3 \perp \mathbf{a}_2, \mathbf{b}_3 \perp \mathbf{a}_1.$$

Thus the absolute value of each of the vectors \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_3 is inversely proportional to direct lattice distances:

$$|\vec{b}_1| = \frac{[\vec{a}_2 \times \vec{a}_3]}{\vec{a}_1 \cdot [\vec{a}_2 \times \vec{a}_3]}; \quad |\vec{b}_2| = \frac{[\vec{a}_3 \times \vec{a}_1]}{\vec{a}_2 \cdot [\vec{a}_3 \times \vec{a}_1]}; \quad |\vec{b}_3| = \frac{[\vec{a}_1 \times \vec{a}_2]}{\vec{a}_3 \cdot [\vec{a}_1 \times \vec{a}_2]}.$$

Direct and reciprocal lattices are mutually connected, that is, a lattice, built on vectors a_1 , a_2 , and a_3 , is the reciprocal lattice relatively b_1 , b_2 , and b_3 , when

$$|b_1| = (a_2 a_3 - \sin \alpha) / V, |b_2| = (a_3 a_1 - \sin \beta) / V, |b_3| = (a_1 a_2 - \sin \gamma) / V,$$

in which connection:

$$\cos \alpha^* = (\cos \beta \cos \gamma - \cos \alpha) / \sin \beta \sin \gamma;$$

$$\cos \beta^* = (\cos \alpha \cos \gamma - \cos \beta) / \sin \alpha \sin \gamma;$$

$$\cos \gamma^* = (\cos \alpha \cos \beta - \cos \alpha) / \sin \alpha \sin \beta,$$

where V is the volume of a unit cell of the direct lattice, whereas α^* , β^* , and γ^* are angles between the axes of the reciprocal lattice.

In physics of crystals, the accurate establishment of rules is very important, that is, rules of orientation of symmetry elements along coordinate axes, because this affects the unambiguous determination of main directions and faces in crystals. The choice of positive directions of axes in crystals is imposed by certain conditions. For example, to study the electrical properties of pyroelectric crystals, the positive direction of the axis (that coincides with polar axis) should be selected as the direction that shows positive electrical charges while heating [7].

The conception of the “reciprocal lattice” is introduced in crystallography mainly to describe the periodic distribution of reflectivity of a crystal relative to X-rays. The reflection of X-rays from planes of crystal is described by the Wulff-Bragg formula

$$2d \sin \theta = n\lambda,$$

where d is the interplanar distance for the family of parallel planes of reflection; θ is an angle, supplementary to the angle of incidence (or angle of reflection, calculated from the atomic plane); n is an integer factor that characterizes the order of the diffraction spectrum; and λ is the wavelength. From the Wulff-Bragg conditions, it follows that, at constant λ , the big d corresponds to a small θ , that is, the greater the interplanar distance, the closer the direction of reflected rays to the direction of the incident beam. The reflection of X-rays from an infinitely extended crystal should be dotted, ideally.

1.3 BASIC STRUCTURES OF CRYSTALS USED IN ELECTRONICS

In any crystal structure, a definite group of bound atoms exists that corresponds to the major structural unit—the *basis*. This is a set of particles, coordinated within an elementary cell; therefore, the whole crystal structure can be obtained by repetition of this basis using translations. An important parameter of a structure is the CN. For a given atom (or ion), this is the number of the nearest (neighboring) atoms or ions in the crystal structure. This number is determined somewhat differently for molecules and crystals. The number of interior atoms is the *bulk CN*, whereas the number of atoms located on the surface of a crystal is the *surface CN*.

If the centers of the nearest atoms or ions connect with each other by straight lines, it generally gives rise to the *coordination polyhedron* [8]. The atom, for which the coordination polyhedron is built, is located in the center of the polyhedron (Fig. 1.18). The coordination polyhedron is not related to the outward form of a crystal.

The size of the structural unit (atom, molecule, or ion) depends on its location in a particular structure. However, when considering different structures, it is important to compare the sizes of structural elements. In part, the term “atomic radius” is used, but it should be noted that an isolated atom has no certain radius, because its electronic cloud, theoretically, extends to an infinite distance from the nucleus, although

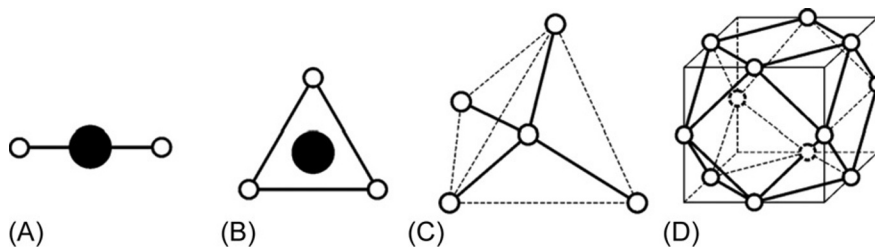


FIG. 1.18

Coordination polyhedra: (A) dumbbell, CN=2; (B) triangle, CN=3; (C) tetrahedron, CN=4; and (D) cuboctahedron, CN=12.

really it becomes very diffuse at a distance of a few angstroms. The only atomic radius that has certain sense is the Bohr radius of the outer orbit.

In case of metals, the radii of the metal ion could be defined by dividing in half the interatomic distance of pure metal. In fact, whenever the atom of any metal is described, three main types of radii are known: ionic (r_i), metallic (r_m), and covalent (r_c). For example, in sodium $r_i = 0.95$, $r_m = 1.57$, and $r_c = 1.59$. It is seen that covalent and metallic radii are very close in size; this is expected because metallic and covalent bonds are related. However, the ionic radius of a cation is smaller, because the outer electrons are removed from the atom, while remaining electrons are located on the levels of internal electron clouds (i.e., located much closer to nucleus).

In all nonmetals as well, three radii can be chosen: ionic (r_i), covalent (r_c), and van der Waals (r_v). It should be noted that the typically covalent radius is much smaller than the ionic and van der Waals radii. For example, in oxygen, $r_c = 0.66$, $r_i = 1.40$, and $r_v = 1.40 \text{ \AA}$. This can be explained as follows: suppose that the radius of the atom is determined by the position of the maximum in the radial distribution of the electronic density curve. When the atom attaches electrons and creates an anion, this maximum shifts to a longer distance due to the enlargement of the number of external electrons and the increased screening from the nucleus. Therefore the curve of radial distribution of electronic density shifts toward a larger radius.

In *ionic* crystals, the shells of cations and anions are completely filled with paired electrons; therefore the overlap of such shells is small. However, in *covalent* crystals, their shells are occupied by paired electrons and therefore the overlap is as large as possible; this interpenetration of electronic orbitals should cause a reduction in the size of the atom. The van der Waals radii, as expected, are very close to ionic radii. In case of a van der Waals connection, as well as for ionic bonding, completely filled electronic shells are situated adjoining each other. As a result of a slight overlap, the interatomic distances get larger values.

During consideration of complex structures, the following circumstances must be considered:

1. In ionic structures, the amount of anions surrounding any cation can be determined by the ratio of their radii. When the covalence part of the bond increases, the coordination of the particle will be determined by hybrid orbitals, and the CN should be less than would be expected from an exclusively ionic model.
2. When there is a choice between “ionic” and “layered” structures, in case of covalent compounds the latter is preferred, especially at low temperatures. If hydroxyl groups are present, their hydrogen bonds also tend to stabilize the layered structure.
3. In complex structures, the charges—attributed to atoms in assumption of an ionic model—aspire for compensation (saturation) within inner surroundings.
4. If stoichiometry allows, multivalent cations are located far away from each other; thus each anion is directly associated with only one cation.
5. The cations, after maximal mutual distancing, seek to form linear bridges through anions. This effect is most typical for multivalent cations with a small radius.

Next, some typical structures of chemical elements and binary compounds in solids are discussed as examples.

Typical structures of metals. Early ideas of the structure of metals lie in the model of “free electronic gas.” According to this model, it is assumed that atoms of a metal are entirely ionized and they are densely packed in the environment of free electrons. With this model (with free electrons and nondirectional bonds), high electrical conductivity and ductility of metals are easily explained. With regard to the structure, electronic gas occupies a small volume (free electron radius is estimated as 10^{-13} cm); therefore, ions of metal form a most densely packing.

For example, the structural type of *copper* corresponds to a dense three-layer packing of identical balls according to a face-centered cubic elementary unit (Fig. 1.19A). This cubic unit cell contains four atoms; its CN = 12, whereas its coordination polyhedron is cuboctahedron. This type of structure is intrinsic for simple metals (gold, silver, nickel, aluminum, etc.) as well as for noble gases in the solid state (e.g., Ne, Ar, Cr, Xe).

The balls shown in Fig. 1.19A are not hard: in crystallography, the outer electronic shells of ions, atoms, or molecules are modeled. These electronic shells can be imagined as a “cloud” of electrons that are resilient, but can partly penetrate into each other. This manner of atoms modeled as “hard balls” is a convenient method to describe structures of crystals.

The structural type of *magnesium* is another example of the simple structures of metals (Fig. 1.19B). The arrangement of atoms in this case corresponds to a hexagonal (two-layered) shape with the most dense packing. All atoms in this case are the same, with CN = 12. Ideally, in a densely packed hexagonal metal, the ratio of the unit cell height to the distance between neighboring atoms is $c/a = 1.633$, although c and a are different in various metals. The structural type of magnesium is typical of many metals: Be, Cd, Mg, Ni, Zn, etc.

In addition, there are other structures of metals (not shown in the figure), namely, the α -tungsten structure. This is a space-centered cubic structure, in which the unit cell contains two atoms. The CN of such a structure is CN = 8, while coordination polyhedron is a cube. A structure of this type is typical for some metals: Ba, Cr, α -Fe, K, Li, Mo, Na, and others.

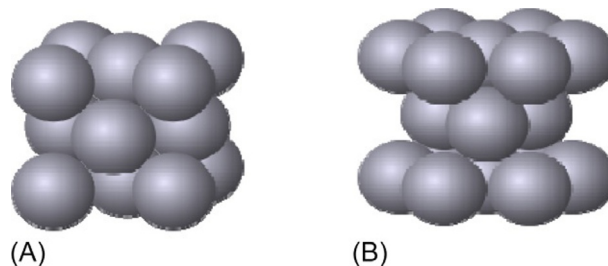


FIG. 1.19

Location ions in the structures: (A) copper and (B) magnesium.

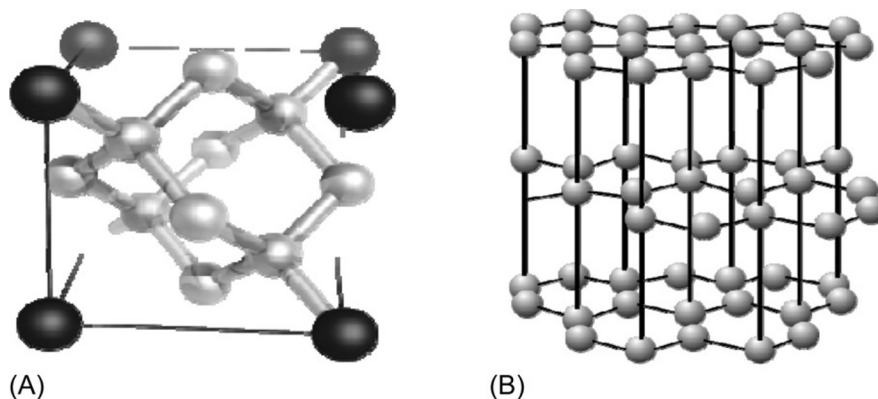


FIG. 1.20

Structural typing of diamond (A) and graphite (B).

Basic structures of semiconductors. The structure of the *diamond* that was previously shown schematically in Fig. 1.5B is shown in greater detail in Fig. 1.20A. This structure is characterized by the manner in which atoms occupy face cells of two units that are inserted into each other, offset by $1/4$ along the diagonal of the cubic unit cell.

The structure of the diamond is peculiar to materials with sp^3 -hybridization of atomic orbitals. Each atom forms four bonds with its neighbors. The basic structural unit cell of the diamond contains eight atoms, with $CN=4$, while coordination polyhedron is regular tetrahedron. The density of structural packing in diamond is much lower than in others. Germanium, silicon, and gray tin also have structures similar to that of the diamond. Similar to this structure is the structure of zinc blende, that is, if two diamond sublattices are occupied by different types of atoms, such as in crystals ZnS or GaAs.

The structure of graphite is characterized by a cleavage. In the hexagonal modification (Fig. 1.20B), graphite layers are placed such that atoms of third layer are located just above the atoms of the first layer at a distance far greater than the distance between the atoms inside the layer. The unit cell contains four atoms, with $CN=3$, and a coordination polyhedron that is an equilateral triangle with a central atom that is located slightly above or below the plane of the triangle.

Like other layered structures, graphite has some polytypic modifications.

The sphalerite and wurtzite structures. Crystals of ZnS are crystallized in a cubic sphalerite structure (also called the zinc blende structure), or they are crystallized in a hexagonal wurtzite structure. In both structures, each zinc ion is tetrahedrally surrounded by sulfur ions; in turn, each ion of sulfur is surrounded by ions of zinc. This structure should be seen as densely packed with sulfur ions and in which the zinc ions occupy half of the tetrahedral voids. Accordingly, the structure of sphalerite has cubic packing; therefore this structure resembles a diamond, in which the unit cell

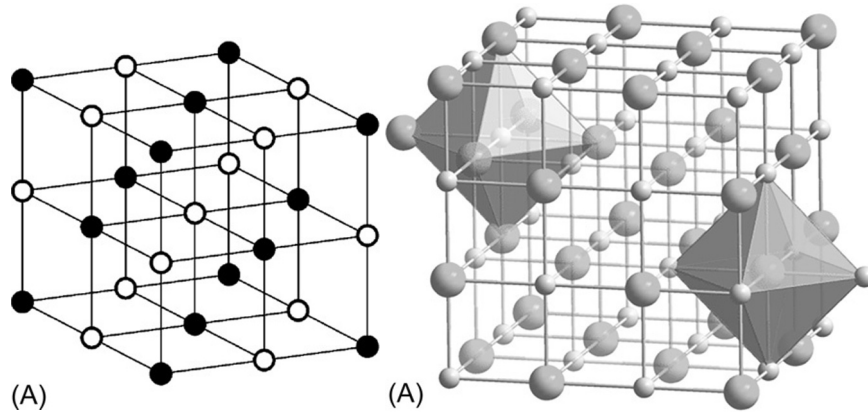


FIG. 1.21

Structure of sodium chloride: (A) centers of ion location, (B) tetrahedral surrounding of ions.

contains four sulfur anions and four zinc cations with $CN=4$. The wurtzite structure has a hexagonal packaging of the unit cell that is schematically shown in Fig. 1.21C. The CN is 4 in both structures.

The sphalerite and wurtzite structures are characteristic of many semiconductor crystals: AgI, AlR, GaR, GaAs, CdS, CdTe, SuI, SuF, NgS, NgSe, ZnS, ZnSe, ZnTe, ZnO, etc.

There are certain differences between diamond-type semiconductors and metals. In semiconductor compounds, each bond has a pair of electrons, and these electrons are localized with these bonds. Metals are characterized by a higher number of bonds than the number of electron pairs, whereas the electrons are not localized, but “smeared” throughout a structure.

The basic structures of dielectrics. The structure of *rock salt* (otherwise halite, sodium chloride, or NaCl) is shown in Fig. 1.21. Chlorine anions form a face-centered cubic structure while sodium cations fill all octahedral voids. The unit cell contains four sodium ions and four chloride ions; the $CN=6$; the coordination polyhedron—an octahedron—is shown in Fig. 1.21B. The structure of halite is characteristic of alkali halide crystals (except cesium haloids) and some oxides of transition metals (MnO, FeO, etc.), as well as for nitrides and carbides of transition groups Ti and V, haloids of silver (AgCl, AgBr, AgF), tin sulfides, and selenides.

The *structure of cesium chloride* is characterized by anions of chlorine that occupy the cubic cell, whereas cesium cations retain voids between them. In the compound CsCl, the radius of the Cs^+ cation is 1.69 Å while anion Cl^- has a radius of 1.81 Å. Smaller Cs^+ ions should determine the CN ; its ionic radii ratio is 0.93 and $CN=8$, which is observed in reality. The coordination polyhedron is cubic; therefore the structure of CsCl is cubic space-centered (Fig. 1.22).

It should be emphasized that these (and many other) images of crystal structures represent only the spatial arrangement of the *centers of atoms* (middle position of

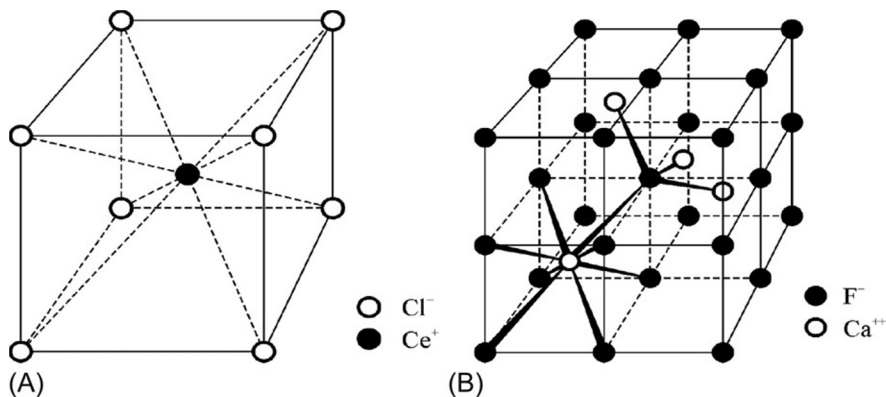


FIG. 1.22

Cubic structure of cesium chloride (A) and fluorite (B).

their nuclei). However, the maximum electronic density is located around nuclei or along areas between adjacent cores. The CsCl type of structure is peculiar for some alkali halide crystals (CsBr, CsI, RbCl, RbBr, RbI, TlCl) and for some metal alloys (FeTi, NiTi, CdAg, AgLi).

The *fluorite* (CaF₂) structure is characteristic in some compounds of the AB^{II} type (Fig. 1.21B). Preserving conditions of neutrality results in the fact that the CNs of cations and anions are different. The calcium ion is surrounded by eight ions of fluoride whereas each fluoride ion is surrounded by four calcium ions. This structure is most favorable for the emergence of Coulomb interaction forces between particles. Compounds that crystallize in the fluorite structural type are SrF₂, ZrO₂, Li₂O, CuF₂, K₂O, CeO₂, Cu₂Se, Na₂O, etc.

The *rutile* (TiO₂) structure. The radius of the Ti⁴⁺ cation equals 0.68 Å while radius of the anion O²⁻ is 1.40 Å. The ratio of the radii is 0.49; therefore, around the titanium ion, there are *six anions* of oxygen. Thus each titanium atom is surrounded by the *octahedral* group of O²⁻, and each oxygen ion is surrounded by three Ti⁴⁺ cations, thereby forming a triangle (Fig. 1.23A). Crystals that belong to the structural group of rutile are MgF₂, MnO₂, CrO₂, PbO₂, ZnF₂, etc.

The *structure of corundum* (Al₂O₃) is typical of some sesquialter oxides, such as Fe₂O₃ (hematite), Ti₂O₃, and Cr₂O₃. In corundum, each aluminum atom is surrounded by distorted octahedral groups of oxygen atoms. In accordance with the requirements of preserving neutrality, every oxygen atom is surrounded by a tetrahedral group of aluminum atoms. Details of the structure of corundum are shown in Fig. 1.23B: oxygen atoms form a hexagonal densely packed structure that comprises the layers O and A1, whereas one-third of the possible locations in aluminum remain unoccupied.

Moreover, there are other, less stable forms of Al₂O₃, demonstrating the “flexibility” of some structures. One of Al₂O₃ modifications crystallizes in the NaCl structural type; however, this structure is faulty: of three cationic places, one remains free. Another Al₂O₃ modification crystallizes in the structural type of spinel.

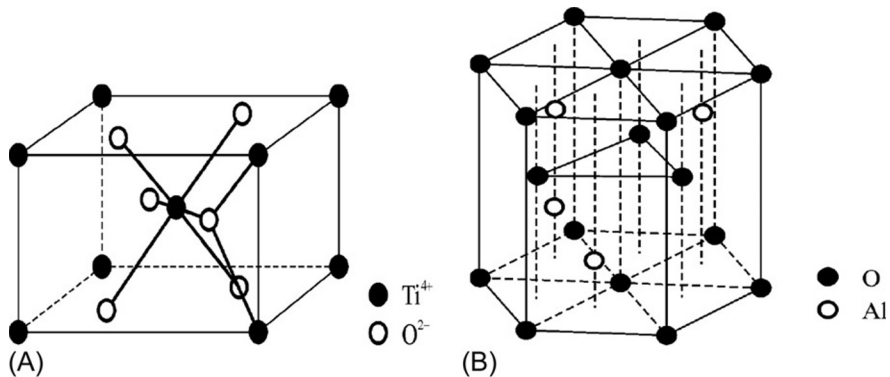


FIG. 1.23

Structures of rutile (A) and corundum (B).

The *spinel structure* is peculiar to many oxides with the formula Me_3O_4 . In this structure, oxygen atoms form a dense cubic packing. The structure is formed by connected structural units (Fig. 1.24). Panel *a* in Fig. 1.24 corresponds to a case where metallic ions are found in octahedral sites, whereas panel *b* shows metallic ions occupying the tetrahedral sites. In addition, there are some vacant positions.

Typical representatives of the spinel structure are MgAl_2O_4 as well as magnetite (Fe_3O_4), hercynite (FeAl_2O_4), and chromite (FeCrO_4). It is possible to assume that they are all folded by one two-valence cation, two trivalent cations, and by four oxygen anions. The correspondent unit cell contains 32 oxygen ions, which is 8 times more than that specified in the formula. This oxygen skeleton is completed by 8 of 64 possible tetrahedral and 16 of 32 possible positions of octahedral cations. In some spinels, the tetrahedral positions are occupied by two-valence ions.

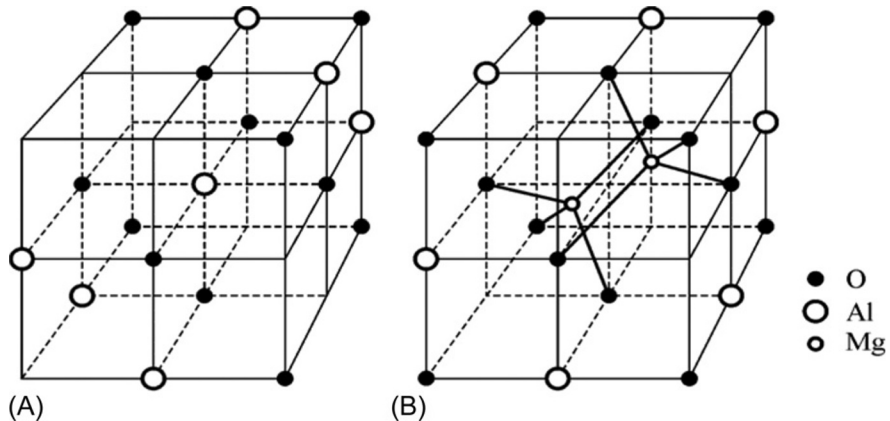


FIG. 1.24

Spinel structure: (A) octahedral coordination in Al, (B) tetrahedral coordination in Mg.

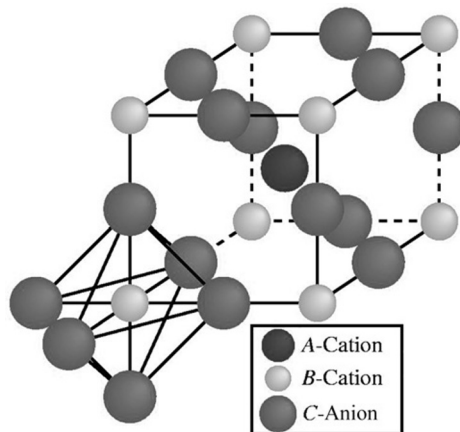


FIG. 1.25

Perovskite structure.

The structure of *perovskite* (CaTiO_3) is peculiar to orthorhombic systems, wherein Ca^{2+} ions (with $\text{CN}=12$) are found inside cuboctahedral cavities, created by octahedrons that are connected to each other by the vertices, whereas Ti^{4+} ions are found inside the octahedron. Many compounds with octahedral location of cations form a perovskite structure.

Typical representatives of such compounds are ferroelectrics of BaTiO_3 type, potassium iodate (KIO_3), and potassium-nickel fluoride (KNiF_3).

The structure of perovskite is shown in Fig. 1.25. At first glance, it is in some respects similar to the structure of cesium chloride, but with one important distinction: each oxygen atom in the group $(\text{TiO}_6)^{8-}$ is located between two titanium atoms; therefore the linkage Ti-O-Ti is linear. Thus it is possible to assume that the structure of perovskite is created from groups of TiO_6 , bound in the vertices, and Ca^{2+} is located above the center of each face of the octahedral group.

A general charge compensation must occur in the structures within inner surroundings. Therefore a structure in which charge compensation could take place only at large distances cannot be steady. In other words, large structural units should not have a residual charge, whereas paired electrons of covalent bonds must not move on long distances.

The structures of molecular crystals. When the particles creating a crystal are whole molecules, they are associated in the crystal by intermolecular forces. Because these forces are many times weaker than forces that bind particles in the ionic, atomic, or metal crystals, molecular crystals have a low hardness, low melting point, and significant volatility.

Inert gas crystals have the simplest cubic structure. Although their lattice is formed by *atoms* of inert gas, the nature of bonds relates to molecular structure, as the valence force plays no role in the formation of these crystals. Due to the

spherical shape and spherical symmetry of the interacting atoms, the crystals of inert gases energetically form the most advantageous structure: a face-centered cubic lattice that is characterized by very dense packing of atoms.

Substances made of diatomic molecules usually form crystals of a more complex structure. Especially complicated structures are formed in materials containing polyatomic molecules. Only the most symmetric and relatively simple molecules such as CH_4 , CBr_4 , and so on crystallize in a cubic system. The tetrahedral angle between them equals $109^\circ 28'$, which corresponds to the lowest energy of electron repulsion.

The most common molecular crystal is *ice* (H_2O). Its crystalline structure resembles the structure of a diamond because each molecule is surrounded by the *four* closest molecules, which are located at the same distance and are placed at the vertices of a regular tetrahedron whose angles are $109^\circ 28'$ (Fig. 1.26). Due to a small CN, the structure of ice looks like netting that shows low density.

At present, there are 14 known crystalline modifications of ice, but the most common is the *hexagonal structure*. In ice, the hydrogen bonds that form between water molecules play an important role. Each oxygen atom is surrounded by four other oxygen atoms, linked via the hydrogen atom. Two of four hydrogen atoms are located closer to the given oxygen atom, and they create a water molecule. Two others are attached with this molecule by hydrogen bonds, and they are a part of other water molecules. The distance between two nuclei of oxygen atoms of neighboring molecules is 2.76 \AA .

Such an arrangement is very far from the dense stacking of molecules: in such cases (when packing balls of radius 1.38 \AA), the molar volume of ice would be approximately two times smaller, because when molecules order in the more dense structure, their mutual orientation cannot be stored, but it is necessary for the emergence of hydrogen bonds. The distortion of a bond's angles requires considerable

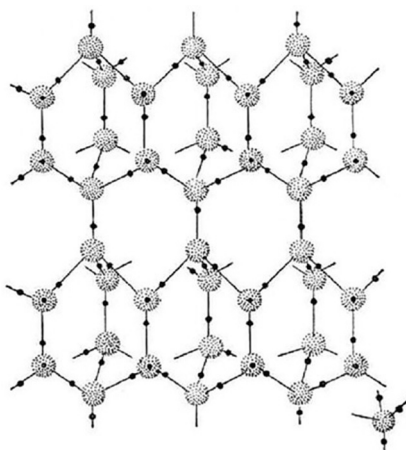


FIG. 1.26

Model of ice.

energy expenditure. This explains the friable structure of ice and why the density of ice is less than that of water.

With modifications too, ice is as crystalline and as amorphous a structure, differing by the relative positions of water molecules and different properties. After ice melts, the ice water partially preserves the structure of ice.

Many types of molecular crystals including, particularly, *macromolecular compounds (polymers)* are complex substances with molecules of high molecular weight. They are constructed from a large number of elementary units that are repeated. Polymers are formed by the interaction of identical simple molecules—*monomers*. These compounds are rubber, artificial fiber, plastics, cellulose, protein, etc. In their properties, the macromolecular compounds resemble colloids because the dimensions of macromolecules are close to the size of colloidal particles. Most polymers have no crystalline structure (polystyrene, polyvinylacetate, rubber, etc.). However, there are some polymers with pronounced crystalline structure, such as polydiacetylene.

Organic substances mainly consist of molecules that have stable structure, such that they can form crystals. Moreover, the concept of intermolecular radii and compact packaging can be introduced for molecular crystals, taking into account the characteristics and geometrical structure of molecules. Molecular crystalline structures tend to have lower symmetry than inorganic structures.

Some macromolecular compounds have such structures wherein crystalline regions alternate with amorphous ones. For example, in natural cellulose, 70% of molecules are well ordered, while in 30% are disordered. The properties of polymers can change quite widely, depending on their molecular and supramolecular (crystalline or amorphous) structure, and they thus find different applications in practice.

The solid solutions. A crystal or polycrystal can consist of several components (e.g., two components in metals that are alloys). Usually, these components cannot chemically interact with each other (forming compound) but have the ability to be *mutually dissolved* (as liquid, but in the crystal state), forming so-called solid solutions (or mixed crystals). Here, atoms of one element are introduced into another lattice, creating solid solutions of intrusion or solid solutions of substitution.

The solid solutions of *intrusion* arise when atoms of an element that dissolves are placed in an empty space in the lattice of solvent. Obviously, the size of atoms of the element that dissolves must be smaller. Usually, it should be less than 0.63 of solvent atom size because, if it is larger, there might be a *distortion* of the lattice.

The solid solutions of the *substitution* are formed by partial substitution of atoms of the solvent by atoms that dissolve. This process can occur without incurring significant stresses in a lattice only in such cases where the size of atoms does not differ greatly. Both types of atoms must be sufficiently close in their chemical properties, and it would be the best if they belong to similar subgroups of the periodic table.

Polytypicism is a property of such structures that are built of identical structural elements but have a different sequence of their location. In the plane layers, the structure of the polytypic lattice usually remains unchanged; however, in directions perpendicular to the layers, the lattice parameters are different, even though they multiple distance between adjacent layers. The phenomenon of polytypicism is usually seen

with densely packed but layered structures. It is associated with a difference in the relative orientation of atoms and results in a change in their identity period.

An example of a compound in which a large number of polytypic structures is found is the semiconductor silicon carbide (SiC). This crystal exists as in the sphalerite cubic modification—in hexagonal modification. The simplest structure for silicon carbide is the six-layer packaging ($n=6$). However, there are other polytypic structures of SiC wherein $n=4$; 15; 21; 33; 65; 192; 270; 400; 594; or 1200. Another semiconductor crystal—ZnS—has approximately 10 modifications. Moreover, polytypic structures are observed in graphite, molybdenite, and other crystals. Polytypicism significantly affects the physical properties of crystals, especially their optical properties.

Isomorphism is a property of chemically closed atoms, ions, or other structural elements to replace each other in a crystal lattice and form a continuously variable composition. Here, atoms with the same valences, bonding type, and polarization that are similar in size (with deviation of not more than 5%–7%) are chemically closed. Isomorphic substances with close but not identical composition crystallize in a similar form.

Both germanium and silicon crystals are examples of isomorphism. The density, lattice parameters, and hardness in an isomorphic row of mixed crystals Ge-Si *vary linearly*. However, as the energy spectra of germanium and silicon are different, the electron's energy bandgap, specific conductivity, and thermoelectromotive power in this series of Ge-Si semiconductors *vary nonlinearly*. By selecting different isomorphic compositions, it is possible to vary the range of operating temperatures and electrical characteristics of semiconductor compounds. Crystalline touchstring can cause crystallization and ordering of another isomorphic substance from a supersaturated solution or melt. The ability of isomorphic substances for mutual growth is used in crystal growth technology.

Polymorphism is the property of certain substances to exist in multiple crystalline phases, differing in symmetry of structure and in physical properties. Polymorphic modifications are called *allotropic elements*. At conformable physicochemical conditions, polymorphic modifications can form stable phases. Each of these phases is stable at a fixed range of temperatures and pressures, and is called the *polymorphic modification*. The relative stability of different phases is determined by the value of free energy and external conditions. Basically, polymorphic modifications differ in their structure, sometimes by the type of chemical bonds.

The change in environmental conditions might influence polymorphous transformation. During these transformations (usually, *phase transitions of type I*), the release or absorption of heat is seen, as well as the jumps of internal energy and entropy. Thus an abrupt change of many physical properties is observed that depends on the arrangement of atoms in a structure: density, specific heat, thermal conductivity, electrical conductivity, etc.

In addition, there are other types of polymorphic modifications that differ very little in their physical properties. The polymorphic transitions between such phases characterize the *phase transitions of type II*, and generally are described as the “order-disorder” phase transitions.

1.4 LATTICE DEFECTS IN CRYSTALS

In solid-state theory, in its first approximation, it is always assumed that the structure of crystals is ideal, that is, the location of atoms in unit cells as well as around all crystals is strictly periodic. However, in practice in real crystals, these ideally perfect structures are impossible.

Defect formation. Defects in the crystals are formed, for instance, during their growth under the influence of thermal, mechanical, and electrical fields (technological defects), as well as under crystal irradiation by neutrons, electrons, X-rays, and ultraviolet radiation (radiation defects). There are *point* defects (zero-dimensional), *linear* defects (one-dimensional; 1D), *planar* defects (2D), and *volume* defects (3D). In case of a 1D defect, its size in one direction is much larger than the distance between neighboring atoms (lattice parameter) whereas in the other two directions, the size of the defect has the order of a lattice parameter. In a 2D defect, its size in two directions is much larger than the distance between the nearest atoms, and so on.

Mechanisms of defect appearance may be quite various. For example, Fig. 1.27 demonstrates a possible mechanism of crystal growth [9]. Atoms are relatively weakly linked to a flat surface of ideal crystal (Fig. 1.27A) but would have better connected near a step formed by two planes (Fig. 1.27B). It is obvious that the atom will be strongly linked in the corner formed by two steps (Fig. 1.27C): this mechanism of crystal growth seems more likely. However, crystal growth becomes even

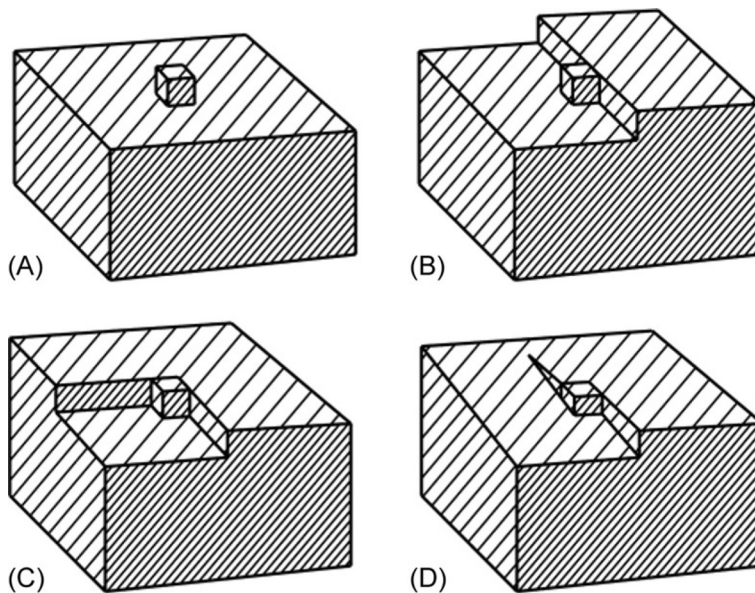


FIG. 1.27

Possible mechanisms of crystal growth.

easier if it happens through *screw dislocation* (Fig. 1.27D). In case of such a structure, the new atoms easily add and form an endless spiral around the disturbance. Crystal growth in this way is much faster, because it does not require the formation of a new embryo, as in the cases shown in Fig. 1.27A and B.

Defect formation in a crystal occurs because of different reasons.

First, atoms sometimes leave their ideal position in the crystal lattice, and therefore create defects of structure through *thermal fluctuations*. With this mechanism of defect formation, a small part of the crystal's own atoms lose their regular places in a lattice (they become the *vacancies*), or squeeze among other regular atoms, creating *interstitial* atoms; in both cases, the ideal structure of a crystal becomes locally broken.

Second, defects can be formed due to other reasons. For example, some atoms come out of the crystal surface; this is the simplest mechanism of point-defect (vacancies) formation near the surface of a crystal. Thus these vacancies—unoccupied sites in crystal lattice—are not accompanied by interstitial atoms. Vacancies of this type are called *Schottky defects*. In most crystals, the energy of vacancy formation is close to 1 eV.

In case of defect formation by the *Frenkel mechanism*, interstitial atoms or ions arise inside a lattice. Due to thermal fluctuations or power external influence (e.g., bombardment of crystal by ions), the foreign atom (or ion) can take root in the regular crystal structure from outside and create “extra” interstitial atoms. It is this very method of introducing foreign atoms into a crystal lattice that is used in modern electronics technology: the point is that a semiconductor should be obviously doped with impurities, whose atoms not only have different sizes but also can have *different valences*. After annealing, these foreign atoms replace the atoms in the crystal lattice, forming a *solid solution*. Thus the nonideal defect structure can be planned specially by technological means.

The need for defect management in structures (quite necessary for semiconductors) is due to fact that defects substantially affect such parameters of crystals as conductivity, dielectric and magnetic energy losses, electrical strength, and other properties of semiconductors, magnetics, and dielectrics, as well as strongly affect the mechanical parameters (strength) of metals.

Therefore many properties of solids are *structurally sensitive*. However, some other properties (e.g., density, specific heat, elastic characteristics) are only slightly dependent on the presence of defects. These properties are *structure-insensitive*, being determined, first of all, by the nature of fundamental atomic bonding as well as by crystal chemical composition.

Defects are very diverse. Sometimes, they are associated with one another, and it is difficult even to assign them to a definite class. However, it is possible to divide the main types of structural defects according to their dimension.

The **zero-dimensional (point) defects** are characterized by structure violation in the nodes or interstitials of the crystal lattice. These defects are caused, primarily, by the disordered location of main atoms in the crystal. Point (zero-dimensional) defects include all defects that are due to the displacement or replacement of individual atom

(or small group of atoms). They arise in the process of crystal growth, but might also be a result of radiation exposure. Moreover, point defects may be made by implantation; these types of defects are most studied, including their motion, interaction, annihilation, or evaporation.

These defects include:

- vacancies—free, unoccupied by atom lattice point;
- impurity atoms—replacing one type of atom by another type of atom by substitution in a lattice;
- intrusion of impurity atom into the interstitial space of the crystal lattice;
- Frenkel pair—vacancy together with interstitial atoms; and
- Schottky defect—vacancy arising due to the release of an atom on the surface.

Usually, Schottky defects are seen in *ionic crystals* as a pair of cationic and anionic vacancies. This defect is often found in the alkali halide crystals. The presence of Schottky defects decreases crystal density, as atoms that create vacancies diffuse to a surface (Fig. 1.28).

The defects generated by Frenkel's mechanism are usually vacancies and interstitial atoms. These defects are typical, for example, for ionic crystals of *silver halides* where *superionic conductivity* exists. Vacancy and interstitial atoms can move within a crystal lattice by the influence of thermal movement. Furthermore, Frenkel's defects are easily formed in structures of diamond-type crystals (silicon and germanium). These defects do not change crystal density.

In general, crystal can have both Frenkel's and Schottky's defects; thus those that dominate that formation require less energy. In ionic crystals formed by two kinds of particles (positive and negative), point defects occur in pairs. Two vacancies of the opposite sign usually form Schottky's defect. The pair consisting of interstitial ions and the vacancies left by them is usually the Frenkel's defect. As already noted, the simplest zero-dimensional defects in crystals are the vacancies and interstitial atoms

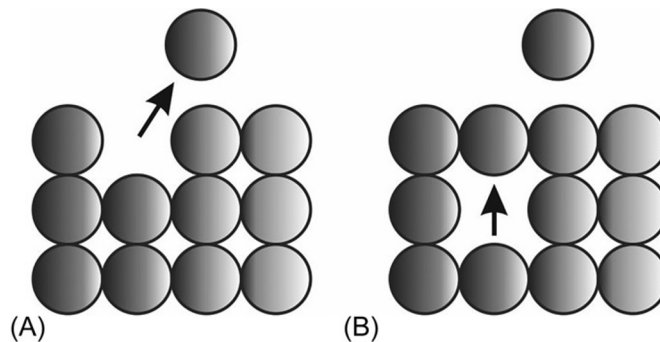


FIG. 1.28

Scheme of Schottky defect formation: (A) atom going out from crystal surface; and (B) shifting another atom onto empty position of first atom [9].

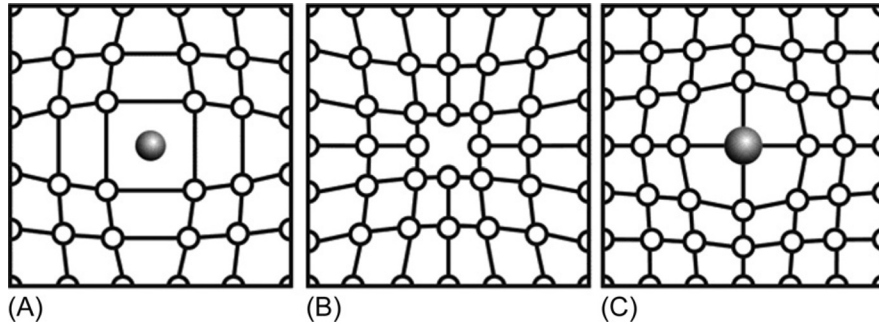


FIG. 1.29

Schematic representation of interstitial atom (A), vacancy (B), and impurity atom (C).

(Fig. 1.29). The displacement of atoms or ions (point defects) causes deformation and elastic fields around defects [9].

According to classification, except for proper point defects, other types of defects are possible, namely, the impurity defects (Fig. 1.29C): if the size of a defect atom or ion is different from the main atoms of crystal. Such defects, for example, are donor or acceptor impurities in semiconductors; similar are the impurities introduced in semiconductors to form the centers of recombination, the charge-carrier-scattering centers, etc. Around such defects, there arise local tension and distortion of the crystal lattice (Fig. 1.29A and B).

If the crystal is *ionic*, the vacancies in it lead not only to lattice distortion, but also to the appearance of effective charges with a sign, opposite to the sign of the charged ion that is missed. However, during defect formation in crystals, the *principle of electroneutrality* is efficient. The electrical interaction is very large and therefore the sum of all charges of defects generated in a crystal should be equal to zero:

$$\sum n_i q_i = 0,$$

where n_i is the concentration and q_i is the charge of the originated defects. Thus, for example, the displacement of an ion from the lattice site into the interstitial position is accompanied not only by charged ion appearance in the interstitial space, but also by charged vacancy in the crystal lattice. As with Schottky's defects, Frenkel's defects in ionic crystals provide local electroneutrality.

In the *atomic crystals* (doped semiconductors), the compensating charges appear due to electrons. The introduction of impurity atoms in semiconductors results in the appearance of donor and acceptor centers. The *donor center* is caused by such an impurity that the valence is higher as compared to the basic atom of the crystal. These centers provide *additional electrons* in the conduction band of a crystal. Donor centers in silicon or germanium crystals very often are formed by phosphorus atoms. The host atom in a lattice has four valence electrons (Si^{+4}). The replacement of some silicon atoms in crystal by phosphorus atoms (P^{+5} , donor impurity) results in “*extra*”

electron ($-e$) as the *compensating charge* that provides common electroneutrality in a crystal. Thus the positively charged ($+e$) ion of a donor generates an electron with a negative charge ($-e$).

The acceptor centers are created in silicon (or germanium) by impurities where the valence is one less than the valence of basic crystal atoms. For example, such an impurity is boron (B^{3+}). Thus when fixed in a lattice, the (*immovable*) impurity ion has a negative charge ($-e$). The lack of one electron is seen as a *hole* ($+e$), that is, the *mobile positive charge*. Therefore the acceptor type of impurity is a lower-valence atom than the own atoms of the crystal, and it gives rise to holes in the valence band.

The *polaron* is a charge carrier, partially bound in a crystal lattice (most often, this is a bound electron or hole). The polaron is not a “static” defect in the crystal, being much more mobile than the vacancies or interstitial ions. However, a polaron is much less mobile than an electron or a hole. As a rule, polarons are peculiar to *ionic crystals* wherein, under the influence of thermal motion or irradiation, some electrons (and holes) appear. In ionic crystals, the appearance of *local deformation* of ionic lattice (i.e., local polarization of lattice) is energetically favorable for electron. Thus the electrical field of the electron (or hole) is partially screened by the polarization that reduces the electrostatic energy of an electron (hole). Being a mobile charged formation, a polaron cannot be fully considered a “point defect,” but as a special state of conductive electron in the ionic crystal.

The *excitons* can also be interpreted as *mobile point defects* in crystals. The presence of excitons is, as a rule, a characteristic feature of semiconductors and dielectrics. In case of exciton appearance, ions (or atoms) in a crystal do not change their location, but become significantly different from their neighbors by infringement of its electronic state. Such a “defect” is *Frenkel’s exciton*. Because the excited state can be found in any ion, and there is strong interaction between the outer electronic shells of ions, the energy infringement can be transmitted from one ion to another. Therefore moving Frenkel’s exciton in a crystal is not related to the change of ion positions; thus it has (as polaron) a much higher mobility than vacancies, interstitial atoms, and impurities of replacement. In general, the exciton cannot be fully considered a localized defect.

The diffusion. In processes of semiconductor device technology, a *heterogeneous* distribution of donors or acceptors is usually necessary to create the *p-n* junctions for diodes or transistors. In addition, during semiconductor device operation, the heterogeneous (in space and in time) distribution of charge carriers often arises. Whenever there is nonuniform concentration, the phenomenon of *diffusion* takes place, and it often plays an important role in the given situation. Therefore diffusion has received a great deal of attention in semiconductor research [9].

Diffusion is the directional movement of molecules, atoms, or charge carries from a region of high concentration to a region of low concentration. Diffusion is caused by the aspiration of any system to reach their equilibrium state, that is, in this case, a *leveling of concentration*. In the first case (technological), it is the smoothing of admixture additive concentration; in the second case (electronic device operation), it is the decrease of excess concentration of charge carriers.

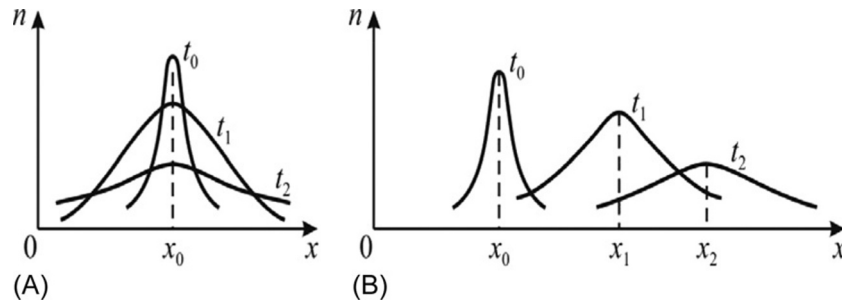


FIG. 1.30

Time dependence of concentration plot in a one-dimensional model of diffusion: (A) degradation with time without external influence; and (B) charge carrier diffusion and simultaneous drift with time in the electrical field ($t_0 < t_1 < t_2$).

In Fig. 1.30A, in a local area with a coordinate x_0 at time t_0 , an excessive concentration of particles n_{\max} is created. This is a nonequilibrium state; therefore, over time (t_1, t_2, \dots) under the influence of thermal chaotic motion, the maximum concentration decreases and the area of increased concentration becomes blurred, aiming for full alignment.

In cases when *donors or acceptors* are locally imbedded in a semiconductor, rather high temperature is needed to smooth down their concentration, and this is an important stage in semiconductor device technology. (At room temperature, a great amount of time is necessary to change the position of admixture additives in the crystal lattice.)

In many semiconductor devices, the locally increased concentration of *charge carriers* is created (usually by injecting of carriers into a specimen from an external circuit). It is obvious that such a concentration of charge carriers is the nonequilibrium state. Therefore electrical current flows from both edges of concentration peak, as shown in a 1D model (Fig. 1.30A). This flow, in which the concentration $n(t, x)$ changes rapidly, is called the *charge carrier diffusion*. The effect of diffusion is eventually to bring the concentration of charge carriers toward their equilibrium situation wherein their concentration is uniform throughout. As time changes, the peak spreads out in both directions and decreases, although the center of the peak remains at the same place x_0 .

A quantitative consideration of these processes reduces to the basic law of diffusion: *Fick's law*, which specifies that, in nonuniform concentration, the *density of particle flow* j' (i.e., the number of particles, crossing unit area per unit time) is given by

$$j' = -D \cdot \partial n / \partial x,$$

where D is a constant called the *diffusion coefficient*. This law states that the flow of diffusion is proportional to the gradient of concentration. Thus the more rapidly n varies, the larger is the flow. The negative sign in a given formula is introduced

for convenience in order to make parameter D a positive quantity. As seen from this equation, j' is opposite to dn/dx .

Fick's law is valid whether particles are neutral or charged. In semiconductors, where moving particles are charged carriers, the flow j' is proportional to the electrical current: $j = e \cdot j'$, because, to obtain the value of the electrical current, one needs to multiply j' by the charge of the carrier.

After the creation of a clot with increased concentration of charge carriers at time t_0 , if the *gradient force field* is switched in (usually, it is an electrical field), the clot of particles will diffuse as before, and the center of the clot will also *drift* because the applied electrical force influences these particles (Fig. 1.35B; this case can be applied also to the movement of defects under a *thermal gradient* influence).

The 1D defects—dislocations—are crystallographic defects or any irregularity within a crystal structure. The presence of dislocations strongly influences many properties of solids.

Dislocation may, furthermore, be interpreted as the linear boundary of a structural violation in a crystal. Mathematically, dislocations can be defined as a type of topological defect, sometimes called the *soliton*. In other words, dislocations are such violations of crystal structure that have greater length (up to macroscopic size), but their lateral dimensions do not exceed several interatomic distances. Therefore, 1D (linear) defects are defects that, in one direction, are much larger than the crystal lattice parameter, whereas, in two other directions, they can be compared to it. There are two primary types of dislocations: edge and screw. Mixed dislocations are the intermediate cases between these two.

The *edge dislocation* is the border of “excess” atomic plane that splits the crystal. It corresponds to a row of convergent atoms along the end of an additional plane of atoms within the crystal. Fig. 1.31A shows the atom arrangement around edge dislocation whereas the right panels of the figure (*b*, *c*, *d*) demonstrate the possible movement of such dislocations in a crystal.

In other words, the edge dislocation is a defect when an extra half-plane of atoms is introduced, distorting the nearby planes of atoms [9]. If an external force is applied

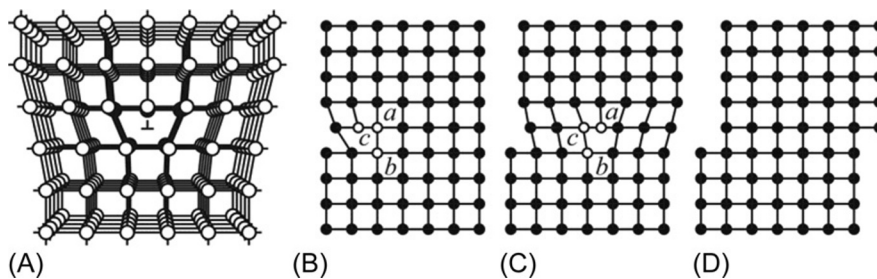


FIG. 1.31

Edge (linear) dislocation (A) and its displacement to left side from one part of the crystal in relation to another (B, C, and D).

to one side of the crystal, this extra plane will *pass through the planes* of atoms, breaking and joining bonds with them, until it reaches the crystal (or grain) boundary. Dislocation is described by two parameters: the *line direction* (i.e., the direction of running along the bottom of the extra half plane) and the *Burgers vector* that describes the magnitude of distortion. In case of edge dislocation, the Burgers vector is *perpendicular* to the line direction.

The formation of dislocations in a semiconductor crystal occurs in a process of crystal growth because crystal cooling is not uniform (the surface cools faster than the volume). As a result of uneven thermal expansion, tensions appear in the crystal lattice. When temperatures are higher than the temperature of ductility, the stressed state of the lattice can be, to some extent, “removed” due to the formation of linear dislocations. Below this temperature, dislocations in crystal become “frozen.”

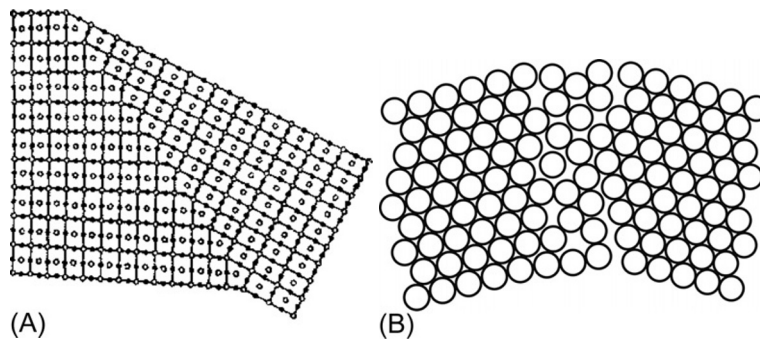
The screw dislocation is a result of change in atom location in one part of a crystal in comparison with another. It corresponds to the axis of the *spiral structure of distortion* that is associated with normally parallel planes (Fig. 1.27D). As shown, screw dislocations are formed during crystal growth and then they remain in the structure. It is possible to say that the problem of crystal growth can be solved by the possibility of screw dislocations rising. If the surface plane crosses only part of the way through the crystal and then stops, the boundary of this cut becomes the screw dislocation. It comprises a structure wherein a helical path is traced around a linear defect (dislocation line) in the crystal lattice. In purely screw dislocations, the Burgers vector is *parallel* to the line direction.

As point defects, dislocations can move through the crystal lattice. However, the movement of dislocations is associated with many limitations, because a 1D dislocation should always be a continuous line. There are two main types of dislocation movements: displacement and sliding.

The dislocation *displacement* is due to the addition (or removal) of atoms from the superfluous half-plane that may occur as a result of thermal diffusion. During this sliding, the extra half-plane of dislocation that takes a definite position in the crystal lattice combines with the atomic plane that is located under the plane of sliding, whereas the neighboring atomic plane becomes the extraneous half-plane. This smooth sliding of the dislocation line can be caused by the *shear stress* applied to the crystal surface.

It is well known that a rod of soft metal, after a series of bends and straightening, stops its bending and eventually breaks. This is the example of *strain hardening*. At each bending, many new dislocations occur in a metal; when their number becomes so large that they cannot move, the crystal loses its ability for plastic deformation and it breaks on any further impact on a crystal.

The 2D and 3D defects. Two-dimensional (planar) defects include intergranular or intercrystallite borders, as well as crystal surfaces and modulated structures inside a crystal. Therefore most common 2D defects are the *boundaries between grains* that are peculiar for polycrystalline materials. They consist of a large number of single-crystal grains that are randomly oriented and tightly interconnected. These structures usually are polycrystalline metals as well as dielectric or magnetic ceramics.

**FIG. 1.32**

Schematic (A) and atomic (B) representation of grain boundaries.

Intergain borders—interfaces of crystallites—are not necessarily flat surfaces. Boundaries between grains (crystallites) in polycrystalline materials might have significant curvature. Layers of atoms near these boundaries are the areas of the disturbed crystal lattice; thus the thickness of defect layers usually equals several atomic layers, providing smooth transition between disordered regions (Fig. 1.32B).

The polycrystalline (block) structure of ceramic materials and metals can significantly affect their electrical, magnetic, and mechanical properties.

The surface of a crystal, by its essence, is also a 2D structural defect. Therefore, each real crystal differs from the ideal crystal due to variations in the structure and properties of a surface. The surface is a special state of crystal, with different sized elementary cells that have other symmetry and other energy. Atoms (ions) located on the surface layer are joined by broken chemical bonds which are not saturated. Located on the surface are unpaired electrons of atoms (ions) that tend to form new connections.

Most often, the state of the surface is characterized by the bonding of neighboring atoms—either in pairs, or in more complex associations. Surface atoms are combined into *larger unit cells*, as compared with the volume of a crystal. For example, on the surface, a silicon unit cell has 7×7 atoms (in germanium 2×8 atoms) whereas their fundamental unit cell contains only two atoms. Thus the elasticity of atomic bonds on the surface is changed and, as a result, the characteristic “melting point” is reduced by 10%–30%. Note that crystal growth occurs just from the surface as well as the melting of the crystal, its evaporation and condensation, and the diffusion of atoms deeper into the crystal.

The electronic energy spectra of the crystal surface significantly differ from the electronic spectra of the crystal volume. As discussed later, the unique properties of *nanocrystalline materials* are due exactly to the fact that in nanoparticles (which have a small number of atoms: 10–1000), the ratio of surface-located atoms to volume-located atoms is 90%–20%.

The physics of modulated structures can be regarded as a boundary area between the physics of nanomaterials and physics of structural defects. In case of nanophysics, the planar (by-layer) modulation of semiconductor structures may result in a specific electronic spectrum—so-called quantum wells. In some types of magnetics, the arrangement of electronic spins in crystals also exhibits periodic complexity in their magnetic ordering, as compared to the usual crystal structure. Modulated structures can be explained as the coexistence of different periodicity in a crystal.

The 3D (volumetric) defects. They are, first of all, the *clusters of vacancies* that form pores and channels; various *imbedded defects*, such as gas bubbles; accumulation of *impurities* in a form of sectors and areas of growth. Three-dimensional defects reduce crystal flexibility, affect its elasticity, and strength as well as change the electrical, optical, and magnetic properties of a crystal. Fig. 1.2 in the Introduction schematically shows 3D structural defects in polycrystalline materials. Inside of each crystallite, many interplanar structural defects can be observed.

Thus 3D defects are solid interstices, liquid or gaseous phases in a crystal, clusters, and other complications with a macroscopic structure. In materials used in electronic technology, 3D defects might also have a fundamental nature, but this case is not considered here.

As a result, some conclusions follow:

- Part of atoms (or ions) of a crystal may be absent in their positions that correspond to the ideal crystal lattice scheme. These defects are vacancies. Foreign (impurity) atoms or ions, replacing basic particles that form a crystal, or inserted between them, can also be seen in crystals. Point defects in a crystal can also be its own atoms or ions that are shifted from their normal positions (internode atoms and ions),
- In the process of crystal growth, as well as during its plastic deformation and in many other cases, dislocations arise. Dislocations are places with ordered accumulation of impurities. The distribution and behavior of dislocations under external influences determine many important mechanical properties of a crystal, including strength, ductility, and other aspects. In particular, the mobility of dislocation determines the plasticity of crystals; dislocations also cause the appearance of internal stress and fracture of crystals. The problem of plastic (i.e., irreversible) flow of metals can be solved by the prevention of dislocation movements. Dislocations impede the process of magnetization and electrical polarization because of their interaction with domain boundaries movement.
- Defects in crystals cause elastic deformation of the structure that leads, in turn, to the appearance of internal mechanical stresses. For example, point defects, interacting with dislocations, can increase or decrease the strength of crystals. Defects in crystals affect absorption spectra and luminescence, light scattering in crystal, etc.; such defects change electrical conductivity, thermal conductivity, ferroelectric properties, magnetic properties, etc.

1.5 STRUCTURE AND SYMMETRY OF QUASICRYSTALS AND NANOMATERIALS

As described earlier, crystal structure is defined as a system with long-range ordering of particles. If the structure of the crystal unit cell is known, 3D periodicity makes it possible to predict the location of atoms of any other cell and the relative positions of atoms of the entire structure as a whole. This means that crystal has *translational symmetry*. The structure of the crystal can be described by the displacement of a single unit cell on three basic vectors of translations.

Translational symmetry results in regular crystallographic planes in crystal, thus making clearly identified narrow peaks of X-ray scattering. This feature of the X-ray diffraction pattern is the distinguishing characteristic of crystals. Polycrystalline bodies, in their structure, are similar to single crystals, because they are composed of small randomly oriented crystals. During X-ray beam scattering in polycrystals, a conical symmetry is formed that also gives distinct diffraction maxima, which can be used to obtain lattice parameters as in a single crystal.

Significant difference occurs in the X-ray spectra of *amorphous* solids that are characterized by the *blurred picture* of diffuse X-ray scattering without clearly identified narrow rings. Such solids in their amorphous state do not show strict 3D periodicity. Thus while defining an amorphous structure, the terms “disordered,” “noncrystalline,” “amorphous,” and “glassy” are synonymous. The arrangement of atoms in amorphous solids, however, is not completely random (as it is in gases). The interactions between atoms in an amorphous body are similar to the forces in crystals and, although there is no long-range ordering, the *short-range ordering*, generally speaking, is preserved [8].

Short-range ordering in the arrangement of atoms is characterized by such parameters as the *length* and the *angles* of bonds as well as by the *number* of their nearest neighbors. It should be noted that in the amorphous state, because of violations of their structure, these options have some statistical dispersion, and their average values may differ slightly from those values in a perfect crystal.

The **quasicrystals** show a new type of symmetry, different from all aforementioned cases. They demonstrate such elements of symmetry that previously were considered as impossible. The translation symmetry of a perfect crystal obeys rigid restrictions as to the *order of rotary symmetry axes*, which describe the symmetry of a crystal.

As shown earlier, the ideal crystal, except with a trivial axis of the first order, can have symmetry axes only of second, third, fourth, and sixth orders. Solely, these axes can provide the parallel transfer of unit cell when it is multiplied to create a crystal. The symmetry of a perfect crystal does not allow the existence of axes of symmetry of the fifth, seventh, or higher orders. Elementary cells that have such axes cannot completely fill even the plane (and, moreover, the volume).

Nevertheless, in 1984, for the first time a metallic alloy was discovered with unusual properties: with the axis of symmetry of the fifth order (Dan Shechtman,

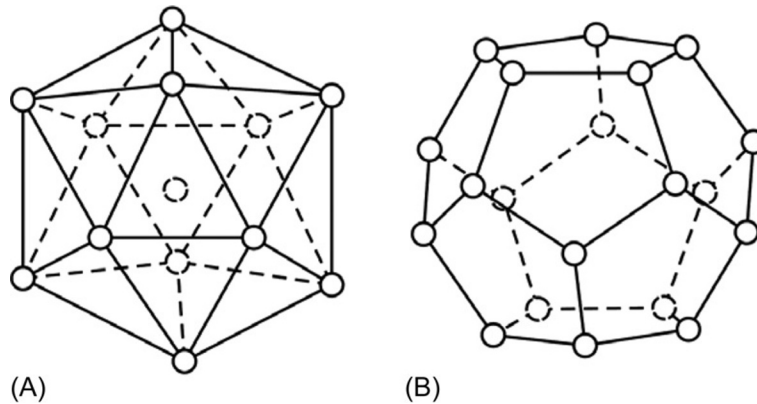


FIG. 1.33

Models of quasicrystals structures: (A) icosahedron and (B) dodecahedron.

Nobel Prize for 2011). This alloy was obtained by a rapid cooling of molten aluminum-manganese (with the speed of cooling near 10^6 degrees Kelvin per second). Grains of this alloy have the form of a dodecahedron with rotary symmetry axes of fifth order.

As known, symmetry axes of the fifth order have two types of “regular convex polyhedra” (Fig. 1.33): icosahedron and dodecahedron (the existence of these regular convex polyhedra was first noted by Euler). The icosahedron is a regular polyhedron consisting of 20 faces—equilateral triangles—and it has 12 vertices and 30 edges (Fig. 1.33A). The dodecahedron is a polyhedron consisting of 12 faces (pentagons) and it has 30 edges and 20 vertices (Fig. 1.33B). The dodecahedron and icosahedron can be inscribed into one another, similar to a cube and an octahedron. It should be noted that the icosahedron and dodecahedron can be described by identical elements of symmetry, including symmetry axes of the fifth order. As in ideal crystals, the symmetry of the axis of the fifth order is prohibited; therefore the icosahedron and dodecahedron are never used in the translational symmetry of classic crystallography.

As mentioned, the diffraction pattern of X-ray scattering for the aluminum-manganese alloy shows regular peaks, corresponding to a structure with a rotary symmetry of the fifth order. This diffraction pattern can be formed only when the atomic structure has the axes of symmetry of the fifth order. This means that icosahedral symmetry characterizes not only grains of metal, but the arrangement of atoms in unit cells as well. The presence of different reflexes in the X-ray spectrum shows the special arrangement of atoms in the structure called *shechtmanite* (a quasiperiodic crystal), whereas the presence of symmetry axes of the order 5 indicates that this material, in the usual sense, cannot be considered a crystal. Some additional research of shechtmanite by methods of electron microscopy confirmed the homogeneity of

this material and the existence of rotational symmetry of the fifth order in the small areas with sizes of a *few tens of a nanometer*.

At present, many alloys of similar structures are discovered and synthesized, and they are called the *quasicrystals*. For example, the quasicrystals can be obtained by a sudden cooling of molten aluminum, copper, and iron that, during solidification, form grains of the dodecahedron type. In most synthesized quasicrystals, using X-ray diffraction studies, the icosahedral symmetry has been identified with point group of symmetry, inherent to the rotary axis of the fifth order. In addition, other quasicrystals are synthesized with rotary axes of symmetry of 8th, 10th, and 12th orders (all these symmetry axes are prohibited in the translational symmetry of ideal crystals).

Quasicrystals usually consist of metal atoms and (sometimes) of silicon, for example, the alloys Al-Li-Cu, Al-Pd-Mn, Zn-Mg-Y, Al-Cu-Co-Si, Al-Ni-Co, and Au-Na-Si. The structure of quasicrystals is characterized by a combination of alternative local symmetry (icosahedral) that is far from ordering, providing sharp peaks in diffraction pattern, observed in experiment. Following the discovery of quasicrystals with fifth-order axis of symmetry, it seems natural to involve the model that can describe structures by regular icosahedrons and dodecahedrons.

For example, the icosahedral clusters can be used as a model, consisting of identical solid spheres that represent atoms. The tetrahedral structure can be formed with four closely linked spheres, limiting their planes passing through the centers of spheres. A compound of 20 tetrahedrons creates a small, distorted icosahedron. A similar structure can be obtained by solid sphere wrapping by 12 equidistant areas. However, between 12 peripheral areas, representing atoms, there are *gaps* that inevitably occur; each atom would be approximately 5% further apart than the distance to the central atom. Compact filling of space by such an icosahedron-type cluster should be quickly broken, that is, the icosahedral packaging cannot spread to the entire crystal [10].

Some structures, which have short-range icosahedral ordering, acquire the term the *metal glasses*. They are formed by a very rapid cooling ($\sim 10^6$ K/s) of the melt of some metals. Such structures have only *short-range ordering*, and, being amorphous, form an X-ray spectrum with *broad diffuse maxima*. In quasicrystals, however, X-ray peaks are expressed *clearly*.

To explain the spectra of quasicrystals, the presence of icosahedral *clusters* with regular distortions on borders is supposed, which could provide long-range ordering in the structure and, therefore, create X-ray diffraction patterns with narrow peaks. Therefore, to describe some complex quasicrystals, the structural units containing a few dozen atoms are proposed. However, a problem arises as to the physical nature of appearance and stability of such complex clusters. Furthermore, X-ray and neutron-diffraction methods showed that, in real structures of quasicrystals, only a small fraction of their atoms have an icosahedral environment.

Thus, for actually existing long-range ordering, all quasicrystal structures should have some “nontranslating” arrangement. In other words, filling of infinite space by atoms in these structures can be determined by such an algorithm when a long-range

order is ensured *without full translational symmetry*. The lack of translational constraints allows the structure to have a quasicrystalline axis of the fifth order. Orderly arrangement of structural units can provide a positive interference of X-ray waves scattered by atoms in some areas, and the formation of narrow and strong diffraction reflexes.

Some ideas as to quasicrystals modeling in 1D and 2D structures are discussed further. Ensuring long-range order in a 1D structure in the absence of translational symmetry is possible in various ways. For example, long-range order in atom distribution can be modeled by the *linear chain of atoms* with constant interatomic distance “ a ” that shifts with the next atom on distance $\Delta = \varepsilon \cdot a \cdot \sin(2\pi j a)$, where j is the serial number of atom while ε and σ are some numbers. If the number σ is irrational, displacements of atoms are different, even if one considers an endless chain of countable atoms. This 1D structure might have translations.

However, the coordinates of all atoms are determined by a definite law, that is, this sequence is totally ordered structure. The lack of translational symmetry in this case is not due to chaotic displacement of atoms (that is typical for amorphous structures), but by imposition of two nondisproportionate periodicity in their arrangement, whereas the ratio of their periods is an irrational number. The lack of random displacements of atoms that leads to the nontranslational arrangement makes an X-ray diffraction pattern, characterized by distinct maxima. Built in such a manner, the chain of atoms is the example of 1D-quasicrystals. This example shows the feasibility of using *irrational numbers* in constructing models of quasicrystals [10].

The “Penrose mosaic” shown in Fig. 1.34 can be used as a mathematical model of 2D quasicrystals. This structure is fundamentally different from the classic “frozen” form of perfect crystals. R. Penrose developed the algorithm on how to fill an infinite plane *with no overlaps and voids* by using figures of only two types. These figures that are needed to build the Penrose mosaic are the rhombuses with the same side. The internal angles of wide rhombus equal 72° and 108° and internal corners of “narrow” rhombus are 36° and 144° .

A mosaic made of rhombuses can fill all “endless” flat surfaces, but only at an aforementioned selection of special corners of rhombuses. Notably, that ratio of “narrow” rhombuses to “broad” rhombuses is exactly equal to the “golden section” (Golden section is a number $(\sqrt{5} - 1)/2 = 0.618\dots$ equals to the ratio of two parts of a whole (Φ and S) that is subject to the following rule: $\Phi/(\Phi + S) = S/\Phi$.)

Because the “golden section” is an irrational number, in the considered mosaic it is impossible to identify any “unit cell” containing a whole number of each type of rhombuses that could fill the plane. Therefore the Penrose mosaic is not a 2D-crystal in the traditional sense, but it is a 2D-quasicrystal.

It is important to pay attention to the following facts:

First of all, it is essential that the construction of mosaics is realized by defined algorithms, which is why this mosaic is not a random, but ordered, structure. Secondly, when calculating the scattering of X-rays for structure, formed by atoms located in vertices of the Penrose mosaic, it is found that the diffraction pattern

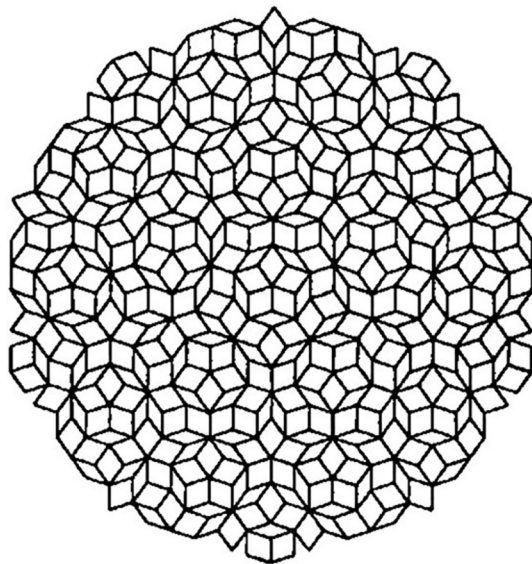


FIG. 1.34

Penrose mosaic as example of two-dimensional quasicrystalline structures.

has a rotary symmetry axis of 10th order. The Penrose filling contains 10 squares with exactly the same orientation. Thirdly, rhombuses of mosaics (with parallel sides) form five families parallel to each other's lines, intersecting at angles that are multiples of the angle 72° .

Thus, the Penrose mosaic has a long-range ordering, providing diffraction pattern of fifth-order rotational symmetry.

After the invention of "shechtmanite," a 3D generalization of the Penrose mosaic is studied that has icosahedral symmetry. Experiments show that in most real quasicrystals, their atoms have nearest neighbors lying in the vertices of a regular dodecahedron. However, the construction of the structure from hard figures with 20 vertices of the dodecahedron by real atoms may engage in no more than eight vertices. Therefore, the first coordination sphere of each atom has a strong volatility. Such structures are characterized by both short-range ordering and long-range ordering (with not usual translating), which can be built only from two types of rhombohedrons. This mosaic is not possible to obtain traditionally by the translations of one elementary cell.

It should be noted that the algorithm of 2D-rhombuses or 3D-rhombohedral in Penrose mosaics consists of several steps, and therefore has alternatives. Although real quasicrystals grow, some failures of its structure are possible because quasicrystals can be formed in regions of their violations. The presence of such "amorphous" inclusions should lead to widening of peaks in X-ray diffraction pattern, as observed in experiments. In addition, an evidence of the presence of disordered local areas is the *low conductivity* of metal alloys of synthesized quasicrystals.

In the melt metals, depending on alloy components, *microsymmetry* is created correspondingly to features of the *electronic structure* of ions that coexist in the melt. Microstructure in melts might be quite diverse; they might have axes of symmetry of the fifth order (as well as axes of symmetry of higher orders), that is, such translational symmetries are forbidden in an ideal crystal. The thermal effect on crystal has the symmetry of a sphere; therefore, it contains any element of symmetry (including axes of any order). Usually, quasicrystals are obtained by a sudden cooling of such alloys, where nontranslational symmetry axes are dominating in the microstructure.

Therefore, in case of imbalance cooling (by the “heat shock”), primarily the structures of short-range order become stabilized (e.g., with the axis of the fifth order), which is typical of the local electronic structure of a given melt and not forbidden by sphere symmetry (according to the Curie principle). The remaining symmetry elements (in considering the case of the axis of the fifth order) become “frozen” after structure heatstroke, providing a sort of “long-range ordering.” These formed clusters may have *enough inner energy* to withstand thermal movement and therefore store the elements of symmetry, unusual for perfect (translational) long-range order. In crystals that experience sharp changes of temperature, these unusual elements of symmetry are stored; therefore the properties of such alloys are not traditional.

The symmetry of nanomaterials. Nanomaterials exhibit short-range ordering of their atoms. Their relatives are, for instance, well-studied amorphous metal alloys (metal glasses). In such substances, their structure is changed quite significantly, allowing the creation, for example, of ferromagnetics with such magnetic properties that, in principle, cannot be obtained in the materials with long-range ordering of atoms.

Topological models of *amorphous materials* are well developed and are based on the random dense packing of both hard and deformable spheres: this is close to that seen in nanostructures. With regard to inorganic glasses with covalent bonds and random packing of atoms, these structures correspond to the model of a random and continuous grid of atoms. All of the said models are characterized by a set of different-size spheres, randomly packed to the largest density [8]. They differ in the rules of packaging, in the interaction potentials, in the relaxation modes, etc. In many configurations of random dense packing, the crystallographic structural elements are allocated, as well as the noncrystalline packing of clusters that can be illustrated by the Bernal polyhedra (Fig. 1.35).

As known, the CN in crystals might be 4, 6, 8, and 12. In the amorphous metallic alloys, the CN for alloys of transition metals with copper remains only *close* to CN = 12 regardless of the compound (in ideal model CN = 12). In reality, for example, in Ni-Li and Cu-Ti alloys, the average CN is 12.8. In the alloys of rare earth metals and transition metals, as usual, CN = 12; however, in the amorphous alloys, CN generally decreases. For example, in DyFe₂ alloy CN = 7.1 ± 1, while in the alloy TbFe₂, CN = 8.4 ± 1.8; thus the environment of iron atoms is approximately the same as in the crystal. Thus the short-range ordering in amorphous and in crystalline states of metallic alloys is different.

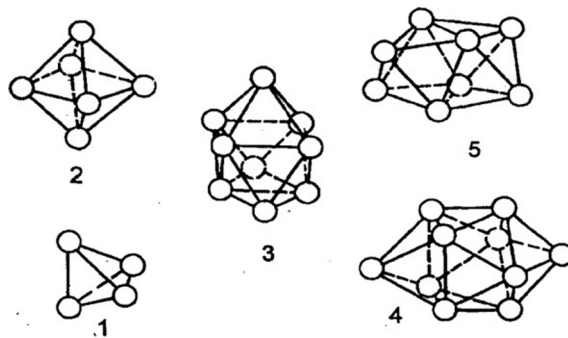


FIG. 1.35

Models of amorphous structures clusters: 1—tetrahedron; 2—octahedron; 3—trigonal prism with three semioctahedrons; 4—Archimedes' antiprism with two semioctahedrons; and 5—tetragonal dodecahedron.

Nanomaterials are small particles of matter (clusters), consisting of 10–1000 atoms, and their properties depend on the *number* of atoms in a cluster as well as on the *relative position* of atoms. The *size* of a nanocluster also has an influence on its shape and symmetry [11].

Consider, for example, the cubic symmetry crystal of magnesium oxide (MgO; Fig. 1.36). An important property of nanoparticles is seen: the difference in the outward form of the same material—*crystal*, *microcrystal*, and *nanoparticle* [12]. In this example one can see a resizing change in the shape of a body. When the size is larger than $100 \times 100 \text{ nm}^2$, long-range ordering prevails, and MgO crystal has this intrinsic to its *cubic* form. However, the microcrystal of MgO tends to have a *hexagonal* shape, whereas the MgO nanosized particle shows a nearly *dodecahedron* form.

Another important example that demonstrates how internal bonding and symmetry influence the properties of materials is of the various forms of carbon. In the periodic table of elements, carbon relates to subgroup 4; the electronic shell of carbon atom has four valence electrons with configuration s^2p^2 , allowing carbon to have valences -4 , $+2$, and $+4$.

The classification of carbon structures is shown in Fig. 1.37. The classic (3D) structures of carbon are diamond and graphite. The **diamond** is a 3D form of carbon; its

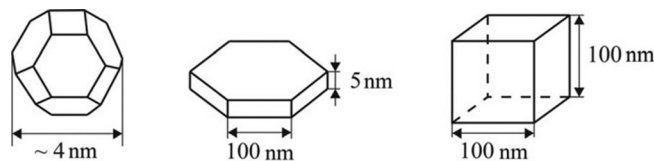


FIG. 1.36

Various forms of MgO structure: 4 nm—nanoparticle; $5 \times 100 \text{ nm}^2$ —microcrystal; $100 \times 100 \text{ nm}^2$ —usual crystal.

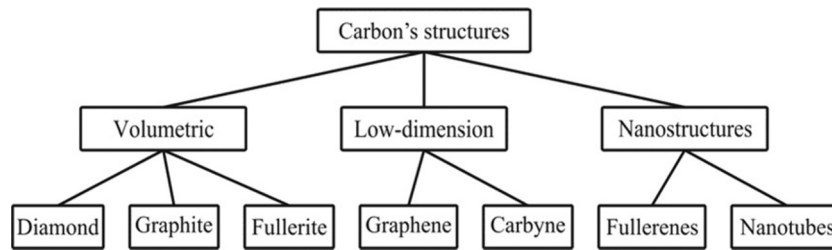


FIG. 1.37

Classification of different forms of carbon.

structure is formed from the electronic state of sp^3 -hybridization. In the diamond crystal, each carbon atom is surrounded by four others that are in the tetrahedral sites; neighboring atoms are combined together by a strong covalent bond that determines the hardness of the diamond. The distance between the atoms in the diamond is 0.154 nm.

In **graphite**, carbon atoms are connected with each other, thus forming the *hexagonal netting*, in which each atom has three neighbors. In such a quasi-2D (plane) form of carbon, its structure originates from the state of sp^2 -hybridization (Fig. 1.38B). The layers of plane nettings of graphite are accommodated one above another. In covalent chemical bond formation, three electrons from each atom take part in creating σ -bonding. The distance between atoms, arranged in hexagonal mesh nodes of graphite, is 0.142 nm—less than in the diamond. Thus neighboring atoms within each layer of graphite are linked by *stronger* covalent bonds.

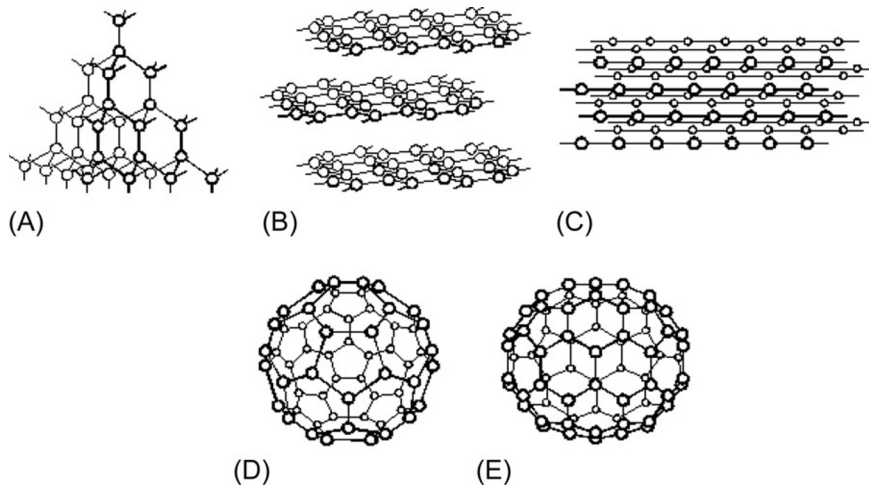


FIG. 1.38

Carbon atom location in various structures: (A) diamond, (B) graphite, (C) carbyne, (D) fullerene C_{60} , and (E) fullerene C_{70} .

However, these layers fit together by the *weak* van der Waals forces, in which four electrons are involved. The hexagonal graphite netting is located at a distance of 0.335 nm from each other, that is, the distance between atoms is more than twice that in the layers. This bonding between layers is the π -bond. A large distance between layers determines the weakness of forces that combine layers. This structure—strong segments, poorly linked—constitutes specific properties of graphite, particularly its flexibility that explains a slight sliding of layers relative to each other, as well as the low hardness of graphite and large anisotropy of its properties.

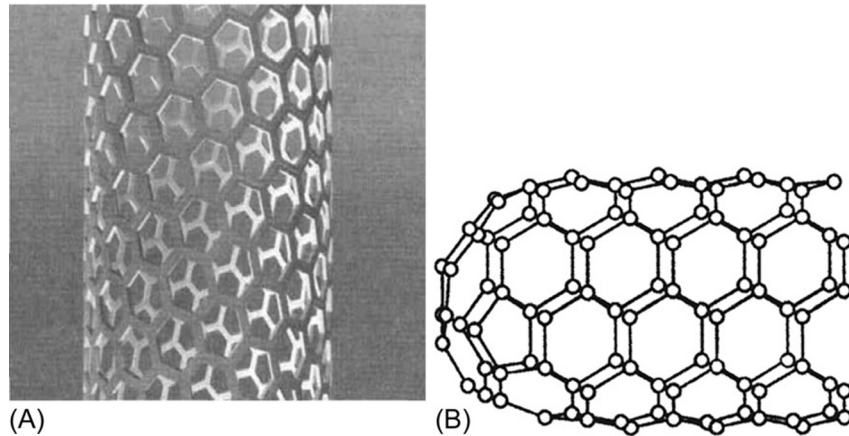
The *carbine* is a linear polymer of carbon that can be obtained in vitro in the form of long chains of carbon atoms, parallel to each other (Fig. 1.38C). The string (linear) structure of carbine is formed by the *sp*-hybridized carbon atoms. In the very long molecule of carbine, carbon atoms are strongly linked in chains by the *triple* bond, as well as by the *double* bonds between them. Carbyne can be obtained in forms of fiber, powder, and films of different structure: disordered long chains, amorphous and quasiamorphous material with microcrystalline inclusions, and bilayer-oriented chains. Crystalline-type samples of carbyne have the shape of plate-form crystals, as well as samples in the form of fiber up to 10 mm in length.

The *graphene* (Nobel Prize for the year 2010) is the plane polymer of carbon: the layer of carbon atoms with a thickness of only one atom is connected by the *sp*²-bonds in the 2D hexagonal crystal lattice. Graphene can be represented as a single plane of graphite, separated from bulk crystal (see Fig. I.3 in Introduction). Graphene is characterized by big mechanical stiffness and large thermal conductivity. The high mobility of charge carriers in the graphene at room temperature makes it a promising material for use in various electronic devices. In particular, graphene can be regarded as an important material for nanoelectronics that allows, in some cases, to replace the silicon in integrated circuits.

The *fullerenes* are molecular compounds belonging to one of the relatively new forms of carbon. They are closed polyhedra composed of carbon atoms that are located on a surface of convex polyhedron (Fig. 1.39D and E). The discovery of fullerenes was also awarded the Nobel Prize. The most stable form of fullerenes is the molecule C₆₀—a polyhedron made of hexagon and pentagon faces.

The *fullerites* are condensed systems consisting of fullerene molecules. In addition, the topical compounds are the *fullerides*—fullerite crystals doped with alkali metal atoms. Some fullerides exhibit high-temperature superconductivity, for example, in the fulleride-superconductor RbCs₂C₆₀, the critical temperature is 33 K.

The *carbon nanotubes* (Fig. 1.39) are lingering cylindrical structures with diameter from one to several tens of nanometers and lengths up to several micrometers. They consist of one or more sheets rolled into a tube hexagonal graphite planes (graphene) and usually terminate in a hemispherical head. There are both metallic and semiconducting carbon nanotubes. *Metallic* nanotubes well conduct electricity even in near-absolute zero temperatures, whereas in the *semiconductor* type of nanotubes at temperature close to absolute zero electrical conductivity is nearly zero but increases with a rise in the temperature.

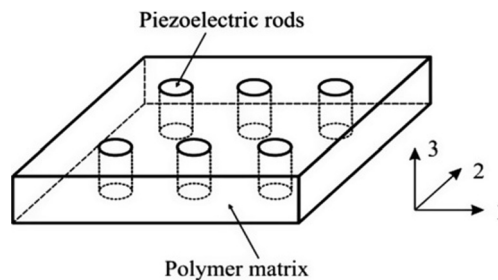
**FIG. 1.39**

Single-walled carbon nanotubes: (A) schematic representation and (B) fullerene-like closed end of the nanotube.

1.6 STRUCTURES OF COMPOSITES AND METAMATERIALS

The term “composite” is implied for multicomponent system in which several materials are combined, being different in composition or form, in order to obtain the specific property of the final material. In this case, individual components of a system retain their individuality and properties to such an extent that they exhibit interphase boundary, and operate the achieving the improved properties that are inaccessible to each component separately. Thus, the properties of composite materials are largely related to the geometric arrangement of components.

As an important example, the piezoelectric composite materials are considered. These composites are used in underwater sonars, for medical ultrasound diagnostic, in some electronic instruments, etc. In the simplest case, an ultrasound composite receiver can be constructed of piezoelectric and polymer components (Fig. 1.40).

**FIG. 1.40**

Element of piezoreceiver consisting of piezoelectric rods in polymer.

Cylindrical rods made of piezoelectric material occupy a relatively small volume of composite, yet they provide practically the same signal in the receiver as would be obtained from a solid piezoelement, because the rods are hard whereas the polymer is very malleable; therefore the mechanical signal almost entirely acts on the rods. Thus the electrical capacity of composite piezoreceiver is tens of times smaller, because the permittivity of polymer is hundreds of times less than that of the piezoelectric element.

Therefore piezoelectric composites are very promising materials because they open the possibility of effective control over their electrical and mechanical parameters. The advantages of such composites are the high coefficient of cohesion, low acoustic impedance (in good agreement with the impedance of water or human tissue), mechanical flexibility combined with low mechanical quality factor. In addition to increased piezoelectric efficiency, some piezoelectric composites can show a magnetoelectrical effect. These composites are composed of magnetostriction ceramics and piezoelectric ceramic and are capable of producing an electrical response (voltage or current) under the influence of an external magnetic field.

The classification of various composite structures is proposed by R. Newnham [13]. Properties of composite can be divided into three major effects: the effect of *sum*, the effect of *combination*, and the effect of *product*.

1. The effect of sum. Assume that one of many physical properties of composite and its components are considered. Suppose that component 1 has a property characterized by parameter Y_1 while component 2 has parameter Y_2 . Then, the composite will have some *intermediate* value of this parameter—a value between Y_1 and Y_2 . In case of a two-component system, the given property is described in the composite with summary function Y^* , shown in Fig. 1.41. In case of sum effect, the obtained dependence of the summary parameter from volume fraction of components may be characterized not only by linear dependence, but also might have a concave or convex shape. Thus, and it is very important, that the average value of Y^* in composite will never be more than Y_1 or less than Y_2 .

This example is commonly used in the microwave range composite material with predetermined permittivity, set within $\epsilon^* = 5 \dots 40$. This composite can be

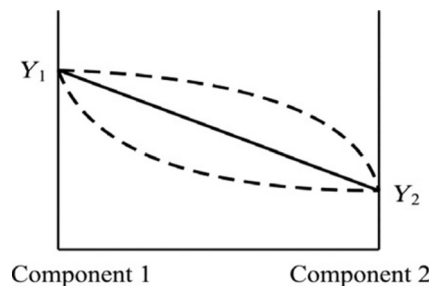


FIG. 1.41

Effect of sum in two-component composite.

prepared from ceramic powder of rutile ($\epsilon_1 = 100$) and polymeric polyethylene ($\epsilon_2 = 2.5$). The dielectric permittivity of such composites depends on the volume fraction of ceramics, but it cannot exceed value ϵ_1 .

2. **The effect of combination.** It is assumed that components of the composite are characterized by two different properties: Y and Z . In this case, sometimes, the average value of obtained certain parameter in a composite *may exceed* the parameters of both components of the composite. The effect of the increase in the output parameter is determined by the ratio Y/Z that depends on both parameters. For example, in some of piezoelectric composites basic properties of components are combined: high piezoelectric modulus of piezoceramics and low permittivity of polymeric matrix (as in Fig. 1.40). As a result, the piezoelectric sensitiveness of composite (that is dependent on ratio of piezoelectric modulus to permittivity) increases substantially. Therefore, this composite has a significant advantage over the properties of components.
3. **The effect of product.** Besides, such a two-component composite is considered, wherein one of components has a significant property Y , which is absent in the second component. However, in the second component, a quite different property Z is present, which does not have the first component. In this case, it is expected to obtain, in the resulting composite, these brand-new features and this is the effect of a product.

For example, based on this concept, the *magnetolectrical* ceramic composite material has been developed, consisting of magnetic components with significant magnetostriction effect (CoFe_2O_4 that is nonpiezoelectric), and the piezoelectric component (BaTiO_3 , exhibiting no magnetic properties) [14]. Under the action of a magnetic field on composite, cobalt ferrite shows magnetostriction, which is transmitted to the grains of barium titanate as the stress and results in generation of electrical charges (of voltage) due to the piezoelectric effect of BaTiO_3 . Thus, due to the composite material, the inexpensive ceramic sensors of magnetic field monitoring are elaborated.

Metamaterials are composites in which the heterogeneous medium contains inclusions; however, in this case, unlike other types of composite materials, inclusions are miniature, sometimes even nanoscale, *radioelements*. Due to these inclusions, metamaterials have unique electrophysical and optical properties, caused by the *resonant interaction* with electromagnetic field.

In metamaterials, a very interesting idea is realized: the possibility to obtain the *negative refractive index* for microwaves or light. In these materials electromagnetic wave, for example, light is not refracted as usual, that is, it deviates not to the right, but to the *left* at the negative angle (Fig. 1.42A). Therefore these materials are often referred to as materials with negative refraction (negative index materials—NIM) or left-handed materials (LHM). V.G. Veselago, who theoretically predicted the existence of metamaterials, called them “left environments.”

In the usual medium, directions of vector of electric field intensity \mathbf{E} , vector of intensity of magnetic field \mathbf{H} , and wave vector \mathbf{k} form the right triplets, that is, they can be described by right-hand fingers (RHM). In contrast, in a metamaterial these vectors form

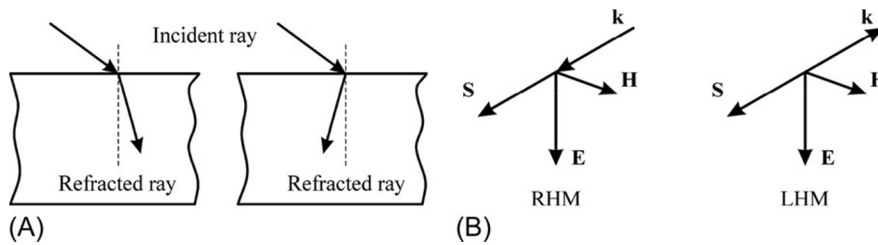


FIG. 1.42

Effect of light refraction in conventional material and metamaterial: (A) direction of rays in medium; (B) orientation of electromagnetic field vectors in ordinary (RHM) material and metamaterial (LHM).

the left-hand triplet (Fig. 1.42B). However, in left-hand material, LHM, the Poynting vector \mathbf{S} that shows direction of energy propagation remains in the right triplet.

It is well known that electromagnetic waves can propagate similarly as in vacuum ($\epsilon = 1$, $\mu = 1$) in a dielectric medium with *positive* permittivity and permeability. Parameters ϵ and μ are fully defined for each particular material due to one or other atomic or molecular structure. In ordinary materials, these parameters are defined by *electrical polarization* (displacement of electrical charges with electrical moment formation) and by *magnetization* (orientation of elementary magnetic moments). Properties of atoms or molecules follow fundamental laws of physics, and they always lead to *positive static values* of permittivity ϵ and permeability μ (at that, in most substances μ is close to one).

However, there are exceptions—in ranges of frequencies where the *own resonant phenomena* are observed: this is possible as for polarization so for magnetization. In the first case, when the phase of dielectric “response” (elastic displacement of charges) lags *behind* the phase of applied field, the response is described by the *negative* value of ϵ . A similar process can occur in case of magnetization, causing negative value of μ . Thus, when resonant response occurs, these *narrow frequency ranges* are characterized by negative ϵ or μ (i.e., however, it is accompanied by a *very large absorption* of electromagnetic waves).

In case of ionic polarization in dielectrics, their lattice resonance occurs in the frequency range of infrared waves ($\sim 10^{13}$ Hz), while for electronic shells polarization—in range of ultraviolet waves (frequencies $> 10^{16}$ Hz): both these ranges are quite far from the frequency range of metamaterials expected applications. Thus, at first glance, there is no basis for hoping to obtain resonant phenomena in *continuous homogeneous medium* as in microwaves so in visible optical range.²

²*Note.* However, it can be noted that in piezoelectrics, the electromechanical resonance is possible that also leads to negative value of ϵ . Usually, this resonance occurs at frequencies of $10^5 \dots 10^7$ Hz (depending on size of piezoelement); to realize this resonance at microwaves, the size of piezoresonators should be only a few micrometers. It is obvious that microelectronic technology is responsible for actualizing this case.

For this reason, the metamaterial can be obtained exclusively from the noncontinuous and inhomogeneous medium: metamaterials always are the composites [15]. Usually, metamaterials are constructed from the *discrete* resonant micro- and nanoelements: “meta-atoms” that mimic electromagnetic reaction of atoms and molecules of natural substances. Meta-atoms are grouped in the form of single or multilayered lattice, and their *small size* (much less than wavelength of radiation) makes it possible to treat the created lattice as a *homogeneous* medium for a given wavelength (by analogy with natural crystals); using the concept of “effective medium” for characteristics calculating.

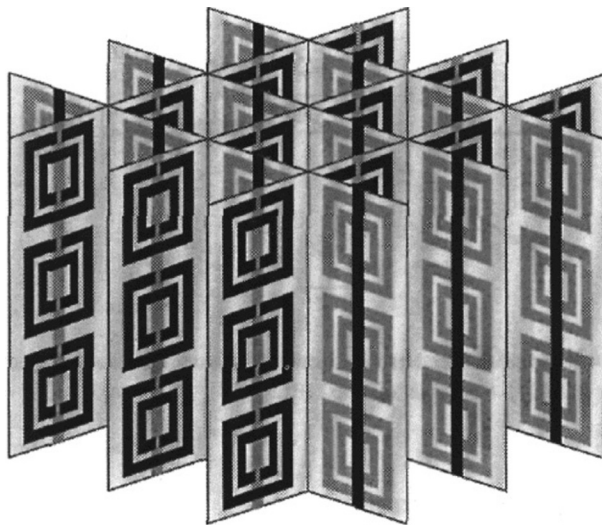
New materials can be used in the development of new types of radioelectronic and photonic functional electronics: devices with negative refraction for controlling radiation in gigahertz and visible ranges that allow obtaining of a clear image of elements with dimensions much shorter than the wavelength without diffraction distortion; systems for electromagnetic invisibility, Stealth technology, and much more.

The negative refractive index n (Fig. 1.42A) is due to a strong spatial dispersion in metamaterial and to negative values of permittivity and permeability: $n = -\sqrt{\epsilon\mu} < 0$. Because $\epsilon(\omega) < 0$ and $\mu(\omega) < 0$, these materials are sometimes called “*doubly negative*.” (Correspondingly, in conventional materials permeability and permittivity have positive sign; therefore ordinary media sometimes are called “*doubly positive*.”) The phase velocity of waves in the metamaterials is directed in the opposite direction relatively to group velocity; therefore these materials are also called “*backward-wave media*.” Metamaterials, interacting with optical frequency radiation, usually are called *photonic* or optical metamaterials.

The main way to obtain metamaterials is based on their “assembly” of huge number of miniature discrete modules, cells, or nanoparticles. These modules (cells and nanoparticles) are sometimes called meta-atoms. It is clear that they are not real atoms, but consist of them, that is, they are made of ordinary matter—mainly metals and dielectrics. Dimensions of meta-atoms greatly exceed atomic dimensions. They form a spatial structure (matrix), for example, an artificial crystal lattice, so that the number of meta-atoms even in a small piece of metamaterial reaches 10^3 – 10^9 (Fig. 1.43).

It should be noted here that meta-atoms do not have any chemical bond with each other, unlike atoms of ordinary materials. Therefore the difference in technologies of conventional materials and metamaterials production is understandable. The former are obtained by chemical synthesis from atoms of chemical elements, the latter are obtained as an assemblage of artificial elements by methods of micro- and nanotechnologies. Moreover, it is important that, for incident radiation, the metamaterial imitates a homogeneous medium; for this, the dimensions of meta-atoms and distances between them should be selected to be less than the working wavelength of radiation; the smaller the dimensions, the better the homogeneity condition.

Externally, meta-atoms are tiny formations of wires, strips, plates, rods, disks, rings, spirals, balls, films, coatings, and multilayer structures. The millimeter-sized high- ϵ dielectric resonators and micrometer-sized piezoelectric resonators can serve as dielectric meta-atoms. Moreover, meta-atoms can be in the form of nanoclusters;

**FIG. 1.43**

Practically implemented metastructure for research in microwave frequencies.

finally, they can be a system of holes in flat elements (e.g., they may resemble a fish net). It is important that the configuration and properties of meta-atoms (capacitors, inductances of oscillatory circuits, or miniature resonators) ensure that they perform functions of simplest capacitors, inductances, oscillatory circuits, or miniature (nano-) resonators.

Thin layers of metamaterials deposited on a substrate are called metafilms or metacoverings. In the simplest case, the metafilm is a patterned single-layer film made of metal, semiconductor, dielectric, or magnetic material that is deposited on a dielectric or semiconductor substrate. The pattern is determined by the configuration of the abovementioned electroradio elements with unique properties due to resonant interaction with an electromagnetic field.

Thus metamaterials are artificial periodic structures with lattice constants much smaller than the wavelength of incident radiation. These are media consisting of resonance elements in which negative propagation of waves takes place. The dimensions of meta-atoms are smaller than the wavelength of radiation interacting with them. They have the ability to simulate homogeneous material, whose properties are absent in natural materials.

It is important to note that metamaterials in the optical wavelength range have already been created, and they opened the door to create a new photonic and quantum-optical technology—optical nanoantennas, nanolasers, nonlinear elements, and other devices for generating and controlling light-transmitted systems developed to overcome the diffraction limit. Metamaterials are the basis for such areas of science and technology as nanoplasmonics and nanophotonics. A new class of

composite materials has become widespread, in which the scale level of individual component sizes reaches the nanometer range.

A **nanocomposite** is defined as the multicomponent solid material in which one of components in the 1, 2, or 3 dimensions has a size not exceeding 100 nanometers; moreover, nanocomposites are understood as structures consisting of a set of repeating component layers (phases), the distance between which is measured in dozens of nanometers [12].

For example, a method for creating anodes from silicon nanospheres and carbon nanoparticles for lithium batteries has been invented. Anodes made from a silicon-carbon nanocomposite are much more closely adhered to the lithium electrolyte, thereby reducing the charging or discharging time of a device. From nanocomposites, consisting of cellulose base and nanotubes, it is possible to produce *conductive paper*. If such a paper is placed in the electrolyte, something like a flexible battery will appear. Furthermore, in the electronics industry, nanocomposites are used to produce thermoelectric materials that demonstrate a combination of high electrical conductivity with low thermal conductivity.

Graphene occupies a special place in the development of nanocomposite materials. Nanocomposites containing graphene and tin can significantly increase the capacity of *lithium-ion batteries* and reduce their weight. Recently, it has been found that the addition of graphene to epoxy composites leads to an increase in rigidity and strength of material compared to composites containing carbon nanotubes. Graphene is better combined with epoxy polymer, more effectively penetrating the structure of composite.

Nanocomposites based on polymeric matrices and nanotubes are able to change their electrical conductivity due to the displacement of nanotubes relative to each other under the influence of external factors. This property can be used to create microscopic sensors that determine the intensity of mechanical action over extremely short periods of time. Moreover, nanotechnology can be used to produce photonic crystals.

Photonic crystals are nanostructured materials in which the periodic change in the refractive index at wavelength scales of visible light creates so-called forbidden bands for photons. These bands influence the propagation of photons of visible light in a material (this effect is similar to how periodic potential in semiconductors affects the determination of the electron flux allowed and the forbidden energy bands). The structure of a photonic crystal can be characterized by a periodic change in the refractive index in 1, 2, or 3 spatial directions. Photonic crystals can be used in the light sources on single crystal, because the pattern of their radiation as well as the direction of beam propagation can be easily controlled.

1.7 SUMMARY

1. Solids are primarily crystals and polycrystals, as well as ceramics, glasses, glass-ceramics, quasicrystals, amorphous substances, composites, and nanocrystalline structures.

2. Crystals are characterized by a near-perfect well-ordered internal structure. Therefore crystals can be described by 3D spatial periodic structure. A peculiar property of crystals is their translational regulation—elementary cell that consists of a few atoms can be translated supposedly “infinitely” in all directions, creating a regular crystal lattice.
3. Polycrystals consist of a large number of small crystals (crystallites). Macroscopic structure of polycrystals, outwardly, seems disordered, but microscopic components of this structure (crystallites units) are small crystals with perfect microscopic structure and similar properties as a large single crystal.
4. The glass-like and amorphous states of solids are characterized by the absence of long-distant (translational) symmetry. However, these materials are characterized by the order in the immediate surroundings adjacent to each atom.
5. In 2D systems, the strictly ordered structure is possible only in a plane. In such a system, if the planar regularity is repeated, the nanodimensional superstructure (artificially created in semiconductor) can have peculiar electronic properties, characterized by the so-called quantum wells (this case relates to 2D nanostructures).
6. The 1D nanostructures might be linear (wire-like) systems, wherein translated ordering is observed along a single direction.
7. There are, furthermore, systems wherein the dimensions along all three directions are commensurate with the distance between atoms. Such zero-dimensional (0D) systems can be “quantum dots,” wherein only $10\text{--}10^3$ atoms have an ordered structure.
8. The creation of ordered crystalline (and other) bodies of atoms is accompanied by a decrease in energy. This corresponds to the certain minimum of a system’s energy when atoms become ordered relative to each other, with significant redistribution in electronic density.
9. According to the electronic theory of valence, the interatomic bond occurs due to the redistribution of valence electronic orbitals, and that results in the stable electronic configuration of noble gas (octet) through the formation of ions as well as by the formation of electron pairs between atoms.
10. Any connections of atoms, molecules, or ions are carried out through electrical interaction. At relatively large distances between particles, the electrical forces of *attraction* dominate whereas, at small distances, the *repulsion* between particles dramatically increases. The balance between long-range attraction and short-range repulsion determines the basic properties of a certain solid. The bond that occurs between atoms (as a result of spatial restructuring of their valence electrons) and which is caused by these electron interactions is the *chemical bond*.
11. At the heart of the classification of solids into metals, dielectrics, and semiconductors is the spatial distribution of valence electrons. In molecular crystals, for instance, electrons are completely locked within their molecules. When crystals are formed from atoms of metal, the orbits of

valence electrons strongly overlap each other. As a result, valence electrons become distributed within the spaces between atoms and can be described by a general wave function. It is believed that, in metals, an *electronic gas* is formed.

12. The *ionic crystals* are chemical compounds that are formed by metallic and nonmetallic elements. The forces of ion attraction are mostly long range: the energy of attraction rather slowly varies with distance. Like molecular crystals, ionic crystals are characterized by such a distribution of electronic charges wherein they are almost completely localized near the ions.
13. The *covalent crystals* have, in principle, a similar nature of connection as the metals-valence electrons become *shared* between atoms. The forces of attraction in case of covalent bonds are not so long range as in the case of ionic bonding. The covalent bond (otherwise called a *homeopolar* bond) is formed by the overlapping (socialization) of pairs of valence electrons. This link is provided with electronic clouds that are called a *mutual* electron pair. At covalent chemical bond formation, the reduction of total energy occurs due to an *exchange interaction* that plays an important role in this process.
14. The *van der Waals bonds* are always present in atomic connections, but they dominate only in the absence of valence bonds; in such cases, these bonds become a principal type of chemical bonding (usually, in *molecular crystals*). The van der Waals forces of attraction are relatively short range and weak as compared with conventional valence forces. In nonpolar molecules, the forces of attraction arise by mutual deformation of electronic shells. Because this mechanism is investigated through optical polarization dispersion, the forces of attraction of this type are *dispersive* ones. In polar molecules, the *orientation interaction* contributes to the energy of bonding. Moreover, there exists an *induction interaction* between the permanent dipole of one molecule and the induced dipole of another molecule.
15. The *hydrogen bonds* are realized when two hydrogen atoms in one molecule interact or when a hydrogen atom in one molecule interacts with an electronegative atom such as P, O, N, Cl, or S of another molecule. The cause of the hydrogen bond is the redistribution of electronic density between atoms, induced by the small size of the hydrogen ion (H^+ ; proton).
16. The *defects* in crystals are formed during their growth (under the influence of thermal, mechanical, and electrical fields), as well as during crystal irradiation by neutrons, electrons, X-rays, and ultraviolet radiation (radiation defects). There are *point* defects (zero-dimensional), *linear* (1D) defects, *plane* defects (2D), and *bulk* (3D) defects.
17. Parts of atoms or ions of a crystal may be missing locally, thereby violating an ideal crystal lattice scheme: these defective places are the *vacancies*. Furthermore, foreign (impurity) atoms or ions can exist in crystals, replacing basic particles that form a crystal or take root between them. Their own atoms (or ions) can serve as point defects in crystals, if they shift from normal positions (interstitial atoms or ions).

18. In the process of crystal growth or during its plastic deformation and in many other cases, *dislocations* can appear. Moreover, dislocations can arise in a crystal during its doping. The distribution and behavior of dislocations under external influences determine many important mechanical properties of a crystal, including strength, ductility, and so on. The mobility of dislocation determines the *plasticity* of crystals; at the location of the greatest internal stress, clusters of dislocation can occur that can cause the *destruction* of the crystal. The problem of plastic flow (i.e., irreversible deformation) in metallic crystals is solved by the association and movement of dislocations. These dislocations impede the process of magnetization and electrical polarization due to their interaction with domain boundaries.
19. Elastic deformations of crystal structure arise in the vicinity of defects that lead, in turn, to the appearance of internal mechanical stresses. For example, point defects interact with dislocations that result in the increase or decrease of crystal strength. Defects in crystals affect the absorption spectra of luminescence, light scattering in a crystal and can change electrical conductivity, thermal conductivity, ferroelectric and ferromagnetic properties, etc.
20. The *vacancies* in crystal lattice usually are *Schottky defects*. The formation of vacancies can be explained by some atoms moving outside from the crystal surface and they being replaced by other atoms from a volume. For most crystals, the energy of vacancy formation is approximately 1 eV. Lattice defects that usually are called the *Frenkel defects* arise by mechanisms that generate interstitial atoms or ions in a crystal.
21. The *polarons* are charge carriers bound in the lattice of an ionic crystal (most often, they are bound electrons). The polaron is not a “static” defect because it is much more mobile than vacancies or interstitial ions. The *excitons* can be interpreted as the mobile point defects in a crystal. In the case of excitons, atoms or ions of crystal do not change their location, but they become significantly different from their neighbors by excited electronic states. The movement of an exciton in a crystal is not connected with the change of atom or ion positions, and therefore excitons (as polarons) have much greater mobility than replacements of vacancies, interstitial atoms, and impurities.
22. The *dislocations* are crystallographic defects or irregularities within crystal structure. The presence of dislocations strongly influences many properties of materials. The *edge dislocation* is a land of “excessive” atomic planes that splits the crystal. It corresponds to the row of ordinary atoms along the edge of an additional part-plane of atoms within the crystal. In other words, edge dislocation is such a defect wherein an extra half-plane of atoms can move through the crystal, distorting the nearby planes of atoms. The *screw dislocation* is a result of changes of one area of crystal with regard to another. It corresponds to the spiral axis of structural distortion, connected to normal parallel planes. It comprises the structure wherein a helical path is traced around a linear defect (dislocation line).

23. The *solid solutions* are widely used in electronic components technology. The presence of two components is possible in crystals or polycrystals (in metals, these are alloys). The solid-state solution is a mixture that remains in a single homogeneous phase. The *interstitial* type of solid solutions is a result of the fact that atoms of the element, which dissolves, are placed in empty spaces of the solvent lattice. The *substitutional* solid solutions are formed by a partial substitution of solvent atoms. This process can occur without incurring significant stresses in structure only when the size of atoms does not differ greatly among themselves.
24. A structure is *polytypic* when it is composed of similar structural elements but with a different sequence of their location. Polytypic lattice parameters in a *plane layer* are unchanged but, in the direction perpendicular to layers, lattice parameters are different, although they are always multiples of the distance between adjacent layers. Polytypism is a special case of polymorphism: 2D translations within layers are essentially preserved.
25. *Isomorphism* and *polymorphism*. The property of chemically closed atoms, ions, or other structural elements to replace each other in the crystal lattice and form continuously variable composition is isomorphism. The ability of certain substances to exist in multiple crystalline phases, differing in symmetry of structure and in physical properties, is polymorphism. The change in environmental conditions may cause polymorphous transformation. During these transformations (that usually are *phase transitions of first order*), heat absorption and internal energy jumps are observed as well as changes in other physical properties of matter. Furthermore, there are such polymorphic modifications that differ by very little changes in physical properties. Polymorphic transitions between states are *phase transitions of second order* and usually are described as “order-disorder” type of transitions.
26. The *symmetry* of crystal structures determines their physical properties. Therefore many properties of solids may be described by the peculiarities of crystal symmetry. The relationship between the geometry of external shape and internal building of crystals, as well as their physical properties, are specified by *physical crystallography*. The physics of crystals formulates some principles that establish a connection between the symmetry of a crystal and physical phenomena; central to these are Neumann principle and Curie principle.
27. The mechanism of how the physical properties of crystals are conditioned with their symmetry was formulated by Neumann: *the symmetry of physical properties of a crystal is not lower than the symmetry of its structure*. This means that the structure of a crystal contains all elements of the symmetry of its properties (but also may have other symmetry elements). Therefore, information about crystal symmetry enables prediction of the possible physical effects in a crystal.
28. In accordance with the Curie principle the crystal, being under external influence, has only those symmetry elements that are common to the crystal in

the absence of influence and impact (in case of lack of crystal)—that is, in the system “crystal-influence,” *only common elements of symmetry remain.*

29. As the *element* of symmetry, an imaginary object can be used that supports the realization of the operation of symmetry. To such elements of symmetry belong the *planes, axis, and center* of symmetry (center of the inversion). The combination of a point (or part of figure) with another point (or part of figure) is called an *operation* of symmetry. Both parts of figures that are combined are symmetric. Operations of *point symmetry* are left in its place, at least, on one point of the final figure. This is the point of intersection of all elements of symmetry.
30. *Rotation* and *mirror rotation* as well as *inverted rotations* and *reflections in the plane* of symmetry are selected as symmetric operations. There are elements of symmetry of the first and second kinds. The former include the symmetry plane, rotary axis, and center of inversion (symmetry); the second include some complex elements of symmetry: inversions and mirror-rotary axes. To analyze symmetry, *screw rotations* and/or *glide reflections* are also used. These are rotations or reflections, together with partial translation. The *Bravais lattices* may be considered the outcome of translational symmetry operations. Combinations of operations with additional symmetry operations produce 230 crystallographic *space groups*.
31. The *plane of symmetry* is a plane of mirror reflection; this is an operation of a similar point combination. To refer to a specific class of symmetry elements, the plane of symmetry can be denoted by P . In the international system, a mirrored plane is denoted by the letter m , it bisects all segments that connect symmetric points that are perpendicular to it (part of the figure).
32. *Rotational symmetry* is symmetry with respect to some or all rotations in the Euclidean space. The rotary axis of symmetry of the n th order is denoted as L_n , that rotates around a certain angle $\alpha = 360^\circ/n$. Moreover, the rotary axes are marked by symbols 1, 2, 3, 4, 5, 6, 7, ..., ∞ , where the numbers indicate the order of axis. The n -fold rotational symmetry operation rotates the object by $360^\circ/n$. Only $n = 1, 2, 3, 4,$ and 6 are permitted in the periodic lattice. The inversion axis is a combination of rotation and the center of symmetry operations.
33. The *center of symmetry* (inversion center, denoted as C) is a special point inside a figure or unit cell; it is characterized by the fact that any line drawn through the center of symmetry falls into the same point of figures on both sides of the center at equal distances.
34. The *class of symmetry* is a set of symmetry elements of the crystal (or any object) that describes its possible symmetric transformations. A unit cell can be selected in any crystal and, on its basis, all crystal lattices can be built using translations. These translations are the displacement of a unit cell within a crystal. The full set of symmetry elements of any material is known as the *group of symmetry*.
35. Crystals and textures that have a center of symmetry cannot show piezoelectric properties. In the absence of external influences, only noncentrosymmetric

structures are capable of being piezoelectrics. Among them, only the crystal with a polar axis might be pyroelectric.

36. The *quasicrystals* exhibit a special, new type of symmetry, different from the usual crystals. They have symmetry elements that previously were considered impossible in crystals: classic crystallography does not allow symmetry axes of fifth, seventh, and higher orders. With these axes, the elementary cell cannot ensure complete filling even on the plane (and, moreover, in volume). However, quasicrystals exist, and they can have axes of symmetry of fifth, eighth, or higher orders.
37. *Nanomaterials*, as a rule, are small particles (clusters) of materials consisting of 10–1000 atoms. Their properties depend on the number of atoms in the cluster and on the relative position of atoms, as well as on the shape and symmetry of clusters.
38. The *composites* consist of different materials united in a single whole, and have important applications in electronic devices. They are used in various active and passive components (e.g., piezoelectric with polymer). The physical and technical properties of composites that ensure their applications, with advantages over crystals, ceramics, and polymeric materials, can be described by three effects: the sum effect, the combinative effect, and the effect of the product.
39. Electromagnetic *metamaterials* are artificially structured in a special way to be mediums that have electrical and magnetic properties, which are significantly different from the original structural materials. For example, a metamaterial can have a *negative refractive index*, which is never observed in natural materials. The internal structure of metamaterials plays an important role in the formation of their characteristics and parameters.
40. *Nanocomposites* are solid formations consisting of a basic matrix and nanosized components that differ in their structural parameters and chemical properties. Mechanical, electrical, thermal, optical, and other characteristics of nanocomposites differ significantly from the properties of ordinary composite materials made of the same basic substances or elements.

REFERENCES

- [1] L.H. Van Vlack, *Materials Science for Engineers*, Addison-Wesley Publishing Co., Reading, MA, 1975.
- [2] A. Holden, *Bonds Between Atoms*, Oxford University Press, New York, 1970.
- [3] C. Kittel, *Introduction to Solid State Physics*, fifth ed., John Wiley, New York, 1976.
- [4] N.W. Ashcroft, N.D. Mermin, *Solid State Physics*, Holt and Winston, New York, 1976.
- [5] Y.M. Poplavko, *Polar Crystals: Physical Nature and New Effects*, Lambert Academic Publishing, Saarbrücken, 2014.
- [6] M.A. Shaskolskaya, *Crystallography*, Vishaya Shkola, Moscow, 1976.
- [7] I.S. Jeludev, *Symmetry and Its Application*, Energoatomizdat, Moscow, 1983.
- [8] I.V. Zolotuhin, Y.E. Kalinin, O.V. Stogney, *New Directions of Physical Materials Science*, Voronej University Ed., Voronej, 2000.

- [9] I.M. Bolesta, *Solid State Physics*, Lvov University Ed, Ukraine, 2003.
- [10] A.F. Kravchenko, V.N. Ovsyuk, *Electronic Processes in Low-Dimensional Solid Systems*, Novosibirsk University Ed, Russia, 2000.
- [11] R. Waser (Ed.), *Nanoelectronics and Information Technology: Advanced Electronic Materials and Novel Devices*, Wiley-VCH, Weinheim, 2005.
- [12] H.S. Nalva (Ed.), *Nanostructured Materials and Nanotechnology*, Academic Press, New York, 2002.
- [13] R.E. Newnham, D.P. Skinner, L.E. Cross, Connectivity and piezoelectric-pyroelectric composites, *Mater. Res. Bull.* 13 (1978) 525.
- [14] K. Uchino, *Ferroelectric Devices*, Marcel Dekker, New York, 2000.
- [15] G.V. Eleftheriades, in: G. Balmain (Ed.), *Negative-Refraction Metamaterials*, Wiley, Hoboken, NJ, 2005.

Mechanical properties of solids

CONTENTS

2.1 Mechanical Stress Tensor	72
2.2 Mechanical Strain Tensor	75
2.3 Elastic Stiffness and Elastic Compliance	78
2.4 Elastic Waves in Crystals	82
2.5 Summary	92
References	93

Mechanical properties, which reflect the internal bonds between the molecules and atoms of a material, are basically *elasticity*, *durability*, *stiffness*, and *toughness*. Because electronic components are sometimes exposed to mechanical impacts during operation, the mechanical strength of materials and, especially, their elasticity (ability to reversibly deform under stress) have considerable practical interest. Information about the mechanical properties of solids is needed for the development of electronic solid-state devices, thus necessitating the study of many characteristics of solids, including an understanding of the nature of chemical bonding.

Knowledge of mechanical properties is necessary when using solid components as structural materials. For example, *durability* characterizes the counteraction to a mechanical load that causes destruction of a solid. However, long before this destruction, the solid body exhibits *elasticity*—the ability to completely recover its shape after the removal of the applied external impact. Besides, if the externally acting force is large enough, *plasticity* is commonly observed before destruction, wherein the change in the form of solid is not instantaneous; it is in the final stage of plastic deformation at which *failure* occurs.

Taking into account the wide application of microelectromechanical systems (MEMS) in modern electronics, many mechanical properties—such as mechanical fatigue and resistance to radiation—should be considered important characteristics of construction materials. It is necessary to note that, in various solids (glasses, ceramics, plastics, etc.), resistance to compressive strength is much greater than that to tension and bending strengths. Furthermore, many materials might have a relatively high resistance to static loads; however, they can be quickly destroyed under dynamic loads, that is, under the action of a suddenly applied force, they are *fragile*.

In this book, only those mechanical properties will be considered that are of high importance for electromechanical and magnetomechanical effects. These are, specifically, *elasticity* and *elastic wave propagation* in different structures, and they define

the practical use of crystals in piezoelectronics, acoustoelectronics, magnetoelectronics, acousto-optics, etc.

When describing the elastic properties of solids, the *discrete* structure of a material may be ignored; therefore a crystal can be considered a *continuous homogeneous medium*—continuum approximation [1]. This approach is justified up to frequencies of 10^{12} Hz, which is much greater than the operational frequencies of conventional electronic devices (up to 10^{11} Hz).

The forced deformation of a solid alters the mutual arrangement of its atoms, with the resultant emergence of an effort within the material to try to restore the body to its initial condition. The forces deforming the body are called *internal*, and the value of such force per unit area is the *stress*. If stress (and its corresponding relative deformation—*strain*), occurring under the action of external forces, quickly reaches zero after the removal of the action, then this is a case of the so-called perfectly elastic body [2]. In such cases, *Hooke's law* holds valid: the relative deformation is proportional to mechanical stress, and the behavior of the body during deformation does not depend on the rate of strain. To apply Hooke's law in anisotropic solids, it is necessary to introduce the concepts of stress and strain *tensors*.

2.1 MECHANICAL STRESS TENSOR

The next model considers the elastic behavior of solids on the assumption that stress is homogeneous and all parts of a body are in a state of *static equilibrium*. The theory of elasticity studies phenomena occurring in a deformable body when a mechanical stress is applied, where interatomic forces are considered as short-range phenomena. In this case, forces acting on one part of a body are directly transferred to the other parts of the body.

Further, a very simplified model is considered: the applied force is proportional to the flattening of a solid, and the force per unit surface area is stress. It is uniform if the force is independent of the location of any selected cell in a body. If this specified condition is not satisfied, stress might be inhomogeneous. Moreover, because a crystal may be anisotropic, mechanical stress may also depend on the chosen direction of force application.

The concept of *stress tensor* for structures of various dimensions is illustrated in Fig. 2.1. Initially, the one-dimensional (1D) structure is considered, followed sequentially by the two-dimensional (2D) and three-dimensional (3D) structures [1]. Fig. 2.1A shows a uniformly elongated elastic rod (1D crystal) that is loaded by forces. Mechanical stress is *not a vector*, and, therefore, it cannot be indicated by a single arrow but instead by a *pair of arrows*—similar in magnitude and opposite in directions. An external power may stimulate the stretching of a rod ($X > 0$) and therefore its compression ($X < 0$). Therefore, mechanical stress, unlike a force vector, does not lead to any movement of the body, making the stressed rod remain in its fixed position. Thus, the unit of measurement of 1D stress is associated with unit of force: $[X] = [\text{N}/\text{m}^2]$.

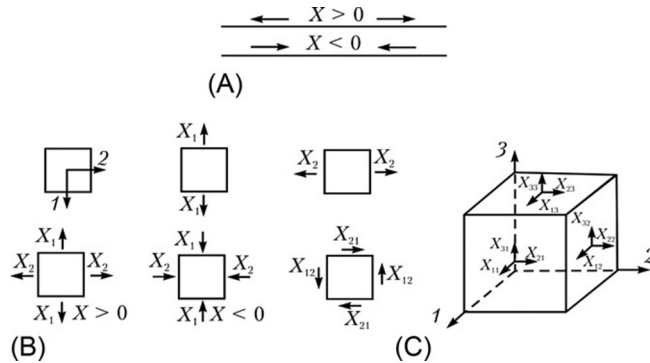


FIG. 2.1

Homogeneous mechanical stresses in solids: (A) one-dimensional model, (B) two-dimensional model, and (C) three-dimensional model.

The 1D model is not only an idealization adopted for ease of understanding. In some microelectronic devices, the metal, dielectric, or semiconductor materials are applied as “quasi-1D” (extensive) small crystals or polycrystals. For example, in the technology of piezoelectric composites, a set of oriented rigid piezoelectric rods placed in a pliable polymer is used (see Fig. 1.40 in Chapter 1). Another example is a set of nanorods (usually, ZnO) located on the substrate (a very promising material in nanoelectronics). Thus, a “quasi-1D” representation of stresses, as shown in Fig. 2.1A, has both theoretical and practical importance.

In the 2D case, the manner of stress application to a flat surface might be different (Fig. 2.1B). A compressive or stretching stress can be presented as independently—along two perpendicular axes, 1 and 2. In the case of an arbitrary direction of the compressive/stretching stress, it should be decomposed into components along two mutually perpendicular axes. In addition to the said stresses, a special type of *shear stress* is possible: twin stresses X_{12} and X_{21} . In this case of a condition of equilibrium, that is, assuming no movement or rotation of a “quasi-2D” crystal, we have $X_{12} = X_{21}$.

The unit of stress in the planar model discussed herein remains the same: $[X] = [\text{N}/\text{m}^2]$. Consideration of 2D crystal structure (similarly as in the 1D structure) is important not only for theory but also in practice, as consistent with real elements—the films (semiconductor, ferroelectric, piezoelectric, pyroelectric, etc.). As an example, a very important 2D crystal semiconductor (graphene) is shown in Introduction, Fig. I.3. Dielectric (piezoelectric) films are used, for example, to excite high-frequency hypersonic waves in a crystal surface, and they are also applied in many technical devices, based on surface acoustic waves (SAW).

For example, by thermal deposition of piezoelectric structures on a substrate (e.g., BaTiO₃ on silicon) at high temperatures, a piezoelectric film can be obtained. During cooling from the temperature of synthesis (~1000 K) to working temperature (~300 K), the film becomes mechanically stressed; therefore the thermal expansion

coefficients of the substrate and piezoelectric film differ. These conditions must be taken into account during the application of piezoelectric films (in acoustoelectronics or in MEMS).

In most practical tasks, there is application of 3D crystals and textures. Mechanical stress, in this case, is determined by the force acting on the surface and has the dimension $[X] = [N/m^2] = [Pa]$. Theoretical analysis suggests that stress is *uniform* (and remains the same throughout in the crystal). Components of this stress (forces acting on the opposite face of a cube) compensate each other. The normal components of stress are indicated by the *same indices*: X_{11} , X_{22} , and X_{33} ; they act perpendicular to the corresponding surface. Obviously, on opposite faces, there would be the same “arrows”—representing stress components (data not shown in Fig. 2.1C). For example, if stress component X_{33} tends to stretch the cube along axis 3, then, on the opposite face of the cube, the same magnitude stress component $-|X_{33}|$ acts by being directed opposite to X_{33} and, thus, balances it.

Besides stresses directed normally to faces, there might also be *shear stresses* directed tangentially to the faces of the considered cube. These include components X_{13} and X_{23} on the upper face of the cube (Fig. 2.1C), components X_{31} and X_{21} on the front face, as well as X_{12} and X_{32} on the right side of the cube. These stresses are counterbalanced and do not induce any rotation of the sample.

Enumerated components form the *stress tensor of second-rank* X_{ij} , which initially appears similar to tensors of permittivity, conductivity, and permeability. However, the mechanical stress tensor, by its physical nature, is quite different from the second-rank tensors ϵ_{ij} , σ_{ij} , or μ_{ij} , which have a structure consistent with the *internal symmetry* of a crystal. Tensors of permittivity, conductivity, and permeability are the *material tensors*, whereas the stress tensor is the *field tensor*, characterizing the structure of forces applied to the crystal *from outside*.

Because shear stress components ($X_{ij} = X_{ji}$) do not create rotary mechanical movements, the full stress tensor can be represented by symmetric matrix:

$$X_{ij} = \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{bmatrix}.$$

Similar to material tensors, the stress tensor can be characterized by the surface of second order:

$$X_{11} \cdot x^2 + X_{22} \cdot y^2 + X_{33} \cdot z^2 = 1,$$

where X_{11} , X_{22} , and X_{33} are the main components of the matrix, reduced into diagonal form. However, depending on the sign of X_{ij} , such a characterizing surface may not only be an ellipsoid but could also be assumed as an *imaginary hyperboloid*, whereas the characteristic surfaces of the material tensors ϵ_{ij} or σ_{ij} are always ellipsoids [3].

When all components of tensor X_{ij} are given with regard to the principal axes, some important cases can be analyzed (Fig. 2.2):

- (a) The line-stressed state (uniaxial stress): a proper matrix is shown in Fig. 2.2A; the example is of a uniform tensile rod.

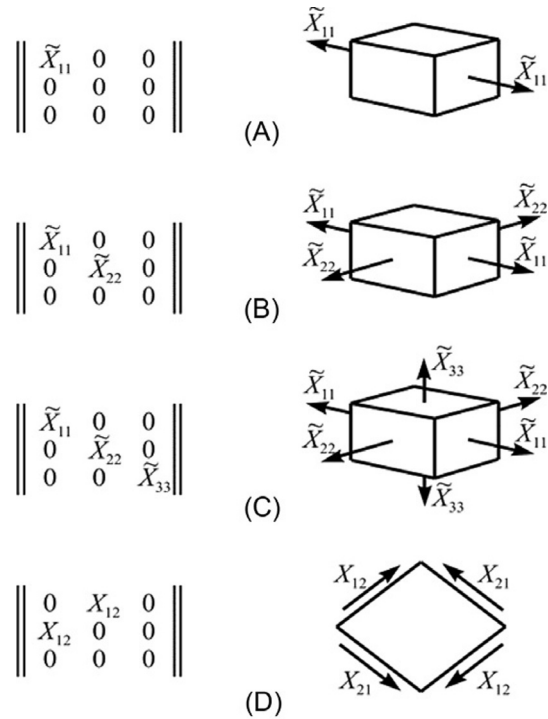


FIG. 2.2

Matrix example for stress tensor with geometric explanation of components.

- (b) The flat-stressed state (biaxial stress): an example and corresponding matrix are shown in Fig. 2.2B.
- (c) The volume-stressed state (three-axial stress): an appropriate matrix X_{ij} and example are shown in Fig. 2.2C.
- (d) The hydrostatic pressure, at which $X_{11} = X_{22} = X_{33} = -p$ (pressure). The corresponding matrix is similar to that in Fig. 2.2C, but the X_{ij} directions in case of a *hydrostatic effect* are opposite to that shown in the figure, and all components have the same value.
- (e) The pure shear stresses (Fig. 2.2D): the shear axis is perpendicular to the plane of a figure.

2.2 MECHANICAL STRAIN TENSOR

Under the influence of mechanical stress, a solid body becomes mechanically deformed. Moreover, the consideration of different strains starts with the *ID model* (Fig. 2.3A) [1]. On an elastic rod OB, the origin O is selected with a big segment OA,

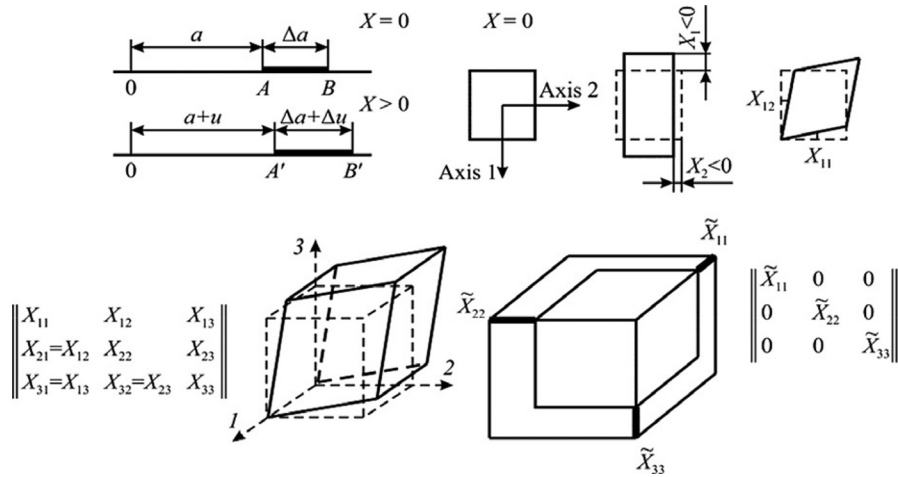


FIG. 2.3

Homogeneous mechanical deformation in a solid: one-dimensional, two-dimensional, and three-dimensional models.

with length a , as well as a small segment AB , with length Δa . When this rod is exposed to positive mechanical stress, it becomes uniformly stretched (Fig. 2.3A). Segment OA acquires length $(a + u)$ whereas the small segment Δa gets an increment Δu . Then, the relative deformation x (i.e., strain) at any point of the rod is defined as

$$x = \lim_{\Delta a \rightarrow 0} \left(\frac{\Delta u}{\Delta a} \right) = \frac{du}{da}.$$

Thus, the physical value of strain is *dimensionless*. In the 1D model, *linear strain* can be indicated as stretching ($x > 0$) or compression ($x < 0$). Under the influence of a large force on a solid, prior to its mechanical destruction, elastic (reversible) strain in the solid can reach $x = 10^{-2} \dots 10^{-4}$. For example, in some dielectrics, when a large external electrical field is applied (increasing up to its breakdown), the relative deformation can reach values of $x \sim 10^{-3} \dots 10^{-4}$.

In Fig. 2.3B, the 2D model is considered. For example, this model can be applied when various *films* are studied (as films are important components in microelectronics). As in the linear model, it is assumed that the deformation of the film is *uniform* over its entire area. This means that, after deformation, the straight lines remain straight (not bent), and the parallel lines remain parallel (not sparked): the lines might be only lengthened (or shortened) to the same extent.

From a planar model consideration, it can be seen that, in addition to linear strain (e.g., x_1 and x_2), angular deformation is possible with shear strains x_{12} and x_{21} . It can be shown that all components of strain constitute the *second-rank tensor* x_{ij} where $i, j = 1, 2$. In the 2D case, the corresponding matrix

$$x_{ij} = \begin{vmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{vmatrix}$$

is symmetric with respect to the main diagonal: $x_{12} = x_{21}$. The symmetric components of the matrix define shear strain, whereas the diagonal components x_{11} and x_{22} represent the deformation of the compressive/stretching type.

In general, $3D$ deformation is the most important for many volumetric effects in solids [1]. This case is shown in the lower panel of Fig. 2.3. Tensor x_{ij} as well as the previously discussed mechanical stress tensor is symmetric, similar to the main diagonal. Diagonal components of this tensor x_{ij} ($i = j$) describe the compressive/stretching type of linear strains, whereas off-diagonal terms characterize different shear strains.

Similar to the stress tensor, the symmetric tensor x_{ij} can be described by the surface of a second-order equation:

$$x_{11} \cdot x^2 + x_{22} \cdot y^2 + x_{33} \cdot z^2 = 1,$$

which, in case of positive factors of x_{ii} , represents an ellipsoid.

After being reduced to the diagonal matrix (when edges of corresponding elementary cube are parallel to the three principal axes of the crystal), the components of strain x_{ij} are shown in Fig. 2.3, in the bottom panel. The main axes are three mutually perpendicular directions of the crystal.

Under a scalar (nondirectional) external influence, for example, when the temperature changes the response (*thermal deformation*), the internal symmetry of a given crystal (or texture) is reflected by the second-rank *material* tensor of *thermal expansion*, α_{ij} . The discriminatory surface (indicatrix) for tensor α_{ij} is a second-rank surface of the *general type* (not only ellipsoid), because components of the thermal expansion tensor might be both positive and negative [3].

The thermal expansion tensor expresses a “connection” between *scalar* influence (temperature) and the *second-rank tensor* response (strain). For comparison, characteristic surfaces of permittivity or conductivity (as well as second-rank material tensors) are always described by ellipsoids, because diagonal components of these tensors are always positive. The point is that such material tensors (ϵ_{ij} or σ_{mn}) define the relationship between *two vectors*: one vector is the “influence” whereas the other is the “response” (e.g., $D_i \sim \epsilon_j E_j$).

Thus, the elastic deformation of a crystal is its reaction to an external action, and it varies on the basis of *scalar* action (temperature δT), *vectorial* influence (electrical field E_i), or *tensor type* of impact (X_{ij}). In all of these cases, the response tensors reflect the *intrinsic property* of a crystal.

As in the case of mechanical stresses (Fig. 2.2), some special cases can be picked out for strains: *linear*, *planar*, and *volumetric* strains (similar to that in Fig. 2.2C). It is necessary to note that “*pure shear*” strain (similar to that in Fig. 2.2D), in contrast to these previously explained simple types of strains (linear and planar), *does not change the volume* of a crystal.

Earlier, only *elastic* deformation was considered, where a linear relationship exists between relative deformation x and mechanical stress X . However, this linearity is maintained only up to a certain value of X_{\max} , which is the *limit of proportionality*. With further increase of stress, linearity is violated because of the appearance of *inelastic deformation* that could mean the beginning of a new state of matter.

Irreversible deformation is a phenomenon wherein deformation becomes *plastic*. In this case, after the removal of the external force, the body does not fully recover to its previous form: a *residual deformation* persists. In case of plastic deformation, Hooke's law cannot be applied. An explanation of plastic deformation is possible by linking it with the sliding or displacement of parts of the crystal lattice in certain planes. Here, the geometrical coordination of atoms usually remains unchanged, because displacement takes place in the whole number of interatomic distances. It was experimentally shown that, in single crystals, consisting of only one element in the unit cell (ions in metal), the *sliding* easily occurs along the direction of greatest linear density on planes with the largest interplanar distance.

The earlier discussion of the case of a *perfectly elastic body* is very simplified. In real solids, mechanical stress may neither exist indefinitely for a long time nor fall instantly (with the removal of the external force). When deformed, the structure of a body continuously varies in a complex way at a particular rate, determined by the nature of the substance. Therefore, all real solids with arbitrary deformations can be characterized by flexible properties. In practice, a body is considered elastic if the external stress does not exceed a certain limit, that is, when the strain is small (usually, $x < 1\%$). Only under this condition, and with sufficient accuracy, is it possible for the linear relationship between strain and stress to be valid.

2.3 ELASTIC STIFFNESS AND ELASTIC COMPLIANCE

An externally applied mechanical stress X can elastically and reversibly alter the shape of a crystal—this is strain x . When the value of strain is small, the following linear relationship holds true:

$$x = sX,$$

where s is the *elastic compliance*. This relationship is exemplified in Hooke's law: deformation x increases (or decreases) in a direct proportion to the applied mechanical stress X . Hooke's law can also be written as

$$X = cx,$$

where c is the *elastic stiffness*, also known as Young's modulus.

Because strain is dimensionless and the unit of stress is $[\text{N}/\text{m}^2]$, the same unit is retained for elastic stiffness: $[c] = [\text{N}/\text{m}^2] = [\text{Pa}]$ (Pascal). Elastic compliance is defined as $[s] = [\text{Pa}^{-1}]$. To measure c and X , sometimes, other (non-SI) units are

used: $1 \text{ kbar} = 10^8 \text{ Pa}$ and $1 \text{ dyne/cm}^2 = 0.1 \text{ Pa}$. As “Pascal” is a very small unit, multiples of the unit gigapascals (GPa) are usually used, where GPa is equal to 10^9 Pa .

As x_{ij} and X_{ij} are second-rank tensors, it can be expected that, in the anisotropic crystals (or textures), each of the nine components of deformations x is induced by the nine stress components of X :

$$x_{ij} = s_{ijkl} X_{kl}.$$

This record, given in the tensor representation, factors in nine equations wherein the right side has nine components. The first of these equations is

$$\begin{aligned} x_{11} = & s_{1111} X_{11} + s_{1112} X_{12} + s_{1113} X_{13} + s_{1121} X_{21} + s_{1122} X_{22} \\ & + s_{1123} X_{23} + s_{1131} X_{31} + s_{1132} X_{32} + s_{1133} X_{33}. \end{aligned}$$

Obviously, as with elastic compliance, elastic stiffness is a tensor of the *fourth rank*, which, in principle, has $3^4 = 81$ components. In reality, however, the number of *independent* components of these tensors is much less, because both stress and strain are symmetric tensors; therefore they contain, even in the most general case, not nine but six components. Accordingly, tensors s_{ijkl} and c_{ijkl} are symmetric tensors toward the first two and last two indices:

$$s_{ijkl} = s_{klij} = s_{ijlk} = s_{jilk}.$$

Consequently, these tensors contain no more than 36 components. In turn, such a tensor with $6 \times 6 = 36$ components is also symmetric toward the diagonal of the corresponding matrix. Therefore even the crystal that falls under the lowest category of symmetry can be described by no more than 21 *independent components* of the elastic compliance (or elastic stiffness) tensor.

To reduce the number of indices, it is acceptable to present elasticity equations not by tensor s_{ijkl} (where $i, j, k, l = 1, 2, 3$) but, instead, by matrix s_{mn} , wherein $m, n = 1, 2, \dots, 6$. The method to go from one type of recording to another is shown in [Table 2.1A](#), whereas components of the elastic stiffness matrix are given in [Table 2.1B](#).

By knowing all components of a tensor, for example, elastic stiffness tensor, it is possible to calculate all components of an inverse tensor (in this case, the elastic compliance tensor):

$$s_{mn} = \frac{(-1)^{m+n} \Delta c_{mn}}{|c_{mn}|},$$

where $|c_{mn}|$ is the determinant and Δc_{mn} is the minor of the matrix without m -row and n -column.

In case of practical calculations, for example, where the piezoelectric or magnetostriction effect influences investigations and applications, other elastic parameters of the crystal or texture, except elastic compliance and stiffness, are important. These parameters are listed in the following paragraphs and can be calculated using the known c_{mn} or s_{mn} .

Table 2.1A Matrix of Elastic Compliances

Tensor indices i, j or k, l	11	22	33	23 and 32	31 and 13	12 and 21
Matrix indices m or n	1	2	3	4	5	6

Notes: $s_{ijkl} = s_{mnn}$ (m or $n = 1, 2, 3$); $2s_{ijkl} = s_{mnn}$ (m or $n = 4, 5, 6$); $4s_{ijkl} = s_{mnn}$ ($m, n = 4, 5, 6$).

Table 2.1B Matrix Components of Elastic Stiffness

	x_1	x_2	x_3	x_4	x_5	x_6
X_1	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}
X_2	C_{21}	C_{22}	C_{23}	C_{24}	C_{25}	C_{26}
X_3	C_{31}	C_{32}	C_{33}	C_{34}	C_{35}	C_{36}
X_4	C_{41}	C_{42}	C_{43}	C_{44}	C_{45}	C_{46}
X_5	C_{51}	C_{52}	C_{53}	C_{54}	C_{55}	C_{56}
X_6	C_{61}	C_{62}	C_{63}	C_{64}	C_{65}	C_{66}

The **density of elastic energy** of the strained (or stressed) crystal can be determined from the expression for elementary mechanical work carried out by force X to create deformation dx : $dW = Xdx$. Depending on the given task and using Hooke's law in two forms, $x = sX$ or $X = cx$, it is possible to obtain the following expression for elastic energy:

$$W_{\text{elast}} = \frac{1}{2}cx^2 = \frac{1}{2}sX^2.$$

The **compressibility** $\langle s \rangle$ is an important parameter of solids, for example, when piezoelectric or ferromagnetic structures are used as emitters and receivers of elastic waves. Moreover, compressibility is an important characteristic of the substance that allows us to judge the dependence of physical properties on interatomic (intermolecular) distances. The greatest compressibility is observed for crystals with long and weak interatomic bonding [2].

Compressibility characterizes the dependence of relative change in volume ΔV of the crystal under hydrostatic pressure p : $\Delta V = -ps$. Parameter $\langle s \rangle$ is formed as an invariant of the elastic compliance tensor:

$$\langle s \rangle = s_{11} + s_{22} + s_{33} + 2(s_{12} + s_{13} + s_{31}).$$

In cubic crystals and other isotropic solids, compressibility equals $\langle s \rangle = 3(s_{11} + 2s_{12})$. It should be noted that compressibility is strongly dependent on the energy of atomic bonds.

The **bulk modulus of elasticity** K is introduced as a parameter that is inverse to compressibility, also called the bulk compression modulus. Bulk modulus can be identified through the elastic stiffness tensor; in cubic crystals, $K = (c_{11} + 2c_{12})/3$. Modulus K is the ratio of stress value to the relative compression value. The bulk

modulus of elasticity describes the ability of a material to resist any change in its volume. Moreover, the bulk modulus K characterizes the capability of an object to change its volume, for example, under hydrostatic pressure. It should be noted that the bulk modulus of a nonviscous liquid is different from zero, whereas for an incompressible fluid, it is infinite.

Poisson's ratio ν is often used to characterize the elastic properties of a material. When a stretching force is applied lengthwise to a solid, the solid starts to stretch. During this stretching, in the vast majority of cases, the cross-section of the material *decreases*. Poisson's ratio shows how the cross-section of a deformable body changes under lengthwise stretching (or compression). Its value is the ratio of the linear contraction of cross-section e' to the elongation e , that is, $\nu = |e'|/e$. In case of an entirely brittle material, the Poisson ratio is zero, whereas for a completely elastic material, $\nu = 0.7$. For example, most steels have $\nu \sim 0.3$; for germanium, $\nu = 0.31$; for quartz glass, Poisson's ratio is small ($\nu = 0.17$), whereas for rubber, Poisson's ratio is large: $\nu \sim 0.6$ (ν is measured in relative units: mm/mm, cm/cm, etc.).

Note. There are materials (polymers) for which Poisson's ratio is negative; these materials are called *auxetics*. In these materials, upon application of a stretching force, the transverse section of the body increases.

The elastic properties of crystals can be considered not only in macroscopic approximation but also in a framework of microscopic theory that takes into account the atomic structure of the crystal lattice and interatomic interactions. In this approximation, it is usually assumed that interactive forces between atoms are central, that is, they *operate along lines* connecting the centers of atoms. Then it is possible to obtain additional relationships between the elasticity coefficients c_{mn} :

$$c_{23} = c_{44}; c_{13} = c_{55}; c_{12} = c_{66}; c_{14} = c_{56}; c_{25} = c_{46}; c_{36} = c_{47}.$$

These ratios that *reduce to six* independent components of elastic stiffness are the **Cauchy relations**.

The **shear modulus**, or modulus of rigidity (abbreviated as G or μ), characterizes the stressed state in case of net shear, that is, the ability of a material to resist any change in its shape while maintaining its volume. Shear modulus is expressed by the ratio of shear stress to shear strain that is defined as the alteration in the right angle between planes, whereon shear stresses are applied to two mutually orthogonal sites.

Young's modulus (E) or modulus of longitudinal elasticity describes a material's resistance to stretching or compression during elastic deformation. The modulus of elasticity is a set of physical quantities that characterize the ability of any solid body to be elastically deformed under conditions where force is applied to it. In simple cases, Young's modulus is defined as the ratio of stress to elongation. In cubic crystals, E modulus equates to three diagonal components of elastic stiffness that are identical: $E = c_{11} = c_{22} = c_{33}$. Young's modulus is measured in GPa—for example, for aluminum, $E = 70$ GPa; for iron, $E = 180$ GPa, but the largest Young's modulus is seen for graphene, where $E = 1000$ GPa.

In case of a homogeneous isotropic body, such as a fine-grained polycrystalline solid (ceramics) with random orientation of grains (i.e., in the absence of textures), both elastic modulus and Poisson's ratio are the same in all directions. Thus values of ν , E , G , and K are related by the following formula:

$$G = \frac{E}{2(1+\nu)}, \quad K = \frac{E}{3(1-2\nu)}.$$

Thus only two of these parameters are independent; therefore elastic properties of an isotropic body can be described by *only two* elastic constants. Those are the **Lame parameters**: μ and λ . They depend only on material properties and are very useful for elasticity research, when stresses are expressed in terms of strains. Lamé constants can be expressed in terms of different elastic moduli by the formula:

$$\mu = G, \quad \lambda = \frac{E\nu}{(1+\nu)(1-2\nu)} = K - \frac{2G}{3}.$$

Here, E is Young's modulus, K is the bulk modulus, ν is Poisson's ratio, and G is the shear modulus. The Lamé constants can be calculated from the experimentally determined elastic modulus.

Therefore homogeneous and isotropic solid materials can be characterized by linear elastic properties, which are fully described by two major elastic components that are any pair of moduli. If a pair of elastic moduli is known, all other moduli can be derived by calculation.

It should be noted that only in the *isotropic* elastic body can the number of independent elastic constants be reduced to two. However, many crystals, such as piezoelectrics, pyroelectrics, ferromagnetics, and ferrimagnetics, are anisotropic. In extreme cases, the number of elastic components of an anisotropic body can reach 21. In solids with some symmetric elements, the number of elasticity moduli reduces. For example, elastic properties of a monoclinic system can be determined by 13 elastic components; for crystals of a rhombic system, this number is nine, and so on.

2.4 ELASTIC WAVES IN CRYSTALS

In connection with the study of dynamic properties in solids, the concept of a *wave* as the space-time periodic process in crystal is considered. An alteration in time is described by an oscillator model, which has parameters of mass m and elastic force F_x dependent on strain: $F_x = -cx$:

$$m \frac{d^2x}{dt^2} = F_x.$$

Wave propagation along a linear chain of elastically coupled atoms (i.e., spatial characteristics of wave) is due to the elastic coupling of atoms. Plane wave propagation along a 1D crystal can be described by the equation

$$x = x_0 \exp[i(\omega t - kx)],$$

where $\omega = 2\pi/T$ is the angular frequency (T is oscillation period) and wave number is $k = 2\pi/\lambda$ (λ is wavelength). The natural frequency of the oscillator is $\omega_0 = \sqrt{c/m}$.

Similarly, it is possible to describe waves in the approximation of *elastic continuum* [4]. Consider the vibrations in the elementary volume, taken within a crystal in the form of a cube $\Delta x \Delta y \Delta z$ (Fig. 2.4). The mass of this cube is equal to the product of its volume and density: $m = \rho \Delta V = \rho \Delta x \Delta y \Delta z$. According to the oscillator model, the acceleration d^2x/dt^2 is determined by a second derivative of strain components: dx_1^2/dt^2 (for simplicity, oscillations along only one direction are considered—along the x -axis).

The elastic force F_x (component of force along the x -axis) can be calculated from a model that compares stress across two faces of a cube: $X_1(x)$ and $X_1(x + \Delta x)$. Their difference can be taken in a series, whereas in linear approximation, it is sufficient to take into account only the first term:

$$X_1(x + \Delta x) - X_1(x) = \frac{\partial X}{\partial x} \Delta x.$$

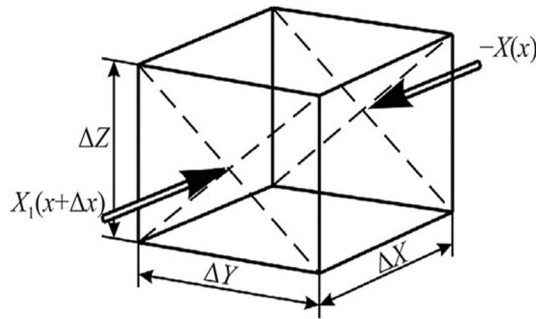


FIG. 2.4

Explanation of elastic wave dynamics in cubic crystal.

	x_1	x_2	x_3	x_4	x_5	x_6
X_1	c_{11}	c_{12}	c_{12}	0	0	0
X_2	c_{12}	c_{11}	c_{12}	0	0	0
X_3	c_{12}	c_{12}	c_{11}	0	0	0
X_4	0	0	0	c_{44}	0	0
X_5	0	0	0	0	c_{44}	0
X_6	0	0	0	0	0	c_{44}

FIG. 2.5

Elastic stiffness matrix in cubic crystal.

The resulting force is equivalent to the difference in stresses; substituting this result into the equation of the oscillator, we get

$$\rho \frac{d^2x}{dt^2} = \frac{\partial X_1}{\partial x} + \frac{\partial X_2}{\partial y} + \frac{\partial X_3}{\partial z}.$$

Similarly, other forces ($\partial X_2/\partial y$ and $\partial X_3/\partial z$) can be also considered in the elementary cube in the direction of displacement x_1 due to changes in the stresses X_2 and X_3 ; however, in Fig. 2.5, these components of force are not shown. Similar equations can be derived for waves of deformations x_2 and x_3 . Solutions of these equations depend on the specific symmetry of crystal or texture, because they are determined by a set of the matrix component c_{mn} (Table 2.1A).

$$F_x = \left[\frac{\partial X}{\partial x} \Delta x \right] \Delta y \Delta z.$$

In a relatively simple case (centrosymmetric cubic crystal), the propagation of a plane wave of deformations along the x -axis (i.e., along [100] direction) when the direction of elastic displacement coincides with wave vector k

$$x = x_0 \exp [i(\omega t - Kx)]$$

yields the following dispersion relation:

$$\omega^2 \rho = c_{11} K^2.$$

In contrast to a similar case of a *discrete* atomic chain, when dispersion law $\omega(k)$ is $\omega = 2\sqrt{c/m} \sin(ka/2)$, in the case of an elastic continuum, and when structure discontinuity is not taken into account, spatial dispersion is absent: the velocity of elastic waves is independent of frequency [4].

The velocity of longitudinal waves along the [100] direction in a cubic crystal depends on the density of the crystal and on one of the elastic stiffness components (Fig. 2.5):

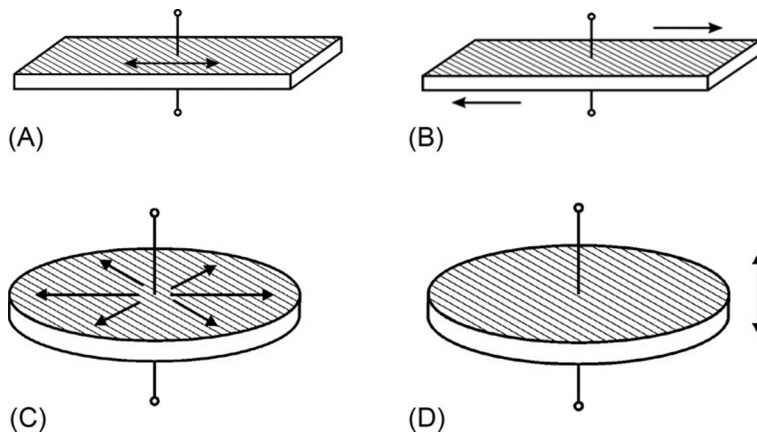
$$v_{LA[100]} = \omega/K = \sqrt{c_{11}/\rho}.$$

In cubic crystals, equally simple expressions for velocity of transverse elastic waves can be obtained, if the strain component x_2 (or x_3) is perpendicular to the direction of wave propagation:

$$v_{TA[100]} = \omega/K = \sqrt{c_{44}/\rho}.$$

The velocity of transverse waves in *cubic* crystals is the same for any orientation of elastic displacement. However, if the wave vector is directed along [110] or [111] axes, the solution of wave equations becomes more complicated. For crystals of low symmetry, including piezoelectrics, the velocity of elastic waves is determined by various combinations of the tensor c_{mn} components.

Thus, in the homogeneous elastic medium, two types of volumetric waves may exist: the *longitudinal wave*, in which particle displacement takes place in the direction of wave propagation, and the *transverse wave*, in which particles undergo

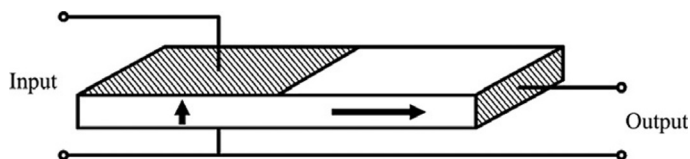
**FIG. 2.6**

Different types of oscillations in bulk piezoelectric elements (*shading* shows electrodes; *arrows* show direction of deformations): (A) transverse oscillation in a piezoelectric plate polarized in thickness; (B) shear oscillations in a piezoelectric plate polarized in thickness; (C) piezoelectric disk polarized in thickness with radial deformations; and (D) disk piezoelement polarized in thickness and having a thickness deformation.

displacement in a plane, perpendicular to the direction of wave propagation. Longitudinal and transverse waves are volumetric oscillations of the elastic medium.

Volumetric elastic waves are used in many electronic devices, in the mode of a traveling wave as well as in the standing wave mode (in the resonance devices), in which longitudinal and transverse waves are elastic medium oscillations [5]. For example, in Fig. 2.6, elements with standing elastic waves are shown that are used in *resonant* piezoelectronic devices; usually, piezoelectric elements are made of polarized ferroelectric ceramics.

Another example of volumetric wave application is the *piezoelectric transformer* with two pairs of electrodes: an exciter and a generator (Fig. 2.7). Using an inverse piezoelectric effect, the exciter part of the plate creates a mechanical deformation involving a resonant wave in the entire volume of the piezoelectric element. In the generating section of the piezoelectric transformer, the output voltage appears

**FIG. 2.7**

Simple layout of a piezoelectric transformer [6].

due to a direct piezoelectric effect; if it is an alternating signal, it is galvanically separated from the input voltage.

Electronics uses volumetric elastic waves that are usually excited piezoelectrically; however, sometimes, when they function by magnetostriction, they are called *piezoelectronics*. In Figs. 2.6 and 2.7, the piezoresonators and piezotransformers are the simplest examples of such devices [7]. However, the most striking example of piezoelectronics application is piezomotors.

The first ultrasonic piezoelectric motors were invented in the Igor Sikorsky Kiev Polytechnic Institute by V.V. Lavrinenko [7]. Thereafter, various piezoengines were

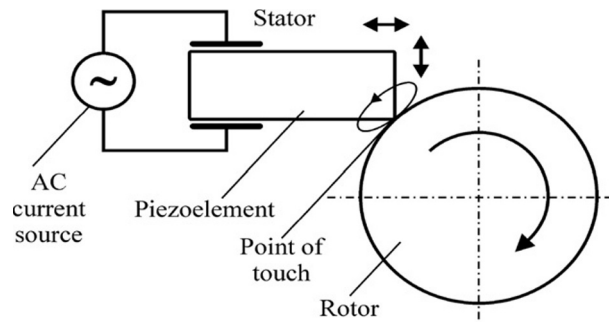


FIG. 2.8

Scheme of the first piezoceramic engine devised by V.V. Lavrinenko.

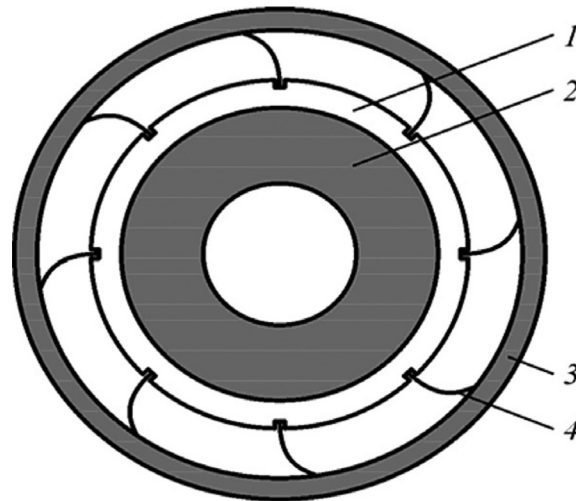


FIG. 2.9

Scheme of piezoelectric motor with oscillator stator: 1—thin steel jacket-bandage; 2—ring piezoelement, 3—rotor; and 4—pushers.

developed: nonreversible and reversible, with a piezoelectric passive rotor and an active stator, with a piezoelectric active rotor and passive stator, with electrical excitation of oscillations of one and two types, etc.

The scheme of the first piezoelectric motor is shown in Fig. 2.8. The drive of a piezoceramic motor is powered by alternating voltage at the resonant frequency of a piezoelectric cell, which is located in the stator of the device. In the interface between the stator and rotor (in case of direct contact), a strong tension arises. Piezoelectric vibrations generate elliptical motion of the stator surface, and the rotor moves due to the friction in the contact area.

Fig. 2.9 shows one of many options for a modern piezoelectric motor, which includes a ring piezoelectric element, embedded in a steel jacket that has pushers mounted inside the rotor. Such a piezoengine works as follows. When a piezoelectric cell is connected to an external excitation source, acoustic radial resonance oscillations arise within it. Due to the coordination of parameters, these oscillations are practically, without weakening, transmitted to the pusher, which then frictionally interacts with the rotor to turn it.

Piezoelectric elements are composed of piezoceramic, but the jacket is made of steel. Pushers are installed in the grooves of a jacket and secured with a compound epoxy resin. Because such a compound is a sound conductor, the presence of grooves can be neglected. The use of such components of oscillator stators only slightly changes the quality of the piezoelement.

Another important example of the application of mechanics in electronics is *microelectromechanical systems* (MEMS). Their technology combines both microelectronic and micromechanical components. MEMS devices are usually made of silicon substrate, similar to that used in integrated one-chip manufacturing technology. Typical dimensions of these micromechanical elements are in the range from 1 to 100 microns, whereas the MEMS chips have dimensions ranging from 20 microns to 1 mm. Miniature integral devices and systems that combine electrical and mechanical components are located on one crystal or substrate. Such a microsystem usually starts from the *sensor* (sensing element) in the input of a circuit, and, then, information enters the *amplifier* and analog-digital *converter*; next, the *microprocessor* follows (in the data-processing path), and the MEMS terminates in an *output device*. All of these stages are realized on one chip—through integral microtechnology.

Therefore microscopic mechanical devices include accelerometers, gyroscopes, and angular velocity sensors. Microactuators are used in medical applications to control instruments and biological objects at the microscopic level. The widespread use of MEMS in medicine is attributable to their application in the microactivation of surgical instruments.

In high-frequency electronics, various resonance elements of MEMS are used—in oscillators, filters, sensors, and so on. In addition to those fixed on one side of *consoles* (cantilevers), *bridges* (fixed on both sides) and *diaphragms* (fixed throughout the periphery) are applied (Fig. 2.10).

The most common element is the cantilever; an example of piezoelectric console implementation is shown in Fig. 2.11. The lower electrode is first applied at the

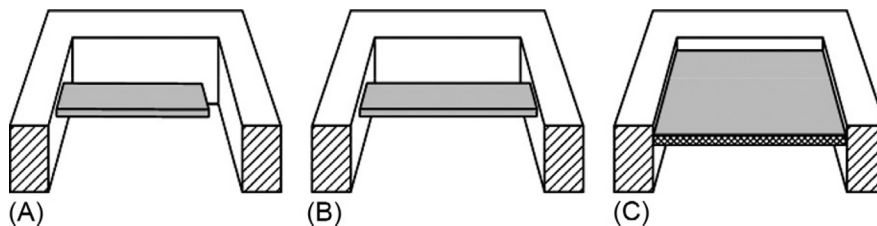


FIG. 2.10

Schemes of elastic piezoelements fastening in silicon MEMS structures obtained by league technology: (A) console (cantilever), (B) bridge, and (C) diaphragm; to simplify, electrodes on piezoelectric elements and their metallic elements are not shown.

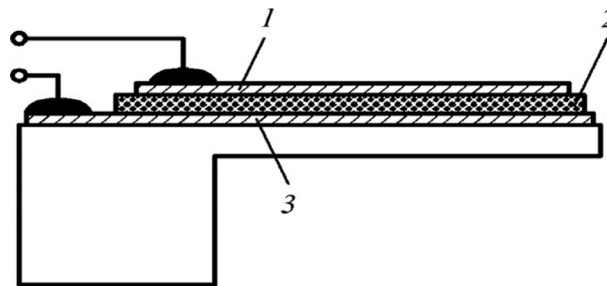


FIG. 2.11

Scheme of piezoconsole deposited on silicon: 1 and 3—electrodes, 2—piezoelectric layer.

silicon surface through a buffer layer of silicon oxide; this is followed by a piezoelectric layer, and, finally, the top electrode. By the method of anisotropic etching of silicon, material under the console is removed.

The first high-frequency MEMS devices were constructed mainly on the basis of thin films of ZnO as well as by using thin films of AlN (with a wurtzite structure). These piezosemiconductors provide low acoustic losses in microwave filters at frequencies of 2–10 GHz. Such devices are based on volumetric acoustic waves that propagate along the thickness of a film. The high acoustic quality and low dielectric permeability of wurtzite films are very attractive properties for microwave devices. At lower frequencies, it is advisable to use ceramic lead-zirconate titanate (PZT) films.

Therefore microelectromechanics produces extremely small and sensitive devices for detecting and measuring displacements, acceleration, pressure, weak electrical signals, ions, and specific biological agents that have utility in medical applications.

In addition to volumetric waves, SAW can be excited in an elastic medium. Electronics pertaining to these surface waves is called *acoustoelectronics*.

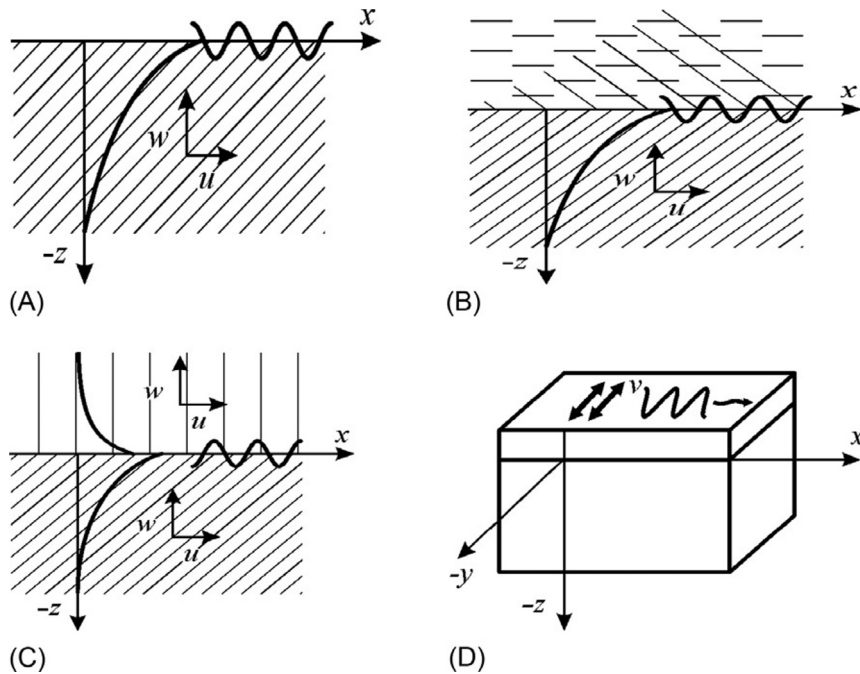


FIG. 2.12

Schematic representation of surface waves: (A) Rayleigh waves on a free rigid body; (B) Rayleigh waves on the interface of solids and liquids; (C) Stoneley wave at the interface between two solids; and (D) Love waves at the interface of the “solid half-space—solid layer.”

Surface waves can easily propagate along the free surface of a solid or along the border of a solid body with other media but have a rather fast damp away from the boundary (Fig. 2.12). Two types of surface waves are used: with *vertical* polarization, when the vector of particle displacement is located in a plane that is perpendicular to the side of a rigid body in contact with other media, and with *horizontal* polarization, when the vector of particle displacement is parallel to the boundary of a rigid body with other media, but remains perpendicular to the direction of wave propagation.

Particular cases where surface waves are used are as follows:

1. The *Rayleigh waves*, extending along the boundary of the elastic half-space (sound conductor) with vacuum or a sufficiently rarefied gas medium (Fig. 2.12A). The phase velocity of Rayleigh waves equals $v_R \approx 0.9v_T$, where v_T is the phase velocity of the shear mode. The velocity vector of these waves is parallel to the surface, whereas oscillating particles can have both transverse (perpendicular to surface) and longitudinal components of the displacement

vector. The vibrations of particles are described by elliptical trajectory in a plane perpendicular to the surface that passes through the direction of phase velocity.

The amplitudes of longitudinal and transverse vibrations are reduced exponentially with increase in distance from a surface into the medium, thus having different attenuation. This results in an effect where, during wave propagation, the ellipse is deformed and, when far from surface polarization, becomes linear. The penetration of a Rayleigh wave into the depth of sound conductor approximately equals the length of the surface wave.

2. The *damped waves of the Rayleigh type*, existing between solid and liquid, are shown in Fig. 2.12B. In a relaxed fluid, the elastic surface waves cannot exist. However, it should be noted that, at an ultrasonic frequency range in an actual liquid, surface waves may still exist that are defined not by elastic forces but by surface tension (so-called capillary waves). If a liquid is bordered with a solid and sound velocity in the liquid is less than the v_{sound} in the solid (this is true for almost all sound velocities in media), the damped wave of Rayleigh type may spread on the interface of the solid and the liquid.

A damped wave, during its propagation, continuously emits energy in the liquid, thereby forming nonuniform waves in it. The phase velocity of a damped Rayleigh wave is almost equal to v_{sound} , whereas its damping at one wavelength is approximately 0.1 such that, on a progression of 10 wavelengths, the wave is damped “e” times. In solids, the depth of stresses and displacements of such a wave is similar to the distribution in a Rayleigh wave.

3. The *nondamping waves (continuous)* with vertical polarization extend along the interface of a liquid and solid. The velocity of sound in a liquid is less than that in a solid, and, therefore, a nondamping wave in a solid is extended together with decaying. It spreads on the interface of the medium with a phase velocity that is less than the velocity of longitudinal and transversal waves. A continuous wave, being vertically polarized, has a completely different structure and velocity than a Rayleigh wave. This wave, in liquids, consists of a slightly inhomogeneous wave with an amplitude that slowly decreases with the distance from the boundary of medium, and, in solids, of two strongly inhomogeneous waves (longitudinal and transversal). Due to this aspect, the energy of the wave and particle motion is localized mainly in a liquid, but not in a solid, body. In practice, this type of wave is rarely used.
4. The *Stoneley waves* propagate along the plane boundary between two solid media, for which the elastic moduli and density are not very different (Fig. 2.12C). The Stoneley wave resembles two Rayleigh waves (one in each environment). The phase velocity of a transverse Stoneley wave is less than the velocity of a longitudinal wave in both the neighboring media. Vertical and horizontal displacements of components in each media are reduced with distance from border such that wave energy is concentrated in two near-boundary layers with thickness similar to the wavelength.
5. The *Love waves* are surface waves with horizontal polarization that can extend in the layered structure: “elastic layer on elastic solid half-space” (Fig. 2.12D).

This is a purely transverse wave, whereas its phase velocity is found in the range between the phase velocities of transverse waves in a layer and in a solid half-space. Love waves propagate with dispersion; because of the small thickness of the layer, the phase velocity approaches the sound velocity in the half-space. In general, the movement of a wave may be such that wave energy is redistributed between the layer and the solid half-space; therefore, phase velocity depends upon the frequency and thickness of the layer.

In *anisotropic crystals*, the same type of surface acoustic waves may exist as in isotropic solids: however, the motion of particles in elastic waves might be more complicated. For example, on some planes of anisotropic crystals that have piezoelectric properties, the Love wave as a Rayleigh wave may extend on the free surface; these waves are called as “*electrosonic*.” Along with the usual Rayleigh waves in peculiar orientations of crystals along the free boundary, a damped wave may extend such that it radiates energy into the crystal (pseudo-Rayleigh wave).

Finally, in a piezoelectric-semiconductor crystal, the surface wave can interact with conduction electrons, resulting in the *amplification* of this wave. In an anisotropic elastic structure, the properties of a Rayleigh wave depend on anisotropy and the direction of wave propagation. These waves can propagate not only in a plane but in the curved free surface of a solid as well. Thus their velocity is changed with depth, and the spectrum of permissible frequencies may become discrete, as in the case of a Rayleigh wave propagating on the surface of a sphere.

The principle of action of acoustoelectronic converters is based on elastic deformations that exist in the piezoelectric crystal due to the piezoelectric effect. Elastic deformation created by the comb electrode generates a traveling surface wave, accompanied by an alternating electric field. By varying converter options, it is possible to manage device parameters (Fig. 2.13).

Ultrasonic and hypersonic surface waves are widely used in technologies for comprehensive nondestructive surface layer testing, as well as for study of surface properties (defectoscopy). If the surface of a solid sample is free, then the usual Rayleigh waves can be applied. In cases where the tested sample is in contact with liquid or other solid samples or a solid layer, Rayleigh waves should be replaced by other suitable types of surface acoustic waves. Hypersonic Rayleigh waves with frequencies between 10^8 and 10^9 Hz are used in acoustoelectronic transducers to create signals, in ultrasonic and hypersonic delay lines, in amplifiers of electromagnetic

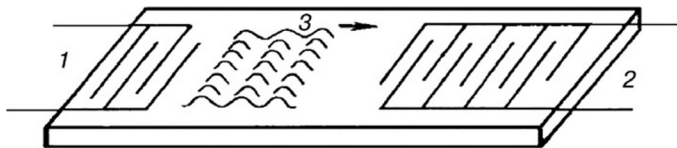


FIG. 2.13

Schematic diagram of the SAW filter; 1—comb activator; 2—output comb; and 3—schematic representation of surface waves [6].

oscillations, and for creating information processing systems. Furthermore, they are applied as sensor displays based on surface acoustic waves.

2.5 SUMMARY

1. Mechanical properties are conditioned by internal bonds among atoms, ions, and molecules of a solid; basically, these properties are *elasticity*, *strength*, *stiffness*, *toughness*, and so on. The most important mechanical properties of electronic devices are *elasticity* (depends on the strength of atomic bonds in crystals) and *velocity* of elastic waves in crystals (which, in addition to elasticity, have specific density impacts).
2. An external mechanical impact on a solid is characterized by the mechanical stress tensor X_{ij} . This is a symmetric ($X_{ij} = X_{ji}$) second-rank *field tensor* that, in its physical nature, is quite different from symmetric second-rank material tensors (e.g., the tensor of permittivity $\epsilon_{mn} = \epsilon_{nm}$, which agrees with the internal symmetry of crystal). The field-type tensor of stress describes the structure of forces applied to a studied sample *from outside*.
3. In different cases of technical application of solids in electronics, five important cases of mechanical stress tensors should be distinguished: linear-stressed state (uniaxial stress), flatness-stressed state (biaxial stress), volumetric-stressed state (three-dimensional stress), and the net shear stress. A separate important case is that of hydrostatic pressure, wherein all components of the stress tensor are same: $X_{11} = X_{22} = X_{33} = -p$, where p is pressure.
4. Depending on the symmetry of the mechanical load and of the crystal, there arises elastic deformation (*strain*) that is also a symmetric second-rank tensor ($x_{kl} = x_{lk}$). It can be classified into one-, two-, and three-dimensional tensors. Two-dimensional stress and strain (stretching/compression) are considered in contemporary planar microelectronic technology.
5. From Hooke's law, which asserts the linear proportionality of strain to stress, two very important tensors for solids follow: *elastic stiffness* tensor c_{ijkl} (also known as Young's modulus) and, inverse to it, the *elastic compliance* tensor s_{ijkl} . Both of them are *material* symmetric tensors of the fourth rank. Special parameters important for applications are *compressibility* $\langle s \rangle$ and the *bulk elastic modulus* K that can be determined by components of the elastic stiffness (or compliance) tensor.
6. The propagation of one-dimensional, surface, and volumetric elastic waves in solids, as well as the resonant properties of solid rods, beams, membranes, and so on, are described by tensors of elastic stiffness. Through excitation in an electrical or magnetic manner, elastic waves are widely used in modern piezoelectronics, acoustoelectronics, acousto-optics, MEMS, and microwaves as well as in many other microelectronic devices.

REFERENCES

- [1] J.F. Nye, *Physical Properties of Crystals*, Oxford Press, Bristol, 1957.
- [2] R.E. Newnham, *Properties of Materials: Anisotropy, Symmetry, Structure*, Oxford University Press, Oxford, 2004.
- [3] Y.M. Poplavko, *Physics of Active Dielectrics*, Lambert Academic Publishing, Saarbrücken, 2015.
- [4] C. Kittel, *Introduction to Solid State Physics*, fifth ed., John Wiley, New York, 1976.
- [5] K. Uchino, *Ferroelectric Devices*, Marcel Dekker, New York, 2000.
- [6] I.S. Rez, Y.M. Poplavko, *Dielectrics: Main Properties and Electronic Applications*, Radio i svyaz, Moscow, 1989.
- [7] V.V. Lavrinenko, *Piezoelectric Motors*, Lambert Academic Publishing, Saarbrücken, 2016.

Thermal properties of solids

3

CONTENTS

3.1 Basic Thermal and Energy Relationships	96
3.2 Thermal Expansion of Solids	98
3.3 Crystal Heat Capacity	104
3.4 Thermal Conductivity of Solids	112
3.5 Summary	118
References	120

Many phenomena and effects of electronic materials could be attributed to their thermal properties—thermal motion determines important features in electrical, magnetic, and other properties. Beginning with the synthesis of crystals and microelectronic as well as nanosized structures and, next, their alloying, annealing, and quenching, there are many other technological operations that occur under special thermal conditions. Thermal energy determines many properties of crystals. For example, the generation and recombination of charge carriers as well as the setting of their equilibrium concentration in *semiconductors* are due to thermal motion in the lattice. In *magnets* and *dielectrics*, phase transitions of dielectric-metal as well as transitions in ferromagnetic or ferroelectric states and observable phenomena such as pyroelectricity, electrocaloric effect, magnetic cooling, thermostriction, and so on are directly related to thermal properties. However, all listed phenomena will be considered in subsequent sections of this book, whereas this chapter is devoted only to three thermal phenomena in solids: *specific heat*, *thermal expansion*, and *thermal conductivity*.

Thermal properties caused by the *internal energy of movement* of molecules, atoms, or electrons are strongly dependent on the internal structure of material: the more stable the bonds between atoms, the greater the energy that must be expended for the displacement of atoms. In other words, more stable interatomic bonds require greater energy for their formation. The stabilization of any physical state in a given system occurs by its tendency to reach minimal energy. A consequence of this law is that electrons occupy orbits with the lowest energies, except in cases when they receive additional energy of excitation from external sources. However, over time, these excited electrons tend to return to more stable unoccupied orbits with lower energy; this happens because of thermal motion in crystals and is described by thermodynamics.

It is worth revisiting these basic concepts of thermodynamics, necessary for describing the thermal properties of solids.

3.1 BASIC THERMAL AND ENERGY RELATIONSHIPS

Potential energy is a part of the energy of a system that depends on the positions of particles and on external force fields [1]. In solids, the source of potential energy is Coulomb forces that cause attraction of opposite charges and repulsion of same-sign charges. *Kinetic energy* (energy of motion) similarly plays an important role for the description of properties of substances. For example, gas pressure is due to the kinetic energy of atoms or molecules. In solids, atoms are not absolutely fixed in a lattice, but continually oscillate as a result of thermal excitation [2]. Such movement significantly affects the basic properties of solids, as discussed in the following sections.

The state of a system is characterized by a thermodynamic function called the *enthalpy* A (i.e., the heat content in a system). As the temperature increases from T_1 to T_2 , enthalpy changes:

$$A_2 = A_1 + \int_{T_1}^{T_2} C_P \partial T,$$

where A_1 is enthalpy at initial temperature T_1 while A_2 is enthalpy at the final temperature T_2 ; and C_P is the *specific heat* under constant pressure P [1].

As a thermodynamic function, enthalpy can be defined in two ways. The first method is based on the determination of the *internal energy*, U , and the *work*, PV , performed by the material:

$$A = U + PV,$$

where P is the pressure and V is the volume of material.

The second method is based on the Helmholtz conception about *free energy* F (or Gibbs free energy G), and on the parameter TS , which is the energy conditioned by *internal disordering* in matter:

$$A = F + TS + PV = G + TS,$$

where T is the absolute temperature and S is the *entropy* of the material. Thus entropy is the measure of a system's internal disorder (chaos). Typically, thermodynamic quantities are given in well-known tables, together with values of entropy, enthalpy, and free energy.

The function $F = U - TS$ (i.e., Helmholtz free energy) is the minimal energy of the equilibrium state of a system [1]. When considering processes in solids, it is more convenient to control the volume of a system (not a pressure), and therefore it is necessary to use another thermodynamic function: Gibbs free energy $G = F + PV$. Here, the minimum value of G characterizes the equilibrium of the system at *constant*

volume and constant temperature. In solids, at atmospheric pressure, the condition of system equilibrium can be assessed using minimum F .

In Fig. 3.1, the enthalpy permanently increases with the temperature, but the contribution of the entropy TS increases *more rapidly*; therefore Gibbs free energy *decreases* with increase in temperature [1]. Because a concept of free energy is widely used in subsequent discussions in the context of properties and stability of solids, it is necessary to draw a few conclusions from Fig. 3.1:

- At zero absolute temperature, free energy equals enthalpy: $A = G$;
- Free energy used to characterize processes of structural change in matter decreases with increasing temperature; and
- The rate of the free energy decrease with temperature is related to entropy.

Furthermore, because the *entropy is always positive and obligatorily increases with a rise in temperature*, the slope of the free energy curve continuously increases with temperature. Thus the value of free energy provides important information about changes in the given phase; therefore the lower the free energy, the more stable the given phase.

In connection with the thermodynamic description and applications of solids in electronics (e.g., in case of active dielectrics or ferromagnetics), some basic concepts need to be elucidated [1]:

Heat is the energy of thermal motion of particles that form a body; in a Gaussian system, it is measured in calories (cal) and, in SI, in joules (J).

The *absolute temperature* is a thermodynamic quantity that characterizes the state of a body at its thermodynamic equilibrium; absolute temperature is denoted by T and measured in degrees of Kelvin (K). The average energy of particles in a body is proportional to the absolute temperature.

The *heat capacity*, denoted as C and measured in (J/deg) or in [cal/(deg mol)], is the heat absorbed from external sources when the temperature increases. In active

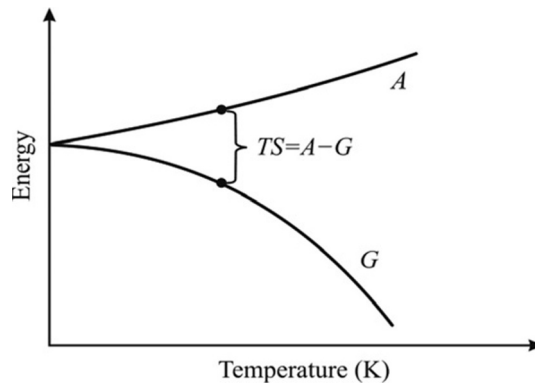


FIG. 3.1

Temperature dependence of enthalpy and free Gibbs energy G .

dielectrics and ordered magnetics, the heat capacity is dependent on the mechanical and electrical boundary conditions of a crystal.

The coefficient of *thermal conductivity*, denoted as λ and measured in [W/(deg m)] or [cal/(deg s cm)], is a characteristic property of a heat-conducting material; numerically, it is equal to the amount of heat passing through a unit area per unit time at a unit temperature gradient.

The coefficient of *thermal expansion*, denoted as α and measured in unit [deg⁻¹] = [K⁻¹], represents the alterations in a solid body's relative dimensions when the temperature changes by 1 K.

The next section presents some examples of the application of thermodynamics in solid-state physics. The focus is on three thermal properties of solids: thermal expansion, heat capacity, and thermal conductivity. These are properties that have the greatest practical importance.

3.2 THERMAL EXPANSION OF SOLIDS

Changes in the dimensions and volume of a crystal with a temperature variation are a result of the *asymmetry* in the interaction of its particles in a crystal lattice. Quantitatively, the degree of a change in the volume is characterized by the *volumetric coefficient* of thermal expansion, α_V . According to general definition, this coefficient is the relative change of volume V in a body on heating by 1° of temperature at constant pressure P , and it can be written as:

$$\alpha_V = (1/V) (\partial V / \partial T)_P.$$

Very often, thermal expansion in crystals is anisotropic and, sometimes, it is negative [2]. This means that when the temperature increases, a crystal can expand differently in various crystallographic directions; moreover, in some directions, the crystal may even be compressed with an increase in the temperature. Therefore, besides the volumetric expansion, the *linear expansion* coefficient α_l is widely used:

$$\alpha_l = (1/l) (\partial l / \partial T)_P,$$

where l is the linear dimension of the tested sample. The coefficient of thermal expansion is a second-rank tensor (i.e., matrix); thus the sum of three diagonal elements is approximately equal to the volumetric expansion coefficient: $\alpha_V \approx \alpha_1 + \alpha_2 + \alpha_3$.

In electronics and microelectronics, knowledge of the thermal expansion coefficient of materials is very important. For the reliability of microelectronic structures wherein semiconductor, dielectric, and metallic layers are integrated into a single monolithic structure, coordination between these components during thermal expansion is apparent. Usually, complex structures are synthesized at rather high temperatures, but they are used at normal conditions. If the thermal expansion coefficients are not matched, the structures obtained would be mechanically stressed and that would affect their properties and even possibly lead to localized destruction.

Nevertheless, there are states (even applications) when the difference in extension allows the properties of these structures to be managed; for example, the temperature of the phase transition in ferroics (crystals with magnetically or/and electrically spontaneously ordered structures) is changed purposefully.

The temperature dependence of the thermal expansion coefficient is shown in Fig. 3.2A, for the most important semiconductors—germanium and silicon (similar dependences are observed in the majority of solids). Parameter α increases significantly in the temperature range of 50–400 K, but thereafter varies very little (if structural phase transitions are absent).

In solids, the coefficient of thermal expansion actually characterizes the internal bonds of atoms, ions, or molecules, in particular, the energy of these bonds. This energy is largely determined by fundamental parameters of a crystal, such as its melting point. In Fig. 3.2B, the expansion coefficient is compared with the bond strength between ions. The inverse proportionality of this relationship corresponds to the nature of thermal expansion. Furthermore, it is noteworthy that the smaller the coefficient α is, the higher is the melting point of the crystal, T_m . There even exists an empirical formula: $\alpha T_m = \text{const}$ [2]; the parameters α and T_m included in this formula are two important thermal properties of crystals, and both can be expressed in terms of the *Debye temperature* (θ_D) of a crystal—this parameter is discussed further in connection with the dynamic properties of the crystal lattice.

Simple model explaining thermal expansion. The change in the volume or shape of a solid body with temperature alteration is attributable to the different nature of forces acting between its atoms. The interaction consists of attractive and repulsive forces. When the distance between the interacting particles changes, these forces vary in different ways.

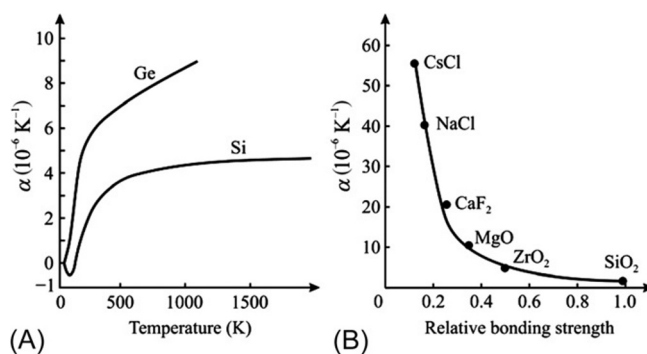


FIG. 3.2

Thermal expansion in crystals: (A) typical temperature dependence; (B) correlation α with strength of interionic bonding [2]. It is remarkable that, in silicon at cryogenic temperatures, the $\alpha(T)$ dependence passes through a negative value of thermal expansion; this peculiarity suggests that interatomic bonds in silicon have a rather complicated structure.

Repulsive forces have a very short range, as their potential energy decreases with the distance r between particles, designated as r^{-9} . The electronic shells of neighboring atoms or ions can only slightly penetrate each other. Conversely, the forces of attraction are long range, as their energy changes with distance, designated as $r^{-1} \dots r^{-6}$, depending on the nature of attraction (i.e., on the type of bonds; ionic, covalent, or molecular). Therefore the total energy versus distance $U(r)$ is characterized by an *asymmetric* minimum (Fig. 3.3A).

To describe the main reason for thermal expansion in a solid body, it is sufficient to consider the simplest diatomic model [3]. In some cases, this simple model is a rather good approximation that is not only qualitative, but also quantitative. The interaction of two atoms in equilibrium can be described by the balancing of the forces of attraction and repulsion (when total energy is minimum; Fig. 3.3A).

Let us suppose that one atom is fixed, being located at the origin. The increase of temperature induces movement that displaces a second atom from the ground equilibrium position under condition of fluctuations. If the temperature is low, thermal oscillations of the particle have a small amplitude x , and this motion can be modeled by a simple *linear* relationship (quasielastic interaction): $f = -cx$, where f is a spring-type force that returns the particle from excited state to its ground equilibrium position (when $x = 0$), while c is the coefficient of elasticity (this equation corresponds to Hooke's law).

Thus in $U(r)$, the dependence potential well (which actually is asymmetric) may be presented for simplicity by a symmetric curve. This means that thermal oscillations are *harmonic* ($x = x_0 \cos \omega t$), and the potential energy is described by a parabolic potential well:

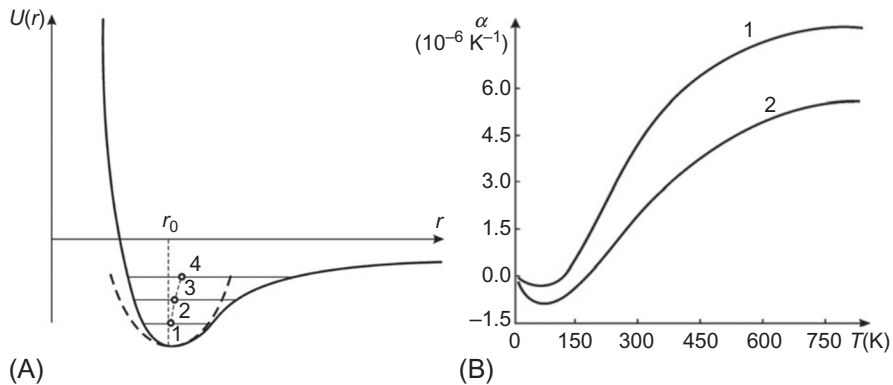


FIG. 3.3

Thermal expansion in solids: (A) dependence of potential energy on distance between atoms; *dashed curve* shows $U(r)$ parabolic approximation; points 1–4 show a thermal expansion curve with a nonparabolic character of real $U(r)$ curve; (B) thermal expansion temperature dependence for zinc oxide crystals: 1—lateral coefficient $\alpha_{\perp} = \alpha_{12}$; 2—longitudinal coefficient $\alpha_{\parallel} = \alpha_{33}$; the negative low-temperature component is due to the internal polarity of zinc oxide.

$$U(x) = \int cx \cdot dx = \frac{1}{2}cx^2.$$

Here, the dependence of energy has a shape, depicted in Fig. 3.3A by a dashed line, and the average position of the oscillating atom *does not depend* on the amplitude of its oscillations, $r=r_0$ (in a crystal, $r_0=a$, where a is the lattice constant). The lower part of the curve of any potential well (close to its minimum) can be approximated with good accuracy by a parabola.

It is obvious that, in case of a *symmetric* $U(r)$, any chaotic thermal oscillations of atoms cannot change the average distance between them; therefore the size of the crystal will be independent of temperature. This explains why the coefficient of thermal expansion in different solids at very low temperature *tends toward zero*.

In reality, however, the energy of the interaction between atoms is characterized by a *pronounced asymmetric well* (solid line in Fig. 3.3A), which is a result of two different summation curves (one is due to attraction and the other corresponds to the repulsion of atoms). With increasing amplitudes of thermal oscillations, the repulsive forces between atoms increase to become much stronger than the force of attraction, that is, the displacement of oscillating atoms to the left becomes much less than their displacement to the right (Fig. 3.3A). Thus the actual force f that acts on an atom becomes a nonlinear function of displacement x (Hooke's law is not met).

Consider the oscillation of one atom relative to another for the given energy in a classic approximation. Different energy levels are depicted in Fig. 3.3A by the horizontal lines 2, 3, and 4. In the position of equilibrium ($r=r_0$), the potential energy of the atom is zero although its kinetic energy reaches the maximum. Moving away from the equilibrium position, the atom acquires potential energy whose peak corresponds to the maximal shift of the atom from its equilibrium position and reaches the level of potential energy, shown by corresponding horizontal line (Fig. 3.3).

With the increasing total energy of the atom (1–4 in Fig. 3.3A), the amplitude of oscillations increases whereby the right shift of the atom will be greater than its left shift. As a result, *middle equilibrium position of the atom shifts* to the right, and this effect becomes stronger with a higher energy of oscillation of the atom. Therefore an increase in energy with rise in temperature leads to a phenomenon where the interatomic distance increases and the crystal enlarges.

In case of small oscillations of atoms around their equilibrium position, the potential energy can be expanded in a Taylor series in terms of atomic displacement with respect to the equilibrium position. To analyze thermal expansion, it is sufficient to limit this expansion by its first two terms:

$$U = \frac{1}{2}cx^2 - \frac{1}{3}bx^3.$$

Coefficient c is ratio of quasielastic bonding, whereas coefficient b is referred to as the coefficient of *anharmonicity*. Accordingly, the force that acts between an oscillating atom and a fixed atom is given by:

$$f = -\frac{\partial U}{\partial x} = -cx + bx^2.$$

In this equation, a nonlinear term “ $+bx^2$ ” is added to the linear term “ $-cx$.” This new term takes into account the asymmetry pertaining to interatomic interaction forces, and it is the *anharmonicity coefficient* b . The role of anharmonicity becomes more significant, the greater the value of displacement x . With this term, the time-dependent displacement of oscillating atom is no longer sinusoidal (i.e., not harmonic); therefore this approximation is called anharmonic. This simple model can explain the thermal expansion of solids [4].

The average potential energy of thermal fluctuations ($\frac{1}{2} cx^2$) at a given temperature equals $\frac{1}{2} k_B T$, where k_B is Boltzmann constant. The average shift of the atom in this model can be shown to be $x_{\text{aver}} = (b/c^2) k_B T$. Thus, in the diatomic model considered here, the thermal expansion coefficient α is defined as the ratio of the average shift x_{aver} to the equilibrium distance r_0 :

$$\alpha = \frac{k b}{r_0 c^2}.$$

It follows that, in the absence of anharmonicity (when $b=0$), the thermal expansion coefficient $\alpha=0$. The asymmetry of the resultant interatomic interaction force in the crystal lattice is considered further as the main cause of the interaction of phonons in a lattice.

The anisotropy of thermal expansion. In cubic crystals, to which the majority of metals and semiconductors belong (including germanium and silicon whose $\alpha(T)$ dependence is shown in Fig. 3.2), thermal expansion is *isotropic*. Therefore the linear coefficient of thermal expansion α_l is independent of the direction in a crystal and equals $\alpha_l = 1/3 \alpha_V$: this is true for metals and most semiconductors. However, many dielectric crystals, especially those that are important in electronic applications (i.e., pyroelectrics, piezoelectrics, etc.) as well as many ferromagnetics, are anisotropic crystals [2].

For instance, hexagonal zinc oxide (ZnO) belongs to the class of anisotropic crystals (Fig. 3.3B). At a very low temperature, the expansion coefficient of ZnO is reduced to zero. However, after temperatures rise, the internal polar bond emerges, and components of expansion coefficients in this pyroelectric (and piezoelectric) structure at low temperature initially become negative. Only with a further temperature increase does the thermal expansion coefficient increase in accordance with the usual cubic power law ($\alpha^{**}_{\perp} \sim \alpha_{\parallel} \sim T^3$) until it reaches saturation, at higher temperatures.

As shown in Fig. 3.3B, the lateral and longitudinal components of thermal expansion are different. Thus the anisotropy of crystal structure leads to the anisotropy of many physical properties, including thermal expansion. Therefore the thermal expansion coefficient is defined not as a scalar value, but as the *second-rank tensor*. Indeed, during *uniform heating*, a crystal is subjected to deformation, which can be described by the strain tensor x_{kl} . This change of temperature is described by the *scalar* value δT , and components of strain tensor are proportional to δT :

$$x_{kl} = \alpha_{kl} \delta T,$$

where α_{kl} denotes components of the thermal expansion coefficient. Therefore α_{kl} is a symmetrical tensor of the second rank because x_{kl} is a symmetrical tensor of the second rank. This equation would be simplified if the strain tensor x_{kl} is reduced to the principal axes of a crystal: x_1 , x_2 , and x_3 ; moreover, these components can be easily determined experimentally. The result is

$$x_1 = \alpha_1 \delta T, \quad x_2 = \alpha_2 \delta T, \quad x_3 = \alpha_3 \delta T,$$

where α_1 , α_2 , and α_3 are the main thermal expansion coefficients that correspond to the diagonal components of tensor x_{kl} [5].

If all thermal expansion coefficients are positive, then a *second-order surface* can portray thermal expansion by the quantities $(1 + \alpha_1 \delta T)$, $(1 + \alpha_2 \delta T)$, and $(1 + \alpha_3 \delta T)$. Moreover, the volumetric coefficient of the thermal expansion of crystal will be equal to the sum of all three linear coefficients $\alpha_V = \alpha_1 + \alpha_2 + \alpha_3$. In some directions of an anisotropic crystal, *compression*—not expansion—in linear dimensions can be seen when heating and, correspondingly, negative coefficients of expansion for these directions are observed. To obtain a full description of the thermal expansion tensor, it is necessary to know the linear coefficients of thermal expansion along the three principal directions of a crystal.

In crystals that belong to a *cubic* symmetry class, coefficient α is the same in any direction, because the second-rank material tensor in this case degenerates into a scalar: $\alpha_1 = \alpha_2 = \alpha_3$. The temperature dependence of α in Fig. 3.2 for semiconductors Ge and Si characterizes the typical case of cubic crystals.

In crystals of *hexagonal* and *trigonal* systems, the expansion coefficient tensor is determined in *two* main directions: parallel and perpendicular to the axis of the sixth (or third) order, whereby $\alpha_{11} = \alpha_{22} = \alpha_{\perp}$ and $\alpha_{33} = \alpha_{\parallel}$. In crystals of the *orthorhombic* system, it is necessary to know the expansion coefficients in three mutually perpendicular directions, parallel to the second-order axes: $\alpha_{11} = \alpha_1$, $\alpha_{22} = \alpha_2$, and $\alpha_{33} = \alpha_3$. The definition of α_{ij} tensor in crystals of lower symmetries (*monoclinic* and *triclinic*) becomes more complicated by the fact that the position of principal axes is not uniquely determined in a crystallographic coordinate system.

Components of thermal expansion coefficient usually might have different temperature dependences: they may be either positive or negative. Their sign depends on the anisotropy of forces that act between atoms in the crystal. The negative coefficient of thermal expansion is a result of long-range bonds in a crystal. It will be shown that such bonds arise during polarization of atoms, and they are the result of decrease in the frequency of acoustic modes in the phonon spectrum near the boundary of the Brillouin zone. Corresponding to these frequencies, the components of an elastic tensor have small positive values for longitudinal oscillations and negative values for lateral oscillations.

Furthermore, this situation is possible for layered and chain structures that are characterized by such interaction between atoms wherein the interaction inside the layer (or chain) is stronger than the interaction of atoms located in different layers

(or chains). In this regard, the thermal expansion coefficient along the chain (or layer) is always less than the coefficient of expansion in the perpendicular direction.

In general—with possibly different signs of the components of thermal expansion tensor—the characteristic surface in Cartesian coordinates is not ellipsoid, nor is it even a surface of the second order. Nevertheless, the knowledge of characteristic surfaces is important for applications of anisotropic crystals. For example, in calcite crystals, the expansion coefficient in the direction of the principal axis of the crystal is positive, but it is negative in directions perpendicular to it. This means that, in some oblique directions, the expansion coefficient should be zero, and therefore, in certain directions, the radius vectors of the indicatory surface should be zero (this case is impossible for an ellipsoid).

Shubnikov [6] considered *all possible forms* of the indicatory surfaces of thermal expansion coefficients in crystals under conditions when linear expansion coefficients α_1 , α_2 , and α_3 differ both in magnitude and in sign (Fig. 3.4). Positive values of α are shown on these figures by *white surfaces* whereas negative values of α are shown by *black surfaces*. As noted earlier, in crystals of cubic symmetry, all three major expansion coefficients are equal, and all three are usually positive. The corresponding surface in this case is obviously the sphere with a positive radius: this is a “white-colored sphere,” but this simplified case is not shown in Fig. 3.4.

When $\alpha_3 \neq \alpha_1 = \alpha_2$ with $\alpha_3 > 0$, the surface describing expansion coefficients is similar to oval; it can be either flattened (at $\alpha_3 < \alpha_1$, Fig. 3.4B) or elongated along axis 3. These surfaces describe simple cases of thermal expansion of optically *uniaxial crystals* that are often found in practice. In the calcite crystal, for example, component α_3 has a positive sign, while components $\alpha_1 = \alpha_2$ have negative signs. The surface that corresponds to such a case is also shown in Fig. 3.4B. It is composed of two egg-shaped positive (white) surfaces and a torus-like negative (black) surface.

Other characteristic surfaces of the thermal expansion tensor shown in Fig. 3.4 exhaust all possible combinations of main components of the α_{kl} tensor.

3.3 CRYSTAL HEAT CAPACITY

The heat capacity of body is a physical quantity defined as a ratio of the amount of heat dQ obtained by the body corresponding to an increase in its temperature dT :

$$C = dQ/dT.$$

The unit of heat capacity in SI is [J/K]. The concept of heat capacity is applicable to substances that are in various states of aggregation (solid, liquid, or gas) as well as to ensembles of particles and even quasiparticles (e.g., the heat capacity of electronic gas in metals or heat capacity of phonons in a crystal lattice). The value of heat capacity depends on the nature of a substance.

Specific heat is the heat capacity per given unit of substance, which can be measured in kilograms, cubic meters, and moles. Depending on the quantification of unit heat applied, the mass, volume, and molar value of specific heat are distinguished.

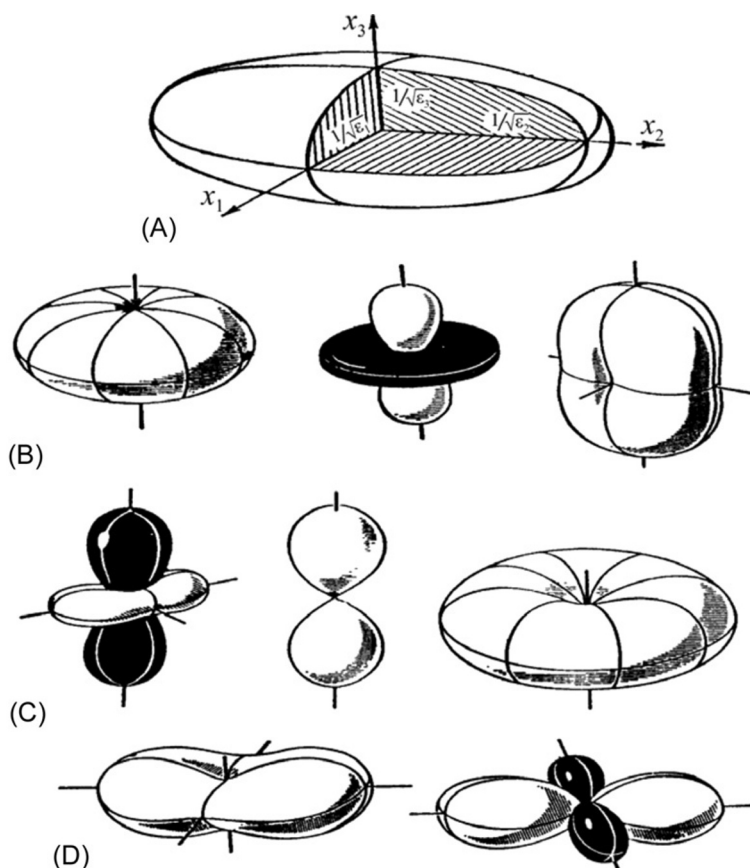


FIG. 3.4

Characteristic surfaces that exhibit anisotropy of physical quantities described by a second-rank material tensor; (A) ellipsoid of permittivity of biaxial crystal symmetry ϵ ; (B–D) figures that describe the thermal expansion coefficient in crystals of different symmetry; *black* shows the negative value of α .

According to A.V. Shubnikov, *Selected Works on Crystallography*, Nauka, Moscow, 1975, p. 551.

The *mass specific heat* is the amount of heat necessary to increase the temperature of a unit mass of material by one temperature unit; in SI, it is $[\text{J kg}^{-1} \text{K}^{-1}]$. The *volumetric specific heat*, C_V , is the amount of heat that is necessary to be applied to a unit volume of material to heat it by one temperature unit; in SI, it is measured in $[\text{J m}^{-3} \text{K}^{-1}]$, that is, joules per cubic meter and Kelvin. The *molar specific heat*, C_μ , is the amount of heat that is necessary for 1 mol of the substance to be heated by 1° ; in SI, it is $[\text{J}/(\text{mol K})]$, while in the Gaussian system, this specific heat is determined in $[\text{cal}/(\text{g}\cdot\text{mol K})]$. The vast majority of solids have specific heat close to $1 \text{ kJ}/(\text{kg K})$; for example, water has a relatively high heat capacity: $4.2 \text{ kJ}/(\text{kg K})$.

In solids, both the crystal lattice and electrons contribute to the specific heat. For reasons that are explained further on, the electronic specific heat in metals at normal conditions is rather small; therefore the mechanisms of the lattice for specific heat are mainly considered (Chapter 4 presents a discussion of the contribution of magnons to specific heat).

The elastic vibrations of atoms in a crystal can arise both in the form of traveling and standing waves. As a simple example of oscillations (phonons), one can consider sound waves that can be excited, for example, via an applied piezoelectric element. As with any waves, lattice vibrations are characterized by wavelength (λ) and frequency (ω). It should be noted that *arbitrary waves* cannot exist in a crystal; only those waves can exist that have a certain relationship between the frequency and wavelength: $\omega = \omega(k)$, where k is the wave vector given by $k = 2\pi/\lambda$. The dependence, $\omega(k)$, is the main characteristic of phonons, that is, atomic vibrations in a crystal. The knowledge of this relationship allows the calculation of many thermal and electrical properties of crystals (e.g., specific heat, thermal expansion coefficient, thermal conductivity, dielectric constant, etc.). However, it should be noted that the concept of phonons is only one of the possible models explaining the thermal properties of solids.

For a detailed consideration of thermal conductivity (as well as the thermal expansion study in crystals), it is necessary to take into account *anharmonicity*, that is, the nonlinearity of lattice vibrations. However, a further simplified explanation of crystal heat capacity that was proposed can be sufficiently limited by the *linear* (harmonic) model of phonons [7].

Historically, several theories of lattice specific heat were developed:

The law of heat capacity constancy (Dulong-Petit law) corresponds to conventional notions and, with some accuracy, are valid at *room temperature and higher temperatures*.

Einstein's quantum theory of heat capacity is the first successful attempt to use laws of quantum mechanics to describe the special properties of specific heat in solids at *low temperatures*.

Debye's quantum theory of heat capacity is based on a model of *constrained* atomic vibrations, and it shows a better association of theory to experimental data than Einstein's theory in the vicinity of low temperatures.

Born's theory of lattice dynamics is the most advanced method to describe crystal lattice dynamics, including the theory of heat capacity.

The law of specific heat constancy states that the *molar* heat capacity of different solids is the same (at room temperature and at increased temperatures):

$$C_{\text{solid}} = 3R$$

where R is the universal gas constant. An important factor is that the molar heat capacity in solids at ambient temperature is more than twice as high in comparison with the heat capacity of an ideal gas: $C_{\text{gaz}} = 3/2 R$ (Fig. 3.5).

It should be noted that 1 mol of *any* substance contains the same number of atoms as determined by Avogadro number: $N_A = 6.02 \cdot 10^{23} \text{ mol}^{-1}$. According to

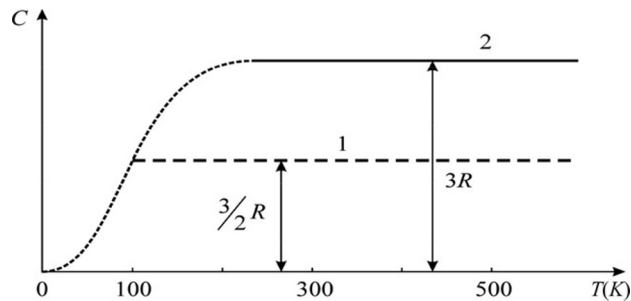


FIG. 3.5

Dependence of molar heat capacity on temperature: 1—ideal gas and 2—molar heat capacity of solids.

conventional statistics (i.e., statistical physics based on conventional mechanics), each degree of freedom of a gas particle makes the same contribution to its molar heat capacity. This rule is the *law of equipartition*. Any particle of monatomic gas has only three degrees of freedom; according to this, the molar heat capacity of gas should be equal to $3/2 R$, that is, approximately $13.5 \text{ J}/(\text{kmol K})$; in other units, this amounts to $3 \text{ cal}/(\text{mol K})$, which is in good agreement with experimental results.

This is because free atoms of gas exclusively have *kinetic energy*; each gas atom has three degrees of freedom, and the contribution to energy of each degree is $(k_B T)/2$ —that is, just $(3/2)k_B T$ in total. Because 1 mol contains N_A atoms, the molar heat capacity of gas equals $3/2 R$ (Fig. 3.5).

Boltzmann constant defines the relationship between temperature and energy: $k_B = 1.4 \cdot 10^{-23} \text{ J/K} = 8.6 \cdot 10^{-5} \text{ eV/K}$.

During its vibration, the atom, being *constrained* in a crystal lattice, possesses not only *kinetic energy*, but also *potential energy* equal to the kinetic energy, on average; that is, each atom in a lattice has twice as much energy in comparison with the same atom in a gas: $3k_B T$. Exactly because of this fact, the law of heat capacity constancy follows. Dulong-Petit law, in the dynamic formulation of a problem, is derived by the assumption that the crystal lattice consists of atoms, and each atom is a *harmonic oscillator in three dimensions* (due to lattice structure) whereby fluctuations in three orthogonal directions are independent. This means that each atom can be associated with a superposition of three oscillators with energy E that satisfies the following formula: $E = k_B T$.

This formula follows from the theorem of energy equipartitioning among degrees of freedom. Each oscillator has one degree of freedom, and therefore its average kinetic energy is equal to $k_B/2$ per temperature unit. Because oscillations are harmonic, the average potential energy is equal to the average kinetic energy, and the total energy is the sum of both. The number of oscillators in 1 mol of substance is $3N_A$, and their total energy per Kelvin equals the specific heat of the solid; from this reasoning itself, the law of constant heat conductivity follows directly. Thus the

classic (and simplest) idea as to thermal motion in a crystal lattice can be reduced to a model of *independent* oscillators.

The **oscillator model and elastic waves**. The dynamic behavior of elastic displacements of atoms (or ions) is described by a model of the harmonious oscillator (Fig. 3.6). In this model, a particle with mass m is elastically connected to a stationary base. In case of forced initial displacement of the particle from its equilibrium position on distance $+x$ (or $-x$), the opposite force occurs because of the elastic connection that seeks to return the particle into an equilibrium position. This force is proportional to displacement x and has the opposite direction: $f = -cx$.

Parameter c is the coefficient of elasticity; here, it describes atomic bonding in a crystal lattice. Upon elastic displacement, force f balances the force of inertia of mobile particles, $m(d^2x/dt^2)$:

$$m \frac{d^2x}{dt^2} = -cx.$$

Therefore the energy of the corresponding oscillator is $U = \int cx dx = \frac{1}{2}cx^2$. This expression is described by the parabolic potential well. The solution of this equation is harmonic oscillations: $x = x_0 \cos \omega_0 t$, or $x = x_0 \sin \omega_0 t$ or a linear combination of these two solutions. It is convenient to represent the general solution in the form: $x = x_0 \exp(i\omega_0 t + \varphi_0)$, where x_0 is the amplitude, φ_0 is the initial phase, and $\omega_0 = \sqrt{\frac{c}{m}}$ is the natural oscillation frequency. If the oscillating particle has an electrical charge q , then, in addition to elastic waves (mechanical), an electromagnetic wave arises too.

The elastic wave of particle vibration in the crystal lattice is the *phonon* (this term resembles *photon*). In phonons, the oscillatory motion of particles of a solid occurs. When it comes to electromagnetic oscillation (photon), the usual classical concepts are not suitable because, according to them, there exists something, which has no mass. The electromagnetic wave (in its simplest form, a plane wave of a certain frequency) is a peculiar form of the existence of matter—the electromagnetic field. An elementary wave is a wave that is infinitely extended in space and time [6].

Returning to the law of molar specific heat constancy in crystals that does not depend on the type of atoms (or ions) of the solid body and does not depend on temperature, it should be noted that even this relatively simple model of equal and independent oscillators is capable of explaining this feature.

However, a low-temperature investigation of specific heat demonstrates the fast decline of the $C_{\text{solid}}(T)$ characteristic (dotted line in Fig. 3.6). Moreover, when the

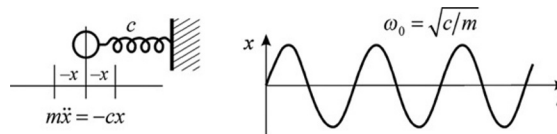


FIG. 3.6

Model oscillator and its wave.

lattice approaches absolute zero, the specific heat vanishes: $C_{\text{solid}} \rightarrow 0$. All this testifies to the shortcomings of the simple model of the classic oscillator.

The temperature dependence of specific heat in solids at low temperatures is explained in the *quantum models* of Einstein and Debye.

Einstein's quantum theory of specific heat. The main assumption of this theory is that the atom oscillator in a crystal lattice is the *quantum object*, not the conventional one. However, as in the previous model, these oscillators are again considered *independent*.

A quantum oscillator with frequency ν can absorb (or emit) energy only in portions—by quanta: $h\nu = \hbar\omega$. This is shown schematically in Fig. 3.7 (see left panel). At a relatively high temperature (T_3) when the thermal motion is rather intense, the average thermal energy of the oscillator ($k_B T_3$) is much greater than the quantum of oscillator energy ($k_B T_3 \gg \hbar\omega$); thus the fact that the oscillator is the *quantum* oscillator is not significant and therefore the classic Dulong-Petit law is satisfactory.

In case of low temperatures, the average energy of thermal motion becomes approximately the same as the energy of the quantum oscillator: $k_B T_1 \sim \hbar\omega$. Nevertheless, the energy distribution between lattice vibrations is chaotic, but when the crystal is cooled, the number of quantum oscillators (which do not accept or radiate energy) increases; therefore specific heat should be reduced with a decrease in the temperature.

This result was obtained by Einstein. His theory is based on the assumption that atoms in a crystal lattice behave as harmonic oscillators that do not interact with each other. The number of oscillators in 1 mol of substance is equal to $3N_A$ and their energy is quantized. According to the model proposed by Einstein, close to the absolute zero of temperature, specific heat tends to zero; however, at high temperatures, the law of Dulong-Petit holds true.

The temperature dependence of C_{lattice} in Einstein's model is described by the *exponential law* (Fig. 3.7, curve 1). However, subsequent experiments have shown that this dependence is described by a *cubic parabola*: $C \sim T^3$. Thus it is necessary to

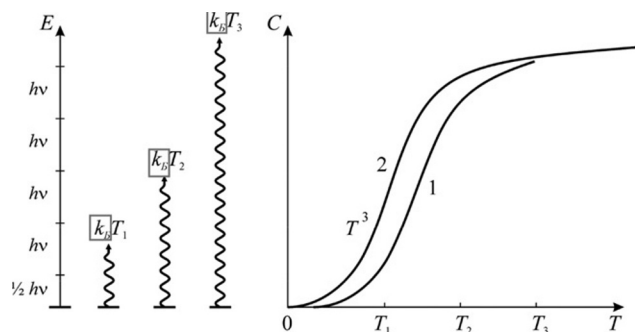


FIG. 3.7

Comparison of specific heat of the quantum model of independent oscillators (curve 1) and model of coupled oscillators (curve 2).

consider the *interaction* between adjacent atoms. Such calculations were made by Debye.

Debye's model of specific heat takes into account the contribution from the *lattice of interacting* atoms to heat capacity. This model correctly predicts the low-temperature specific heat proportionality to T^3 , and considers that atom oscillators in the crystal lattice are *elastically connected* to each other; therefore their vibrations are interdependent.

To explain the influence of interaction of atoms on the frequencies of their oscillations, two models are shown in Fig. 3.8A: the free and bound pendulum. In case of the free pendulum, the eigen frequency of oscillations ω is dependent only on the length of the pendulum—this model corresponds to the case of independent oscillators discussed earlier.

The constrained pendulums (Fig. 3.8A) can serve as a model to explain the simplest two-atom bond. In case of two resiliently connected pendulums, the oscillation process becomes more complicated, as each pendulum has the same eigen frequency ω , but there is also an additional *combinational* frequency Ω . If there are three pendulums, then such a system would have three characteristic frequencies. Obviously, for n pendulums (which mimic the crystal lattice of n atoms), the number of characteristic vibration frequencies will be $n + 1$.

To illustrate Debye's model, it is possible to consider the oscillations of a *string* with length l that is attached at the ends (Fig. 3.8B). The main tone has a frequency 0 that corresponds to a wavelength $\lambda = 2l$ of the elastic string. The overtones are $2\omega_0$, $3\omega_0$, ..., and they are located on the same line $\omega(k)$ with wavelengths l , $2/3l$, The dependence of oscillation frequencies ω on the reverse wavelength (wave vector) $k = 2\pi/\lambda$ is shown in Fig. 3.8C.

In Debye's model, the movement of the *center of masses* of an interconnected lattice with N elements is considered. It is assumed that this complex movement

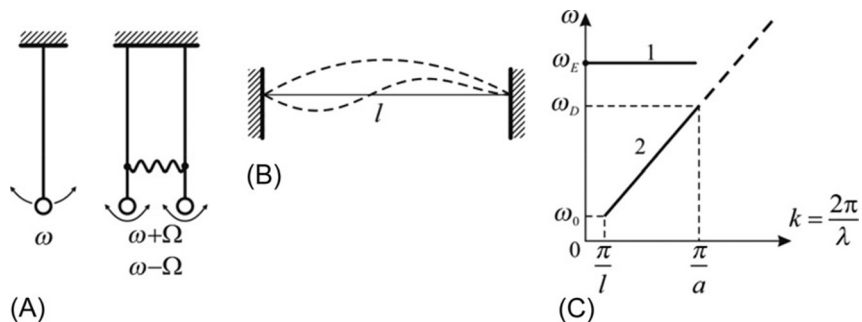


FIG. 3.8

Explanations related to the Debye model: (A) single and two connected pendulums; (B) string oscillations (primary tone and first overtone); (C) $\omega(k)$ dependence of the oscillator frequency of string on its length (*dotted line*); 1—Einstein's mode ω_E of free oscillators; 2—Debye's mode of bounded oscillators with maximal frequency ω_D .

(lattice vibrations) is equivalent to $3N$ harmonic oscillators. The coordinates of harmonic oscillators are the *normal coordinates*, and their fluctuations are termed the *normal modes*.

The internal energy and heat capacity of a solid consists of additive contributions of energy from individual normal vibrations. To derive a formula that describes the dependence of specific heat on temperature, it is necessary to know the frequency spectrum of normal vibrations. This spectrum can be calculated theoretically whereby, in case of the simplest lattice, this solution contains three *acoustic modes*, with $\omega(k)$ dependence that corresponds to three possible independent orientations of polarization vectors of waves (two transverse modes and one longitudinal mode).

The relationship $\omega(k)$ constitutes the *dispersion law*. In case of Einstein's model, the frequency ω_E does not depend on the wave vector k (see line 1 in Fig. 3.8C). In contrast, according to Debye's model, this relationship exists and is characterized by a *sloping line*—line 3. In Debye's model, the dependence $\omega(k)$ is *linear* (similarly as for string); however, there is one important *restriction*: this line ends at the abscissa value π/a . This means the limiting of wavelength ($\lambda = 2a$) because there is no physical carrier for shorter wavelengths.

At low temperatures, the energy of a crystal increases with the temperature due to two factors: firstly, due to an *increase in the average energy* $k_B T$ of normal vibrations (i.e., proportional to T) and, secondly, due to the *number of excited oscillations* that increases similarly as T^3 . Therefore, the total energy of a crystal increases with temperature proportionally to the fourth power of temperature:

$$E_{\text{lattice}} \sim T^4$$

Accordingly, the heat capacity of the lattice that is determined as a derivative ($C \sim dE_{\text{lattice}}/dT$) is proportional to the temperature in a cubic power:

$$C \sim T^3$$

which is in good agreement with experiments.

At high temperatures, all normal lattice vibrations are already excited and therefore a further temperature increase does not result in an increase in the number of phonons. Consequently, at relatively high temperatures, the growth of the energy in solids can only take place due to the increase in the degree of excitation of normal vibrations, which proportionally causes an increase of their average energy to temperature ($k_B T$); thus the energy increase in a solid must be proportional to T :

$$E_{\text{lattice}} \sim T$$

whereas the heat capacity of the lattice ($C \sim dE_{\text{lattice}}/dT$) should not depend on temperature:

$$C = \text{const.}$$

Thus at increased temperatures, the specific heat tends to a constant value $3R$ —according to Dulong-Petit law—where the characteristic temperature θ_D exists, the *Debye temperature*. Below θ_D , the quantum nature of lattice vibrations becomes

decisive. Thus the θ_D approximately indicates a temperature limit below which the quantum effects become *non-negligible*. On the basis of other fundamental constants (Planck constant h and Boltzmann constant k_B), Debye temperature can be expressed in terms of Debye frequency: $\omega_D = 2\pi\nu_D$. Indeed, by analogy with equation $k_B T = h\nu$, it is possible to define a similar equation $k_B \theta_D = h\nu_D$, therefore

$$\theta_D = (h/k_B)\nu_D, \text{ or } \theta_D = (\hbar/k_B)\omega_D.$$

In different crystals, the value of Debye frequency is located in the range of $\nu_D = 10^{13} - 10^{14}$ Hz. These frequencies of elastic vibrations correspond to the far-infrared range of the electromagnetic spectrum. It is assumed that, at Debye temperature, almost all oscillatory modes (types of oscillations) in the crystal are excited. During a further increase of temperature, new oscillatory modes do not persist, but, instead, the existing modes increase their amplitude, that is, the average energy increases linearly with the increasing temperature.

In different crystals, values of Debye temperature are diverse, but typically $\theta_D \sim 200 - 400$ K. For most of the important crystals in electronics, these temperatures are in silicon, $\theta_D = 650$ K; in germanium, $\theta_D = 380$ K; and in quartz, $\theta_D = 250$ K. In alkali halide crystals, the θ_D varies from $\theta_D = 730$ K in the LiF crystal up to $\theta_D = 100$ K in the RbJ crystal; the highest Debye temperature $\theta_D = 1860$ K is seen in the diamond.

Debye's theory is therefore in good agreement with experiments at low temperature. Moreover, the Debye temperature characterizes not only specific heat, but also some other thermal properties of a solid (e.g., thermal conductivity, thermal expansion, melting points).

The *dynamic theory of Born* is considered in further detail in [Chapter 4](#). This theory gives a chance to calculate specific heat and other parameters of solids more accurately than with Debye's theory by using peculiarities of the atomic structure of crystals. The solid body is treated as a lattice composed of elastically interconnected point masses. Not only are the forces closest to a given atom taken into account but also forces, acting between atoms located at larger distances [8]. Even in the case of the simplest model, that is, a one-dimensional model (i.e., series of elastically joint atoms), it may be shown that Debye's result of *linear* dependence $\omega(k)$ should be corrected: in Born's dynamic theory, the *dispersion* of elastic waves is predicted (in good agreement with the experimental results). However, in case of low temperatures, only low-energy phonons can be excited; thus Born's $\omega(k)$ dispersion is negligible. Therefore the low-temperature dependence of the lattice's specific heat in Born's theory is also cubic: $C \sim T^3$.

3.4 THERMAL CONDUCTIVITY OF SOLIDS

Thermal conductivity is heat transfer by structural particles of a material (molecules, atoms, and electrons) in course of their thermal movement. The transfer of heat is caused by the tendency of the system to be closer to thermodynamic equilibrium,

which is established by temperature equalization. The heat spreads from the hotter part of a material to a cooler part. Heat exchange can occur in any substances in case of nonuniform distribution of temperature; however, mechanisms of heat transfer depend on the physical state of a matter.

The coefficient of thermal conductivity is a quantitative assessment of the ability of a particular substance to conduct heat. In a steady state, the flow of thermal energy, transferred by heat conduction, is proportional to the temperature gradient:

$$\Delta Q = -\lambda \text{ grad } T.$$

This relation is known as *Fourier's heat conduction law*, where ΔQ is the heat flux vector whose magnitude is the amount of energy that passes in unit time through a unit area, oriented perpendicular to the direction of heat transfer; T is temperature, and λ is the coefficient of thermal conductivity (sometimes, referred simply as *thermoconductivity*).

In a simplified model, the steady flow of heat from one side of the parallelepiped to the opposite side is considered, and the formula of heat transfer can be written as:

$$P_{\text{therm}} = -\frac{\lambda S \Delta T}{h}$$

where P_{therm} is the power of heat flux, S is the cross-section of the parallelepiped, ΔT is the temperature difference between its sides, and h is the length of the parallelepiped, that is, the distance between its sides.

As in the case of electrical charge transfer phenomena when, together with electrical conductivity σ , the reciprocal value, resistivity $\rho = 1/\sigma$, is widely used, it is sometimes possible to use a corresponding reciprocal parameter—the *thermal resistance*, $R = 1/\lambda$ —for heat transfer.

It should, furthermore, be noted that in anisotropic crystals, thermal conductivity λ_{ij} as well as thermal resistance R_{ij} are symmetric material tensors of the second rank and can be described by a second-order surface (usually in the form of an ellipsoid). Thermal energy can be transferred both by electrons and by lattice elastic waves (phonons). Various solids can have quite different thermal conductivities that can vary 1000-fold.

In *metals*, thermal conductivity is usually large, and the electronic nature of heat transfer dominates (>90%). At normal temperatures (300 K), the largest thermal conductivity among metals is observed in silver: $\lambda = 430 \text{ W/(m K)}$. Thermal conductivities are somewhat lower for copper $\lambda = 390 \text{ W/(m K)}$, gold $\lambda = 320 \text{ W/(m K)}$, and aluminum $\lambda = 230 \text{ W/(m K)}$. In other metals and alloys, thermal conductivity is $< 100 \text{ W/(m K)}$.

In *semiconductors*, heat transfer is predominantly obscured by phonon processes (i.e., lattice thermal conductivity). Under normal conditions, the contribution of phonon mechanisms dominates electronic thermal conductivity (dominating in metals). However, as a rule, the thermal conductivity of phonons in semiconductors is inferior to the electronic thermal conductivity of metals. For example, at temperature 300 K, the thermal conductivity of silicon is $\lambda = 150 \text{ W/(m K)}$, germanium $\lambda = 70 \text{ W/(m K)}$,

and gallium arsenide $\lambda = 40 \text{ W}/(\text{m K})$. Moreover, the phonon mechanism of thermal conductivity in semiconductors and dielectrics greatly depends on the temperature: correspondent data for the three most important crystals in microelectronics (silicon, germanium, and quartz) are shown in Fig. 3.9.

In *dielectrics*, thermal conductivity has an exclusively phonon character; however, sometimes, thermal conductivity can be rather high: in beryllium oxide (BeO) $\lambda = 80 \text{ W}/(\text{m K})$, in magnesium oxide (MgO) $\lambda = 60 \text{ W}/(\text{m K})$, in sapphire ($\alpha\text{-Al}_2\text{O}_3$) $\lambda = 40 \text{ W}/(\text{m K})$, and in polycorundum (Al_2O_3) $\lambda = 30 \text{ W}/(\text{m K})$. It is very remarkable that a newly developed ceramic *aluminum nitride* (AlN) has the highest thermal conductivity of all available technical dielectrics at $\lambda = 180 \text{ W}/(\text{m K})$. All named dielectrics are used in electronic devices when it is necessary to ensure high thermal conductivity of the dielectric substrate [4].

In the majority of dielectrics, the phonon thermal conductivity at normal temperature is tens of times smaller than the electronic thermal conductivity of metal. For example, sodium chloride crystal (NaCl) shows $\lambda = 6 \text{ W}/(\text{m K})$, crystalline quartz is characterized by $\lambda = 14 \text{ W}/(\text{m K})$, while the thermal conductivity of quartz glass (fused silica) is only $\lambda = 1 \text{ W}/(\text{m K})$.

However, the highest values among solids (the largest coefficient of thermal conductivity at *normal temperature*) is not a metal, but a dielectric—the diamond: at a temperature of 300 K, it has $\lambda > 1000 \text{ W}/(\text{m K})$. This feature is explained by peculiarities of vibrations of the diamond lattice (very large Debye temperature). A detailed examination of lattice thermal conductivity needs more careful investigations of phonon processes, although the basic experimental facts are as follows:

1. Similar to the diamond (C), compounds of light elements AlN, BeO, or MgO are characterized by high thermal conductivity because they have a relatively low

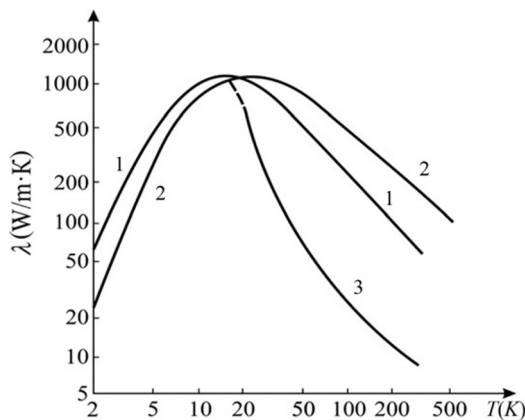


FIG. 3.9

Temperature dependence of thermal conductivity in crystals: 1—germanium, 2—silicon, and 3—quartz.

atomic mass and increased elasticity modulus. These factors lead to high-velocity elastic wave propagation that determines increased thermal conductivity by phonon mechanism.

2. Glasses and quasiamorphous materials have low thermal conductivity as compared with their crystalline modifications, because elastic waves propagate more poorly in disordered structures than in structures of high regularity.
3. The increase in temperature results in the reduction of thermal conductivity in all crystalline solids because it leads to an increase in the intensity of chaotic thermal vibrations in the crystal lattice that scatters elastic waves.

Mechanisms of lattice thermal conductivity. The thermal energy of a solid body mainly consists of elastic vibrations of its particles. For *long waves* (whose length significantly exceeds the lattice constant), the propagation velocity equals the sound velocity; therefore it has the order of magnitude of a few kilometers per second.¹

For *short elastic waves* (heat waves), the velocity of propagation in crystal is significantly reduced, although in absolute terms it would still be considerable if one would assume the “ideal conditions” for propagation of thermal vibrations. However, the thermal resistance in most real dielectric crystals is large. Debye attributed this discrepancy to the scattering of thermal waves. Much of the energy in the thermal wave spectrum falls on waves whose length is comparable with the magnitude of interatomic distances. The smaller the wavelength, the stronger the wave scattering by static defects (inhomogeneous structure), which are caused by impurities, mosaic structure, and mechanical deformations.

Ordinary (long) sound waves propagate in solids without noticeable scattering, because their length is much greater than that of the atomic and microscopic structural defects (sound waves reflect only from the surface of a body). Thus, for ordinary acoustic wavelengths, solids are a *good transparent medium*, whereas waves that correspond to a high-frequency range (in particular, the range of thermal fluctuations) are intensely scattered by irregularities of structure of microscopic and atomic scales, whose number in actual crystals increases with a decrease in the scale of defects. Therefore crystals that are well transparent for long elastic waves prove to be a *turbid medium* in case of short elastic waves, which have a strong diffuse-type scattering. This reduces the effective velocity of short-wave propagation, just as particle collisions reduce the rate of diffusion in gases, although the absolute velocity of the translational motion of particles can remain significant.

However, only this scattering of thermal waves on the *static* structural defects (and their reflections from the crystal surface) cannot explain the large thermal resistance in dielectrics. At normal and elevated temperatures, the main dissipation of

¹**Note.** Usually, the velocity of sound v_s in solids is measured with the sample placed between two piezoelectric elements—one is an ultrasonic emitter and the other is a receiver of the transmitted signal. As a rule, $v_{\text{sound}} = 4\text{--}6$ km/s and depends on crystal orientation. Sound velocity in quartz, for example, is 5 km/s; in silicon it is 9 km/s; and in germanium $v_s = 5$ km/s. In diamond, longitudinal sound wave velocity reaches 18 km/s.

heat occurs on the *dynamic inhomogeneities* of the crystal and is caused by thermal movement. In other words, there is *mutual scattering* of thermal waves by thermal vibrations—phonons.

The excited state of the crystal lattice is traditionally described by the existence of the “ideal” phonon gas. Therefore the results obtained in the kinetic theory of gases can be used for heat conductivity as:

$$\lambda = \frac{1}{3} C v l = \frac{1}{3} C v^2 \tau.$$

where C is the specific heat, v is the average velocity of particles, l is the mean free path (before collision with another particle), and τ is the free path time.

Phonons are often compared with gas. However, in contrast to gas, wherein the number of molecules is constant in time (because molecules cannot pass through vessel walls), phonons may both appear and disappear on the sample surface. Thus phonons can be either reflected from surfaces of a sample or absorbed (or emitted) on its faces, transferring its energy to the environment.

In *harmonic approximation*, the thermal chaotic motion of elastic waves means that phonons propagate without interaction; therefore the principle of linear superposition of fields is applicable. Elastic waves are associated with elastic shifts of particles that have to move *independent* of each other. Therefore, in harmonic approximation, the thermal expansion of crystals cannot exist ($\alpha=0$) and thermal resistance in an ideally infinite crystal should be absent ($R_t=0$, i.e., $\lambda \rightarrow \infty$). The heat flux in a crystal in the absence of the interaction of phonons is similar (in the context of a gas) to the convective-type heat transfer that passes through a cylinder open at both ends.

To take into account the possibility of power redistribution between different waves, and the possibility of establishing thermal equilibrium in the crystal, it is necessary to assume the *anharmonicity*, that is, the nonlinearity in thermal vibrations of atoms. Thus the assumption of direct proportionality between particle displacement and the force that tends to return the particle to equilibrium is unfair (Hooke’s law is not valid).

Thus, in case of phonon collisions, anharmonicity should obviously be taken into account. Two mechanisms of phonon collisions are considered: normal processes (N -processes) and flip-over processes (U -processes). The *normal process* of phonon scattering means there are such collisions of phonons when the initial and final quasimomenta of phonons are equal. During the collision of two phonons, a new phonon can be formed with preservation of total energy and total quasimomentum. Thus the direction of heat transfer is preserved and therefore thermal resistance does not occur.

Thermal resistance (incurring significant limitation of heat transfer) is due to another scattering mechanism—a *flip-over process*—when the initial and final quasimomenta after a collision of phonons differ by the nonzero vector of the reciprocal lattice. During such collisions, the energy is preserved, but the assumption of quasimomentum conservation becomes specific—due to the change in the direction of phonon movement. These so-called U -processes are the cause of the thermal resistance of crystals.

However, at low temperatures, the normal scattering process is particularly pronounced; therefore the lower the temperature, the larger the lattice thermal conductivity. The effect of low temperatures with the assumption of quasimomentum conservation is sometimes expressed by the assertion that, at sufficiently low temperatures (when dominant scattering processes are normal processes), the lattice thermal conductivity tends to infinity.

With a temperature decrease, the number of phonons that can participate in the flip-over process decreases exponentially. Fig. 3.10 shows the typical experimental dependence of thermal conductivity on temperature for pure dielectric crystals of different sizes. Below temperatures of approximately 15 K, thermal conductivity is limited by elastic wave scattering *on the surface* of the crystal; therefore the larger the cross-section of a sample, the higher is its thermal conductivity. However, at very low temperatures $\lambda(T)$, dependence is attributable solely to *specific heat*, which is proportional to T^3 and therefore thermal conductivity vanishes with decrease in temperature. As the temperature increases to more than approximately 15 K, the effect of *U*-processes becomes noticeable, and thermal conductivity, after peaking, begins to decrease.

The maximum $\lambda(T)$ occurs when the average free path of phonon-phonon scattering becomes comparable with the average free path of scattering on the surface. On further temperature increase, thermal conductivity rapidly (exponentially) decreases, because the probability of phonon-phonon scattering increases rapidly. The higher the probability of scattering of individual phonons (which contribute to the heat flux), the greater is the number of phonons, and therefore the relaxation time decreases with increase in temperature.

Furthermore, because specific heat at increased temperatures is almost independent of temperature (Dulong-Petit law), it is expected that thermal conductivity at a higher temperature range will decrease with a temperature increase. This fact has been

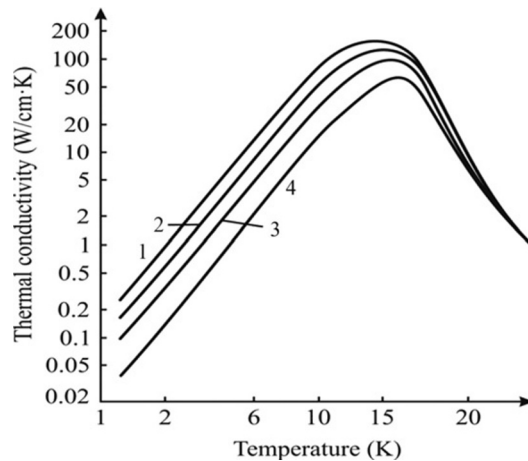


FIG. 3.10

Low-temperature lattice heat transfer of LiF pure crystals of different countermeasures, mm²: 1-8 × 6; 2-4 × 4; 3-2 × 2; and 4-1 × 1.

confirmed by many experiments: thermal conductivity above a temperature of approximately 100 K decreases with increasing temperature according to a power law:

$$\lambda \sim 1/T^x,$$

where $1 < x < 8$. The uncertainty of the degree of this empirical law depends on the competition between different processes of phonon scattering.

Thus the temperature dependence of thermal conductivity in the wide temperature range can be briefly described as follows. In case of very low temperatures, the thermal conductivity is limited by temperature-independent scattering processes, caused by the geometry of the sample and the purity of the crystal. Therefore, when temperature increases, thermal conductivity increases in proportion to T^3 , in accordance with specific heat dependence. This increase occurs until the temperature reaches a limit at which flip-over processes (U -processes) become so intense that the average free path of thermal waves becomes small. Here, thermal conductivity reaches a maximum and then starts to fall off rapidly, exponentially, at first, due to the increasing probability of flip-over processes with increasing temperature. Then, this sharp (exponential) decrease of thermal conductivity is replaced by a slower decrease due to the fact that, at increased temperatures, there are a very large number of phonons that can participate by U -processes.

3.5 SUMMARY

1. The thermal properties of materials are attributable to the internal energy of the lattice (formed by atoms, ions, and electrons), and these properties are *specific heat*, *thermal expansion*, and *thermal conductivity*. *Potential* energy is a part of the energy of the system or body that depends on the positions of particles with respect to external force fields. In solids, the sources of potential energy are Coulomb forces that cause attraction of opposite-sign charges and repulsion of same-sign charges. *Kinetic* energy is the energy of motion in solids that appears due to continuous oscillations of atoms (or ions) caused by thermal excitations.
2. An important thermodynamic function is *enthalpy* (heat content), which characterizes the energy state of a system or material. Enthalpy increases with increasing temperature. Another thermodynamic function of great importance is *entropy*, a measure of the internal disorder (chaos) in system. Thermodynamic function, termed the *Helmholtz free energy*, is minimum in the equilibrium state of a system that corresponds to a certain volume and temperature.
3. The *thermal expansion* coefficient is a characteristic feature of the internal connections of atoms, ions, or molecules and depends on the energy of these bonds. This energy is largely determined by such fundamental parameters of the crystal as its melting point. In cubic crystals (which include most metals and many semiconductors), thermal expansion is an isotropic parameter; however, in anisotropic crystals (e.g., pyroelectrics, piezoelectrics, magnetics) thermal expansion has a pronounced anisotropic character.

4. *Heat capacity* (and closely related to it specific heat) is the ability to store thermal energy in a material when it is heated. Numerically, specific heat is the energy that must be entered into a unit volume of material to heat it up by 1° . Heat capacity depends on temperature; near zero Kelvin, it is extremely small and then increases as T^3 ; however, in the range of normal and elevated temperatures, specific heat varies only slightly with temperature change. Any jump in specific heat is associated with crystal restructuring.
5. Several theories of specific heat in a lattice are considered. The first is the law of heat capacity *constancy* (Dulong-Petit law), derived from classic ideas and, with some accuracy, valid at normal and higher temperatures. Einstein's quantum theory of heat capacity was the first successful attempt to use *quantum laws* to describe the low-temperature dependence of heat capacity. Debye's theory of heat capacity is based on the model of *connected oscillators* and shows better agreement with low-temperature experiments than Einstein's theory. Born's theory of lattice dynamics is the most perfect description of crystal lattice dynamics that also includes the theory of heat capacity.
6. Heat capacity depends on the motion of atoms, but atoms in a crystal are not isolated from each other. Therefore each atom cannot oscillate independently, but moves together with adjacent atoms; therefore, when excited, the elastic wave propagates in a crystal. In addition, each wave is characterized by certain wave vector k and has a frequency ω . Thus this wave can be represented by the oscillator that oscillates with frequency $\omega(k)$. Such an oscillator model represents the *elementary form of motion* of atoms in a crystal. Although the motion of each oscillator is elementary, it involves many atoms of a solid. Ideally, each oscillator exists independently. Therefore the energy of vibrational motions of atoms is a sum of energies of individual oscillators.
7. If one relies on the positions of classic mechanics, the value of oscillator energy could be *anything*: the greater the vibration amplitude, the higher the energy U . However, in quantum mechanics, the energy of the oscillator can assume only discrete values: $U = \hbar\omega (n + 1/2)$, $n = 0, 1, 2, 3, \dots$. The quantum properties of oscillators should be considered only for *microscopic* objects because, when studying the movement of a macroscopic body, it is not necessary to consider the discontinuity of energy levels as permissible energy levels are located so close that their discreteness can be neglected.
8. In the gas of classic particles, the average energy of each particle equals $3/2k_B T$, where T is the temperature and k_B is the Boltzmann constant. A solid can be represented as a "vessel filled with oscillators" while energy of any oscillator equals the sum of kinetic and potential energies (that, on average, equals). The energy of each oscillator, according to laws of conventional physics, equals $k_B T$. This makes it possible to determine whether one can use the formulas of conventional mechanics or would it be necessary to involve quantum physics. The difference between the energy levels of oscillator $\hbar\omega$

should be compared with thermal energy $k_B T$. At normal temperatures, $k_B T \gg \hbar \omega$; consequently, the applicability of conventional mechanics is obvious, and it can be used in normal and high-temperature studies of solids.

9. The physical quantity $\hbar \omega_D$ is Debye's energy. This equates to thermal energy $k_B T$ at a certain temperature, called the *Debye's temperature*, denoted by θ_D . Thus $\hbar \omega_D = k_B \theta_D$; therefore $\theta_D = \hbar \omega_D / k_B$. Important characteristics for crystals include *Debye's frequency* $\omega_D = 2\pi \nu_D$ and Debye temperature θ_D that are interconnected by fundamental constants: Planck constant \hbar and Boltzmann constant k_B .
10. In most solids, Debye temperature does not exceed "normal" temperature (usually, $\theta_D < 300$ K). Therefore almost all solids at normal conditions (300 K) do not exhibit quantum characteristics. However, there are some exceptions, which are very interesting for applications (e.g., diamond, aluminum nitride, beryllium oxide, magnesium oxide), when Debye's temperature is large (> 1000 K). Such crystals, being dielectrics, nevertheless have considerable thermal conductivity under normal conditions and, consequently, they are very important as substrates in electronics. At low temperatures, the main contribution to the vibrational energy of a crystal is produced by acoustic waves. The energy of the corresponding oscillators is small; therefore mostly they are easily excited.
11. Thermal conductivity determines the ability to transfer thermal energy through matter. Heat transfer is due to the thermal motion of structural particles of matter (molecules, atoms, and electrons). Heat spreads from more heated parts of the material to less heated parts. Heat transfer can occur in bodies with nonuniform distribution of temperature, whereby mechanisms of heat transfer are dependent on the physical state of a crystal. Thermal conductivity is characterized by coefficient λ , which is numerically equal to the heat flow that passes through a unit area at a temperature difference of 1 K.

REFERENCES

- [1] K.A. Putilov, *Thermodynamics*, Nauka, Moscow, 1971.
- [2] R.E. Newnham, *Properties of Materials: Anisotropy, Symmetry, Structure*, Oxford University Press, Oxford, UK, 2004.
- [3] C. Kittel, *Introduction to Solid State Physics*, John Wiley and Sons, New York, 1976.
- [4] Y.M. Poplavko, *Physics of Active Dielectrics*, Lambert Academic Publishing, Saarbrücken, 2015.
- [5] J.F. Nye, *Physical Properties of Crystals*, Oxford Press, Bristol, 1957.
- [6] A.V. Shubnikov, *Selected Works on Crystallography*, Nauka, Moscow, 1975, p. 551.
- [7] L.H. Van Vlack, *Elements of Materials Science for Engineers*, sixth ed., Addison-Wesley Publishing Company, Massachusetts, 1989.
- [8] M. Born, K. Huang, *Dynamical Theory of Crystal Lattices*, Clarendon Press, Oxford, 1988.

Quasiparticles in solids

CONTENTS

4.1 Different Elementary Movements in Solids	121
4.2 Quasiparticle Statistics	125
4.3 Photons	128
4.4 Phonons	131
4.5 Magnons	141
4.6 Electrons in Atoms and in Crystals	145
4.7 Electrons in Metals, Dielectrics, and Semiconductors	154
4.8 Summary	160
References	164

4.1 DIFFERENT ELEMENTARY MOVEMENTS IN SOLIDS

A solid is characterized by strength, hardness, and rigidity that seemingly exclude the possibility of any internal movement. However, there are many different types of microscopic motions and displacements in solids.

Firstly, the *movement of structural defects* is possible—through the displacement of interstitial atoms, dislocations, and vacancies (see [Section 1.5](#)). The energy of a crystal is increased in the vicinity of defects so that defects can move (very slowly) in order to find an energetically more favorable configuration.

Secondly, *diffusion transfer* is another type of motion of atoms or ions in solids. This mechanism is the result of thermal fluctuations: the kinetic energy of some particles due to fluctuations can increase such that the particle can overcome a potential barrier that separates one particle from another; this displacement disrupts the equilibrium of the neighboring particle. In most crystals, the probability of such processes at a temperature close to 300 K is small, but it increases significantly when approaching the melting point of the crystal. According to this, diffusion is a classic example of the motion of atoms in solids [1].

Thirdly, *electrons* can move in solids, and their movement alone determines many electrical and magnetic properties of a matter. It is obvious that *cations* and *anions* show directional movement in a crystal in case of ionic conductivity. Compared with the high velocity of electrons, the velocity of ions is very slow; thus in investigations of electronic motion, ions or atoms can be considered as immobile

particles (*adiabatic approximation*). The accuracy of this approximation is determined by the parameter $(m_e/M)^{1/2}$ —the ratio of electron mass m_e to the mass of ion M [2].

However, elementary movements in solids are not limited by the listed mechanisms. To explain the very diverse characteristics of solids, it should be imagined that there are some other (“hidden”) dynamic changes that resemble the properties of other aggregate states of matter—*quasiparticles*—which can behave as a *gas* (vibrations of atoms in the lattice), as a quantum *fluid* (electrons in metal), and even as electron-hole *plasma* (in semiconductors at certain conditions).

Formally, quantum (wave) mechanics describe microscopic objects only mathematically and, certainly, any conventional model of quasiparticles is inadequate [3]. However, quantum mechanics allows retention of the idea of quasiparticles as some mobile “clusters” within the crystal; moreover, they might be described by the overall picture of waves that appear as “wave clots” or “wave packets” (Fig. 4.1).

For example, the free movement of electrons in a crystal can be imagined as the spreading of a wave packet (Fig. 4.1B). The actual electron might be located at any point within this packet and the probability of finding the electron in any definite *point* is close to zero. In Fig. 4.1, the wave amplitude describes only a probability of finding a particle at some point; more precisely, this probability is proportional to the square of the amplitude of a wave. This simple model only promotes the understanding that moving quasiparticle is accompanied by a wave.

The quasiparticle might be interpreted both as *collective motions* of some particles in solids and as *local vibrations* of the atomic group in a crystal lattice. Although this

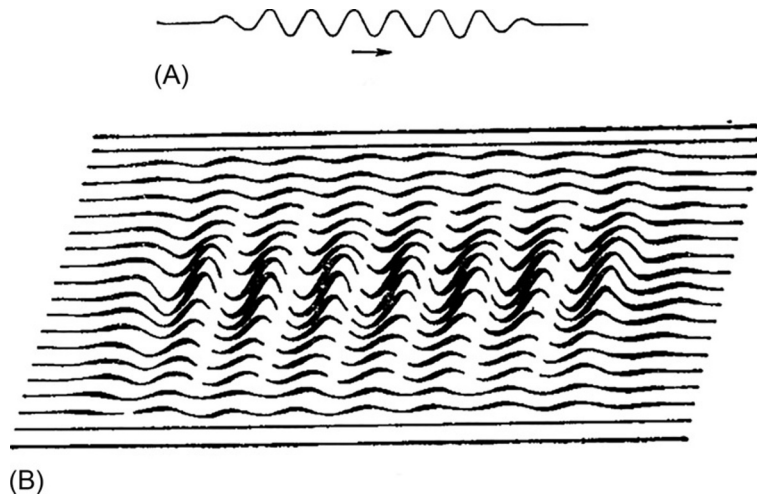


FIG. 4.1

Wave packet: one-dimensional (A) and two-dimensional (B) representation.

oscillation involves many atoms, this movement nonetheless has an *atomic scale*, because the average energy of each oscillation (phonons) is approximately $k_B T$.

Another example of collective motion is the *electronic excitation* of atoms or molecules that, for example, arises when the crystal lattice absorbs photons. This collective excitation is not localized within a particular cell of the crystal, but moves from cell to cell in the form of the *exciton*. The average energy of excitons has the same order of magnitude as the energy of the excited state of individual atoms.

There are some phenomena in solids that involve several quasiparticles. For a description of magnetic properties in ordered magnetics, *magnons* are used with the assumption that the magnon is the quantum fluctuation of electronic spins [4]. The electrical charge transfer is described mostly by *electrons* (in dielectrics—by the *polarons*), whereas heat transfer is attributed to phonons, electrons, and magnons.

In accordance with classic laws, the average energy of thermal motion of particles equals $k_B T$ and, therefore, the internal heat in a solid is $E \sim N k_B T$, where N is the number of particles. However, with decreasing temperature, this simple linear dependence of energy on temperature $E(T)$ is violated, because the internal energy of solids tends to become zero much faster than would occur linearly (see Section 2.7). This fact can be explained by the discrete (quantum) nature of the energy spectrum of solids. Thus, with decreasing temperature, a part of the collective excitations of atoms (or ions) *freezes out*. This process is initiated near Debye temperature (usually, 200–300 K); however, in some crystals, the nonlinearity in $E(T)$ dependence is observed at much higher temperatures. The greater the difference between energy levels, the higher the temperature of freezing out of appropriate motion. Therefore quantum motion in solids may occur at different temperatures. With the exception of electrons, phonons, excitons, and magnons, the quanta of electromagnetic field—photons—can be excited and may spread in solid dielectrics and semiconductors.

In summary, it can be concluded that:

- Materials consist of three kinds of *elementary particles*: electrons, protons, and neutrons. *Quasiparticles* represent the convenient theoretical model of solids that is used to explain the majority of crystals properties; it is obvious that quasiparticles can exist only inside a solid.
- Movements in solids might be very complicated; thus some simple *classic motion* in solids exists as transfer of structural defects, diffusion of atoms and ions, and movement of electrons. However, only these cases are insufficient to describe the electrical and thermal properties of solids, because more complicated *collective movements* need to be considered. This is precisely the motivation for the concept of quasiparticles and collective excitations. Thus the complicated motion of actual particles in solid can be artificially described by the simple motion of imagined quasiparticles, which behave more like noninteracting particles.
- Strictly stated, elementary excitations might be regarded as “quasiparticles” if they are *fermions*, and as “collective excitation” if they are *bosons*. However, in further discussions, both are united under the term “quasiparticle” without any

precise distinction. For example, the free electron is a particle with definite value of weight (rest mass), but in a crystal it behaves as if it has another “effective mass” because it is affected by the environment; in both cases, the electron always is the fermion. Another example is the phonon that characterizes oscillatory motions of neighboring atoms in a crystal; it is the collective excitation in a lattice because it has no “rest mass,” being the boson.

- *Quasiparticles* are a mathematical tool for simplifying the description of many properties of solids. Instead of an inconceivable difficult account of “how a large number of electrons and atoms moves in a specific coordinated way,” the simplified concept of quasiparticles is used.
- In most solids, elementary excitations (quasiparticles) are treated as free (independent) but, in reality, they are only *very close* to being understood as independent. In many cases, it is necessary to take into account their interaction, for example, when explaining the electrical resistivity by electron scattering on phonons or the thermal resistivity by phonon-phonon scattering.
- Using the concept of “quasiparticles/collective excitations,” it is possible to deal only with a handful of somewhat-independent elementary excitations, instead of analyzing interactions of a very large amount of particles in solids ($\sim 10^{23} \text{ cm}^{-3}$). Therefore, this is a very effective approach to simplify many-body problems in quantum mechanics.
- The *electron in solids* is a quasiparticle because it is affected by forces and interactions. The “quasiparticle-electron” has the same charge and same spin as “elementary particle-electron,” and both are fermions. However, in a crystal, the mass of the “quasiparticle-electron” can differ substantially from a normal electron: it has an *effective mass* that might even be anisotropic.
- The *hole* is a quasiparticle consisting of a lack of electron in a crystal cell; the hole has the opposite sign of charge to the electron, has an effective mass, and belongs to the classification of fermions. This concept is commonly used in the context of empty states in the valence band of a semiconductor.
- The *exciton* is a complex of an electron and a hole bound together.
- The *polaron* is a quasiparticle that describes an electron interacting with surrounding ions by local polarization of the dielectric; polarons have increased *effective mass* and belong to the class fermions [5].
- The *phonon* is a collective excitation associated with collective oscillation of atoms (or ions) in a crystalline structure. It is a quantum of the elastic wave and belongs to bosons, with a rest mass of zero.
- The *magnon* is a collective excitation associated with electronic spin structure in the ordered magnetic lattice. It is the quantum of a spin wave; its rest mass is zero; and it belongs to the classification of bosons.
- The *photon inside a crystal* is a quasiparticle because it is dependent on interactions with material. In particular, a “photon-quasiparticle” has a modified relation between energy and impulse (dispersion relation) that is described by the index of refraction of the material.

- The *polariton* is a special form of the photon in crystal, especially seen near its resonance with the lattice vibrational mode. For example, an excited polariton is a superposition of a photon on a phonon.
- The *plasmon* is a collective excitation that is the quantum of plasma-type oscillations (wherein electrons simultaneously oscillate with respect to the ionic lattice).

4.2 QUASIPARTICLE STATISTICS

The fundamental law of statistical physics is the Gibbs distribution that determines the probability of the microscopic state of a system that consists of a large number of particles with specific values of position and momentum. If a large number of *non-interacting* particles moves, the *Maxwell-Boltzmann statistic* is the determining method of the physical system that describes its behavior according to laws of classic mechanics.

In case of quantum system statistics, the energy distributions for fermions and bosons have different peculiarities. In the event that particles, which are unlimited in any state, constitute a special case of statistics, the Bose-Einstein distribution is used and such particles are known as *bosons*. If particles are subjected to the Pauli principle proclaiming that only one particle can exist in a certain state, this case corresponds to the Fermi-Dirac distribution and the particles are *fermions*. The macroscopic system that, at given temperature, is found in thermodynamic equilibrium has such energy and other parameters that almost coincide with their mean values. At high temperatures, when the probability of finding a particle in any state is much smaller than one (because number of energy states is much bigger than the particles), similar to the Fermi-Dirac distribution the Bose-Einstein distribution turns into the classic Maxwell-Boltzmann distribution [6].

The Maxwell-Boltzmann distribution, which determines the probability n_k of the fact that the particle is found in a state with energy E_k at temperature T , is given by a formula:

$$n_k = e^{(\mu - E_k)/k_B T},$$

where μ is the chemical potential, T is the temperature, and k_B is the Boltzmann constant, showing the relationship of temperature to energy. The Boltzmann constant is the ratio of the universal gas constant R to the Avogadro number N_A : $k_B = R/N_A = 1.38 \times 10^{-23}$ J/K.

Electrochemical potential is free energy that falls to one particle in a state of thermodynamic equilibrium with the environment. In turn, *free energy* is a part of the total energy through which the system interacts with the environment. Statistical thermodynamics determines electrochemical potential as $\mu = (E - TS + PdV)/N$, where E is the total energy of the system, P is the pressure, V is the volume, S is the entropy, and N is the number of particles in the system. The Boltzmann

distribution is valid only in cases when $n_k \ll 1$. In solids, this condition can be realized at normal and higher temperatures.

In classic statistics, the particles of an ideal gas have only kinetic energy. The number of particles, whose impulses are found in the interval of $(p, p + dp)$, is determined by the formula:

$$dn_p = \frac{N}{V(2\pi mk_B T)^{3/2}} e^{-p^2/2mk_B T} dp_x dp_y dp_z, \quad (4.1)$$

where m is the particle's mass, V is the volume, and N is the number of particles in a system. When this formula is written in terms of *velocities*, it becomes the *Maxwell distribution*:

$$dn_v = \frac{N}{V} \left(\frac{m}{2\pi k_B T} \right)^{3/2} e^{-mv^2/2mk_B T} dv_x dv_y dv_z, \quad (4.2)$$

Fig. 4.2A shows the distribution of particles by velocities according to Maxwell-Boltzmann statistics; the dotted curve corresponds to a higher temperature.

When charged particles are the ideal gas located in the gradient of an external field with potential $U(r)$, free energy changes. In this case, the Boltzmann distribution is dependent on coordinate r and density n_0 of the particles:

$$n(r) = n_0 e^{-U(r)/k_B T}. \quad (4.3)$$

Formulas similar to Eq. (4.3) are valid in semiconductors and dielectrics to determine the distribution of charge carrier density (electrons or holes) in the electrical field. If free energy has components of both kinetic and potential energy, the *electrochemical potential* should be introduced in the Maxwell-Boltzmann formula as: $\mu - U(r)$.

Quantum statistics explores systems that consist of a large number of particles that obey the laws of quantum mechanics. The main purpose of quantum statistics is

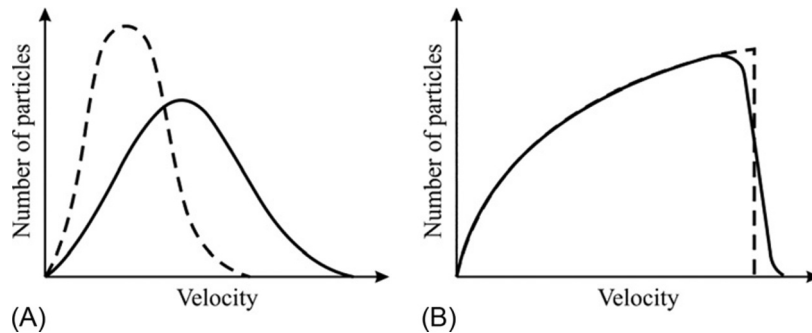


FIG. 4.2

Comparison of classic (A) and quantum (B) distribution of particle by velocities, A—distribution of gas molecules according to Maxwell-Boltzmann statistics; B—quantum Fermi-Dirac distribution for electronic gas in metal; the dashed line corresponds to low, and the solid line to elevated, temperature.

to find the distribution function of particles in a system by various parameters—coordinates, momentums, energy, and so on—as well as to provide calculations of the average values of parameters, characterizing the macroscopic state of the whole system of particles. This system is *degenerate* if its properties are quite different from that of the classic system. The behavior of both Bose and Fermi gas of quasiparticles differs from that of the classic gas because they are degenerate gases.

The degeneration of gas of quasiparticles is significant when the temperature is rather low (phonons in crystals) as well as in the event of very high densities (electrons in metals). The temperature of degeneracy is θ_D , below which quantum properties manifest themselves. In case of $T \gg \theta_D$, the behavior of particles is described by classic laws [7].

Quantum statistics is based on the principle of *identity* (i.e., the principle of indistinguishability) of microparticles: for example, all electrons in metal do not differ from each other. Besides, the Pauli principle should be applied for fermions: in each quantum state, only one particle can exist. The function of full distribution $N(E)dE$ is introduced, reflecting the number of particles, which have energies in the interval from E to $E + dE$. This function is served as a product of the number of states $g(E)dE$ and the distribution function $f(E)$ attributable to the energy interval dE :

$$N(E)dE = f(E)g(E)dE. \quad (4.4)$$

The distribution function $f(E)$ determines the probability of filling states by particles, attributable to the energy interval dE , that is, the average number of particles that are in this state. Therefore, to find a full distribution function, the functions $g(E)dE$ and $f(E)$ should be calculated. Depending on wave function symmetry, all elementary particles are divided into two classes:

- particles with half-integral spin \Rightarrow fermions;
- particles with an integral spin \Rightarrow bosons.

Fermions and bosons show differing behavior toward microstates: in any cell (i.e., in each quantum state), no more than one fermion with a definite set of quantum numbers can exist, whereas the number of bosons with the same parameters may be arbitrary.

Fermi-Dirac statistics for ideal gas of fermions (Fermi gas) is described by a function of energy distribution as:

$$f(E) = \{ \exp[(E - \mu)/k_B T] + 1 \}^{-1}. \quad (4.5)$$

The electrochemical potential μ determines the change of internal energy in a system when one particle is added, under the assumption that all other parameters (that affect internal energy) are fixed. According to function (4.5), the probability that the particle is in state with energy $E = \mu$ equals $1/2$. As any probability must be positive, the value of the electrochemical potential μ is always less than the energy E of the ground state of quasiparticles.

In their main state, fermions occupy the *lowest* possible energy levels. The imposition of the Pauli principle results in all of the lowest levels of the fermions being

occupied at zero temperature (when ground state would be realized). Thus the *highest* occupied level is the *Fermi level*, and the distribution function has a stepped form (Fig. 4.2B). When the temperature increases, the probability emerges that some fermions of the system may have energy greater than the energy of the Fermi level. Therefore the probability that the Fermi energy level is free is nonzero.

Bose-Einstein statistics describe the ideal gas of bosons (the Bose gas of quasiparticles). The “quantum particle-boson” differs from particles of classic physics because it cannot be distinguished (again, the principle of particle indistinguishability is valid). In addition, the wave function of bosons is always *symmetric* as to particle permutations. The energy distribution of bosons follows from Gibbs canonical distribution (but with a *variable number* of particles), presuming that the number of identical bosons in a given quantum state can be arbitrary:

$$\langle N(E) \rangle = f(E) = \{ \exp[(E - \mu)/k_B T] - 1 \}^{-1}. \quad (4.6)$$

This function is Bose-Einstein distribution, which determines the probability of the quantum-mechanical *many-boson system* that exists in a single quantum state. The application of Bose-Einstein statistics makes it possible to explain the specific heat temperature dependence of solids at low temperatures (see Section 2.7, Debye’s temperature). A consequence of quantum Bose-Einstein statistics at low temperatures is the ability to exist in a system as a special phase of matter consisting of bosons—the *Bose condensate*.

The value of the *electrochemical potential* μ can be found in a condition when the sum $\sum \langle N(E) \rangle$ is the total number of particles in a system:

$$\sum \langle N(E) \rangle = N.$$

If expression $\exp[(E - \mu)/k_B T] \gg 1$, both the Bose-Einstein and the Fermi-Dirac distributions turn into the classic Maxwell-Boltzmann distribution:

$$\langle N(E) \rangle = A \exp(-E/k_B T), \quad (4.7)$$

where $A = \exp(\mu/k_B T)$. Thus, at high temperatures, the “quantum gas” of quasiparticles behaves as a classic gas.

From bosons statistics, it follows that bosons tend to *collectivization*—to gather (to condense) in one state. This property of bosons is the basis of quantum light generators (lasers), and it is the cause of such physical phenomena as superconductivity and superfluidity (in quantum liquids). Moreover, Bose-Einstein statistics enable an explanation of electromagnetic radiation when it is found at thermal equilibrium with a body. Precisely, the application of this statistic explains the radiation of a black-body. Besides this, the quanta of light—photons—are examples of Bose particles.

4.3 PHOTONS

When developing the theory of external photoelectric effect, Einstein showed that light is not only emitted and absorbed by the quanta, but also is a stream of peculiar particles (photons) that extend with a discrete portion of energy $h\nu$, where ν is the

light frequency. Based on the idea of the quantum nature of light, Einstein explained not only photoelectric effect but also many other phenomena that cannot be explained in terms of the previous electromagnetic theory of light.

The *duality in the nature* of light was established much before the wave properties of electrons were discovered. The first idea of a relatively *corpuscular* (discrete) structure of light was proposed because it was consistent with experimental facts. However, discussion among scientists continued for a long time between supporters of the *corpuscular theory* of light and proponents of the *wave theory* of light. Finally, using wave theory, a rectilinear propagation of light and laws of refraction and reflection were explained. After the development of electromagnetism theory, doubts about the wave nature of light disappeared [3].

However, the only possible explanation of the law of “black-body” radiation (and explanation of the photoelectric effect) can be explained on the basis of corpuscular properties of light: it can be considered as *photons*—unusual particles with *no rest mass*. It can be shown that Coulomb’s law (a relatively slow decrease of electrical interaction with distance) owes to the zero rest mass of a photon.

It is known that the electrostatic (Coulomb) interaction of charged particles causes a very large force (as compared, e.g., to gravitational interaction). Consider the interaction between charged particles q_1 and q_2 . If the second particle would be removed to “infinity,” the first particle will create an *electrical field*, which has potential ϕ that is proportional to q_1/r . If one would bring the removed particle q_2 to distance r , then a force will act proportionally to q_1q_2/r^2 and be directed from q_1 (if electrical charges have the same sign), or to q_2 (if charges have the opposite sign). In the theory of electricity, the concept of *potential* is introduced. The existence of an electrical field (gradient of potential) indicates that the point where the electrical charge is located has a peculiarity. It was precisely to explain the fundamental interactions of electrical charges that the concept of the *electrical field* was introduced, in order to avoid the idea of “long-range interaction” (i.e., *immediate* power effect on a distance), which is contrary to the relativistic theory.

In classic physics, charged particles interact by a scheme:

particle \rightarrow electromagnetic field \rightarrow particle.

The corresponding quantum scheme is as follows:

particle \rightarrow photon \rightarrow particle.

Therefore a charged particle, *during its movement*, creates a photon, which can be absorbed by another particle, and this determines strength of particle interaction [4]. Thus it should be taken into account that a photon, having an electromagnetic nature, is electrically and magnetically a *neutral particle*.

Furthermore, it is interesting to note that Coulomb’s law (when the force of interaction of charges is inversely proportional to the *square* of the distance between charged particles) is a consequence of the fact that the mass of a photon is zero ($m_\gamma=0$). Precisely because the rest mass of the photon is zero, it has light velocity in vacuum.

The photon, being an electromagnetic wave, also *demonstrates duality*, similarly as an electron demonstrates the particle-wave dualism. The *corpuscular* characteristic of any object is its *impulse*, whereas the characteristic property of the wave is the *wave vector*. They are related by the *de Broglie ratio*,

$$p = \hbar k, \quad (4.8)$$

which can be written inversely as: $\hbar k = p$. The corpuscular properties of electromagnetic waves become apparent, for example, when a wave with frequency ω cannot have energy less than $\hbar\omega$ (in contrast, according to classic concepts, wave energy is proportional to the square of its amplitude and can be arbitrarily small).

The *spin of a photon* is an integer value that equates to unity because photons (unlike electron) belong to the class of Bose particles (bosons). It should be noted that the photon can be found only in one of two spin states: +1 and -1. These two states of the photon means the right and the left circular polarization of a wave, respectively; this is important for understanding many electro-optical, acousto-optical, and magneto-optical effects.

Thus the peculiarities of the photon are *zero mass*, and speed c equals the *speed of light*. The greater the *energy* of a photon, the greater its *momentum* p , that is, the smaller the length of the electromagnetic wave λ , inasmuch as the impulse is $p = \hbar k = 2\pi\hbar/\lambda$. The *dispersion law* of photon, that is, the relationship between its energy and impulse, is expressed by the simple formula

$$E = cp.$$

In Section 2 and Fig. 2.17, this dispersion law is shown in another form—as the *linear* dependence of angular frequency $\omega = E/\hbar$ on the wave number $k = p/\hbar$. This linear dependence of $E(p)$ is quite different from the $E(p)$ dependence for particles that have a determined rest mass ($m \neq 0$) and *quadratic* dispersion law: $E = p^2/2m$. However, from the *relativistic formula* for energy $E^2 = c^2p^2 + m^2c^4$, one can see that, in case of a very large value of impulse (i.e., when $p \gg mc$) any quasiparticle can show linear dependence “energy—impulse”: $E \approx cp$.

Furthermore, it should be noted that technologies using ultrahigh frequencies and *waveguides* can convert electromagnetic waves (in this case, microwaves) into “slower waves.” Thus the “*microwave photon*” really obeys the general law of dispersion: $\hbar^2\omega^2 = E + c^2p^2$, where E is the energy of the “rest photon” that is inversely proportional to the square of the waveguide with radius R (in circular waveguide). This allows the introduction of the concept of the “heavyweight photon” $m^* = \beta\hbar^2/R^2c^2$ that provides additional proof of the corpuscular properties of photons.

When photons penetrate into transparent media (dielectric), their velocity v is reduced in proportion to the refractive index n : $v = c/n$ (this is equivalent to a corresponding reduction of the light wavelength in a crystal). At optical frequencies ($\sim 10^{15}$ Hz), movement in dielectric photons excites valence electrons in the lattice. Below the optical range (the far-infrared frequency range, $\nu < 10^{13}$ Hz), photons excite optical phonons and can move in a mixed photon-phonon state (*polariton*).

Therefore the photon is an elementary particle, usually designated by the symbol γ ; its rest mass is 0, its spin is even (± 1); it has no charge, and it obeys the Bose-Einstein statistic. The photon is a quantum of light (as well as all other forms of electromagnetic radiation). Because the photon has zero rest mass, it allows long-distance interactions. The photon exhibits wave-particle duality, demonstrating properties as a wave and as a particle.

4.4 PHONONS

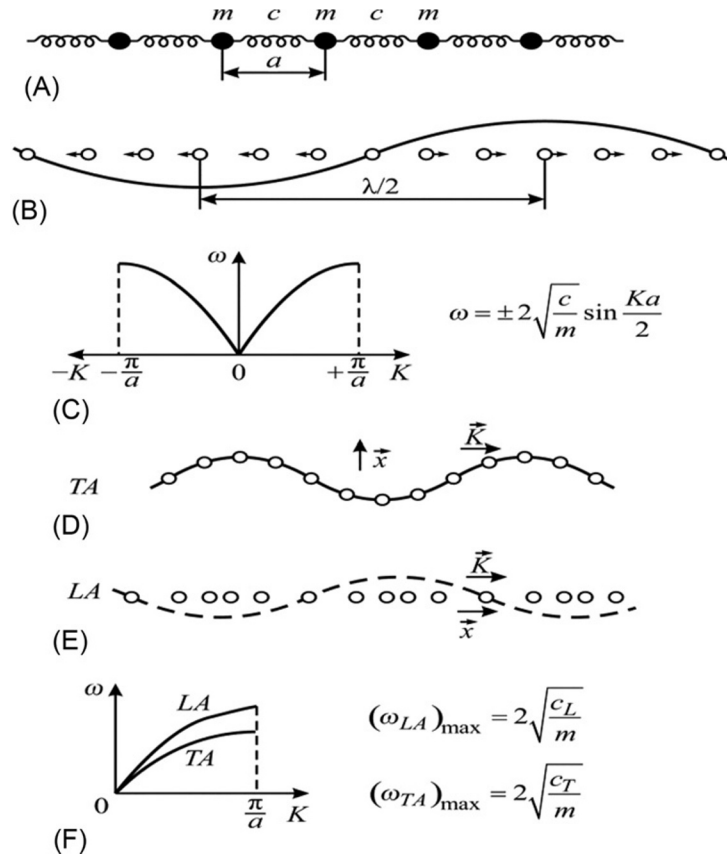
The local microscopic vibrations in the crystal lattice—phonons—are likened by analogy with photons. Unlike electrons and atomic nuclei, phonons are not real particles but only quasiparticles (“like particles”)—convenient objects used to describe many electrical, magnetic, thermal, optical, and mechanical properties of a crystal. The crystal itself, in this case, can be considered a medium for the dissemination, interaction, and transformation of the “gas of quasiparticles.”

Phonons have already been mentioned earlier in connection with Debye’s theory of heat capacity. Further on, the highly simplified lattice-dynamics Born’s theory is considered, which constitutes the foundation of crystal physics [8]. This theory, as set out in Section 3.3 when describing Debye’s and Einstein’s theories, also explains the temperature dependence of specific heat at low temperatures by cubic law ($C \sim T^3$). However, in addition, this theory makes it possible to link lattice vibrations (phonons) not only with the thermal properties of crystals, but also with *electrical and magnetic* properties—electrical conductivity, electrical polarization, energy losses, and electrical breakdown—which are important for electronics as well as to explain the magnetic properties of ferromagnetic crystals, phase transitions in crystals, and so on. Therefore, some aspects of Born’s dynamics theory are discussed across different chapters of this book.

Further, in simplified form, only the basic ideas of Born’s theory, which describe lattice vibrations (phonons), are presented. In Fig. 4.3, not a single oscillator is considered, as was done in Chapter 3 with Fig. 3.6; however, the model of the *one-dimensional monoatomic* crystal is depicted in the form of a linear chain of elastically coupled atoms that are in equilibrium under forces of attraction and repulsion. It is believed that potential relief, describing the position of each atom, is a parabolic well; therefore oscillations of atoms can be described by a *harmonic oscillator* model.

Firstly, it is assumed that the mass of atoms in a one-dimensional chain is the same and they are *not charged* (this corresponds to a homeopolar or molecular crystal). For simplicity, it is assumed that elastic displacements (oscillations) are possible only *along* the longitudinal axis of a chain, and the interaction is taken into account only between the *nearest* neighboring atoms.

In contrast to the previously discussed oscillator with “fixed bearing” (Fig. 3.6), in the model shown in Fig. 4.3A, the displacement of each atom influences displacements of two neighboring atoms; therefore the *wave of displacements* should spread


FIG. 4.3

Elastic waves in a one-dimensional *atomic* crystal: (A) chain of elastically coupled atoms; (B) longitudinal elastic wave; (C) wave's dispersion in first Brillouin zone; (D) transversal acoustic wave; (E) longitudinal acoustic waves; and (F) dispersion law ("branches" of acoustic waves).

through the entire one-dimensional chain as an *elastic wave* (Fig. 4.3B). This periodic process (in space and in time) can be described by the equation

$$x = x_0 \exp[i(\omega t - kx)], \quad (4.9)$$

where $\omega = 2\pi/T$ is the circular frequency that characterizes wave periodicity *in time*, whereas $k = 2\pi/\lambda$ is the wave vector modulus, characterizing the *spatial* periodicity of a wave.

The *phase velocity* of such a wave process $x = x_0 \cos(\omega t - kx)$ is determined by the ratio $v_{\text{ph}} = \lambda/T = \omega/k$, whereas the *group velocity* that describes the propagating of energy is determined by the ratio $v_{\text{gr}} = d\omega/dk$. The feature of elastic wave propagation in a *discrete* chain of atoms is the impossibility of propagation of such a

wave, which has a length less than $2a$: no physical barrier exists between atoms. The correspondent dispersion relation that describes the connection between frequency ω and wave number k (in other words, the relationship of elastic oscillation energy $E = \hbar\omega$ to quasi-impulse $p = \hbar k$) is given by the expression:

$$\omega = \pm 2\sqrt{\frac{c}{m}} \sin \frac{ka}{2}. \quad (4.10)$$

This dependence is shown in Fig. 4.3C in the range of the wave vector in interval $-\pi/a \leq k \leq +\pi/a$. The positive value of k corresponds to waves propagating in the positive x -direction, whereas negative k corresponds to waves moving in the negative x -direction. The restriction of a wave vector space by an interval $(-\pi/a \dots +\pi/a)$ is due to the *discreteness of the oscillating system* (in which no waves with length less than $2a$ are possible). The indicated range of the wave vector values is the *first Brillouin zone*.

If the displacements of atoms are *perpendicular* to the direction of wave propagation, that is, $x \perp k$ (Fig. 4.3D), the *transverse* wave will spread. These waves have the same dispersion law as *longitudinal* waves, but their frequency is lower. The maximal frequencies of longitudinal and transverse waves are located on the *boundary* of the Brillouin zone: $\omega_L = 2\sqrt{\frac{c_L}{m}}$, $\omega_T = 2\sqrt{\frac{c_T}{m}}$, where c_L and c_T are the stretching and bending elasticity, respectively. In most crystals, these frequencies are located in the range of 10^{12} – 10^{14} Hz. The cutoff frequency of elastic oscillations of atoms in a crystal is the *Debye frequency*.

In case of *small-wave* vectors, that is, in the long-wave approximation, when $k \rightarrow 0$ and $\lambda \rightarrow \infty$ (near the center of the Brillouin zone) $\sin(ka/2) \rightarrow ka/2$, the *phase velocity* of waves is almost equal to their *group velocity*; consequently, spatial dispersion is practically absent:

$$v_{\text{ph}} = \omega/k = a\sqrt{c/m} = d\omega/dk = v_{\text{gr}}.$$

In the event of *large values* of wave vectors k (short-wave approximation, near the boundary of the Brillouin zone), these velocities vary considerably. This means that spatial dispersion takes place on the boundary of zone $v_{\text{gr}} \rightarrow 0$.

It is important in crystal physics and some applications that the elastic displacements of atoms determine the *propagation of sound waves*; therefore v_{gr} corresponds to sound (or ultrasound) wave velocity. Those crystal oscillation modes that are characterized by the dispersion law (4.10) are *acoustic modes*. Acoustic waves in Fig. 4.3 are denoted as: *LA*—longitudinal acoustic wave and *TA*—transverse acoustic wave. The quantization of elastic waves is associated with concept of quasiparticles—longitudinal and transverse *phonons*. Thus wave packets of elastic oscillations in the crystal lattice are phonons by analogy with photons—the quanta of electromagnetic waves.

Curves *LA* and *TA* (in Fig. 4.3F) correspond to *branches* of phonon modes in the first Brillouin zone, and they show the relationship between the frequency ω and wavelength $\lambda = 2\pi/k$ in a crystal. *Monochromatic* acoustic waves can be excited in crystals experimentally, for instance, by a piezoelectric vibrator. However, in a

crystal at any nonzero temperature, many *chaotic* acoustic waves exist (with wide distribution of their frequencies and wave vectors); these phonons represent the *thermal reservoir* of the crystal lattice [9].

As mentioned earlier, the maximal frequency of acoustic phonons is located on the boundary of the Brillouin zone; according to Born's model, this frequency ω_B is less than ω_D —the maximal frequency in Debye's model (Fig. 4.4A). In addition, for both models as a phase, the group velocities are characterized by *dispersion* (Fig. 4.4B). Thus group velocity is zero at the boundary of the Brillouin zone, whereas phase velocity is minimal.

Acoustic vibrations of atomic lattices are “electrically inactive” (but only in centrosymmetric crystals), because elastic displacements in this case occur with an electrically neutral (uncharged) center of mass of the unit cell. Accordingly, at acoustic oscillations, any electrical polarization does not arise (if the crystal is nonpolar).

The electrically active components are not acoustic but are *optical phonons*. In the simplest model of monoatomic crystal (considered in connection with Figs. 4.3 and 4.4), the elementary cell consists of only one atom; therefore optical phonons are absent: the simplest one-atomic (or one-molecule) crystal holds *only acoustic* (longitudinal and transverse) phonons.

However, in more complicated structures of crystals (when the unit cell contains two or more atoms), the displacement of particles *within a unit cell* occurs. Due to the very high elastic forces when there are such “counter” displacements, the frequency of these movements lies in the optical range (more precisely—in the *far infrared part* of the optical range) [9]. Coordinated in many neighboring elementary cells, these “internal” oscillations are optical phonons.

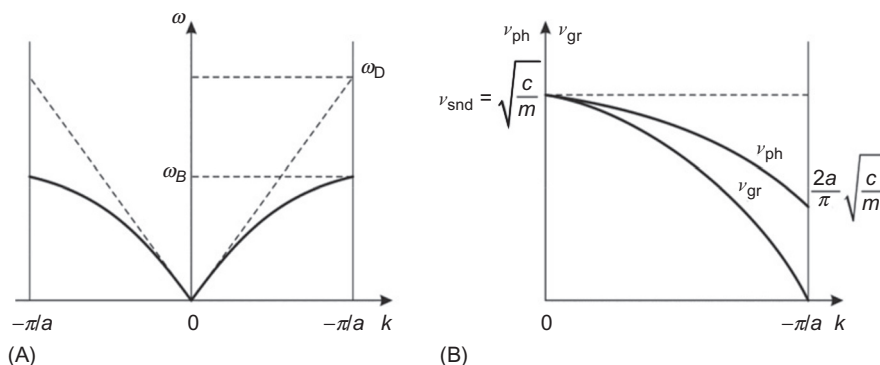


FIG. 4.4

Phonon characteristics in a one-dimensional chain of atoms in Born's model: (A) branches $\omega(k)$ for waves propagating in positive ($k > 0$) and negative ($k < 0$) directions, in comparison with the dotted line that shows $\omega(k)$ for Debye's model; and (B) dependence of phase and group velocities on wave number.

Fig. 4.5 shows a one-dimensional model of the simplest *ionic crystal*—a linear chain of cations and anions with the unit cell lattice parameter a . In this model, as in a model shown before (see Fig. 4.3), the acoustic vibrations LA and TA exist.

However, together with these another type of phonons appears, namely, the optical phonons. In case of *longitudinal optical waves (LO)*, the displacement of ions is parallel to the direction of wave propagation, that is, $x \parallel k$ (Fig. 4.5B). If displacements of ions are perpendicular to the direction of wave propagation, that is, $x \perp k$, the waves are *transverse optical (TO)* (Fig. 4.5C). These waves have a similar dispersion law as for longitudinal waves, but their frequencies are lower (Fig. 4.5D; because the bending elasticity is less than the stretching-compression elasticity). In this model, the acoustic oscillations LA and TA also exist, but are not shown in Fig. 4.5.

Unlike acoustic frequencies, in case of optical oscillation modes (LO and TO) the *spatial dispersion* is the obvious characteristic in the whole Brillouin zone (Fig. 4.5D). In this case, ionic oscillations are determined by the elastic force acting between *neighboring ions*, and therefore their frequency is not strongly dependent on the wavelength. The frequency of optical oscillations always corresponds to the far-

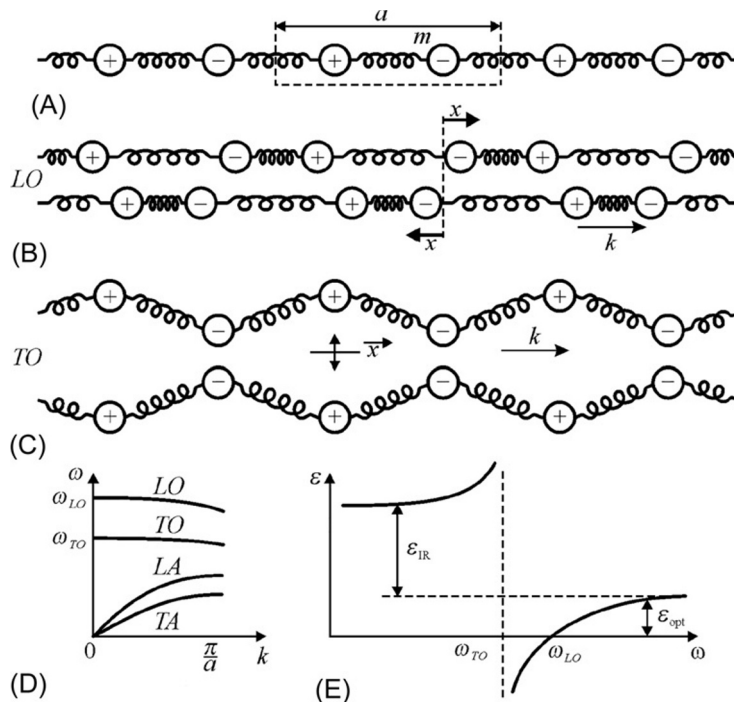


FIG. 4.5

Elastic waves in the one-dimensional ionic crystal: (A) chain of elastically bound ions; (B) longitudinal optical wave; (C) transverse optical wave; (D) dispersion law (“branches”) of optical and acoustic waves; and (E) frequency dispersion of dielectric permittivity.

infrared optical range (10^{12} – 10^{14} Hz). When the wave vector is reduced ($k \rightarrow 0$ that means $\lambda \rightarrow \infty$), frequencies of the optical branches LO and TO do not decrease (as in the case of acoustic phonons); moreover, these frequencies even increase up to the limits of ω_{LO} and ω_{TO} .

In the range of the far-infrared spectrum, the *dispersion* of the crystal dielectric permittivity should be seen: at first, $\epsilon(\omega)$ increases and then, at frequency ω_{TO} , permittivity falls sharply to negative values (Fig. 4.5E) that correspond to the model of the oscillator:

$$\epsilon(\omega) = \epsilon_{\text{opt}} + \frac{\epsilon_{\text{ir}}}{1 - (\omega/\omega_{TO})^2}. \quad (4.11)$$

The resonant frequency of this oscillator is ω_{TO} (the frequency of transverse optical phonons), whereas the longitudinal optical frequency ω_{LO} corresponds to $\epsilon(\omega_{LO}) = 0$ (Fig. 4.5E).

To describe optical-range vibrations by oscillator, besides the ion inertia force $m(d^2x/dt^2)$ and the elastic returning force cx , it is necessary to consider the *electrical force* of ionic interaction qF , where q is the charge and F is the microscopic Lorentz field:

$$m(d^2x/dt^2) = -cx + qF.$$

In the polarized environment, the molecular Lorentz field differs from the average macroscopic field E : $F = E + P/(3\epsilon_0)$. In case of transverse optical oscillations, the vector of elastic displacement is perpendicular to the direction of wave propagation ($x \perp k$); therefore, on average, the *macroscopic field* $E = 0$, because adjacent “combs” of polarization waves are different in their polarity (Fig. 4.6A). Consequently, in Eq. (4.11), written for the transverse mode, it is necessary to make a substitution: $F = \frac{1}{3\epsilon_0}P$. Polarization, as usual, can be expressed in terms of induced dipoles that have a density N and dipole moment $p = qx$, so $P = Nqx$. As a result, the oscillator’s equation takes the form

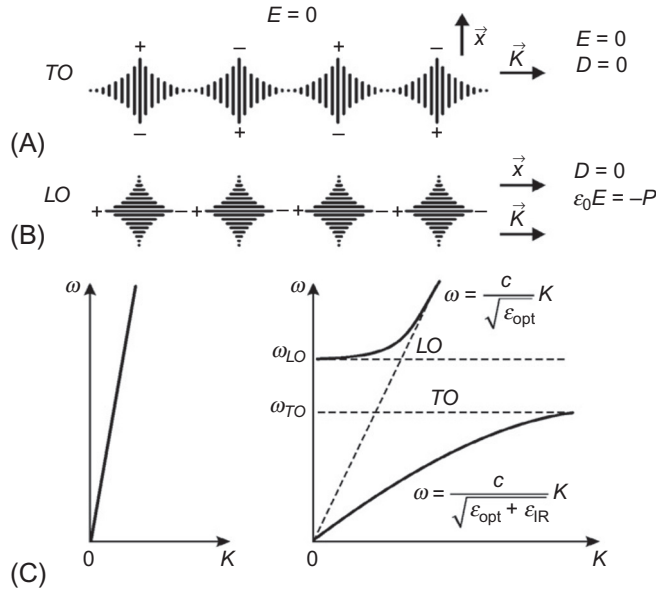
$$m \frac{d^2x}{dt^2} + \left(c - \frac{Nq^2}{3\epsilon_0} \right) x = 0,$$

where the intrinsic frequency of the oscillator that corresponds to the frequency of *transverse* optical phonons is

$$\omega_{TO}^2 = \frac{1}{m} \left(c - \frac{Nq^2}{3\epsilon_0} \right). \quad (4.12)$$

Thus the Lorentz field F *reduces* the elastic constraint and accordingly *decreases* the oscillator frequency $\omega_0 = (c/m)^{1/2}$ to the frequency of *transverse* optical phonons, that is, promotes “softening” of vibrations (reducing frequency ω_0 of oscillator frequency: $\omega_{TO} < \omega_0$ [5]. This event is associated with the polarization of the “short-circuited” crystal, when $D = P$ (because $D = \epsilon_0 E + P$ and $E = 0$).

In case of *longitudinal* oscillations, the local Lorentz field is significantly different (Fig. 4.6B) because the electrical field E is directed *opposite* to polarization P : $\epsilon_0 E = -P$. In the macroscopic theory of polarization, this case corresponds to the


FIG. 4.6

Lattice elastic waves and dispersion in a one-dimensional model of ionic crystal.

“open-circuited” crystal with induction $D = \epsilon_0 E + P = 0$. Taking into account the Lorentz field in Eq. (4.11), it is possible to obtain for longitudinal waves:

$$F = E + \frac{P}{3\epsilon_0} = -\frac{P}{\epsilon_0} + \frac{P}{3\epsilon_0} = -\frac{2P}{3\epsilon_0}.$$

The corresponding equation of the oscillator, taking into consideration $P = Nqx$, acquires the following form:

$$m \frac{d^2 x}{dt^2} + \left(c + \frac{2P}{3\epsilon_0} \right) x = 0.$$

The intrinsic frequency of the oscillator, which corresponds to the longitudinal vibrations, is

$$\omega_{LO}^2 = \frac{1}{m} \left(c + \frac{2Nq^2}{3\epsilon_0} \right). \quad (4.13)$$

Therefore the frequency of the oscillator that characterizes longitudinal optical vibrations in the polarized medium is *higher* than the frequency of the isolated oscillator ($\omega_0 = (c/m)^{1/2}$). These results, obtained in Eqs. (4.12), (4.13), explain the location of phonon branches: LO lies over TO, as well as the location of two characteristic frequencies ω_{LO} and ω_{TO} (Fig. 4.6C).

The dielectric permittivity of ionic crystals depends on a *difference* in the frequencies of longitudinal and transverse optical oscillations ω_{LO} and ω_{TO} in the *center*

of the Brillouin zone. The equation that describes far-infrared polarization of ionic crystals maintains the frequency of transverse optical phonons in its *long-wave limit*:

$$\varepsilon(\omega) = \varepsilon(\infty) + \frac{\varepsilon(0) - \varepsilon(\infty)}{1 - \left(\frac{\omega}{\omega_{TO}}\right)^2},$$

with the corresponding dielectric contribution of ionic polarization being

$$\varepsilon(0) - \varepsilon(\infty) = \frac{Nq^2}{\varepsilon_0 m \omega_{TO}^2} = \frac{Nq^2}{c - \frac{Nq^2}{3\varepsilon_0}}. \quad (4.14)$$

This equation implies that the stronger the ionic polarization influences the dielectric properties of crystals, the higher ionic charge q and the less elastic the coupling coefficient of ions, c [5].

Eq. (4.14) allows quantitative calculation of the infrared contribution to permittivity. Indeed, the concentration of ions N can be found according to the density of a crystal: m is the reduced mass of oscillating ions, q is the ionic charge, and ω_{TO} is the frequency of “residual” rays (determined experimentally by infrared wave reflections from the surface of the studied crystal). The coefficient c that describes the elastic coupling of ions can be calculated from macroscopic elastic properties.

For example, experimental data on phonon dispersion in a simple two-ion crystal NaI are shown in Fig. 4.7A. These data are compared with dispersion in the diatomic crystal diamond (Fig. 4.7B). It is noteworthy that, in the ionic crystal of sodium iodide in the center of the Brillouin zone ($q=0$), frequencies of longitudinal and transverse optical modes are different, $\omega_{LO} > \omega_{TO}$, whereas, in case of a diatomic crystal, they are the same: $\omega_{LO} = \omega_{TO}$.

Comparison of heat capacity in the theories of Einstein and Debye shows that the Debye approximation better describes temperature dependence of specific heat $C(T)$,

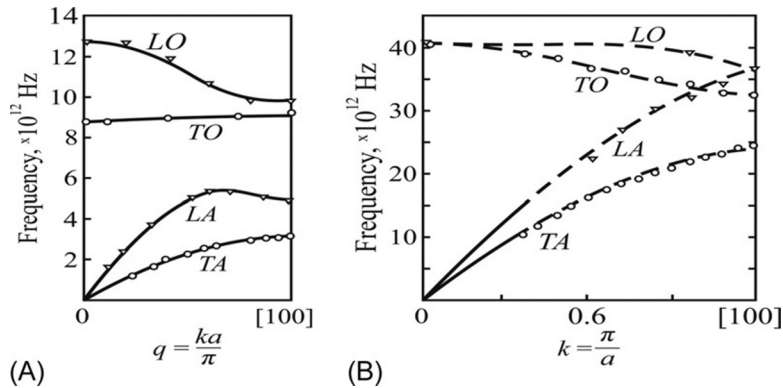


FIG. 4.7

Dependence of phonon frequency on wave vector in direction [100] for the two-ion crystal NaI (A) and diatomic crystal diamond (B); $q = ka/\pi$, the normalized wave vector.

especially at low temperatures (see Chapter 3). This can be explained by taking into account that the atomic spectrum of atomic oscillations in crystals consists of optical and acoustic branches. It is obvious that Einstein's theory describes exactly the *optical* branches of crystal oscillations, wherein frequency dependence on the temperature is observed only in a narrow interval of wave vectors (in contrast to acoustic oscillations) [7]. In addition, in case of low temperatures, mainly *acoustic* phonons are excited as in the spectrum satisfactorily described in Debye's theory.

It should be noted that, even at low temperatures, there are enough phonons in crystals: for example, at a temperature close to one-tenth of Debye's temperatures (20–50 K), 1 cm^3 of the crystal contains approximately 10^{20} phonons (recall that 1 cm^3 accommodates $\sim 10^{23}$ atoms). As the temperature increases, the concentration of phonons increases hundreds of times with a simultaneous increase of their average energy.

Elastic waves in crystals are more diverse and complex, as compared to those in other environments. In gas (e.g., in air) or in liquid, only *fluctuations of density* (or pressure) are possible, that is, only *longitudinal* sound waves can spread. However, in solids, in addition to waves of density fluctuations, shear (*transverse*) waves can move. In case of density waves, atoms oscillate *along the* wave vector k (longitudinal waves), whereas in case of shear waves, atoms can also oscillate in a plane, perpendicular to the wave vector k (transverse waves). Note that, in low-symmetry crystals, two transverse waves are different.

In general, in a crystal, 3ξ types of oscillation modes can propagate, where ξ is the number of atoms (or ions) in the unit cell of the crystal [9]. For example, in the NaCl crystal, a unit cell contains two ions ($\xi = 2$); therefore the $3 \times 2 = 6$ modes of elastic waves can propagate, whereas in metallic sodium, where unit cell has only one atom Na ($\xi = 1$), only three modes exist. Of the 3ξ types of waves, three are *acoustic waves*. A distinctive feature of these waves is that, in case of small-wave vectors (i.e., large wavelengths), acoustic waves have a small frequency. At very small values of wave vector k (when it goes to zero), the frequency of acoustic oscillation tends to zero. The other ($3\xi - 3$) types of waves are *optical waves* (they were first detected by optical methods of investigation). As mentioned earlier, the optical wave frequencies ω_{LO} and ω_{TO} are maximal when the wave vector k is zero.

Each of the 3ξ dependences of the $\omega_j(k)$ type (j enumerates indexes: $j = 1, 2, \dots, 3\xi$) is a periodic function of the arguments. This periodicity is a manifestation of the general properties of a crystal and reflects periodicity in the arrangement of atoms in real space; this, furthermore, leads to a periodic arrangement of cells in the “inverse” k -space (the dimension of k is $[\text{m}^{-1}]$). In summary, all considerations of crystal lattice vibrations can be limited by only one unit cell in the first Brillouin's zone.

Quantization of elastic waves corresponds to the concept of quasiparticles, which are longitudinal and transverse *collective* displacements of lattice atoms. Knowledge of the phonon spectra is necessary to analyze and calculate many physical properties of solids—optical, thermal, electrical, and so on. In experiments, the dispersion curves of longitudinal and transverse waves usually are determined in directions of highest symmetry. This information can be used to calculate the numerical density of states. A very important step to interpret the spectra of oscillations is the analysis of critical points.

An example of the density of states calculations of a spectrum is shown in Fig. 4.8 for aluminum [10]. The sharp maximum in the total dependence of states density $D(\omega)$ corresponds to the maximal frequency of certain types of phonons. In crystals with complex many-atom lattices, the peculiarities in spectrum may also be associated with optical branches of vibrations.

When phonons are compared with real particles, it should be noted that the number of particles at their collisions remains unchanged, but the number of phonons at collisions *is not saved*. However, the main difference between phonon collisions and real particle collisions is that, during collision of phonons, the *impulse is not preserved*.

Because the behavior of phonons determines the thermal properties of solids, in investigations of heat capacity and heat conductivity it is possible to obtain information about the main properties of phonons. The study of acoustic phonons at a low value of pulse is a relatively simple procedure, as they are ordinary sound waves. Experimental data concerning velocity and attenuation of sound in crystals make it possible to obtain the characteristics of long-wave acoustic phonons. To study long-wave *optical phonons*, investigation of resonant absorption of light by crystals is used (where, upon falling on the crystal, a photon is converted into a phonon). As the light velocity is very large, optical phonons are born with very small impulses. Therefore, by optical research, experimental data can be obtained only for phonons *near the center of the Brillouin zone*.

Nevertheless, a method of *inelastic scattering of neutrons* in crystals exists that gives the most detailed spectrum of phonon branches. Flying through a crystal, a neutron excites oscillations of atoms. Thus it might be said that the neutron generates phonons. The larger the path of the neutron in a crystal, the greater the probability of phonon generation. By an inelastic neutron scattering study, it is possible to directly derive the law of phonon dispersion in the entire Brillouin zone.

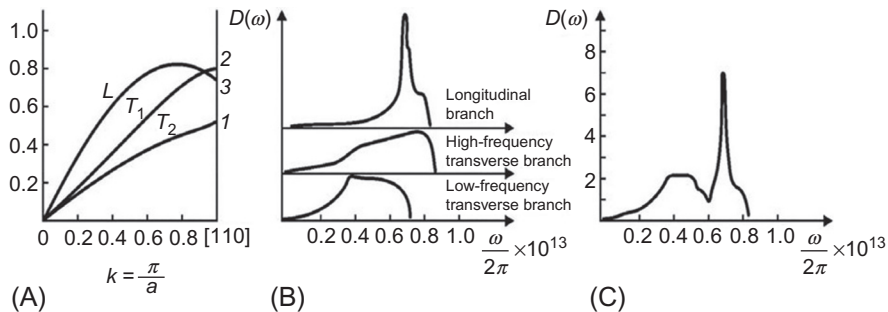


FIG. 4.8

Oscillatory spectrum of aluminum: (A) phonons branches in Brillouin zone; (B) function $D(\omega)$ for longitudinal and transverse branches; and (C) total view of $D(\omega)$ function.

4.5 MAGNONS

Besides charge and mass, some elementary particles (electrons, protons, or neutrons) are characterized by a spin that determines their magnetic properties (word *spin* means “spindle” or “rotation”). Thus the particles cannot be imagined as “fixed balls.” If classic concepts are applied, some particles may be represented as “rotating balls.” However, the rate of this rotation *cannot be changed*, because what is interpreted simplistically as a “rotation” is really an *intrinsic property* of the particle itself. An electron or proton cannot change the value of its spins nor the value of their mass or their charge.

The classic idea of “spin” is an extreme simplification, and this concept contradicts the theory of relativity. In fact, spin is not a consequence of spatial rotation, but rather is a specific property of an elementary particle that determines, in particular, its behavior in the “collective” of surrounding particles. In one kind of particles, spin can only be integer-valued, whereas in others it can exactly be the half-integer value. Zero spin refers to integer spin.

The magnitude of the spin is denoted by the letter s ; a particle with spin s has an angular impulse: $[s(s+1)]^{1/2}\hbar$. The electron has a spin equal to $1/2$. Because the electron has an electrical charge e , it is the *source of the electrical field*. In connection with its “rotation” (which, in classic physics, may conventionally be considered the circular current), the electron is also the *source of a magnetic field*. Thus the particle with spin $1/2$ and electrical charge e has a magnetic moment:

$$\mu_B = e\hbar/2mc.$$

This value, which is called *Bohr’s magneton* for electron, is equal to approximately 10^{-20} erg/Gs.

When discussing the magnetic properties of solids, it should be noted that the magnetic moment of the electron is an unusual vector, because it can be oriented in space only in *two ways*: lengthwise along the external magnetic field or against it. Accordingly, the angular momentum of a particle can always be orientated by a $g = 2s + 1$ manner; inasmuch as the electron’s spin is $s = 1/2$, only two of these manners are possible. Spins characterize not only elementary particles, but quasiparticles as well. Photons and phonons are characterized by the *integer* spin (they are bosons); bosons also include the magnons.

The *magnon* (spin wave) is a quasiparticle, introduced theoretically to describe the system of *collective excitations* of interacting spins in the *ordered magnetic crystals* (ferromagnetics, antiferromagnetics, and ferrimagnetics). As with thermal motion, the magnetic field can influence the magnetic moments of electrons. However, in ferromagnetics, the localized single inverted spin cannot exist—this is prevented by exchange interaction. Thus elementary excitations in ferromagnetics (as in other magnetically ordered substances) are the inverted spins *distributed in a certain area* of a crystal [4].

These magnetic violations have the nature of waves that are characterized by a certain wave vector k and frequency ω . Such violations are called *spin waves* (magnons). They exist in ferromagnetics at any temperature that is lower than the Curie temperature (and, in antiferromagnetics, below Neel temperature); thus the closer the temperature is to phase transition, the greater the intensity of magnons.

A single magnon corresponds to the wave of precessing of neighboring spins. An idealized picture of spins excitation in a one-dimensional crystal is shown in Fig. 4.9. The model of an upturned spin among other oriented spins (Fig. 4.9B) is less likely because such a situation requires significant energy cost. There would be much smaller energy needs if all spins are predominantly directed in parallel (Fig. 4.11C). Therefore, this is a more realistic model according to which the ends of the spin vector precesses on the surface of the cone, whereas each subsequent spin is shifted in phase as with the previous spin (the angle between them remains constant).

This wave is formed due to a strong exchange interaction between atoms; as a result, the deviation in the magnetic moment of atoms from their equilibrium position is not localized, but is distributed along a chain. A spin wave may occur, mainly, in magnetically ordered solids—ferromagnetics, antiferromagnetics, and ferrimagnetics. Thus, in crystals with multiple submagnetic lattices (i.e., in antiferromagnetics), there can exist several types of magnons with different energy spectra [9].

Fig. 4.10 shows a more detailed model of a magnon—a spin wave whose structure resembles the wave of an acoustic phonon (shown in Fig. 4.3). A series of atoms is shown, and the distance between them is the crystal lattice parameter. In magnetic field H , all spins precess with frequency ω_0 (homogeneous precession). In real systems, small oscillations of the magnetic moments of atoms are seen in the form of waves with inhomogeneous precession.

It should be noted that magnons, being waves of *electronic spin precession*, differ significantly from phonons, which are the elastic displacements of *atoms*. The difference is seen in the comparison of dispersion law for magnons and phonons: in the dependence of energy $E = \hbar\omega$ on impulse $p = \hbar k$ (or, equivalently, in the dependence of frequency $\omega = 2\pi/T$ on wave vector $k = 2\pi/\lambda$). For example, the dispersion law for spin waves in a one-dimensional model of magnons is expressed as follows: $\omega = 8JS/\hbar \sin^2(ka/2)$, where J is an exchange integral; S is the spin moment; and a is the

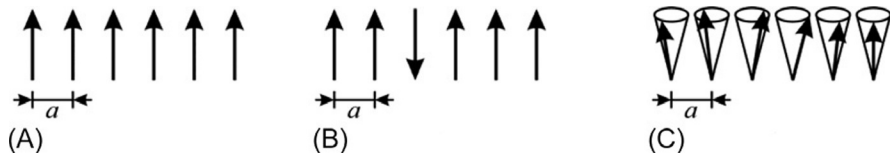


FIG. 4.9

Different ideas about spin waves in a one-dimensional lattice with parameter a : (A) classical scheme of the ground state of a simple ferromagnetic—all spins are parallel and directed in one direction; (B) the simplest idea of an excited magnetic state, an inverted spin; and (C) spin wave in ferromagnetics.

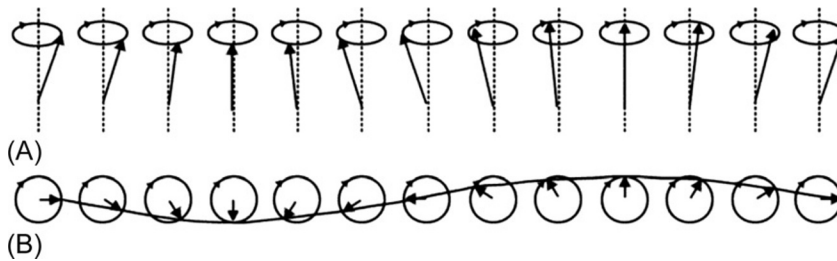


FIG. 4.10

A spin wave in a linear series of spins: (A) series of spins shown from side; (B) series of spins shown from above (a wave is shown as a line that runs through the end of spin vectors).

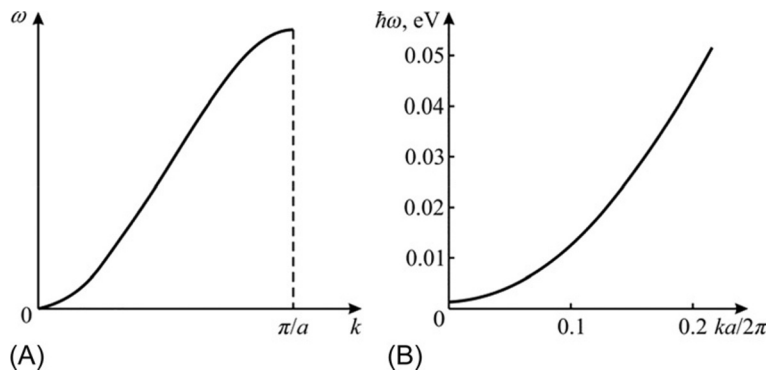


FIG. 4.11

The law of dispersion for spin waves in a one-dimensional ferromagnetic: (A) theoretical calculation; and (B) magnon spectrum measured in alloy $\text{Co}_{0.92}\text{Fe}_{0.08}$.

crystal lattices parameter. Graphically, this dispersion of magnons is shown in Fig. 4.11A. As shown earlier, for *long-wave acoustic phonons* ($k \rightarrow 0$), their frequency is proportional to the wave number: $\omega \sim k$. However, for *long-wave magnons* ($k \rightarrow 0$), the law of dispersion is parabolic: $\omega \sim k^2$. This frequency dependence of magnons is observed in experiments by using neutron scattering in magnetic environments (Fig. 4.11B).

Thus, magnons characterize the movement of elementary magnetic moments in the magnetic. Similarly as phonons, magnons are *excited by the thermal motion* of atoms or ions. In addition, long-wave magnons can be excited by electromagnetic fields of ultrahigh frequency. Magnons behave like weakly interacting quasiparticles; they are characterized by *integer spin* (equal unity) and therefore obey Bose-Einstein statistics. Calculations show that, at high temperatures, the concentration of magnons in a ferromagnetic crystal can be significant. Their density in

ferromagnetics n_{fer} proportionally increases with the temperature and depends on their remoteness from the Curie point: $n_{\text{fer}} \sim (T/T_C)^{3/2}$. In antiferromagnetics, this dependence is different: $n_{\text{anti}} \sim (T/T_N)^3$, where T_N is Neel point.

At very low temperatures (near absolute zero), a ferromagnetic reaches its lowest energy state wherein all atomic spins are oriented in one direction. Therefore when the temperature is lowered, magnons become frozen; thus near absolute zero magnons should be practically absent (this effect of *Bose-Einstein condensation* is confirmed experimentally). The growth of the quantity of magnons is caused by a temperature increase and thus magnons *reduce the magnetic ordering* in a crystal.

In *antiferromagnetic* the number of magnons is proportional to T^3 that reminds Debye law for temperature dependence of phonon concentration (T^4). However, in the *ferromagnetic* concentration of magnons increases proportional to $T^{3/2}$. The point is that increase of magnons quantity decreases spontaneous magnetization of ferromagnetic, in which connection change of magnetization is proportional to $T^{3/2}$ (Bloch law). Correspondingly, dispersion law for magnons in the antiferromagnetics differs from magnon dispersion law in the ferromagnetics, Fig. 4.12. In antiferromagnetics, the variance of magnons $\omega(k)$ is similar to the dispersion of phonons (see Fig. 4.3C).

Therefore it is possible to describe the properties of ferromagnetics below Curie point (as well as of antiferromagnetics, below the Neel point), assuming that spin waves exist that can be represented by the nearly degenerate gas of magnons. The electrochemical potential of this “gas” is zero, and therefore the number of magnons is not saved. The Bose distribution function for magnon energy permits the calculation of the temperature dependence of magnetic thermodynamic properties (i.e., magnetization, heat capacity, magnetic susceptibility). The more accurate the resulting expressions, the closer the gas of magnons is to the ideal Bose gas. Deviations from theory are the result of *interactions* of magnons with each other as well as their interactions with other quasiparticles (phonons or electrons). As the temperature

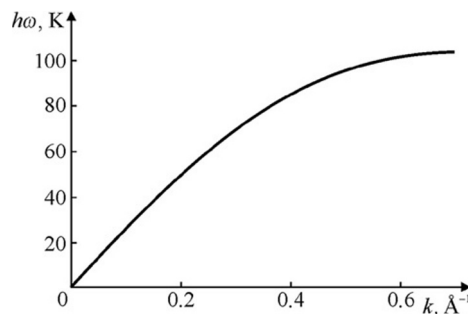


FIG. 4.12

Spectrum of spin waves in antiferromagnetic RbMnF_3 , obtained by inelastic neutron scattering.

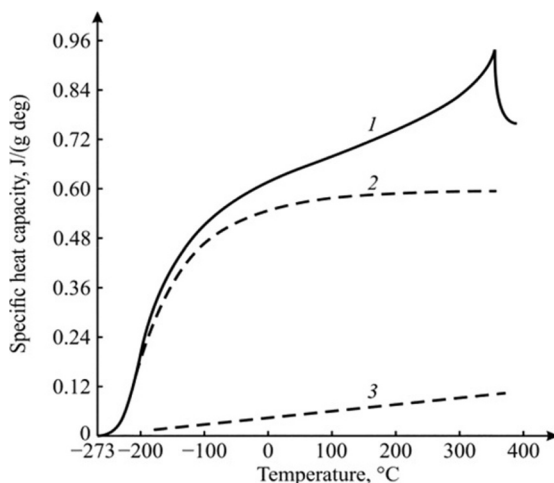


FIG. 4.13

Temperature dependence of different contributions to the heat capacity of nickel: 1—total heat capacity; 2—lattice (phonons) heat capacity; and 3—electronic heat capacity.

increases, the number of quasiparticles grows, and therefore their interaction becomes more significant than that of the “ideal gas of magnons.”

Representation of magnons allows the description of many properties of magnetics—not only their thermodynamic (equilibrium) properties, but also their kinetic and resonance properties. For example, magnons play a significant part in the heat capacity of magnetics, together with phonons and electrons (Fig. 4.13).

It is appropriate to note that the concept of “spin waves” is broader than that of “magnons.” Spin waves can exist in nonmagnetic metals as well; they represent spin-density oscillations of *conductive electrons* due to energy exchange and interactions between them. The existence of spin waves in nonmagnetic metals is found in electronic paramagnetic resonance.

4.6 ELECTRONS IN ATOMS AND IN CRYSTALS

The consideration of electronic states in crystals originated from the electronic spectra of atoms. Atoms can be characterized by two complementary models: a spatial model and an energy model. The spatial model of an atom reflects its volumetric three-dimensional structure, and, within this structure, the location of electrons in the atom is described by the *probability density*. Electrons that are distributed near the nucleus form an *electronic cloud*. In the simplest case, this cloud is spherical (in a hydrogen atom in the nonexcited state); however, in most cases, the electronic cloud has a complex configuration.

A conventional image of the external shape of an electronic cloud is known for its different quantum states. Schrödinger's equation provides an opportunity to undertake a rigorous mathematical description of electronic clouds: their geometric features in atoms and ions. However, sometimes, their visual representation is impossible because it might be quite difficult to find the probability of electron distribution in a cloud. Therefore the model of Bohr—a *simplified model*—is often used to describe the configuration of an atom. This model allows the atom to be represented as a central positively charged nucleus with electrons moving in their orbits around it.

The number of electrons determines the position of the atom in Mendeleev's periodic table, and it is exactly equal to the number of protons in the nucleus of atom. From experiments and theory, it is known that the radius of an atom equals $a \approx 10^{-8}$ cm. Therefore the radius of the nucleus is estimated at a size of approximately 10^{-13} cm, and is roughly the same as the size of an electron; thus the size of atoms is 100,000 times greater than the size of their nucleus. Therefore, on the face of it, the volume of an atom looks "empty"; however, in solid-state physics, the atom is usually represented by a solid ball, and this is a "good working" model. The fact here is that this "ball" is "filled" by a very strong electromagnetic field [3].

Next, to simplify further consideration, the simplest atom is discussed, namely, the hydrogen atom consisting of one proton and one electron. In this atom, the positively charged nucleus holds a negatively charged electron by the Coulomb force of attraction:

$$F_{\text{Coul}} = e^2/a,$$

where e is the electron's charge (the proton has the same charge). To ensure the stability of the atom, the force of attraction must be balanced by the force of repulsion. This force is the centrifugal force:

$$F_{\text{centr}} = mv^2/a,$$

where m is mass of electron and v is its velocity. The equality of F_{Coul} and F_{centr} makes it possible to determine the velocity of an electron's movement in its circular orbit:

$$v = (e^2/ma)^{1/2}.$$

Both the charge e and the mass m of an electron are *fundamental constants*. By substituting the values of these constants in a given formula, it is possible to find the velocity of the electron's rotation in its orbit: $v = 10^8$ cm/s. In these calculations, the relativistic effects are negligible because $v/c \approx 1/300$. However, if the atom has a size close to that of its nucleus (10^{-13} cm), the velocity of the electron's rotation would be close to the velocity of light (obviously, this is impossible).

The total energy of the electron in the field of the nucleus (sum of its kinetic and potential energy) is:

$$E = -e^2/2a.$$

The “minus sign” means that for zero energy electrons that have been sent an “infinite” distance away from the nucleus should be considered (with decreasing distance, the energy decreases).

According to a simple model, the electron moves in an atom with a velocity $v \approx 10^8$ cm/s *in a circle*; therefore the vector of velocity constantly changes its direction. It is reasonable to believe that $\Delta v \approx v$, which means that the uncertainty of velocity Δv is equal to the velocity. From the indeterminacy principle (i.e., the uncertainty relation), it follows that $\Delta x \cdot \Delta p \geq \frac{1}{2}\hbar$. Taking into account that the impulse is $p = mv$, the uncertainty in an electron’s coordinates is $\Delta x \geq \hbar/2mv$. From the mass of electron $m \approx 10^{-27}$ g, its velocity is $\sim 10^8$ cm/s and, using the Planck constant \hbar , it is possible to find the uncertainty of the electron’s location $\Delta x \geq 10^{-8}$ cm that exactly corresponds to the *size of the atom*.

This means that the sphere of radius a represents the volume containing the electron; however, to clarify its position in this volume is impossible. The quantum indeterminacy principle (the Heisenberg principle) allows the estimation of the size of an atom, namely, atomic radius is determined by the uncertainty of the orbital position of electrons: $a \approx \Delta x \approx \hbar/mv$. Using this expression for the orbital velocity of the electron: $v = (e^2/ma)^{1/2}$, it is possible to get:

$$a = a_0 = \hbar^2/me^2.$$

Thus the atomic radius a_0 can be expressed through *fundamental parameters*: the Planck constant \hbar , the mass of electron m , and the charge of the electron e . This radius approximately equals 0.5×10^{-8} cm and it is the *Bohr radius*; it coincides with the radius of the hydrogen atom in its ground state.

According to quantum mechanics, not *all* states are allowed but only states with certain energy are permitted; thus there is one state (ground state) in which electron *does not radiate energy*. In addition, besides the ground state with Bohr radius a_0 , there are a number of *excited states*; the emerging transitions between them result in the emission (or absorption) of light quanta.

Inasmuch as the electronic waves in an atom propagate in three dimensions, it is possible to depict them graphically through intersections (Fig. 4.14). A section shows two types of permitted waves (a and b) in three quantum systems: an electron in a hydrogen atom (1), a one-dimensional particle in limited space (2), and a quantum oscillator (3) [3].

It is a fact that all electronic waves in an atom have “tails” that extend to large distances (infinitely); Fig. 4.14(1) shows that the electron has a slight chance to extend to a large distance, but it is most likely to have its location near the nucleus. Thus the energy levels of an electron that correspond to its possible natural waves in a hydrogen atom (Fig. 4.14 shows only two of them) can be placed in a series that converges as shown in Fig. 4.15B.

Unlike energy levels in the *quantum oscillator*, in which the distance from each other is always $h\nu$, the distance between an electron’s energy levels *in an atom* decreases with the increase in energy [3]. Therefore by acquiring adequate energy,

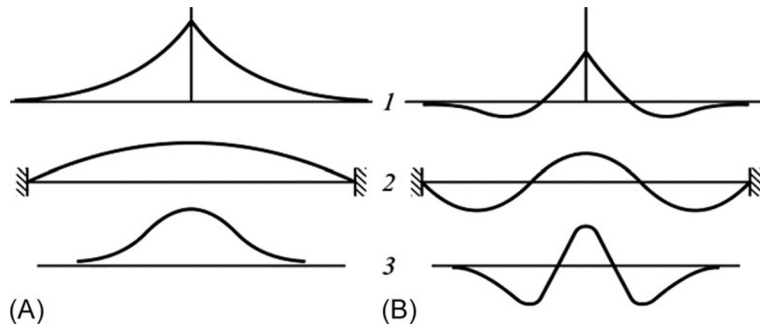


FIG. 4.14

Waveform probabilities for two permitted states of electrons in atoms (1); for a particle that moves in a straight line (2), and for a harmonic oscillator (3).

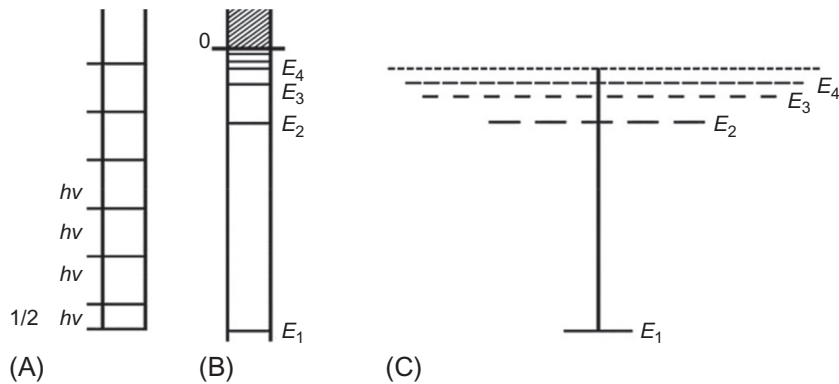


FIG. 4.15

The energy spectrum of the quantum oscillator and the hydrogen atom: (A) permitted energy levels of the quantum oscillator; (B) permitted energy levels of the electron in the hydrogen atom; and (C) correspondence to levels of permitted states (number of strokes).

an electron can finally leave the atom; next, its energy changes *continuously*, as is shown at the top of Fig. 4.15B by a continuous energy spectrum. Exactly the state at which an electron is found very far from its nucleus is selected as the “zero point” of energy.

When the approach of each electron falls under the influence of the electromagnetic field of adjacent electrons (without outside electromagnetic field), electrons interact with each other as if they are two small interacting magnets. The Pauli principle proclaims: if two electrons are in one of the stationary states (e.g., on a single orbit), they cannot have spins oriented in one direction, but necessarily must orient their spins in the *opposite* direction.

In the *helium atom*, both electrons at normal conditions are authorized by a state with the lowest energy. Because both of them are in the same state, their spins, according to the Pauli principle, are opposed and form a *complete s-shell* (Fig. 4.16A).

The *lithium atom* has three electrons; two of them choose a complete *s-shell*, while third electron would also acquire a condition with minimal energy. However, this case is prohibited by the Pauli principle, because the main (*s*-) state is already fully occupied by two electrons with opposite spin directions. Therefore, the third electron in the lithium atom reluctantly takes one of four following states, characterized by higher energy than the *s*-state (Fig. 4.16B).

Lithium starts a new row of elements in Mendeleev's periodic table. The state that takes the third electron is *one of four* possible levels in the electronic *p-shell*, following the *s-shell* (Fig. 4.16B). Each item in this series can be filled by electrons, and there might be eight electrons in the *p*-state district. They are all completely filled in the neon atom (Fig. 4.16C): electrons occupy all four levels that are in the *p-shell*, and each of them contains two electrons with spins in opposite directions [3].

The condition of electrons in atoms of a given element determines its physical and chemical properties. For example, the chemically neutral inert gas argon has 18 electrons, but adding only one electron to a shell (and one proton to the nucleus) transforms the inert argon atom into a chemically very active potassium atom.

Electrons in crystals. The energy model of a crystal is considered at the elementary level and in close connection with the previously explained concepts of the energy spectrum of the atom (see Figs. 4.15 and 4.16). Without knowledge of the main features of the energy spectrum of electrons in a crystal, it is impossible to understand the principles of operation of microelectronic devices (most of which are based on semiconductors).

The spectrum of electronic energy of a crystal is directly related to the energy spectrum of atoms entering the crystal structure. Specific examples of energy band formation as well as the creation of *overlapping areas* are considered further with a relatively simple example—metallic sodium. The energy diagram of the sodium

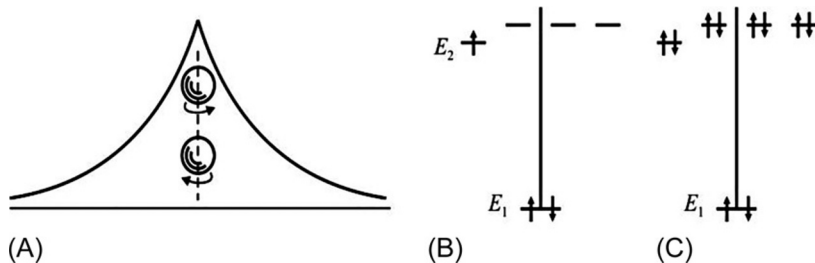


FIG. 4.16

Electrons and their spins in atoms: (A) wave function for two electrons with opposite spins in helium; (B) basic (E_1 and E_2) levels in lithium atom (*dotted line* shows the next higher allowed states); and (C) five levels in the atom of neon, occupied by electrons.

atom with 11 orbital electrons (balancing the same number of positive charges in the nucleus) is shown in Fig. 4.17. The first two electrons occupy the lowest level of energy $1s$ in the K -shell. Next, the third and fourth electrons occupy the lowest energy level $2s$ (shell L_I), the fifth electron is located in the lowest remaining level (shell L_{II}), and so on. From Fig. 4.17, one may conclude that, on the third level, six electrons are set. However, there are three levels that differ only by a little energy (this is peculiar in atoms with a low atomic number); therefore, they cannot be depicted with boundaries.

The correspondent energy levels for all 11 electrons are shown as characteristic of the *neutral* atom of sodium. With the addition of each subsequent electron, the form of the potential energy curve E_p changes (Fig. 4.17), while location of energy levels becomes different. Each outer electron can approach the atom with lower velocity because it is subject to not only attraction from the nucleus, but also repulsion from other, deeper electron shells.

At the highest occupied level M_I (Fig. 4.17), the nonexcited sodium atom has only one valence electron (in the state $3s$), and it is this electron which determines most of the chemical, electrical, and optical properties of sodium. The remaining 10 electrons are located so deep in the well of potential energy that they cannot participate in chemical, electrical, or thermal processes.

In case of solid-state formation from individual atoms, the energy description in the first approximation is necessary for constructing potential energy curves for a *series* of atoms, located at a distance equal to the crystal lattice constant (Fig. 4.18). Because atoms in the crystal lattice are located close to each other, the potential curve between them cannot rise to the level $E=0$, as it happens in an atom located outside the crystal (Fig. 4.17). The maxims of potential energy between atoms cannot reach even the energy of a single valence electron of the atom. Therefore, nothing prevents valence electrons (which originally belonged to the

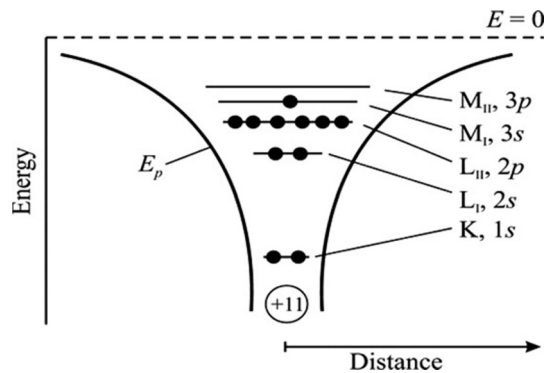


FIG. 4.17

Electron configuration in the sodium atom: the valence electron is located on level $3s$ in the M_I shell; K , L , and M —designations of electron shells, E_p —potential energy.

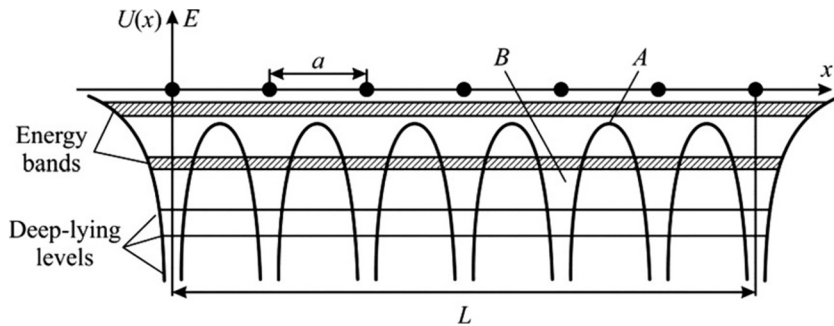


FIG. 4.18

One-dimensional energy model of a crystal: a is the interatomic distance; L is the overall size of the crystal; A is the potential barrier that limits electron transition from one atom to its neighbor; and B is the potential well.

atom) from leaving their atom and starting to move freely through a crystal under the influence of heat or other impacts.

As shown in the simplified diagram in Fig. 4.18, valence electrons belong to the whole crystal and have the same energy. At first glance, this contradicts the Pauli principle. However, experiments show that the *emission spectra* of metal are not discrete (as it is seen for atomic spectra), but they are continuous.

Thus discrete energy levels of atoms become split, and they form the *band* (or *zone*) consisting of the same number of separated levels as there are atoms in the crystal. It is expected that the crystal is characterized by as many energy bands as the energy levels that have isolated an atom of substance (see Fig. 4.17). In this example, in 1 cm^3 of sodium crystal, the number of electronic levels in any band equals 3×10^{22} ; therefore all valence electrons occupy different levels, but in the same area $3s$.

As can be seen from Fig. 4.19, the valence band holds N with narrowly located energy levels; in accordance with Pauli principle, this band can accommodate $2N$ electrons. Therefore, levels in the valence band are only *half filled*, because the separated sodium atom has only one valence electron. In addition, specifically for sodium, the width of the highly placed bands corresponds to the number of $3s$ and $3p$ levels that overlap each other. Therefore some electrons move out from the $3s$ zone to lower levels of the $3p$ zone such that both zones are filled together until the entire stock of electrons is exhausted (other energy bands, located above areas that overlap, are not shown in Fig. 4.19 in order to simplify the figure). Thus the valence band of sodium crystal is not fully occupied by electrons. The top energy level, which in metals is occupied by electrons at temperatures $T=0$, is the *Fermi level*, E_F [9].

The valence electrons of metal are not located near their individual atoms, but move freely around a crystal similar to gas molecules in a certain vessel. This system of electrons in metals is the *electronic gas* (or quantum electronic liquid). The Fermi

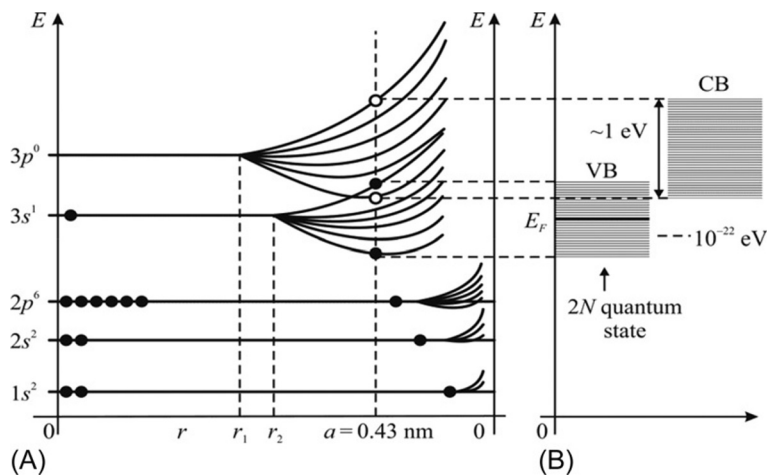


FIG. 4.19

Electronic energy spectra of system of N atoms of sodium (Na) depending on the distance between them; CB, conduction band; VB, valence band; E_F , Fermi level.

level in metal plays the same role for “electronic liquids” as the level of fluid in communicating vessels. If two crystals with different Fermi levels touch each other, electrons will “flow” from one crystal to another until their Fermi levels are aligned. A clearer definition of the Fermi level position is presented in thermodynamics.

If an electrical field is applied to a metal, electrons can easily change their energy states, going from one level to another (located very close). Electrons, in addition to their random thermal motion, move in opposite directions to the electrical field that causes an electrical current. The monovalent metal sodium is the simplest case of the location of electronic levels in the energy spectrum of metals. Sodium demonstrates a purely *electronic* conductivity that is verified experimentally by Hall’s effect study and by the definition of a sign of thermoelectromotive effect.

The electronic spectrum of *copper* is not as simple as that in sodium (however, it is not as complicated as the spectra of some rare earth metals). However, as already in the case of copper, the contribution to the conductivity is made not only by free electrons, but also by electronic vacancies—holes. The energy diagram of a copper crystal is shown in Fig. 4.20, where not only a band diagram with overlapping energy levels, but also the formation of energy bands in case of individual atoms coming into contact is shown.

The energy levels of Cu in Fig. 4.20B appear discrete and narrow. However, as atoms converge, the interaction between electrons of outer shells begins and overall energy levels become split, thereby creating the band. With subsequent convergence of atoms, the splitting amplifies, and energy levels become deeper. When the interatomic distance of copper becomes equal to the lattice constant a_0 , the bands $3d$, $4s$, and $4p$ become so extended that they overlap with each other, as shown in

Fig. 4.20A. Deeper energy bands (only the $3p$ level is shown) expand considerably less. In case of the entirely separated copper atom, its energy levels are completely filled up to the $3d$ level; on the $4s$ level (that can accommodate two electrons), only one valence electron of each atom is located. In the copper crystal, the three top areas are united and completed.

In the described cases, the valence band is found to be partially filled with electrons (Na), or to have some overlapping energy bands with the formation of a broader band of levels that remain partially unfilled (Cu). In metals that can be characterized by the discussed energy diagrams, the electrons are free; therefore these materials are good conductors of electrical current.

Thus the interaction of atoms in a solid significantly changes the electronic energy spectrum. Highly located discrete energy levels of isolated atoms are changed into wide energy bands (zones) when atoms are interconnected in a crystal. The dependence of potential energy on coordinates $U(x,y,z)$ radically changes: it becomes periodic. The neighboring atoms of a crystal change each other's potential such that it turns into a periodical set of potential barriers and potential wells (Fig. 4.18). Furthermore, the interaction between atoms causes changes in the initial position of discrete quantum states and splits them into separate closely located energy bands. The permitted band, in which valence electrons are located, is the *valence band*. In the sodium crystal, this band is formed as a result of $3s$ -level splitting.

Consequently, during crystal lattice formation, all peculiar electronic levels for a given type of atoms (as filled by electrons, or therefore unfilled) are displaced as a result of the neighboring atoms' influence on each other. Due to the convergence of atoms, the electronic energy levels of individual atoms become separated into the bands of energy levels of electrons.

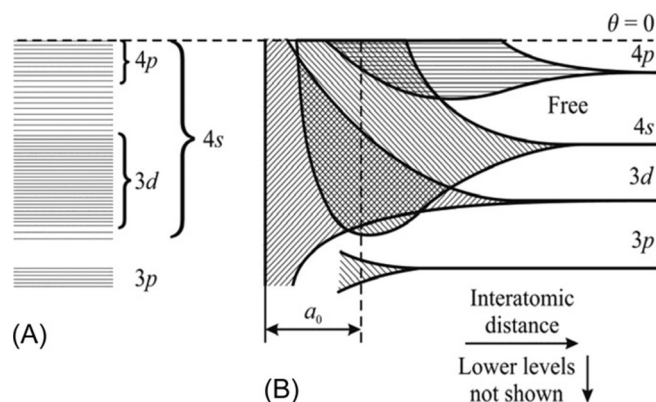


FIG. 4.20

Energy bands of copper overlapping (A) and splitting of energy levels in case of copper atoms coming into contact (B); a_0 —lattice constant (bands are arbitrary shading, lower levels are not shown).

4.7 ELECTRONS IN METALS, DIELECTRICS, AND SEMICONDUCTORS

The foregoing analysis of electron behavior in simple monatomic crystals opens a possibility to draw a preliminary conclusion as to the *cardinal types of solids*. The fact is that features of the energy electronic spectrum cause a large difference in the electrical, optical, thermal, and mechanical properties of materials.

Metals and dielectrics, due to the fundamental distinction in the nature of their atomic connections, differ significantly from each other in thermal and mechanical properties* as in their electrical and optical properties.

**Note.* However, it should be noted that, very seldom, it is possible to encounter crystals wherein the energy barrier between dielectric and metallic states is not large, and these materials can exist in both states. Moreover, some solid materials undergo phase transition of the “dielectric-metal” type. At these transitions, the conductivity jumps by thousands and millions of times, and this property is used in electronic devices [5].

Initially, it is better to compare *electrical properties*: conduction and polarization. The temperature dependences of conductivity σ in dielectrics and metals are shown in the Introduction (Fig. 1.8). These dependencies are opposite: while in dielectrics, σ increases with temperature according to the exponential law (because thermal motion in a crystal generates new charge carriers), in metals, owing to the charge carriers scattering on the thermal vibrations of the crystal lattice, conductivity decreases approximately as $1/T$.

Therefore when metal is cooled to a low temperature, its conductivity greatly increases, tending to infinity (superconductors really have $\sigma = \infty$). In dielectrics, on the contrary, σ value is close to zero at very low temperatures, because free charge carriers are not generated in dielectrics if the intensity of thermal motion is small (and there is no radiation exposure). Similarly as in dielectrics, the conductivity of semiconductors at low temperatures tends to be zero.

Electrical polarization (which is the most important phenomenon for dielectrics) does not occur in metals due to the high concentration of free electrons, which form an almost free “electronic gas” around positively charged ions. The electronic gas in metals gives rise to an almost complete screening of the electrical field. Only at very high frequencies, much higher than the frequency of visible light (i.e., $> 10^{16}$ Hz), the electronic gas in metals demonstrates its sluggishness: it has no time to interact with the extremely fast change of the electromagnetic field, and $\sigma \Rightarrow 0$. Thereby, it is possible to notice the polarization of deep electronic shells, which are located closer to ion nuclei. Such polarization, occurring at frequencies higher than the optical range, determines the specific permittivity in metals.

Comparing *optical properties* of metals and dielectrics, it should be noted that free electrons in metals cause almost a complete reflection of electromagnetic waves from the surface of metals, which explains their metallic shine. In contrast, electromagnetic waves of optical frequency can easily penetrate into dielectric substances, and the majority of them are optically transparent (the color and opacity of some

dielectrics are due to the presence of impurities that absorb or scatter light by their inhomogeneous structure).

A significant discrepancy between *dielectrics and semiconductors* can be seen in the frequency dependence of absorption of electromagnetic waves in these materials. Dielectrics are transparent in their optical wavelength range: their fundamental absorption is observed solely in the ultraviolet wavelength region. Only at a very high frequency (10^{16} Hz) does the energy of photons exceed the bandgap in the electronic spectrum of a dielectric, whereby both photoconductivity and light absorption appear. In semiconductors, the absorption and reflection of electromagnetic wave start at approximately 10^{14} Hz (in the near-infrared region); however, semiconductors, unlike dielectrics, have good transparency in the *far-infrared* wavelength range.

The ***thermal properties*** of dielectrics and metals differ mainly in the value of their thermal conductivity. The very high thermal conductivity of metals is due to the participation of free electrons in heat transfer, whereas in solid dielectrics, heat passes mainly through crystal lattice vibrations (phonons). The magnitude of the *thermal expansion* and *heat capacity* of metals and dielectrics are not very different: due to quantum effects, the specific heat of electronic gas in metals is very small as compared with the specific heat conditioned by lattice vibrations.

With regard to ***mechanical properties***, crystalline dielectrics are more fragile, while metals are usually pliant. This is also due to the impact of free electrons on the properties of metals, which crystallize in simple, densely packed lattices, where the overwhelming strength of interaction is the metallic bond (other types of electrical bonds between atoms in metals are shielded by free electrons). In contrast, dielectrics have complicated polyatomic structures with different physical natures of interaction in their structural elements.

Several investigations of dielectrics and metals have shown that the main differences in their properties are conditioned by the presence of free electrons in metals and the complicated atomic bonding in dielectrics. A more rigorous deduction of the difference between the properties of metals and dielectrics is explained on the basis of the energy-band theory.

The ***energy-band structure*** of electrons in crystalline dielectrics and metals is qualitatively different. As atoms approach each other and form a crystal, many levels of electronic energy appear. Due to the interaction of electrons, the splitting of energy levels takes place, forming zones (bands; Figs. 4.19 and 4.20). This cleavage occurs mainly in those energy levels that correspond to the outer (valence) electrons as they have much stronger interactions with each other than electrons of the deep shells of an atom. The type of electronic spectra of crystals depends on the peculiarities of atomic wave functions and on the degree of overlap of those functions when atoms approach each other during crystal formation.

In the theory of electronic energy spectra, the *one-electron approximation* is typically used: it is assumed that each electron moves in a force field of ions and electrons, while individual (pair) interactions are not taken into account even between the nearest neighboring electrons. The interactions are taken into account as a so-called *middle field*. In this case, the solution of Schrödinger's Equation in the periodic

potential of the crystal lattice is the Bloch function, and the eigenvalue spectrum of electrons forms the energy bands (Figs. 4.19 and 4.20).

The number of levels in each band is determined by the number of atoms in the lattice, thereby forming *quasicontinuous* energy bands. According to the Pauli principle, only two electrons (with opposite spin values) can coexist in each level of a zone; at $T=0\text{K}$, electrons occupy states with minimal energy in each energy band.

The electronic energy spectrum of crystals, that is, electron energy distribution in permitted bands, is usually described in the quasimomentum space, that is, in the reciprocal lattice. The dispersion law $W(p)$ for free electrons is the dependence of electron energy W from their momentum $p = \hbar k$, where k is the wave number. In case of free electrons, the function $W(p)$ is a simple parabolic function:

$$W = \frac{\hbar^2 k^2}{2m} = \frac{p^2}{2m},$$

where m is the mass of the electron. Accounting for the periodic potential of the crystal lattice (Bloch method) complicates this relationship, resulting in breaches of parabolic dependence $W(p)$ in the area of the forbidden energy band (Fig. 4.21). The function $W(p)$ is continuous only in definite intervals of momentum space, namely, in the Brillouin zones (the first zone corresponds to $\pi/a \leq k \leq \pi/a$). During transition from one to another Brillouin zone, this function is terminated.

The one-electron band theory with Bloch wave functions perfectly agrees with and is justified in crystals with *s*- and *p*-electrons that have a *big orbital space* with *significant overlap*. In crystals with *d*- and *f*-orbitals, this band theory might be applied with caution.

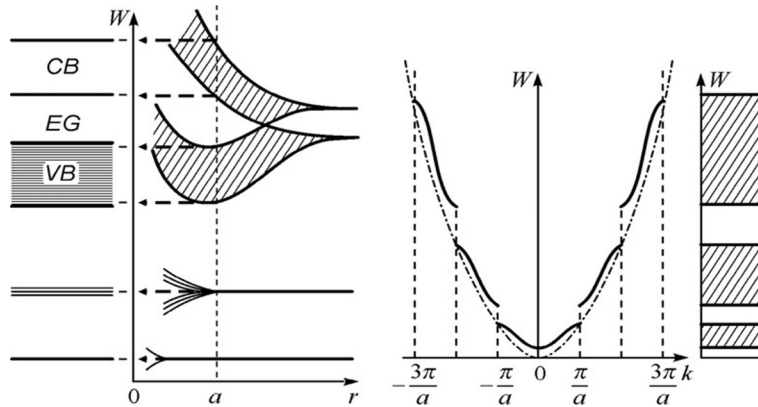


FIG. 4.21

Splitting of energy levels of electrons of isolated atoms, energy band formation due to atom convergence: *CB*, conduction band; *EG*, energy gap; *VB*, valence band; a , lattice constant; W , electrons energy; r , distance between atoms.

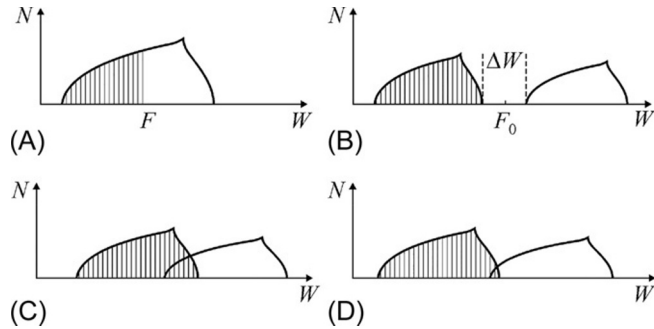


FIG. 4.22

Electronic levels in spectra (filled levels are shaded): (A) “true” metal with odd number of electrons in unit cell; (B) dielectric or semiconductor with gap ΔW between the valence band and the conduction band; (C) metal with even number of electrons in the unit cell; and (D) semimetal.

The energy-band structure of the electronic spectrum allows the construction of models of different variants of electronic spectra of crystals. There are three main cases:

1. Energy bands of the electronic spectrum do not overlap (Fig. 4.22A and B).
 - a. Crystals with an *odd number* of electrons per unit cell of crystal have an upper energy band filled to exactly half (Fig. 4.22A). These crystals are metals; in each energy level, two electrons can be placed (according to the exclusions principle). Thus the energy band has $2N$ vacancies, half of which is occupied by electrons: electrons occupy the lowest energy levels. In the ground state (when $T=0\text{K}$), the boundary of filling that separates in the impulse space that is the filled part from the unfilled part of the valence band is the *Fermi level* F (in the three-dimensional model, F corresponds to the Fermi surface). If $T>0\text{K}$, the boundary of the Fermi surface becomes smeared as a result of thermal perturbations (phonons), and part of the electrons goes on to levels above F (therefore some levels below F are released). Because the distance between the levels in band is extremely small ($\sim 10^{-23}\text{eV}$), even a very small external electric field can increase the energy of electrons and cause electrical conduction in metals (limited only by electrons scattering due to lattice vibrations). With decreasing temperature, the conductivity of metals increases: if temperature $T \rightarrow 0$, then conductivity $\sigma \rightarrow \infty$.
 - b. Crystals with an *even number* of electrons per unit cell are dielectrics or semiconductors (Fig. 4.22B). In the ground state (at $T=0\text{K}$), their energy bands are completely filled or empty. Therefore the electrical field cannot change the energy of electrons in the filled bands (because all levels are filled), while in the empty bands there are no charge carriers. Consequently,

if the temperature is critically reduced ($T \rightarrow 0$ K) in dielectrics or semiconductors, conductivity is absent ($\sigma \rightarrow 0$). The upper filled band (valence) and the nearest empty band (conduction band) are separated by the energy gap ΔW (forbidden band; Fig. 4.22B).

In crystals with energy gaps, the Fermi *surface* in the electronic spectrum is absent; however, in the middle of a gap (when there are no impurities and local levels), there exists the Fermi *level* F_0 (Fig. 4.22B). To excite electrical conductivity in these crystals by thermal vibrations or by other factors, it is required that the valence band is partially released from electrons (*holes* mechanism of electrical conductivity) or that the conduction band is partially filled by electrons (*electronic* conductivity mechanism).

2. Bands of electronic spectrum overlap (Fig. 4.22C and D).

Such crystals, similar with even or odd numbers of electrons per lattice site, are referred to as metals. Significant overlap of two bands (Fig. 4.22C) results in a situation that is not very different from the case shown in Fig. 4.22A. In the event of a *small overlap* of bands, the crystals belong to the class *semimetals* (Fig. 4.22D). The Fermi surface for semimetals has discontinuities, and their conductivity by several orders of magnitude is lower than the conductivity of metals. For example, in the semimetal bismuth, the number of filled states in a conduction band is 10^4 times smaller than in conventional metals, and, consequently, bismuth shows much lower conductivity. Other examples of semimetals are antimony and graphite.

3. Bands of energy spectrum are in a contact without overlapping.

Crystals of this rare class are *gapless semiconductors*. The Fermi surface of such semiconductors is a line or a point in the impulse space (whereas, in semiconductors, such a surface does not exist and, in semimetals, this surface has discontinuities). In the semimetal under the influence of an electrical field, electrons move within their area, but the lower density of states reduces their *mobility*. In the gapless semiconductor, electrons relatively easily (as compared with conventional semiconductor) come into the conduction band, but the dynamic properties of charge carriers in these materials are significantly modified.

Therefore crystals that, in the ground state, have no partially filled bands belong to the class of dielectrics or semiconductors. Metals and semimetals, in contrast, are characterized by an electron spectrum with *partially filled* bands.

A comparison of the electronic spectra of metals, semimetals, semiconductors, and dielectrics is shown in (Fig. 4.23), which demonstrates the energy spectra of electrons in these materials. In metals, there is no energy gap between the valence and conduction bands; therefore electrons can easily change their energy, moving from level to level; thus they are free. Electrons in metal are not localized—they belong to the entire crystal and do not form spatially directed bonds between ions.

In all other crystals, most of the electrons are, to some extent, localized. In semimetals, however, excitation energy is almost zero; therefore, even at temperature $T \rightarrow 0\text{K}$, mobile electrons exist and can provide essential conductivity. However, some electrons in semimetals are localized between atoms and form spatially directed linkages.

Valence electrons in semiconductors (which are mainly covalent crystals) form the directed orbitals to link atoms, and their excitation energy (energy gap ΔW) usually exceeds the thermal energy ($\Delta W > k_B T$). However, in semiconductors, this energy gap is smaller than the energy of visible light ($\Delta W < 3\text{eV}$).

Valence electrons in dielectrics (which are predominantly ionic and molecular crystals) are localized much stronger than in semiconductors. Thus they are localized not at the bonds between atoms (as in the case of semiconductors) but near individual molecules or anions. The binding energy of electrons in dielectrics far exceeds not only their thermal energy ($\Delta W \gg k_B T$), but also the energy of the visible light quantum: $\Delta W > \hbar\nu$. Therefore the probability of electron excitation in dielectrics by thermal motion and even by light is very small. Moreover, the small curvature of the band frontiers in the vicinity of their extremes in dielectrics (Fig. 4.23A) gives rise to increased effective masses of charge carriers, which result in the low mobility of electrons.

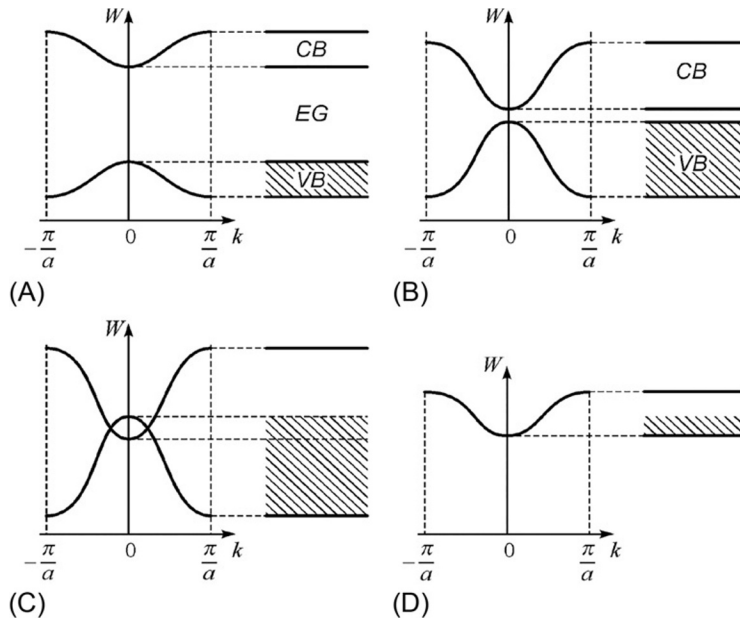


FIG. 4.23

Comparison of energy bands: (A) dielectric, (B) semiconductor, (C) semimetal, and (D) metal. *CB*, conduction band; *VB*, valence band; *EG*, energy gap.

Dielectrics and semiconductors are qualitatively similar: they both have an energy gap in the spectrum of their electronic states. However, in semiconductors, this band gap is much smaller. Therefore the conductivity in semiconductors takes a wide interval, separating the value of conductivity of metals and dielectrics. For example, pure lead at temperature 300 K has a conductivity $\sigma = 5 \times 10^6$ S/m whereas in pure germanium conductivity is $\sigma = 2.5$ S/m. That is, the conductivity of semiconductors is approximately a million times lower than the conductivity of metals, but the conductivity of semiconductors is greater than in insulators.

In semiconductors, the $\sigma(T)$ dependence can acquire a “metallic character” only in exceptional cases and in a narrow temperature range; in general, the temperature dependence of conductivity in semiconductors and dielectrics is similar. The width of the energy gap of germanium is 0.72 eV, in silicon it is 1.12 eV, while in the diamond (a dielectric of the same crystal structure as silicon and germanium), the energy gap is approximately 5 eV. If the band gap $\Delta W \leq 3$ eV, the crystal can be regarded as a semiconductor, while with larger values of ΔW , it is a dielectric.

The qualitative difference in the band gap and conductivity results in significant differences between the optical, magnetic, and electrical properties of dielectrics and semiconductors. In the visible optical range, dielectrics are light transparent and only a few reflect light, while semiconductors have an almost metallic reflection but a dull sheen. The reason for this lies in the fact that the narrow energy gap of semiconductors allows light quanta with energy of approximately 2.5 eV to excite free electrons, which results in light reflection. In dielectrics, such reflection is possible only in the eye-invisible ultraviolet part of the spectrum.

The covalent crystals of semiconductors (e.g., silicon), however, in contrast to ionic dielectrics, have good transparency in the infrared region of the spectrum, as the energy of photons of this frequency (10^{12} – 10^{14} Hz) is insufficient to excite free electrons. Therefore in the far-infrared electronic devices, silicon and germanium can be used as a transparent material for optical elements (lenses). Consequently, typical silicon and germanium semiconductors in the far-infrared range play the role of “perfect dielectrics.” However, glasses and ionic crystals, commonly used in visible optics, cannot be used in the far-infrared range because they strongly absorb and intensively reflect these electromagnetic waves. Thus in the far-infrared range, the own oscillation frequencies of the ionic crystal lattice are located, and this causes absorption of these waves.

Thus it would not only roughly divide materials into dielectrics and semiconductors, but rather would distinguish the semiconducting and dielectric *properties* of crystals that have an energy gap in the spectrum of electronic states.

4.8 SUMMARY

1. A crystal is the aggregate of regularly spaced and *strongly interacting* particles. However, any oscillations and other excitations of these particles can extend through the crystal in the form of *weakly interacting* waves with

wave vectors k_j and frequencies $\omega_j(k_j)$. Each wave can be associated with an oscillator that has a certain frequency.

2. To explain the main characteristics of solids, one should imagine that they contain some “hidden” states that resemble the properties of different aggregate states of a matter, namely, the *gas* of quasiparticles (atomic oscillations), the quantum *fluid* (electrons in metal), and even the electron-hole *plasma* (in semiconductors).
3. The *quasiparticle* is the collective movement (disturbance) of many closely located particles of solids, such as local vibrations of neighboring atoms in the crystal lattice. Although many atoms are involved in each excitation, this movement, nevertheless, has an atomic scale, as the mean energy of each excitation (phonon) is approximately $k_B T$.
4. The energy distribution in quantum systems is expressed as a function of energy, the degree of degeneration, and the number of particles in system. For particles, whose number in any state is unlimited, a special case of quantum statistics—the Bose-Einstein distribution—is valid (such particles are *bosons*). If particles are subject to the Pauli principle (i.e., in a certain state, only one particle can exist), the Fermi-Dirac distribution is applicable, and particles are *fermions*.
5. The *photon* is a typical example of bosons—that is, an electromagnetic wave that can extend both in vacuum and in dielectrics. The photon, as well as the electron, shows dualism, sometimes revealing the properties of the particles. The corpuscular property of the phonon is its *impulse*, whereas its wave property is a *wave vector*. Both of them are related by the de Broglie ratio: $p = \hbar k$, and this ratio can be read inversely as: $\hbar k = p$. The spin of the photon is integer-valued: the photon can have only two states of spin: +1 and –1. The two spin states of a photon means a right and left circular polarization of the electromagnetic wave, respectively; this fact is important for an understanding of some electro-optical and magneto-optical effects in solids.
6. The *phonon* is a quantum-mechanical description of elementary oscillation movement in the crystal lattice, wherein some adjoining atoms oscillate with a single frequency. The comparison “*wave* → *quantum oscillator* → *phonon*” is arranged in such a way that the energy of the excited state of each quantum oscillator E_n is an integer quantity of $\hbar\omega_j(k)$: $E_n = n\hbar\omega_j(n + 1/2)$, where $n = 0, 1, 2, 3, \dots$ (here, n is the number of a certain type of phonons with impulse $p = \hbar k$ and energy $E = \hbar\omega_j$). The average number of phonons with impulse p and energy E is determined by the Bose distribution and is proportional to $[\exp(\hbar\omega/k_B T) - 1]^{-1}$. Thus the ensemble of interconnected harmonic oscillators that describe atomic oscillation in crystals can be expressed by a set of so-called normal (not interconnected) oscillators, whose number equals the number of degrees of freedom of a system.
7. A model of phonons, in many cases, allows consideration of any solid as a “vessel” containing in it the “gas of phonons.” Similar to conventional gas particles, phonons can move from “wall to wall,” facing each other. The gas of

phonons is the main *heat reservoir* of a solid; however, it differs from the conventional gas: the number of phonons in a crystal is not a constant (whereas, in normal gas, the number of molecules in the vessel is invariable). The greater the number of phonons, the more intense the thermal motion of atoms—the higher is the temperature. At high temperatures, the energy of phonons increases in proportion to temperature T ; however, at low temperatures (closer to absolute zero), their number tends to zero in proportion T^4 —the fourth degree of temperature.

8. There are *acoustical* phonons and *optical* phonons. It is obvious that phonons cannot leave the crystal because they are only the motion of the atoms of a crystal. At low temperatures, when the quantum description of crystal properties is necessary, the number of thermally excited optical phonons is very small because the heat energy is inadequate for their formation. Therefore the *acoustic phonons* determine the heat capacity and thermal conductivity of a crystal.
9. Phonons differ from photons, firstly, by a relatively *low velocity*: their velocity corresponds to the rate of sound, being in four orders of magnitude less than light velocity. Secondly, phonons differ from photons by their distinction among the types of waves. Namely, the elastic wave (phonon) has one of *three types* of polarizations ($L+2T$), whereas light wave (photons) has one of *two types* of polarizations ($2T$).
10. In the space of impulses at low temperatures ($T < \theta_D$), thermally excited phonons occupy only a small area near the center of the Brillouin zone. As temperature increases, the number of phonons increases as well; as the temperature becomes higher, the space of impulses gets more uniformly filled with phonons. In the range of low temperatures, the main contribution to the oscillation energy of a crystal is given by *long acoustic waves*. The energy of the corresponding oscillators is small; therefore they are easily excited. On the contrary, short acoustic waves and optical waves at low temperatures are practically not excited: there is not enough heat for their stimulation in the temperature range $T < \theta_D$.
11. The *magnon* or *spin wave* is quasiparticle that can be ascribed to the properties of such crystals that have an orderly arrangement of spins: ferromagnetics, antiferromagnetics, and ferrimagnetics. Magnons are generated by thermal motion that excites elementary magnetic excitations in a crystal, when the spins of some electrons do not coincide with their magnetized ground state. This magnetic excitement can move in a crystal lattice from one place to another due to exchange interaction, and it is characterized by quasi-impulse and by energy. The properties of magnons can be described in Bose statistics. In crystals that have many atoms in a unit cell, there are *several magnon branches* (by analogy with relevant branches of phonons).
12. Magnons interact with each other and with other quasiparticles. The existence of magnons is confirmed by experiments using neutron scattering, as well as from electrons and light scattering in magnetic states, during

which spin waves are excited. The model of magnons is used in solid-state physics to explain not only magnetic properties, but also to explain some thermal and high-frequency properties of magnetics. For example, the increased heat capacity (C_v) in ferromagnetic crystals below Curie temperature (including C_v maximum in Curie point) is due to the fact that (in addition to phonon contribution to heat capacity) a similar contribution is made by magnons.

13. Free electrons in space can have any energy: they have a *continuous* energy spectrum. However, electrons in an isolated atom, according to quantum mechanics, have *discrete* values of energy. According to the Bohr postulate, in an isolated atom, the energy of the electron can take only strictly discrete values (respectively, one can assume, that the electron occupies one of some possible orbitals). In several atoms combined by chemical bonds (i.e., in molecule), electronic orbitals split in an amount proportional to the number of atoms, forming the so-called molecular orbital. A similar discrete electronic spectrum is characteristic of nanoparticles, inclusive of dozens of atoms.
14. In a macroscopic crystal—a solid body with tightly bound atoms—the number of possible electronic orbitals becomes very large such that the electronic energy spectrum consists of a large number of levels, joined in the *permitted* energy bands that are separated by the *forbidden* energy bands. Because the difference in energies of electrons for adjacent orbitals is very small, the energy levels are split up almost continuously and make discrete sets (energy bands). One of them is the *valence band*: at temperatures close to zero in dielectrics and semiconductors, electrons occupy all their energy states. In metals, the highest allowed band is the *conduction* band, wherein levels of conduction electrons are located.
15. At the heart of band theory, there are some approximations: firstly, it is believed that a solid is a perfectly periodic crystal; secondly, it is assumed that the equilibrium position of crystal lattice is fixed (nuclei during fast motion of electrons are practically immobile—this is an *adiabatic approximation*); finally, the many-electrons system is reduced to a one-electron task (impact on a given electron from all other electrons is accounted by an averaged periodic field).
16. The *band theory* is the foundation of the modern theory of solids. It allows an understanding of the physical nature and explains important properties of conductors, semiconductors, and insulators. The value of band gap E_g (energy gap between valence and conduction bands) is a key value in band theory; it predetermines the electrical and optical properties of dielectrics and semiconductors.
17. In various crystals, as well as in different forms of same crystal, energy bands are different. With the relative position of these bands, all substances are divided into three groups:
 - the *conductors*—in which the conduction band and valence band overlap, forming a zone that is called the conduction band; thus electrons

can move freely occupying higher levels in this band when received even at low energy (in case of a potential difference applying to conductors, electrons are free to move from a point of lesser potential to a point with higher potential, creating an electrical current);

- the *dielectrics*—in which electronic energy bands do not overlap, and the distance between them is more than $\sim 3\text{eV}$; to transfer an electron from the valence band into the conduction band, a considerable energy is required; therefore dielectric-insulators practically cannot conduct electricity;
 - the *semiconductors*—their bands do not overlap, but the distance between them is smaller than $\sim 3\text{eV}$; to transfer an electron from the valence band into the conduction band, much less energy (than in dielectric) is required; therefore in chemically pure semiconductors, only weak electrical current can pass.
18. In a more general approach of solid-state theory, it turns out that the prediction of various physical effects by band theory methods is much wider from the initial approximations. For example, small fluctuations of atoms around their equilibrium positions (which can be described as phonons) can create perturbations in the electronic energy spectrum. However, a bunch of many-electron physical phenomena, such as ferromagnetism, superconductivity, and others, where the role of excitons offer benefits, cannot be consistently reviewed as part of band theory.

REFERENCES

- [1] L.H. Van Vlack, *Materials Science for Engineers*, Addison-Wesley Publishing Co., Massachusetts, 1975.
- [2] C. Kittel, *Introduction to Solid State Physics*, John Wiley and Sons, New York, 1976.
- [3] A. Holden, *The Nature of Solids*, Dover Publications, New York, 1968.
- [4] M.I. Kaganov, V.M. Tsukirnik, *The Nature of Magnetism*, Nauka, Moscow, 1952.
- [5] Y.M. Poplavko, *Physics of Active Dielectrics*, LAMBERT Academic Publishing, Saarbruken, 2015.
- [6] J.S. Blackmore, *Solid State Physics*, Cambridge University Press, Cambridge, 1985.
- [7] J.C. Sletter, *Insulators, Semiconductors and Metals*, McGraw-Hill, London, 1967.
- [8] M. Born, K. Huang, *Dynamical Theory of Crystal Lattices*, Clarendon Press, Oxford, 1988.
- [9] J.M. Ziman, *Electrons and Phonons*, Clarendon Press, Oxford, 1960.
- [10] M. Ali Omar, *Elementary Solid State Physics: Principles and Applications*, Addison-Wesley Publishing Co., New York, 1993.

Metals

CONTENTS

5.1 Defining Features of Metals	166
5.2 Electrical Conductivity of Metals	168
5.3 Thermal Properties of Metals	176
5.4 Electronic Properties of Metals and Fermi Surface	180
5.5 Electron Scattering in Metals	187
5.6 Special Electronic States in Metals	192
5.7 Superconductivity in Metals and Alloys	204
5.8 Summary	215
References	220

The most important property of metals is their large electrical conductivity. However, it should be noted that in addition to metals good conductors of electrical current might be also other solids, liquids, and even ionized gas (plasma). Among nonmetal solid conductors, there are some modifications of carbon and metal-oxides (the latter are usually used at very high temperatures). However, metals and their alloys, of course, are most important conductive materials that are applied in electrical engineering, electronics, and instrumentation. Due to high conductivity, metals are used in chips (as joining), in wires and cables, windings of transformers, microwave waveguides, generator tubes, etc.

In some cases, it is necessary to employ the *very low resistive* metals—hyperconductors and superconductors. On the other hand, metals with *high resistance* are also widely applied: in resistors and electrical heating elements. Sometimes, *liquid conductors* are also of technical interest: they are various electrolytes and molten metals. However, for most metals, rather high melting point is peculiar; only mercury and some special alloys (e.g., indium-gallium alloy) can be applied as liquid conductors at conventional temperatures.

The mechanism of current flowing in metals—as in both solid and liquid phases—is due to the movement of electrons; therefore, they are called as conductors with *electronic conductivity*.

5.1 DEFINING FEATURES OF METALS

The term “metal” originated from the Greek word “metallon,” which means “mine.” Distinctive properties of metals are high electrical conductivity, ability to reflect light (shine), mechanical plasticity and flexibility, as well as large thermal conductivity.

Most chemical elements (simple substances) are metals, and many alloys of these elements and their compounds are also metals. Sometimes, other substances can be referred to as metals, having one or other of metallic properties, and they are called “synthetic metals” (intercalated), “organic metals,” and others. Of 119 elements of Mendeleev’s periodic table, 92 are metals. The boundary between metals and nonmetals in this periodic table has a diagonal from B to At. Some elements, such as germanium (Ge) and antimony (Sb), are difficult to be qualified; however, Ge is considered as a semiconductor, while Sb is a semimetal. It is interesting to note that tin can exist in metallic modification (β -Sn), so in a semiconducting phase (α -Sn).

However, in Ge, Si, P, and some other “nonmetals” another modifications can be obtained under *increased pressure* that exhibit properties of metals. Moreover, at super high pressure all substances must acquire properties of metals [1]. To find out whether any material is metal or nonmetal, not only physical properties but also *chemical* properties should be taken into account. Sometimes, for elements that lie on the border between metals and nonmetals the term *semimetal* is used.

Earlier, mostly specific shine, plasticity, and malleability were considered as characteristic features of metals. However, metallic shine might be seen in some nonmetals and semiconductors as well. Plasticity also cannot be a reliable defining feature of metals, as many brittle metals are known. Therefore the *negative temperature coefficient* of electrical conductivity should be considered as the main feature of metals (i.e., electrical conductivity decrease with temperature rise) [2].

Metals are characterized by a special type of bonding—*metal type connection* (see Section 1.1), in which crystalline lattice is formed by the positive ions, while valence electrons are delocalized throughout a lattice space. Therefore metals can be presented as the lattice of positive ions crowded by “gas of electrons” that compensates forces of mutual repulsion of positive ions.

According to structure of electron shells, metals can be divided into four groups: *s*-metals (all *s*-elements, except H and He); *p*-metals (elements of third group, except B) and Sn, Pb, Sb, Bi, Ro; *d*-metals (transition elements); and *f*-metals (also transition elements but of lanthanide group) [3]. Metals of the first two groups are sometimes called as “simple metals.” In these groups, some narrower *subgroups* can be highlighted. Among *s*-metals, there are *alkaline* metals and *alkaline earth* elements; among *d*-metals, we find *platinum* subgroup of metals. The group of *rare earth elements* includes *f*-metals, Sc-subgroup, and lanthanides.

Most metals are crystallized in one of three structural types, namely, cubic, hexagonal dense packing, and space-centered cubic lattice. In case of dense packing, each ion of metal at equal distances has 12 nearest neighbors. In the space-centered

cubic lattice, each ion has eight equidistant neighbors, while another six neighbors are located at slightly more distance (15%). Therefore the coordination number of this structure is considered equal to 14 (8+6). Interatomic distances in metal structure are characterized by the “metallic ionic radius.”

After melting, metals, basically, retain their electrical, thermal, and optical properties (this indicates the importance of *short-range order* in arrangement of atoms). Near melting temperature in liquid metals approximately the same short-range ordering is preserved as in crystalline metals. However, with increasing temperature, this short-range ordering is disrupted until complete disorder.

Physical properties of metals vary widely. For example, *melting temperature* can be found between -39°C (Hg) and $+3380^{\circ}\text{C}$ (W), while metal *density* might be from 0.53 g/cm^3 (Li) to 22.5 g/cm^3 (Os). Specific electrical *resistance* ρ of metals at normal temperature has a magnitude between $1.6\text{ }\mu\text{Om sm}$ (Ag) and $140\text{ }\mu\text{Om sm}$ (Mn). However, temperature coefficient of resistance does not vary much: from $4 \times 10^{-3}\text{ K}^{-1}$ (Hg) to $9 \times 10^{-3}\text{ K}^{-1}$ (Be).

As for the special effects, in metals *thermoionic emission* can be observed, that is, the ability to emit ions at high temperature. Electronic emission also occurs under the influence of electromagnetic radiation in visible and ultraviolet regions of spectrum (*photoelectronic emission*). Under the influence of external electric fields of high intensity, *autoelectronic emission* is possible. During metal surface bombardment by electrons, the *secondary electron emission* occurs, while ionic bombardment results in the *ion-electron emission*. Finally, when metal surface interacts with plasma, the *explosive electron emission* can be observed. *Thermal-EMF* (electromotive force) is caused in metals under temperature influence.

Optical range radiation is almost entirely reflected from metal surface such that metals are *opaque* and have a peculiar *metallic luster*. Being reflected from metal surface, the plane-polarized light becomes elliptically polarized. Some metals, such as gold (Au) in a form of very thin foil, can be light translucent.

To use metals as *constructive materials*, a combination of mechanical properties (plasticity and viscosity) with considerable strength, hardness, and elasticity is essential. These properties depend not only on the chemical composition and purity of a metal, but also on the perfection of its crystal lattice (presence of defects), as well as on other features of structure, obtained during previous thermal and mechanical processing [4].

In practice, most mechanical properties of metals are determined by the presence of defects, firstly by *dislocations* (see Section 1.2); therefore movement of dislocations in crystal lattice is the main mechanism of plastic deformation of metal. Interaction of dislocations with other defects increases metal resistance to plastic deformation. In the process of deformation, the number of dislocations increases, and correspondingly, resistance of metal to deformation increases (strain hardening). However, stressed state and slander after deformation can be eliminated by metal annealing. Increased tension in the places of dislocations—“thickening”—causes nucleation of cracks that promotes destruction. The most important characteristic

of mechanical properties of metals is the modulus of elasticity (Young's modulus), that is, mechanical tension, which corresponds to a unit of mechanical deformation.

5.2 ELECTRICAL CONDUCTIVITY OF METALS

Regarding electrical field, the main property of any matter is electrical charge transfer, that is, the conductivity—the ability of material to conduct electrical current under the influence of constant voltage (not changing in time). If the substance is placed in an electrical field E [V/m], free charged particles—the *carriers*—under the force $F = qE$ get acceleration, where q is charge of particle; in metals, this is charge of electron: $q = e$. The acceleration of charges is directed toward the vector E for carriers with positive charge $+q$ (i.e., for electronic holes or positive ions), or in the opposite direction for charge carriers with negative charge $-q$. Directed in space motion of electrical charges is *electrical current*.

With regard to electronic conductivity ($q = e$), when only one sign of free charge carriers exists, the *current density* j , that is, electrical charge that flows per unit time through unit area (oriented perpendicular to vector E) equals to:

$$j = nev, \quad (5.1)$$

where n [m^{-3}] is the number of charge carriers per unit volume of substance (carrier concentration); v [m/s] is the *drift velocity*, that is, average velocity of *ordered movement* of charge carriers that arises under electrical field influence. This velocity usually is proportional to the field strength E :

$$v = uE, \quad (5.2)$$

where u is the proportionality factor called the *mobility* of charge carriers, measured in [$\text{m}^2/(\text{V s})$].

With expression (5.2), Eq. (5.1) can be represented as

$$j = \sigma E = E/\rho, \quad (5.3)$$

where σ [S/m] is *specific electrical conductivity*, $\rho = 1/\sigma$ [Ohm m] = [Ω m] is *electrical resistivity* ([S] = Siemens is SI unit of conductivity). Eq. (5.3) is *Ohm's law*. Specific conductivity σ or resistivity ρ defines current density in material at a given electrical field; at that, the phenomenon of electrical conductivity is the *electrical charge transfer*.

Parameter ρ or σ also determines the scattering process (losses) of electrical power in a matter. According to Joule-Lenz law, the *density of thermal energy* p , [W/m^3], that is, electrical energy that is converted into a heat per unit time and in unit volume, is

$$p = E^2/\rho = \sigma E^2. \quad (5.4)$$

From formulas (5.3) and (5.4) it is possible to pass into the formulas for material *conductance* G , *resistance* R , and *power* P dissipated in a sample of *any size and shape*:

$$P = U^2 G = U^2 / R. \quad (5.5)$$

In commonly used practice, to measure resistivity ρ some outboard units are customary. While unit of specific resistance in SI is $[\Omega \text{ m}]$, for *metals* another unit $[\Omega \text{ mm}^2/\text{m}]$ often is applied, because the cross-section of conducting wire usually is measured in square millimeters (mm^2) and the whole length of a wire l is measured in meters $[\text{m}]$:

$$1 \Omega \text{ m} = 10^6 \mu\Omega \text{ m} = 10^6 \Omega \text{ mm}^2/\text{m}.$$

This unit is very convenient because in commonly used conductors comfortable numeric values are kept. At temperatures near 300 K, the range of ρ for metals is from $0.016 \mu\Omega \text{ m}$ (silver) up to $10 \mu\Omega \text{ m}$ (resistive alloys); this means that ρ in metals covers three orders of magnitude.

The temperature dependence of conductivity. Electrical conductivity of metals varies significantly with temperature (Fig. 5.1). Temperature dependence of conductivity can be described by $\sigma(T) \sim T^{-1}$, but at very low temperatures this dependence is another. The point is that in case of deep cooling dependence $\sigma(T)$ reaches saturation, the level of which depends on the concentration of “static” defects. When cooling metals that have ferromagnetic impurities, at a certain interval of temperature $\sigma(T)$ dependence can even show a decrease (*Kondo effect*).

In electrical engineering and electronic equipment, in addition to high conductivity, *high thermal conductivity* λ_e , $[\text{W}/(\text{K m})]$ (Fig. 5.1), of metals is very important. Thermal conductivity of metals is conditioned mainly by the presence of high-mobility electrons, and, therefore, λ_e is proportional to conductivity. The identity of λ_e/σ ratio for different metals is the *Wiedemann-Franz law*: the ratio $\lambda_e/\sigma T$ is

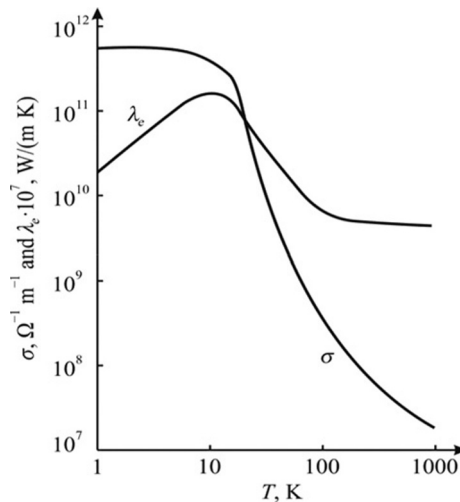


FIG. 5.1

Temperature dependence of conductivity σ and thermal conductivity λ_e in copper.

weakly dependent on temperature (because λ_e is practically independent on temperature). This ratio has the same value for many metals and represents the *Lorenz number*: $L = (\lambda_e/\sigma T)$. For most metals at temperature 300 K, this parameter equals $L \approx 2.4 \times 10^{-8} \text{ W } \Omega \text{ K}^{-2}$.

Frequency dependence of conductivity. Due to large electrical conductivity of metals, they almost entirely reflect electromagnetic waves: their reflection coefficient is $R = [(n - 1)/(n + 1)]^2 \approx 1$, where refractive index is $n = (\epsilon\mu)^{1/2}$. Up to the optical range of electromagnetic waves, conductivity of metals practically does not change with frequency, because electrons have very low inertia (due to small mass) [5].

Since in the range of optical frequencies magnetic permeability $\mu = 1$, very large optical reflection of metals ($R \approx 1$) means that *effective* dielectric permittivity $\epsilon_{\text{ef}} \approx n^2$ is large (and negative). However, in the ultraviolet range of spectrum inertia of “electrons ensemble” (which is associated with ionic lattice) shows *plasma resonance* at frequency ω_{pl} . Frequency of this resonance (located approximately at 10^{16} Hz) is inversely proportional to relaxation time of electrons in plasma: $\omega_{\text{pl}} = 1/\tau_{\text{rel}}$. As a result, in the range of ultraviolet light, conductivity of metals decreases with increasing frequency (Fig. 5.2), and metal gradually becomes transparent for harsh electromagnetic waves (x-rays).

Charge transfer description. Valence electrons in ionic lattice of metals are practically free, because ions form energetically favorable lattice for electron movement. Concentration of free electrons in metals is large (the number of atoms per unit volume is approximately equal to 10^{23} cm^{-3}). Electrons can be treated as particles that weakly interact with each other; virtually they have no volume and move randomly through a crystal. The assumption that electrons practically do not interact with each other, seemingly, contradicts Coulomb repulsion between them. However, the Coulomb *attraction* of electrons to positively charged ions of crystal lattice should be also taken into account. As quantum mechanical analysis shows, these joint actions in the *strongly periodical structure* of crystal make a reasonable assumption about “practically free electrons” (although this supposition has approximate nature).

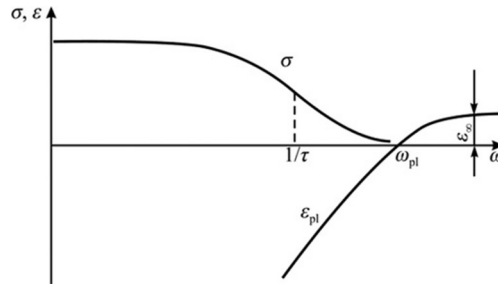


FIG. 5.2

Frequency dependence of conductivity σ and plasma contribution to effective permittivity ϵ_{pl} in the vicinity of plasma resonance in metals (in ultraviolet part of spectrum).

Electronic gas exists in the thermodynamic equilibrium with crystal lattice that is established through the *collisions* of moving electrons and ions in lattice. Electrons in metal are always found in a movement, and they move even at lowest temperatures (near absolute zero) [6]. This quantum motion of electrons is chaotic, and different electrons move with different velocities. Most electrons in metals move with *Fermi velocity* $v_F \sim 10^6$ m/s. This value can be defined from *Fermi energy*, taking into account kinetic energy of electron: $(m_e v_F^2)/2 = E_F$.

In the *ideal* metal with *infinite* conductivity, an electrical field cannot exist. To a *real* metal only very small electrical field can be applied (otherwise, huge electrical current would appear and melt the metal). Under the influence of external electrical field, a current flows in metal, that is, the chaotic movement of electrons becomes *partially directed*: on disordered motion of electrons, their drift is superimposed. To calculate corresponding current, one needs to take into account the *average* velocity v_{av} of electron drift (at that, velocity of random motion of electrons is independent on electrical field).

As current density j is the amount of electricity passing per second through unit area of a conductor, then, according to this definition: $j = -n_e e v_{av}$. If current density is large enough, for example, $j = 1$ A/cm², calculations show that average drift velocity of electrons is only $v_{dv} \leq 10^{-3}$ cm/s. Thus directional movement of electrons in metals is very slow as compared not only with velocity of their chaotic movement but even in the macroscopic scale. The smallness of drift velocity is because only a very weak electrical field can be applied to the metal.

Charge carrier mobility. There exists a direct proportionality between difference of potential applied to metal and caused by its current:

$$j = \sigma E. \quad (5.6)$$

Using [expression \(5.6\)](#) for current density, it is possible to establish that average drift velocity of electrons in a conductor is proportional to the force acting on them:

$$v_{dv} = (\sigma / en_e) E = u E. \quad (5.7)$$

Parameter $u = \sigma / en_e$, that is, drift velocity caused by influence of unit of field, is the *mobility*. Its unit can be clarified from the formula $u = v/E$; as a result, in SI the dimension of mobility is $[u] = [m^2/(sV)]$. It can be seen that such unit of mobility is difficult to relate with the physical meaning of this phenomenon [7]. If one continues to find physical sense of mobility unit using the SI system, it can be written as $[u] = T^{-1}$, that is, return value of “*Tesla*” that is unit of magnetic induction in SI, because $Vs = Wb$ (Weber) and $Wb/m^2 = T$. Another possibility to find mobility unit in SI is also inconvenient for easy interpretation: $[u] = A s^2 kg^{-1}$.

On the contrary, in the *Gauss system* of units (CGSE), the unit of mobility is estimated as [s/g] (second/gram), that is much more simple, hence making the content more understandable. In fact, mobility characterizes the increase of velocity (m/s) under force influence of $v = u \cdot f$. The force is $f = ma$ and in SI has unit [N] = Newton; therefore in the Gauss system force is $1 N = 10^3 g \cdot 10^2 sm/s^2$; then mobility has a simple unit: $u = v/f$, therefore $[u] = [s/g]$.

In fact, mobility indirectly characterizes the opposition of a medium through which electrons drift under influence of electrical field. If any braking force is absent, then the electron under electrical field will move with constant *acceleration* (such as in vacuum), but not with constant *velocity* as they move in crystals. Mobility characterizes how charge carriers, being forced by electrical field to the directed ordered motion, can overcome their thermal chaotic motion that is characterized by continuous collisions with phonons and defects. Therefore mobility is the *degree of freedom* of electron's *directed* motion in a media [7] (while almost nothing prevents the electrons in their chaotic quantum Fermi motion).

Analogy of a current, flowing in a conductor, with a liquid flowing through the pipe indicates that electrons in a conductor move with some similarity of "friction." Thus there are some reasons that violate *field-induced directional* movement of electrons inside a metal. Analyzing these reasons and taking into consideration that mobility should be expressed through specific conductivity, it is possible to obtain

$$\sigma = n_e e u. \quad (5.8)$$

The necessity to express metal conductivity σ by means of two other parameters n_e and u is because each of them can be found in *independent experiments*.

Indeed, the *concentration* of free electrons n_e never changes with temperature: it is a peculiar property of the given metal. Conversely, another parameter, the *mobility* u , can vary with the change of temperature 100 times, even 1000 times. In addition, by cleaning a metal from impurities it is possible to increase electron mobility many fold. For this reason, it is important that two characteristics—number of electrons per unit volume n_e and their mobility u —allow independent measurement.

Hall's effect. One of the most suitable methods to find concentration of electrons n_e is Hall's effect, that is, potential differences across plate-shaped conductor with flowing current, which arises when a conductor (semiconductor) is placed in the transverse magnetic field (Fig. 5.3).

The cause of Hall's effect is Lorentz force that acts on electrons under combined influence of electrical and magnetic fields:

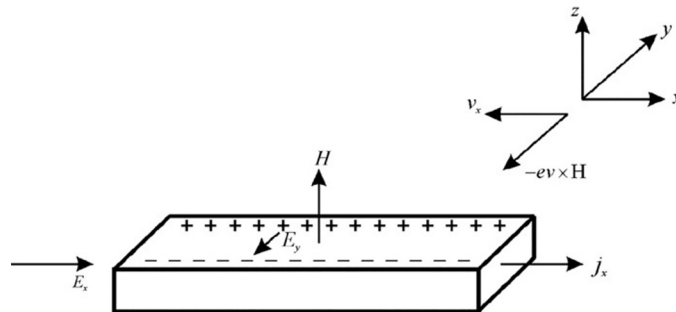


FIG. 5.3

Experimental scheme for investigating Hall effect.

$$\mathbf{F}_{\text{Lor}} = e(\mathbf{E} + [\mathbf{v} \times \mathbf{B}]), \quad (5.9)$$

A moving electron forcedly shifts in direction, perpendicular to both vectors \mathbf{E} and \mathbf{B} . At the condition when the circuit is opened, the charge redistribution leads to inhibitory force: electrical field directed of y -axis (axes are marked in Fig. 5.3).

After transformations it is possible to obtain the next expression for y -component of electrical field:

$$E_y = (1/n_e e c) j \cdot H. \quad (5.10)$$

As charge of electron e and light speed c are known, while current j and magnetic field H are measured directly, expression (5.10) determines the number of electrons per unit volume of a conductor. These measurements show that in metals—good conductors—the density of electrons is close to the value $n_e \sim 10^{23} \text{ cm}^{-3}$. The greatest density of electrons is seen in beryllium ($n_e \sim 2.5 \times 10^{23} \text{ cm}^{-3}$) and in aluminum ($0.8 \times 10^{23} \text{ cm}^{-3}$), and the smallest density is observed in cesium ($0.09 \times 10^{23} \text{ cm}^{-3}$) and rubidium ($0.1 \times 10^{23} \text{ cm}^{-3}$). Also, the metal that has the highest electrical conductivity at temperature 300K is silver; however, its density of electrons is $n_e = 0.6 \times 10^{23} \text{ cm}^{-3}$. Therefore the conductivity depends not only on the concentration of charge carriers.

The conformity between density of electrons and density of ions validates the foregoing assumption: from each metallic atom one or more electrons come off and then can move freely through a crystal. This is another proof of relative “freedom” of electrons in metals, especially given data on electronic density are consistent not only with electrical, but also with other properties (e.g., with electron heat capacity).

If the number of electrons per unit volume is found, it is possible to calculate the mobility of electrons in a metal using formula (5.8). In copper, for instance, this mobility at room temperature equals $u = 2 \times 10^{13} \text{ s/g}$ (electrons in copper are most mobile among other metals under normal conditions). Near absolute zero of temperature, electron mobility becomes greater thousands of times [7].

The effect of magnetoresistance is the change of electrical resistance in the magnetic field. In general, in case of magnetic field impact the change of electrical current is observed. Therefore, any substance, to some extent, exhibits the effect of magnetoresistance, not only metals. Also, in the semiconductors, the relative change of resistance in magnetic field might be greater than in metals [2].

Magnetoresistance of a substance depends on the *orientation* of the studied sample relative to the magnetic field. Magnetic field does not change the *projection* of electron’s velocity on the direction of magnetic field, but because of Lorentz force the magnetic field distorts the *trajectory* of a moving electron in a plane perpendicular to this field. This effect explains the reasons for the transverse magnetic field acting on the resistivity being stronger than the longitudinal magnetic field.

The effect of magnetoresistance can be explained while examining the trajectory of charged particles in the magnetic field. Consider passing along a sample current j_x along x -axis (Fig. 5.3). Electronic gas is degenerated; therefore average velocity of

electrons is Fermi velocity (near Fermi level), which significantly exceeds the velocity of directional movement (drift velocity). Charge carrier, between its collisions, without magnetic field influence moves in a straight line. However, in the external magnetic field H (applied perpendicular to electrical field) free path of electron will have another length: $l_x \approx l \cos \varphi$ (where l is average free path). Therefore, during the time between two collisions *along* electrical field direction, the electron overcomes the path smaller than l . This results in drift velocity decrease, hence, in the decrease of conductivity, that is, *resistance increases*. Therefore *magnetoresistance* is the relative difference between resistance measured at the presence of magnetic field and resistance, measured without the magnetic field. Based on this effect some electronic devices are elaborated and used as *magnetic field sensors*. However, except for metals, there are other much more sensitive materials: semiconductors and, especially, ferromagnetic and nanomagnetic materials.

Response time and free path of electrons. With the knowledge mobility, it is possible to estimate two very important characteristics of electrons in a metal. As already indicated, during mobility analysis, it seems more convenient to express the unit of mobility in CGSE system: [s/g], because from this follows that average free time $\tau = m_e u$ has the dimension of seconds. The important parameter τ [s] can be defined in many ways, but in solid-state physics the most often used term is *response time* (or time constant); it should be noted that an analogous term *free run time* is used in molecular-kinetic theory of gases, while in the theory of scattering the term *relaxation time* is more familiar [8].

Analysis of particle movement “with a friction” shows that as soon as the acting force is an electrical field the electrons start to move with acceleration, and their velocity increases; however, at the same time, frictional force also increases proportional to velocity. During time $\tau = m_e u$, the inhibitory force completely compensates the external accelerating strength, and particle moves with constant drift velocity. Thus response time describes the interval at which steady state of motion is installed (i.e., movement with constant drift velocity).

Response time for electrons moving in metal is very small: $\tau \approx 10^{-14}$ s. During electrical field switching, it looks impossible to notice the start of free movement of electron: as soon as metal is placed in the electrical field, Ohm’s law in a circuit begins to operate *immediately*. However, despite extreme smallness of response time, dynamics of electron “free” movement can still be estimated by τ indirect measurement. In this case, one needs to explore behavior of metals at very high frequency fields, as already shown in Fig. 5.2. By studying the variance of $\sigma(\omega)$, that is, frequency dependence of conductivity, it is experimentally possible to evaluate τ for different metals.

Thus, from a microscopic point of view, electron movement “with a friction” can be represented as the “flight” under force influence that accelerates the electron up to its collision, wherein the electron returns energy, gained by electrical force, to a lattice. At that, parameter τ is the average time between two collisions, while the product of this time on Fermi velocity v_F gives the *average free path* δ that is the distance between electron collision: $\tau v_F = \delta$.

To explain the nature of a conductor's resistance, why an electron in a crystal has *finite* average free path and what is its numerical value have to be elucidated. In the ideal crystal (without defects), in which each ion is fixed in its place, the electron moves quite freely ignoring surrounding ions. In an ideal case (when $T \rightarrow 0$ K and no defects) the electron, being accelerated by applied field, would increase its energy until facing the border of a sample. However, a real crystal exists in conditions of lattice *thermal fluctuations*; moreover, *static defects* of a structure also prevent free movement of an electron. From the expression for conductivity ($\sigma = n_e e u$) and using response time ($\tau = m_e u$), we get:

$$\sigma = n_e e \tau / m_e. \quad (5.11)$$

If both numerator and denominator in this formula are multiplied by v_F , the expression for conductivity will take another form:

$$\sigma = n_e e \delta / m_e v_F. \quad (5.12)$$

The skin effect. At very high frequencies (10^8 – 10^{11} Hz) electromagnetic field can penetrate in a conductor only to a small depth, which becomes less when the frequency is higher and conductivity and magnetic permeability are greater. The result is the *uneven distribution* of current density in the cross section of a conductor—this is the *surface effect* (or skin effect) [5].

The depth of penetration (skin) δ_{skin} is such a distance inside a conductor, at which the amplitude of electromagnetic wave reduces in “ e ” times. Using Maxwell equations, the following expression for the depth of penetration can be obtained:

$$\delta_{\text{skin}} = (2 / \omega \sigma \mu \mu_0)^{1/2}, \quad (5.13)$$

where ω is circular frequency, σ is conductivity, and μ is relative permeability of conductor. Relative permeability in most good conducting metals is close to unity ($\mu \approx 1$). However, in ferromagnetic materials where $\mu \sim 10^3$ the penetration depth is much lower than in nonmagnetic metals. The minimal penetration depth ($\delta_{\text{skin}} \rightarrow 0$) is observed in superconductors, in which $\sigma \rightarrow \infty$.

In radio engineering, especially in ultrahigh frequency (UHF) range, for skin effect description the concept of *surface resistance* R_s (measured in ohms per square) is used:

$$R_s = 1 / \sigma \delta_{\text{skin}}. \quad (5.14)$$

In microwave transmission lines, wave oscillations extend by both electrical and magnetic fields. Wave oscillations cannot pass through metallic walls of transmission lines, and, therefore, are distributed in the dielectric between waveguide walls. If these walls were made of ideal conductors (with $\sigma = \infty$), then microwave signals will not penetrate in walls of the guiding conductor (closely related to this case are superconducting materials that sometimes are used in microwave devices, and they essentially decrease wave attenuation).

In normal cases, waveguide walls are not perfect conductors; thus the microwave field can penetrate the waveguide walls. The depth of penetration, as shown in

formula (5.13), depends on microwave frequency and conductivity of the metal from which the transmission line (waveguide) is made. For example, in copper at a frequency of 10 GHz the penetration depth is only 0.5 microns. This implies that just a *surface* of conductors significantly affects quality of microwave transmission line. Imperfect surface leads to losses—attenuation of signal, propagating through the microwave waveguide. To adequately transfer signals, the microwave power transmission line should have a wall thickness that equals approximately 10 layers of skin thickness (at a frequency of 10 GHz it is less than 10 microns).

Thus the main cause of microwave energy losses is the skin effect, which is not so large in high-conductive metals. Skin effect is absent in case of superconductivity.

Hyperconductivity (cryoconductivity) in metals. High conductivity is a favorable factor to reduce attenuation of waves in waveguides and microwave resonators. Currently these devices have expanded their use up to millimeter waves that need to increase the quality factor of resonant microwave structures. In case of low temperatures (see Fig. 5.1), conductivity of metals increases significantly. Moreover, some metals at low temperatures become superconductors. The electrical resistance of superconductors below their critical temperature T_c (phase transition temperature) at low frequency becomes close to zero ($\rho \approx 0$).

Superconductivity is used in *cryogenic electronics*. However, in *metals* superconductivity is possible only by using helium temperature ($T \sim 4$ K), but liquid helium is very expensive in practical use. Nevertheless, in electrical engineering and electronics *hyperconductivity* might be successfully applied at another cryogenic temperature (77 K, liquid nitrogen that is much cheaper than helium). In some metals, at temperature 77 K, it is possible to obtain very small resistance (thousands of times lower than at normal temperatures).

Metals with favorable characteristics in the range of cryogenic temperatures are the *hyperconductors* (or cryoconductors). The phenomenon of hyperconductivity is not similar to superconductivity. Application of cryoconductivity in metallic microwave resonators and other microwave devices significantly increases their operating parameters. It is necessary to mention that there are many metals with small resistance at nitrogen temperature. However, a significant advantage at liquid nitrogen temperature is *beryllium*: exactly it has the smallest possible ρ value. In contrast to superconductivity, hyperconductivity is not destroyed by magnetic field. At that, hyperconductive metals must be well cleaned to have perfect structure [5].

5.3 THERMAL PROPERTIES OF METALS

According to classic electronic theory of metals, a solid conductor may be represented as a system, consisting of ionic lattice that contains inside “gas” of collectivized (free) electrons. Assuming the metal as a crystal, in which positive ions form stable lattice with mobile electrons between them can explain many basic properties of metals: ductility, malleability, high thermal conductivity, and large electrical conductivity. Similar to a solid state, in a liquid state of metal a large number of free

electrons exist: $n_e = (0.5-25) \times 10^{-22} \text{ cm}^{-3}$; they are the charge carriers providing electrical current passage through a metal. At normal temperature, electron mobility in metals is $u = (2-7) \times 10^{-5} \text{ m}^2/(\text{V s})$.

Thermal conductivity of metals. Heat transmission through a metal occurs by the same free electrons that determine electrical conductivity. Thermal conductivity of metals is high due to a large number of electrons per unit volume of metal. At the same time, the coefficient of thermal conductivity by electrons λ_e in metals exceeds thermal conductivity λ_{ph} in dielectrics, where heat transferred mainly has phonon nature.

Obviously, if other things being equal, the higher the specific electrical conductivity σ in a metal the greater the metal thermal conductivity λ_e . As temperature increases, the mobility of electrons in metal and, therefore, its electrical conductivity σ reduces; at that, ratio λ_e/σ has to grow. Mathematically, this law is expressed by the Wiedemann-Franz-Lorenz expression:

$$\lambda_e/\sigma = LT, \quad (5.15)$$

where T is absolute temperature and L is Lorenz number that equals:

$$L = (\pi^2/3)(k_B/e)^2. \quad (5.16)$$

Substituting Boltzmann constant $k_B = 1.38 \times 10^{-23} \text{ J/K}$ and charge of electron $e = -1.6 \times 10^{-19} \text{ C}$ in Eq. (5.16), it is possible to obtain Lorenz number $L = 2.45 \times 10^{-8} \text{ W } \Omega \text{ K}^{-2}$.

In most metals, Wiedemann-Franz-Lorenz law is well evidenced at temperatures close to normal temperature or at slightly elevated temperatures. For example, for copper at temperature $T = 293 \text{ K}$, by substituting conductivity $\sigma = 57 \times 10^6 \text{ S/m}$ and $\lambda_e = 390 \text{ W/(mK)}$ in formula (5.16), it is possible to obtain Lorenz parameter $L = 2.54 \times 10^{-8} \text{ W } \Omega \text{ K}^{-2}$ that is very close to theoretical value. At normal temperature, in aluminum $L = 2.1 \times 10^{-8}$, in lead $L = 2.5 \times 10^{-8}$, and in iron $2.9 \times 10^{-8} \text{ W } \Omega \text{ K}^{-2}$. However, at low temperatures Lorenz number might be changed; for example, while cooling it passes through minimum (in copper), but approaching absolute zero Lorenz factor again becomes close to theoretical value of L .

Thermal capacity of metals. Despite large and almost independent of temperature electronic conductivity, in metals the electronic contribution to specific heat C_{elec} at normal conditions ($T \sim 300 \text{ K}$) is *small* (Fig. 5.4). This feature of metals should be considered while elucidating those electronic devices that have to work at increased power [6].

Heat capacity of metal, predominantly, is formed by thermal fluctuations in crystal lattice (phonons) and at low temperature $C_{\text{lattice}} \sim T^3$. At that, electronic contribution to heat capacity of metal increases in direct proportion to absolute temperature: $C_{\text{elec}} = \xi T$. That is why electronic heat capacity in metals becomes significant at very low (cryogenic) temperatures ($T \ll \theta_D$), because lattice (phonons) contribution to specific heat tends to zero much faster in comparison with electronic contribution C_{elec} . Sometimes, below temperature $T \sim 10 \text{ K}$ electronic C_{elec} can exceed lattice (phonon) contribution C_{lattice} .

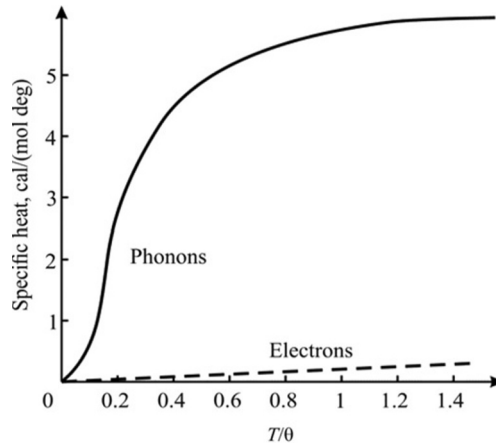


FIG. 5.4

Temperature dependence of phonons and electron contribution to specific heat in metals (θ is Debye temperature).

Similarly, at very high temperature, when $T \gg \theta_D$ and metal still remains solid (not molten), electronic contribution to heat capacity can be compared with lattice contribution.

Thermoelectromotive properties (thermal EMF). When two different metals (or semiconductors) are in contact, the difference in contact potentials occurs between them, caused by the difference in electronic work function and by distinction in free electron concentration in relevant metals (or semiconductors).

If temperature of two distant contacts (junctions) of different metals (entering into a closed circuit) is equal, the distinction in potentials in these metals is zero, and no current in circuit can be seen. However, if one junction of metals A and B has temperature T_1 while another junction has temperature T_2 ($T_1 \neq T_2$), thermoelectromotive potential occurs:

$$U = (k_B/e)(T_1 - T_2) \ln(n_A/n_B),$$

where n_A and n_B are concentrations of free electrons in metals A and B, respectively, k_B is Boltzmann constant, and e is charge of electron. This formula, referring to phenomena of thermocouple, can be also written also $U = \alpha(T_1 - T_2)$, where α is constant factor for given pair of conductors—the *thermoelectric coefficient*. Thus thermal EMF must be proportional to temperature difference between junctions (Fig. 5.5).

Thermocouples, composed of two different metals or alloys, are widely used for measuring temperatures. As a thermocouple wire, the metal with large and stable coefficient of thermal EMF should be applied [1]. On the contrary, in high-quality measurement systems and in reference resistors contacting metals and alloys that have the lowest thermal EMF should be used to avoid any interference from unwanted thermocouples in measuring.

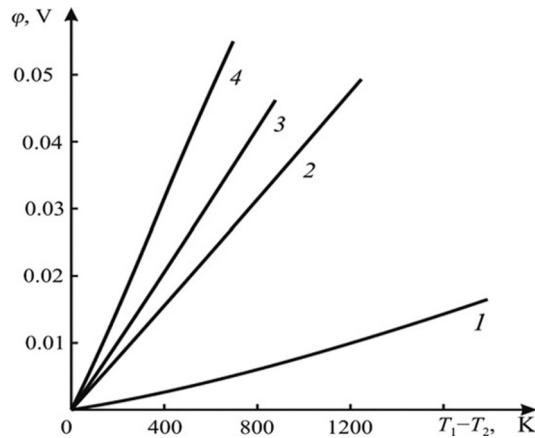


FIG. 5.5

Thermal EMF dependencies on temperature difference between two junctions of thermocouples: 1—platinum-rhodium-platinum; 2—chromel-alumel; 3—copper-copel, and 4—chromel-copel.

The linear thermal expansion coefficient is defined as

$$TC_l = l^{-1} dl/dT,$$

where l is arbitrary linear dimension of the studied sample. The knowledge of this coefficient is necessary to coordinate joining parts of devices, when in their design various combinations of materials are used (metals, dielectrics, semiconductors). Ignoring the coordination of thermal expansion may result in stresses or even in cracking, for example, in case of vacuum-tight connections of metals with glasses, and in the event of temperature change, this connection might be broken [4].

Using TC_l , it is possible to calculate the temperature coefficient of electrical resistance of a wire:

$$TC_R = TC_\rho - TC_l. \quad (5.17)$$

In pure metals usually $TC_\rho < TC_l$, that is, it can be approximately considered that $TC_R \approx TC_\rho$. However, in case of alloys with low TC_ρ formula (5.17) has practical value.

The value of TC_l of metals increases with temperature rise, especially when approaching melting point of a metal. Therefore, the *fusible metals* typically have relatively high TC_l , while in the *refractory metals* their TC_l has relatively small value.

Tensosensitivity of metals is used for strain measurements. Electrical resistance in metals changes significantly in case of metal deformation. This phenomenon is applied in *strain-sensing elements*. Tensiometric (strain-gage) alloys are applied in various devices used for strain measurements under mechanical influence (usually, the stretching is studied).

The principle of operation of these sensors is based on the change in resistance during metallic element stretching. At that, strain-sensitivity factor is determined by expression $C_{\text{tens}} = (\Delta R/R) : (\Delta l/l)$, where $\Delta R/R$ is change of resistance due to deformation Δl of element with length l . In most cases $C_{\text{tens}} = 1.6 \dots 2.2$, with the exception of *nickel* that has this ratio $C_{\text{tens}} = 10$ (this feature of nickel is due to peculiar structure of its Fermi surface).

Basic material for strain measurement sensors that operates at relatively low temperatures is *constantan* (Cu-Ni-Mg alloy). In the high-temperature range, sensors with Fe-Cr-Ni alloys are usually used [5].

The mechanical properties of metals are characterized by *limiting strength* during stretching, by *relative elongation* at break, by *brittleness*, *hardness*, and other parameters. Mechanical properties of metallic conductors are strongly dependent on the mechanical and thermal processing of a metal, as well as on the presence of impurities in a metal, and so on. The annealing leads to significant decrease in strength but increases elongation.

5.4 ELECTRONIC PROPERTIES OF METALS AND FERMI SURFACE

Classic electronic theory of metals (*Paul Drude's theory*). High electrical and thermal conductivity of metals as well as typical metallic luster indicate that electrons in a metal can be treated as free. Analysis of Ohm's law leads to the same conclusion: when even a very small voltage is applied to metal it is always seen that *current is proportional to voltage*, and proportionality factor ($1/R$) is the same as it is for increased values of voltage and current. If electrons in a metal are linked to specific nodes in a crystal lattice, then the *threshold electrical field* will exist, from which "normal" Ohm's law will begin to operate. In other words, conductivity of metal would be less in a smaller electrical field than at a larger electrical field. The fact that such phenomenon is not observed testifies the model of free electrons [1].

A study was conducted to determine the ratio of charge to mass for charge carriers in metals. In these experiments, the coil with a metal wire (copper, aluminum, or silver) is exposed to a rapid rotation and then abruptly pulled up. Under these conditions, free charge carriers would have to move by their inertia. Indeed, at the moment of sudden stop of coil the electrical current is registered; corresponding calculation gives the value of ratio e/m_e close to ratio of charge to mass for free electrons (1.76×10^{11}). Therefore this experiment supports the assumption that electrical current in metals is caused by free electron directional movement.

Drude's theory, supposing chaotic (thermal) motion of electrons and their drift under the influence of directional electric field, makes possible to substantiate Ohm's law. In case of electron collision with imperfections in crystal lattice, the energy, accumulated during electron acceleration in the electrical field, passes into crystal lattice such that it becomes heated (Joule-Lenz law). Thus classic electronic theory of metals can analytically describe and experimentally explain obtained basic laws of

conductivity and power losses in metals. It can also explain the relationship between electrical conductivity and thermal conductivity of metals. Moreover, some other experiments confirm the hypothesis of electronic gas existing in metal, for example, the curvature of electron trajectory in the transverse magnetic field: electromotive force changes electrical resistance of a conductor.

Therefore metals are different from other solid bodies because of free electron existence that practically are not connected with atoms but nearly freely move inside a metal. Using the concept of free electrons not only electrical properties of metals, but also other peculiarities can be explained, such as flexibility.

Assuming that electrons in a metal represent the *classic gas*, Drude's model offered metal as a "vessel" containing "gas" of freely circulating electrons that makes possible to get formulas for high-frequency conductivity $\sigma(\omega)$ and for electronic contribution to thermal conductivity λ_e :

$$\sigma = \sigma_0 / (1 - i\omega\tau); \quad \sigma = n_e e^2 \tau / m; \quad \lambda_e = L\sigma_0,$$

where n_e is number of electrons in 1 cm^3 ; $\omega = 2\pi\nu$ is frequency of electrical field; σ_0 is conductivity at very low frequency; τ is free path of electron; and L is universal constant (Lorentz number). Frequency dependence of σ is shown in Fig. 5.2, while Wiedemann-Franz law that implies Lorentz number is discussed in Section 5.3. Thus, by introducing metal as a system, in which positive ions are fastened by means of freely mobile electrons, it is possible to explain basic properties of metals: elasticity, ductility, high thermal conductivity, and large amount of electrical conductivity.

However, there are some *contradictions* between conclusions of Drude's theory and experimental data [3]. These contradictions are as follows:

- (1) disagreements in experimental and theoretical data in the temperature dependence of resistivity;
- (2) discrepancies between theoretically predicted and experimentally observed specific heat of metals.

Namely, in metals observed specific heat is much less than predictions of Drude's theory; it looks like electronic gas hardly absorbs heat during metal heating. Experiments show that required energy is much less than expectations of Drude's theory. The main drawback of Drude's model is the assumption that free electrons in metal are as free as molecules in an ideal gas. Also, electron-to-electron interaction is completely neglected. These contradictions can be overcome by considering the main standpoint of quantum mechanics.

Quantum distribution of electronic gas. The quantum theory of electronic gas in metal helps to explain all electrical and thermal properties of electronic gas, particularly, low heat capacity of electronic gas that is not explained in Drude's model.

In the process of metal heating, electrons located in the *lower energy levels* cannot increase their velocity (by perceiving heat energy) because higher energy levels are already occupied. Occupied levels and free levels in the electronic energy spectrum of metal are divided by the Fermi level. Thus only those electrons that are

located near the Fermi level can be thermally excited—only they can rise to the *unoccupied states* located just above the Fermi level.

Quantum distribution of electronic gas velocities is significantly different from classic distribution of molecule velocity in the normal gas, described by Maxwell-Boltzmann function. The point is that classic distribution is greatly dependent on temperature: at lower temperature, the maximum of distribution is narrower than at higher temperature. Moreover, this maximum becomes more blurred and noticeably *shifts toward higher velocities* (see Fig. 5.6A). In contrast, in case of *quantum* electronic gas distribution, the density of states at room temperature (Fig. 5.6B, dotted line) differs only a little from the density of states at absolute zero (Fig. 5.6B, solid curve).

However, during the study of electrical and thermal properties of metals, usually the distribution of *energy* but not the velocity is used.

Fermi energy level. Main ideas of electron quantum statistics were considered previously (Section 4.6). To determine the number of free charge carriers in metal the quantity of energy levels (states) of electrons in that conduction band, which is actually occupied needs to be known. Dependent on temperature and energy, the probability $w(T, E)$ of electron existing on the energy level E is determined by distribution function of Fermi-Dirac:

$$w(T, E) = \{1 + \exp[(E - E_F)/k_B T]\}^{-1},$$

where k_B is Boltzmann constant, T is absolute temperature, and energy E_F is Fermi level. As seen from this formula, at $T \neq 0$ distribution function for level $E = E_F$ is $w = 1/2$.

In the *ground state*, that is, at temperature $T = 0$ K:

- For energy levels that are located below Fermi level distribution function is unity ($w = 1$), because $E < E_F$ and $\exp[(E - E_F)/k_B T] \rightarrow 0$. This means that all levels that lie below Fermi level at absolute zero are occupied by electrons.

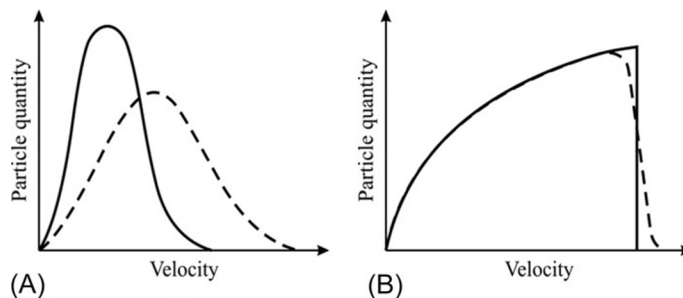


FIG. 5.6

Comparison of classical Maxwell-Boltzmann distribution of gas molecule velocity (A) and quantum Fermi-Dirac distribution of electronic gas velocities (B); dotted lines corresponds to higher temperatures.

- For energy levels that are located above Fermi level the Fermi-Dirac distribution function equals zero ($w = 0$), because if energy $E > E_F$ and $T \rightarrow 0$ the function $\exp[(E - E_F)/k_B T] \rightarrow \infty$, so the probability $w(T, E) \rightarrow 0$, that is, at absolute zero all levels, lying above Fermi energy, are empty.

Thus, in metals at temperature $T = 0$ Fermi level divides conduction band by a half: entirely occupied part of a band and entirely empty part of a band *without any energy gap* between these parts.

Fig. 5.7 shows the difference between Maxwell-Boltzmann classical statistics and Fermi-Dirac statistics. However, Fermi-Dirac statistics should be used only when the quantum effects are considered, and particles (in this case, electrons) do not differ from each other. Quantum effects are found when concentration of particles n is greater than quantum concentration n_q , while the distance between particles is close to de Broglie wavelength, that is, if wave functions of particles are *touching but not overlapping*.

This situation corresponds to metals. As a result of electrostatic repulsion, free electrons in a metal never come close to one another: each electron is surrounded by a free cavity, into which another electron cannot enter. However, this cavity is partially filled by positive charge of lattice; therefore this positive charge screens given electron from all others. Thus, due to electrical repulsion force, electrons move inside a metal with very rarely collisions between them.

Typically, Fermi-Dirac statistics is used at low temperatures, but in case of *metals* this statistics should be applied at normal temperature. The point is that temperature ~ 300 K with correspondent energy $k_B T \sim 0.03$ eV for metals seems very low, because Fermi energy in metals is large ($E_F \sim 5 \dots 10$ eV). As shown in Fig. 5.7, Fermi function at temperature 300 K for metal varies only in the narrow range near Fermi energy.

In contrast to the ideal gas, in which additional energy is absorbed by *all molecules*, in metal absorbed heat can excite only a relatively *small amount* of electrons located near Fermi level. For this reason, to raise temperature of electronic gas much less heat is required (see Fig. 5.4, electronic thermal capacity C_{elec}).

Fermi surface. Unlike classic electronic theory, quantum mechanics shows that gas of electrons in a metal under normal temperature is found in a state of

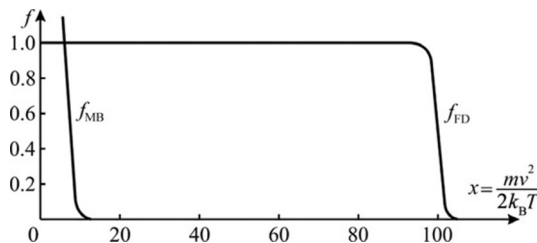


FIG. 5.7

Maxwell-Boltzmann f_{MB} and Fermi-Dirac f_{FD} distributions near room temperature.

degeneration. In this state, energy of electronic gas is practically independent of temperature, that is, thermal motion changes electron energy only a little. That is why thermal energy practically cannot heat electronic gas, and this is clearly detected by measurement of thermal capacity. In this state, similar to conventional gas, electronic gas appears as if its temperature were several thousand degrees [6].

According to quantum statistics, gas of electrons cannot have more than two electrons with same quantum parameters. Electrons occupy all allowed states of impulses, but not higher than those limited by Fermi level. The Fermi surface (Fig. 5.8) is a boundary between occupied and unoccupied states of electronic gas at absolute zero. Therefore Fermi surface is the isoenergetic surface in a space of quasimpulses (p -space) that corresponds to Fermi energy E_F :

$$E_s(p) = E_F.$$

Here $E_s(p)$ is dispersion law of conduction electrons; s is number of energy band; therefore Fermi surface at temperature $T=0\text{K}$ separates occupied electronic levels from empty levels.

Sometimes, in Fermi-surface representation, it is possible to restrict the task by using only one p -space, namely, the 1st Brillouin zone, located just at the ends of vectors p that describes all nonequivalent states. If the Fermi surface is completely housed in one cell of p -space, then this surface is *closed* (Fig. 5.8A). In this case, Fermi surface is a sphere with radius $k_F = (2mE_F)^{1/2}\hbar^{-1}$ that is determined by the concentration of valence electrons. If Fermi surface intersects the boundaries of cell in p -space, it is called *opened* (Fig. 5.8B). When extended p -space is used, closed Fermi surface is endlessly repeated from cell to cell.

In the impulse space, all states inside Fermi sphere of radius p_F are filled. The linear response of metal to electrical and magnetic fields or to thermal gradient is

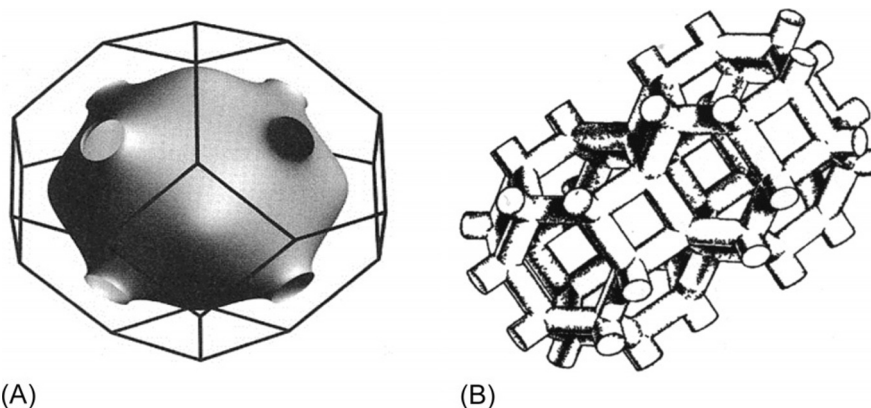


FIG. 5.8

Closed (A) and opened (B) Fermi surfaces for copper and lead, respectively.

determined by a shape of Fermi surface, because electrical current is due to changes in the occupancy of states near Fermi energy.

Thus Fermi surface is the abstract boundary in the reciprocal space, which is useful to predict thermal, electrical, magnetic, and optical properties of metals and semimetals [1]. The shape of Fermi surface is derived from the periodicity and symmetry of crystal lattice, as well as from occupation of electronic energy bands. The success of Fermi surface model is direct confirmation of Pauli principle that allows maximum one electron per one quantum state. Most electrons are placed inside Fermi surface, and only some of them can be outside of it (in the strip of $k_B T$, see Fig. 5.7)

Remarks on band theory of metals. Band theory uses the fact that an electron has both properties of particle and wave; therefore in case of close location to ions in a lattice, wave properties of electrons necessarily manifest themselves. The Bloch wave functions, used in *one-electron band theory*, is true for metals with *s*- and *p*-electronic orbitals, which have a large length of orbital space with considerable mutual overlap. However, in case of metals with *d*- and *f*-orbitals, band theory should be used with caution. Conduction electrons in these metals (and in their alloys) have increased effective mass; moreover, other physical properties of these crystals might be very specific [7].

Electronic energy bands of metals with *s*- and *p*-electrons are discussed in more detail in Section 4.6, where specific examples are shown: energy spectra of sodium that has very simple electronic energy spectrum (Fig. 4.19), and more complicated band structure of well-known conductor—copper (Fig. 4.20). The wavelength of electron in crystal depends on its impulse; simply stated, not a wavelength (λ) is used, but the wave number (k) that is expressed as $k = 2\pi/\lambda$. Obviously, wave number is *directly proportional* to wave velocity. Movement of electron can be described, using the relationship between wave number of electron and its energy; this dependence looks similar to energy dependence on velocity. Therefore Fermi surface can be shown in the space of wave numbers.

The curvature of Fermi surface is dependent on density of allowed states near it, and this curvature affects electrical, thermal, and magnetic properties of metal. The more electrons are located near Fermi surface, the more electrons can increase their energy during metal heating as well as the more electrons can orient their spins in magnetic field (which results in paramagnetism of electronic gas in metal).

Metals always have some electrons on the blank levels above Fermi level (for this reason, metals are good conductors of electricity), implying that metals have such Fermi surface from which electrons can easily be transferred to the blank (allowed) higher energy levels. Current in conductor is carried by those free electrons that can be easily accelerated, moving through allowed unfilled states.

Decrease of electrical conductivity with increasing temperature is a typical property of metals. This dependence is due to local thermal vibrations in crystal lattice. Accelerated by electrical field, electrons are scattered on these vibrations, and, therefore, the velocity of electrons reduces. Raising temperature increases thermal vibrations intensity, thereby decreasing free path of electrons between two collisions.

Conversely, with lowering temperature, the interval between collisions (and relaxation time) increases and conductivity increases. Electrical conductivity of any metal can be calculated, if the shape of Fermi surface and the relaxation time for these electrons are known. Fermi surface also permits to evaluate electrical and thermal conductivity of metal at different conditions. Finally, knowledge of Fermi surface shape is necessary to explain peculiarities of metal structure and their absorption and reflection properties, as well as to describe superconductivity in some metals at low temperatures, and many other physical properties.

Charge carriers in metals are free electrons that are also called as conduction electrons. They are typical *quasiparticles*. Their properties are substantially different from “normal” electrons existing in a free space, although electrical charge of conduction electron coincides with the charge of electron in a vacuum. However, the *energy* of “electron = quasiparticle” is a complex periodic function of impulse. When an electron is found in a free space, the surface of its equal energy is a sphere. For conduction electron in metal, the surface of equal energy might be complicated, being correspondent to the *surface of equal energy* in the impulse space. The form of this surface and its size depend on energy value that, its turn, is dependent on quasiparticle dispersion law [7].

Fermi surfaces of different metals can be quite various. In some metals, they resemble billiard balls (K, Na, Rb, Cs), and in others, they are complex designs of various shapes (Au, Ag, Cu, Zn, Cd, and others). One of ordinary Fermi surfaces—for copper crystal—is shown in Fig. 5.8A. This example is chosen because copper has one of the highest values of conductivity at 300 K and copper is widely used in electrical engineering and electronics.

The change of particle energy is equivalent to its impulse change multiplied by velocity. Therefore, to explain complicated energy spectrum in a given metal, one needs to know the shape of Fermi surface and superpose electron velocity with this surface. Increasing energy of any crystal (particularly, metal) can be described as generation of new quasiparticles. The increase of conducting electrons' energy in metal is possible by moving at least one electron under the Fermi surface into external impulse space. At that, the appearance of electron beyond Fermi surface is always accompanied by *unoccupied state* in filled part of a band—the *hole*, which can also be interpreted as a kind of quasiparticle—the antiparticle to the electron that emerges from under the Fermi surface. Therefore increasing energy of free electrons in metal is always accompanied by the birth of two quasiparticles. Calling hole as an “antiparticle,” there is the possibility of recombination: when electron returns “into its place” (under Fermi surface), metal is returned again closer to its ground state, because both quasiparticles—electron and hole—disappear.

Therefore from full Fermi sphere only those electrons that are located directly beneath the surface and in small distance from it should be selected. For this reason, it is important to identify and explore just nearest neighborhood of Fermi surface—the portion of impulse space in which quasiparticles—electrons and holes—can coexist. The mechanism of current flow through conductor might be rather intricate.

Due to applied electrical field, electrons by infinitely small portions increase their impulses (and energy); next, by collisions with foreign atoms, dislocations, boundaries of crystallites, and phonons (i.e., with any kinds of defects in crystal lattice), electrons return its impulse, gained from electrical field, to the lattice. For electrons located in fully filled areas *deep* under Fermi surface, such a process is impossible: Pauli principle of exclusion prevents them to leave their energy levels; electrons can only move from their level to the releasing level.

The quantum model of nearly free electrons is successful in theoretical description of many properties of metals. It has been found that in some cases the main cause of Fermi surface complications is the *interference* effects, arising due to crystal periodicity. This understanding allows building models of Fermi surfaces for polyvalent metals, and, with the support of a variety of methods, specifies their quantitative characteristics.

5.5 ELECTRON SCATTERING IN METALS

While referring to *average* free path or *average* free run time, the term “average” has two senses: all electrons are averaged (although one is scattered in a certain manner, and another a little differently), or the obstacles, in which electrons are scattered, are averaged.

An ideal crystal does not exist, with the absence of such a thing in nature. At that, each violation of lattice periodicity is perceived as the barrier to electrons. These obstacles can be the impurities of atoms that accidentally fall into crystal, the boundaries of crystallites (usually metal consists of many small crystals—crystallites), the vacancies that appear in crystal when regular sequence of ions is disturbed, and so on.

Thus, in the real crystal, there are many varieties of defects. However, at normal (room) temperature, the main cause of electron scattering (hence, the cause of which electrical resistance) predominantly depends not on the static defects in crystal but also on the *thermal motion* in lattice. Only at very low temperatures (near absolute zero), when thermal motion is almost absent, the structural defects become main centers of scattering.

Mechanisms of electron scattering can be divided into three classes: collisions of electrons with phonons (with thermal motion in crystal lattice); collisions of electrons with static violations of crystal periodicity; and collisions of electrons with each other [7].

Scattering on crystal defects. If it would be possible to create a perfect infinite crystal, even in this case electrons will have a limited free path—due to the thermal motion. However, free path in an ideal case *infinitely increases* with decreasing temperature; the length of free path in a perfect crystal is the *ideal* free path l_{id} , while resistance of ideal crystal is the *ideal* resistance ρ_{id} .

In a real crystal, near-zero-temperature electrons collide mostly with *static* structural defects. The length of free pass that depends on these collisions is the *residual*

free pass l_{res} , while corresponding resistance is the *residual* resistance ρ_{res} . This resistance remains even if temperature is very close to absolute zero.

There is a simple rule, according to which the resistance ρ of a real metal is considered as the sum of ideal and residual resistances (Matthiessen rule):

$$\rho = \rho_{\text{id}} + \rho_{\text{res}}$$

Analysis of experimental data supports this rule formulation. It has been experimentally shown that in samples of the same metal (but with different additives) the temperature dependence of resistance is quite *similar*. As shown in Fig. 5.9, one $\rho(T)$ curve is shifted relatively to another on value $\rho_{\text{res}1} - \rho_{\text{res}2}$ (the number indicates the number of sample).

In order to determine using Matthiessen rule what is the ideal resistance of metal, there is no necessity to create a perfect sample (usually this is not possible). Temperature dependence of real sample resistance should be extrapolated to absolute zero. Accordingly, it is possible to find the ρ_{id} for given metal, studying its nonideal samples: $\rho_{\text{res}} = \rho(T \rightarrow 0)$ by subtracting from $\rho(T)$ the part ρ_{res} .

Resistance caused by scattering. Average free path and average free time have a simple meaning: the *probability of collisions* (scattering) $w \sim 1/\tau$. If in a crystal there are several possible reasons for scattering, the probability of collisions is the sum of probabilities of each obstacle:

$$w = w_{\text{res}} + w_{\text{id}}$$

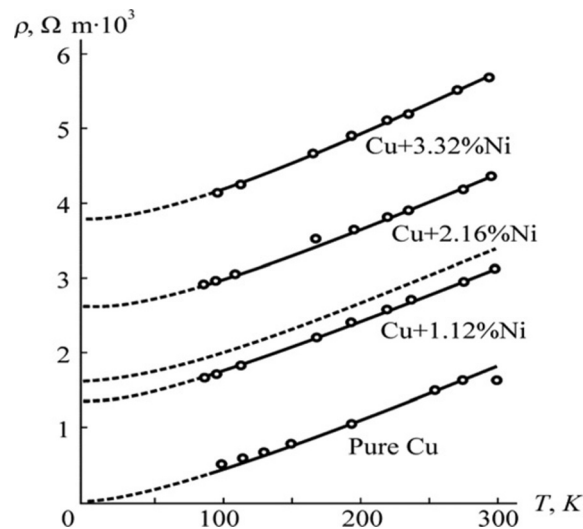


FIG. 5.9

Experimental verification of Matthiessen rule for copper and its alloys; dotted line shows copper with distorted structure.

A comparison of w_{res} , probability of collisions, correspondent to residual free pass, and w_{id} , probability of collisions in perfect crystal, gives

$$1/l = 1/l_{\text{res}} + 1/l_{\text{id}},$$

that is equivalent to Matthiessen rule, because resistance $\rho = 1/\sigma$, while coefficient of proportionality between ρ and $1/l$ is independent of scattering mechanism. Thus it is possible to divide various mechanisms of resistance, studying each of them separately [7].

Electron scattering on phonons. The term “collision” needs clarification—it should not be understood mechanically. In fact, just as electrons so also phonons are waves, and interaction between waves should take place as between *quasiparticles*; therefore the laws of energy and impulse *conservation* must be implemented (Fig. 5.10).

Solid lines in the figure depict electrons, while dashed lines depict phonons; near these lines, *pulses* and *energies* of electron and phonon are written, before and after collisions. In the first case, collision does not occur, but electrons *generate* phonons. In the second case, there is electron and phonon collision. However, here phonon “dies” (being absorbed by electron). It is possible to determine whether such a process can really take place, by applying conservation laws to impulse and energy [7]. Neglecting indexes, in both cases the conservation laws can be recorded by the same way:

$$E(p) + \hbar\omega = E(p + \hbar k). \quad (5.18)$$

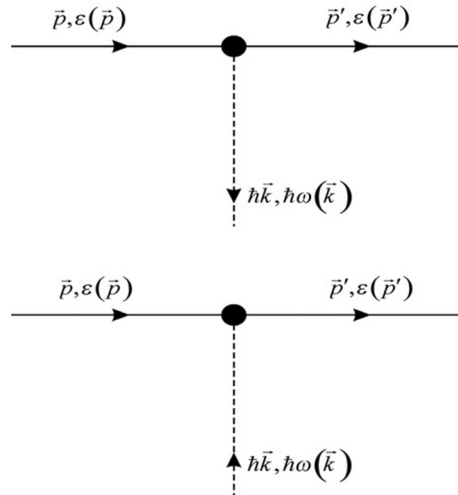


FIG. 5.10

Symbolic description of electron scattering on phonon mechanisms.

It is pertinent to recall that at electrical and thermal conductivity only those electrons whose energy is close to Fermi level are primarily involved, wherein it should be noted that Fermi energy is much higher than energy of phonon: $E_F \gg \hbar\omega = \hbar\omega_{\max} = k_B\theta_D$. Indeed, Debye temperature usually equals $\theta_D \sim 10^2$ K, while Fermi temperature of electronic gas is $T_F \sim 10^5$ K. Therefore in formula (5.18) the value $\hbar\omega$ can be neglected; therefore conditions of birth (or death) of phonon with absorption of its pulse can be simplified:

$$E_F(p) = E_F(p + \hbar k).$$

This equation shows that this process is allowed, that is, just as the “birth” (generation), so also the “death” (doom) of phonon is possible. Turning to classic description, it is better not to speak about “birth or death of phonon” but about *sound wave* radiation (or absorption). Radiation or absorption of sound waves is possible, because the Fermi velocity much exceeds sound velocity in crystal: $v_F \gg v_{\text{sound}}$. This condition always holds true: $v_F \sim 5 \times 10^5$ m/s, $v_{\text{sound}} \sim 5 \times 10^3$ m/s.

In order for collisions to take place, implementation of conservation laws is not adequate: the effective *interaction* between colliding quasiparticles is also necessary. Phonon is a quantum of lattice vibrations, while electron moves in the field of oscillation ions. When ions are shifted from their equilibrium positions, interaction energy between electron and ions varies; therefore, energy of electron-phonon interaction changes due to the shift of ions.

Calculations show that only that part of electrical resistance, which nature owes to the collisions with phonons, turns to zero at absolute temperature. The point is that collective motion (drift) of electrons is described by the average velocity that is very low. An “average” electron cannot emit sound waves, that is, generate phonon, but it can be dissipated on the oscillating ions. However, when such a condition is created, under which average velocity of *directional* motion of electrons becomes greater, then conductor’s resistance increases dramatically—“average” electrons begin to excite phonons.

Scattering of “average” electrons can be also described without using the understanding about phonons. Such an approach can be used not only to understand the nature of electrical resistance, but also to estimate the ideal free path. Thermal motion disturbs strong periodicity of lattice: ions oscillate in disorder and create disordered scattering. The probability of scattering w is a value inversely proportional to average free pass: $w = v_F/l_{\text{ph}}$, and this value becomes greater the larger the square of amplitude of ion oscillation: x^2 . Correspondingly, the average free path of phonon l_{ph} is determined by

$$1/l_{\text{ph}} = N_i x^2,$$

where N_i is number of ions per unit volume. This formula means that the probability of dissipation is proportional to the area that oscillating ion occupies in a plane, perpendicular to electron’s velocity. Calculations show that at all temperatures $x^2 \ll a^2$; therefore in the case of scattering on phonons, free path is much higher [7].

The *inefficiency* of electron collisions on oscillating ions is an important phenomenon, which is a main cause of deviations from Wiedemann-Franz law. In the event of collision, an electron changes its energy on phonon energy, while its impulse changes on a value of phonon impulse. Estimations show that the module of impulse, which is changed by phonon absorption, is very small compared with Fermi impulse, and this confirms the fact that for a significant deviation of electrons a large number of collisions are required.

When studying thermal conductivity, one needs to know how an electron loses its heat. With each collision, an electron changes its energy by amount close to $k_B T$. Although an electron only slightly deviates from its path, nevertheless, it loses energy. When colliding with phonon, the length of an electron's free path l_T (relatively to heat loss) is much less than its common free path l for loss of direction. The ratio of $l_T/l \ll (T/\theta)^2$ is a measure of collision inefficiency at low temperatures; therefore deviation from Wiedemann-Franz law is possible. At higher temperatures ($T \gg \theta_D$), the majority of phonons have energy $k_B T$; therefore each collision completely knocks electron out of its way, and, hence, if $l_T \approx l$ the Wiedemann-Franz law is performed.

Electrons collision with each other. The term “gas” corresponds to the image of many particles moving in different directions with different velocities; they collide with each other and change their direction, exchanging by energy and impulses. Electrons in metal also face each other and this affects resistance of metals, but only in such a case when these processes are accompanied by the *transference from one band to another* [7].

Both before and after collision, all energy states that are less than Fermi energy are occupied. However, what kind of electrons exactly occupies these levels is impossible to determine, because all electrons are indistinguishable (*indistinguishability* of electrons is one of basic principles of quantum mechanics). The consequence of this principle is the infinite length of electron free path at absolute zero temperature.

As the resistance is proportional to $1/l$, then a part of resistance, caused by electrons collision, is proportional to square of temperature ($\rho_{el} \sim T^2$). However, this term (in Matthiessen rule) can be reliably observed only in the *transition metals*; in other metals this mechanism is not seen on a background of more significant mechanisms.

Thus at *high* temperatures ($T > \theta_D$), the main reason for resistance in metals is the scattering on thermal vibrations of ionic lattice (phonons), and, therefore, Wiedemann-Franz law is executable. At *low* temperatures ($T < \theta_D$), the resistance of metals can be submitted as a sum of three items:

$$\rho = \rho_{\text{add}} + \rho_{\text{ph}} + \rho_{\text{el}}. \quad (5.19)$$

The first item ρ_{add} is conditioned by scattering on impurities and it is *independent* on temperature; the second item ρ_{ph} that corresponds to scattering on phonons is *proportional to T^5* ; the third item ρ_{el} , being responsible for scattering electrons on electrons, is *proportional to T^2* .

The resistance in an *ideal metal* is determined both by phonon item ρ_{ph} and by electron item ρ_{el} . True properties of metal (i.e., peculiarities of given metals ρ_{id}) that are not associated with technology of preparation of the studied sample can be determined by examining the “ideal part” of resistance using the Matthiessen method [5].

As a part of resistance $\rho_{\text{id}} = \rho_{\text{ph}} + \rho_{\text{el}}$ is the *fundamental parameter* of given metal (and it falls with decreasing temperature); therefore the desire to reduce resistance by decrease of ρ_{add} is natural. The art of growing pure metal samples at present is so advanced that the average free path, in fact, is no more a microscopic parameter. In most pure metals near absolute zero, the average free path reaches several millimeters. At that, in these pure metals at room temperature the free path of electrons is approximately 10^{-6} cm. Note that in atomic scales this free path is not so small: it is about 100 times that of atomic distances. Therefore collisions are not a restricting factor for freedom of electron movement in a metal.

5.6 SPECIAL ELECTRONIC STATES IN METALS

Based on quantum mechanics, band theory of solids successfully explains most properties of metals. However, in some cases, experiments are confronted with habitual simple explanations, and these cases are not only important in themselves but also allow to enhance understanding of some peculiarities of band theory. In this section, highly unusual properties of some metallic alloys are discussed that extend the application possibilities of band theory.

There is a special class of metals and alloys with strong anomalies in many properties: compounds of rare-earth metals, characterized by incomplete $4f$ -shell. Their electronic properties are difficult to explain using existing concepts. The point is that these substances are the intermediate materials between magnetic and nonmagnetic materials as well as between metals and dielectrics, while valence electrons in them are found between the localized and free states. Investigation of these compounds helps to understand their metallic and magnetic properties, specify the conditions of “energy band arrangement” in metal and dielectric states, and understand some peculiarities of electronic states in crystals [9].

Metals with intermediate valence. During investigation of rare earth metal properties, the main attention is focused on a phenomenon known as “intermediate valence” or “heavy fermions.” It is appropriate to bear in mind that all electrons of atoms that form a solid can be divided into two groups: electrons strongly bounded inside atom (in the residue) and electrons that can leave its atom—they either move to another atom (i.e., from atom Na to atom Cl during formation of ionic rock salt crystal, NaCl) or form covalent bonds (such as in germanium crystal). Electrons also might be generalized within crystals, and this occurs with conduction electrons in metals. In all these cases, the conception of atom *valence* is used, that is, a number of electrons that can be detached and moved away from the atom in the process of solid formation. For example, valence of Na is “+1” as in ionic crystal (NaCl) so also in metal (Na).

However, there are some known substances in which outward electrons demonstrate a binary, *ambivalent* nature: keeping partly localized in “native” atom, they also can demonstrate the intention to collectivization. Regarding the systems with *unstable valence* (or intermediary valence), some compounds of rare-earth metals can be included (those elements that have *unfinished 4f*-electron shells). These compounds have unique physical properties and anomalous characteristics that can be accounted for formation of *heavy fermions*: extraordinary electrons whose effective mass is 10^2 – 10^3 times greater than the mass of free electron.

Historically, the study of this phenomenon began when peculiar α - γ *phase transition* in the metallic cerium was discovered. Most phase transitions in solids are accompanied by a change in crystal symmetry, which means change in atom packing in a lattice (as well as the spin ordering in ferromagnetics or dipole ordering in ferroelectrics). However, the α - γ phase transition in Ce appears to be an exception: when it occurs, the ordering in lattice symmetry does not change, but transition is accompanied by an essential jump in crystal *volume*, reaching 15%.

Primary explanation of this phenomenon is reconstruction of electron structure of *ionic residue* during α - γ transition in cerium. This concept, with some additions and changes, is still preserved, and even applied to many other systems. Recently, a rather wide class of compounds have been found, in which valence instability is accompanied by strong anomalies of almost all physical properties of correspondent substances.

Among these systems, many features are close to the *ordered magnetic* states, and some substances have been found to be ready to transform into the *superconducting* state. This is quite unexpected: the *antagonism* between magnetism and superconductivity is well known (see next [Section 5.7](#)). However, an important fact is that not only normal, but also superconducting properties of these substances are quite unusual, which has led to speculation that these are a new type of superconductors that differ from all available data.

Compounds with intermediate valence and heavy fermions, where anomalous properties are most severe, include UBe_{13} , in which uranium atoms form cubic lattice while Be_{13} atoms are placed between them, creating an almost regular polyhedron—*icosahedron* Be_{12} with another “extra” Be in the center of this icosahedron. There are also some cerium compounds: CeAl_3 with hexagonal lattice, CeCu_6 with orthorhombic structure, CeCu_2Si_2 with tetragonal lattice, and some others. Besides, there are many systems with less-heavy electrons; they are compounds with the *intermediate* valence.

Specific heat in metals with heavy fermions. Heat capacity is a well-known *defining characteristic* of solids. In ordinary metals, electronic contribution to specific heat *linearly* increases with temperature: $C = \gamma T$. At that, temperature dependence of lattice specific heat corresponds to Debye law: $C \sim T^3$. In sum, the specific heat of normal metals is $C = \gamma T + \xi T^3$, but using scale $(C/T)(T)$ more simple for analyses dependence can be seen: $C/T = \gamma + \xi T^2$ ([Fig. 5.11A](#)). In [Fig. 5.11B](#), to compare the behavior of ordinary metals (curve 1) and metals with heavy fermions (curve 2) another comfortable coordinates can be also used: $(C/T)(T^2)$.

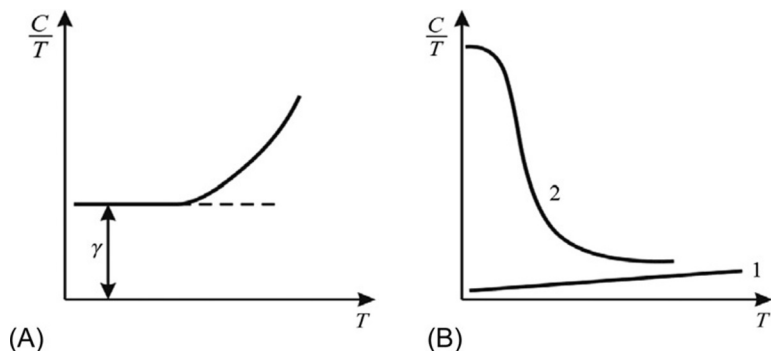


FIG. 5.11

Specific heat temperature dependence in various scales (A, B); comparison of normal metals (1) and systems with heavy fermions (2).

Table 5.1 Proportionality Factor $\gamma = C/T$ and Magnetic Susceptibility χ in Different Metals

Crystal	Cu	Li	CePd ₃	CeAl ₃	CeCu ₆	CeCu ₂ Si ₂	UBe ₁₃
γ , mJ/molK ²	0.695	1.63	35	1620	1500	1000	1100
χ ($T \rightarrow 0$) 10^{-3} CGSE/ mol	0.008	0.03	1.5	36	27	8	15

Temperature coefficients of specific heat in normal metals and heavy-fermion materials at low temperatures are given in Table 5.1. In *conventional* metals (Cu, Li), the proportionality coefficient $\gamma = C/T$ is about 1 mJ/molK², while in the *transition* metals this ratio is greater by order of magnitude $\gamma \sim 10$ mJ/molK²: this indicates a particular behavior of electronic subsystem already present in transition metals.

It is noteworthy that heavy-fermion systems have coefficient γ even greater by 2–3 orders of magnitude (and the same applies to their magnetic susceptibility). At that, CePd₃ is the *compound* with intermediate valence, while CeAl₃, CeCu₆, CeCu₂Si₂, and UBe₁₃ are the *systems* with heavy fermions. If specific heat were to be estimated as in normal metals, the observed values of γ will correspond to the effective mass of electron 10^2 – 10^3 times greater than the mass of free electron. Unusual is also the *temperature dependence* of specific heat (Fig. 5.11B).

Magnetic properties of heavy-fermion metals. In *conventional* metals with collectivized electrons, magnetic susceptibility χ of free electrons is almost independent on temperature: $\chi \cong \text{const}$ (Fig. 5.12A, curve 1). This is because χ (as well as the temperature coefficient γ of specific heat) is *proportional to the effective mass* of electron. However, in *paramagnetic* metals (Fig. 5.11A, curve 2), temperature

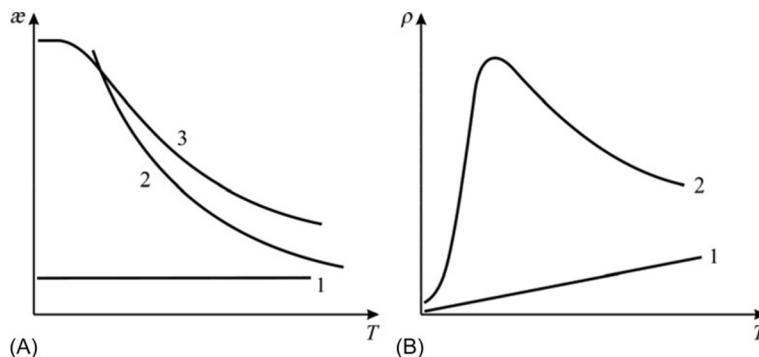


FIG. 5.12

Comparison of normal metals and heavy-fermion metals: (A) temperature dependence of magnetic susceptibility: 1—nonmagnetic metal, 2—paramagnetic with localized magnetic moments, 3—heavy-fermion systems, this susceptibility really is situated much above than curve 2; (B) temperature dependence of resistivity: 1—normal metal, 2—metal with heavy fermions.

dependence of susceptibility obeys Curie law: $\chi \sim 1/T$, that is, parameter χ increases as temperature decreases (sometimes, such substances at low temperatures can even come to a state with magnetic ordering—ferromagnetic or antiferromagnetic).

In the systems with *heavy fermions* at low temperatures, increase of $\chi(T)$ with decreasing temperature is also seen, but, unlike paramagnetics, the increase reaches *saturation* (stops at fixed temperature, Fig. 5.11A, curve 3). At that, any magnetic ordering in systems with heavy fermions is not established (except for some special cases, such as U_2Zn_{17}), but temperature change of susceptibility $\chi(T)$ enters onto a mode, peculiar in conventional metals, when $\chi = const$ (such as in curve 1). It is necessary to note that the quantity of this constant looks abnormally large (Table 5.1). When $T \rightarrow 0$ magnetic susceptibility in systems with heavy fermions may exceed values of paramagnetic susceptibility of conventional metals more than 1000 times.

Previously, such a large value of paramagnetic susceptibility was supposed as specific property of the ferromagnetics. Among “nonferromagnetic” substances the record shows metal Pd, in which at low temperature $\chi(T \rightarrow 0) = 0.7 \times 10^{-3}$ CGSE/mol. However, it is necessary to note that palladium is nearly a ferromagnetic metal: adding to Pd only a few percent of iron makes it a typical ferromagnetic. Nevertheless, in $CeAl_3$, for instance, paramagnetic susceptibility is 50 times greater than in Pd, although no ferromagnetism in this alloy can be observed. Despite this, some heavy-fermions systems, instead of magnetic ordering, become superconductors (we recall that in classical cases superconductivity is incompatible with magnetism).

Electrical conductivity of heavy-fermion systems. It should be also recalled that in conventional metals specific electrical resistance ρ decreases with temperature fall, at that, approximately linearly. The resistance of pure metals, ideally, in case of $T \rightarrow 0$ tends to zero (Fig. 5.12B, curve 1), while in the presence of impurities small

residual resistance can be observed. As stated earlier, positive temperature coefficient of resistance (which means negative temperature coefficient conductivity) is the hallmark of metal.

By contrast, in semiconductors or dielectrics electrical resistance *increases* with decreasing temperature and at $T \rightarrow 0$ tends to infinity. Similarly, at a sufficiently high temperature in the heavy-fermions metals $\rho(T)$ dependence resembles the behavior of dielectrics or semiconductors that is quite atypical for conventional metals (Fig. 5.12B, curve 2). However, at a certain temperature $\rho(T)$ reaches a maximum and then, as in metals, it decreases practically linearly in the low-temperature region.

Theoretical explanation of this unusual combination of properties of heavy-fermions materials will be given later.

Band theory and heavy fermions. According to band theory, it is possible to make a conclusion that all properties of solids depend on the ratio of free electron number to the number of states in bands. If the quantity of electrons is just enough to fill a certain band *completely* (while the next band that has higher energy remains *empty*), those crystals belong to dielectrics or semiconductors (see Section 4.6). Inasmuch as electronic states are localized, under the influence of electrical field, no electrical current can flow in such crystals. At zero temperature in dielectrics and semiconductors, electrons in conduction band are absent, while valence band is completely filled, in accordance with Pauli principle: every energy level can hold only two electrons with opposite spins.

Another possible behavior of electrons is realized in metals, in which *only lower energy levels* of the conduction band are filled, while the nearest energy levels are available for electrons—without any energy gap. In this case, even arbitrarily small electrical field can easily move electrons; therefore crystal exhibits metallic conductivity. It is appropriate to recall that highest energy level, occupied in metal at zero temperature, is the Fermi level.

Therefore for most solids the states of valence electrons are clearly described: in metals (such as Na, Al, Pb), semiconductors (such as Ge and Si), and typical dielectrics, such as NaCl. Metal type of electron behavior is symbolically illustrated in Fig. 5.13A: it shows trajectories of collectivized (delocalized) electrons and electrons localized near atoms. Conception of delocalized states supposes that internal electrons of atomic residue have relatively small radius of their orbits and, of course, these orbits vary only slightly in case of associations of atoms in a crystal.

However, quite another behavior is possible for electrons, based on the external shells of atoms: they can be only *partially tied* with their native atoms, so only for a while revolve around residue, but from time to time they move to neighboring atoms like “free” electrons (Fig. 5.13B), and then return to the partially connected state. These electrons are *almost delocalized*.

According to band theory, developed for crystal with *entirely* delocalized electrons, they move in the periodical field of ionic cores: the result is formation of allowed and forbidden energy bands. Electrons have effective mass m^* that is described by energy near the bottom of energy band: $E = p^2/m^*$, that is, by the same equation as for free electron, but with effective mass m^* instead of free electron mass m_e .

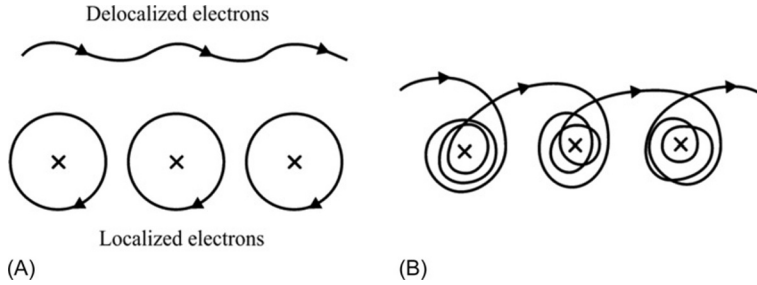


FIG. 5.13

Schematic representation of electron movement: (A) in metals with delocalized electronic states, (B) in metals with heavy-fermion state.

The value of effective mass m^* is closely connected with the *width* of conduction band ΔE : $m^* \sim 1/\Delta E$. If impulse of electron runs a certain value (actually, in crystal $|p| \leq \hbar/a$, where “ a ” is interatomic distance), the width of band is $\Delta E \sim \hbar^2/2ma^2$, that is, *small* effective mass meets the *wide band*, while *large* effective mass corresponds to the *narrow band*.

However, another description of electron properties in solids is considered: the delocalized electronic states, the collectivized electronic states, and the states of electrons localized in atoms. Despite the fact that this discussion concerns metals, it makes sense to consider whether all dielectrics are similar. Each substance, including dielectric, is individually different from another substance by a specific set of attributes: color change, hardness, electrical conductivity, and so on. Hence, it is not evident whether the nature of dielectric state is the same in various substances.

Corresponding to standard band theory, dielectric is a substance, in which valence energy band is entirely occupied, while located above it conduction band is empty and separated by the essential energy gap. However, not all dielectrics are arranged exactly this way—*another nature* of dielectric state is possible. Description of electronic structure of solids is based on the conception that an electron moves in the lattice, created by regularly situated atoms or ions. If there are many electrons, it is assumed that they *do not interfere* with each other, and they are allowed to occupy energy levels in accordance with Pauli principle. However, in reality all electrons *interact* with each other; therefore it is necessary to compare their *interaction energy* U with their *kinetic energy*, which is characterized by the width of energy band ΔE . If $U < \Delta E$, then simple band theory is applicable. If, on the contrary, $U > \Delta E$, the situation changes cardinally.

Formally, energy band can be completed only partially, but the movement of electrons (that is required for charge transfer) is prevented by other electrons—electrons of neighboring atoms. By their influence, they can “lock” each electron in the atom and make the *dielectric* crystal, although based on energy band characteristics it would be a metal. These substances are *Mott’s dielectrics*, named after English physicist H. Mott. The width of band ΔE depends strongly on interatomic

distance a , more precisely, on ratio between a and corresponding radius of orbital a_q : the smaller the a_q/a , the narrower the energy band.

Therefore, to obtain a status of “dielectric crystals” there is not only a single way, but at least two: (1) *complete filling* of energy bands (in usual band diagram), and (2) *strong repulsion of electrons* that leads to Mott type of dielectrics. Metals with heavy fermions sometimes behave *like Mott-type dielectrics* (it suffices to recall dependence $\rho(T)$ shown in Fig. 5.12B, curve 2).

Specificity of rare-earth metals. The electronic structure of rare-earth compounds is as follows: typical valence of rare-earth metals is “+3,” that is, three outer electrons can be detached from the rare-earth atom and directed into the conduction band (or form chemical bonds in the compound). As a result, ion R^{3+} is formed, which usually keeps *incomplete 4f-shell*. For example, in gadolinium ion (Gd^{3+}) instead of 14 electrons permitted in the 4f-shell, only partial filling of shell is observed: $4f^7$, while in dysprosium ion (Dy^{3+}) only 9 from 14 electrons exists: $4f^9$ (Table 5.2).

As the f -states are located rather close to atom’s nucleus, they have a small radius: $a_0 \sim 0.4 \text{ \AA}$. It is much smaller than the distance between atoms in solids that is usually about 3 \AA . Therefore it looks like f -electrons are not involved in chemical bonding, and it would seem that their state can be considered as localized; therefore they

Table 5.2 Electrons Spins Allocation in the Orbitals of Lanthanides

Z		4f	5s	5p	5d	6s
55	Cs	□ □ □ □ □ □ □	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □	↑↓
56	Ba	□ □ □ □ □ □ □	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □	↑↓
57	La	□ □ □ □ □ □ □	↑↓	↑↓ ↑↓ ↑↓	↑ □ □ □ □	↑↓
58	Ce	↑ ↑ □ □ □ □ □	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □	↑↓
59	Pr	↑ ↑ ↑ □ □ □ □	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □	↑↓
60	Nd	↑ ↑ ↑ ↑ □ □ □	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □	↑↓
61	Pm	↑ ↑ ↑ ↑ ↑ □ □	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □	↑↓
62	Sm	↑ ↑ ↑ ↑ ↑ ↑ □	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □	↑↓
63	Eu	↑ ↑ ↑ ↑ ↑ ↑ ↑	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □	↑↓
64	Gd	↑ ↑ ↑ ↑ ↑ ↑ ↑	↑↓	↑↓ ↑↓ ↑↓	↑ □ □ □ □	↑↓
65	Tb	↑↓ ↑ ↑ ↑ ↑ ↑ ↑	↑↓	↑↓ ↑↓ ↑↓	↑ □ □ □ □	↑↓
66	Dy	↑↓ ↑↓ ↑ ↑ ↑ ↑ ↑	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □	↑↓
67	Ho	↑↓ ↑↓ ↑↓ ↑ ↑ ↑ ↑	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □	↑↓
68	Er	↑↓ ↑↓ ↑↓ ↑↓ ↑ ↑ ↑	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □	↑↓
69	Tm	↑↓ ↑↓ ↑↓ ↑↓ ↑↓ ↑ ↑	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □	↑↓
70	Yb	↑↓ ↑↓ ↑↓ ↑↓ ↑↓ ↑↓ ↑	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □	↑↓
71	Lu	↑↓ ↑↓ ↑↓ ↑↓ ↑↓ ↑↓ ↑↓	↑↓	↑↓ ↑↓ ↑↓	↑ □ □ □ □	↑↓
72	Hf	↑↓ ↑↓ ↑↓ ↑↓ ↑↓ ↑↓ ↑↓	↑↓	↑↓ ↑↓ ↑↓	↑ ↑ □ □ □	↑↓
73	Ta	↑↓ ↑↓ ↑↓ ↑↓ ↑↓ ↑↓ ↑↓	↑↓	↑↓ ↑↓ ↑↓	↑ ↑ ↑ □ □	↑↓

belong to the ionic residue. By analogy with Mott type of dielectrics, f -electrons can be regarded as being located far on the “dielectric side” of Mott-type transition. The nature of chemical bonding and the type of crystal lattice (which determines metallic or dielectric properties) should have been identified only by *three valence electrons*.

Therefore electrons of f -state strongly influence the *magnetic properties* of correspondent crystals. If f -shell is only partially filled, magnetic moments of electrons are *not compensated*, so total magnetic moment is nonzero. Thus such ions are similar to elementary magnets. It is clear that the presence of these ions in crystal results in the fact that crystal is paramagnetic with localized magnetic moments, and at low temperatures these moments might be spontaneously ordered, making crystal either ferromagnetic or antiferromagnetic, or might acquire a more complicated ferrimagnetic structure.

Valence instability of rare-earth elements. This instability is caused by the *fluctuations* of a valence. In conventional rare-earth compounds the valence of rare earth ion is “+3” while f -shell is filled only partially. However, sometimes, the rare-earth elements in their compounds exhibit the *anomalous* valence: “+4” or “+2.” For example, among crystals, based on cerium, there is ionic compound CeF_4 , in which cerium is clearly tetravalent, while the rare-earth metal Eu (europium) even in its *metallic modification* can show properties with valence “+2.” Among other things, it should be noted that europium is one of basic materials of magnetic semiconductors: EuO and EuS , where europium is bivalent.

Anomalous valence in the compounds is peculiar for rare-earth elements located in the *beginning*, in the *end*, and just in the *middle* of rare-earth elements group (Table 5.2). At the beginning of this period, there are Ce and Pr; in the end of the period, there are Tm and Yb, and in the middle of the period, Sm and Eu are located. The quantum theory offers a convincing explanation of this situation, which implies that such electronic states have anomalies in valence stability, when $4f$ -shell appears empty, or completely filled, or filled in exactly half.

Aforesaid can explain the nature of anomalous valence in compounds of these elements. As cerium would have normal “+3” valence, its f -shell will have only one f -electron. However, this configuration is *competing* with the empty configuration of f -shell, that is beneficial to pull away from ion still one (the fourth) electron. The result is a state of *tetravalent* cerium, in which its f -shell is empty. This situation is quite competitive, because its energy is close to the energy of conventional trivalent state.

Similar arguments show, that, for example, compounds based on ytterbium ion instead of valence “+3” (when $4f$ -shell has 13 electrons) might have an advantageous condition for Yb^{2+} in which f -shell is completely filled, that is, $n_f = 14$ (Table 5.2). Similarly, for europium the valence of Eu^{2+} state is competitive (and often preferred) before expected Eu^{3+} , because in a state Eu^{2+} has exactly half-filled f -shell, $n_f = 7$.

The presence of similar energy states in rare-earth ions with different valences *significantly affects their properties*, changing fundamentally all characteristics of

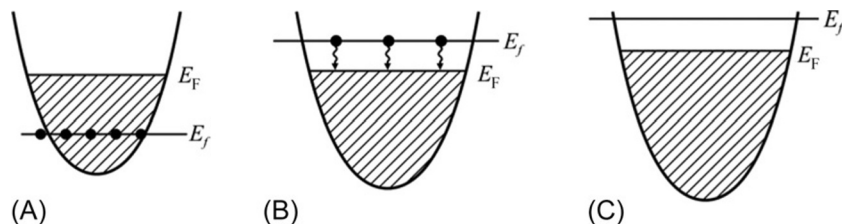


FIG. 5.14

Electronic structure of f -metal, illustrating nature of transitions with changing valence; conduction band is filled up to Fermi level E_F , while f -level marked as E_f .

correspondent substances. The *instability* of ionic valence state leads to the emergence of a *special class of systems* with intermediate valence, and, in some special cases, heavy fermions.

A possible situation for *metals* is demonstrated in Fig. 5.14 (while the situation for *semiconductors* is shown in Fig. 5.15) [9].

The *normal state* of a rare-earth metal with whole valence (i.e., with whole filling of f -levels) corresponds to a picture shown in Fig. 5.14A. It is evident that for such a rare-earth metal the energy E_f of f -level is much lower than Fermi level E_F . Accordingly, in the semiconductor (Fig. 5.15A), f -level is located in the forbidden band (energy gap) *below* the bottom of empty conduction band.

In metals with *unstable valence*, such a situation is realized, when level E_f is located *near* the Fermi level E_F (Fig. 5.14B; or in the semiconductor E_f is located close to bottom of conduction band, Fig. 5.15B). Complex configuration of f -orbital leads to the fact that some external conditions can move E_f level, for example, this level floats up when pressure increases or temperature changes (Figs. 5.14B and 5.15B).

If, due to this displacement, the E_f level rises sufficiently to *cross* the Fermi level of metal (Fig. 5.14B; or cross the bottom of conduction band in semiconductor, Fig. 5.15B), the energy of those electrons that occupy E_f level becomes greater than the energy of other states in conduction band, located between E_f and E_F (most of

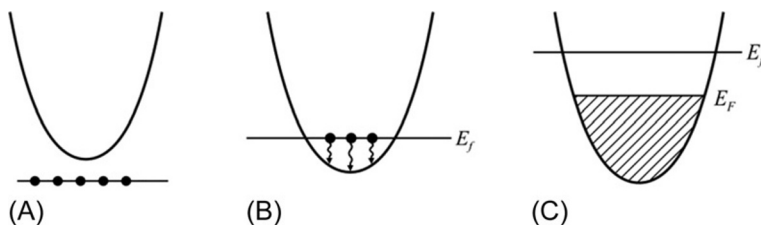


FIG. 5.15

Transition with a change in valence for compound that in initial phase is a semiconductor with whole number valence.

them previously were empty). It is obvious that this is not profitable energetically; therefore the electrons will move from E_f level to empty states of conduction band (Fig. 5.15B).

If the number of these empty states is large enough, this situation leads to devastation of the E_f level (Figs. 5.14B and 5.15B). However, in a more general case, the electron filling of E_f level is arranged in such a way that *only one* of E_f electrons leaves each ion. Because of this, in the conduction band only one electron per ion can be found (Fig. 5.14C; in semiconductors with rare-earth element—Fig. 5.15C). This corresponds to the *increase by one* the valence of rare-earth ion.

This change in electronic state results in a fundamental modification of crystal properties. Very often, a possibility occurs for transition from nonmagnetic to magnetic state. Indeed, if the initial situation (Fig. 5.14A) with localized electron on the f -level remained unchanged, the system with localized magnetic moments can become ordered at low temperatures (leading, e.g., to ferromagnetism). At the same time, the situation with the f -level devastation (Fig. 5.14C) would lead to the unpaired electrons leaving f -level, making localized in ion magnetic moments disappear (Fig. 5.14C) that corresponds to normal nonferromagnetic metal.

Thus the position of f -level relative to Fermi level determines the presence or absence of magnetic properties in a system: at that, a possible shift of the f -level (e.g., under temperature action) may cause phase transition from a magnetic to nonmagnetic state.

In the case of semiconductors, in conduction band the initial state is empty (Fig. 5.15A), but after possible transition (Fig. 5.15C) some electrons appear in this band and then can move freely in a crystal. In other words, the valence transition will be also the transition of metal-dielectric type (such transitions are described in Chapter 10).

There are some systems in which a new state of electrons can be implemented as described earlier; however, the transition with a change in valence might be not completed in full, but *stopped “halfway.”* There are some compounds, in reality, that in normal conditions (standard temperature and pressure) exhibit the *intermediate* situation. Many compounds with unstable f -shell are known, in which the intermediate phase is realized; therefore this state demonstrates *specific physical properties*. Also, in this case special electronic states are formed, and compounds with intermediate valence are identified as possessing of “heavy-weight” fermions [9].

Thus, when such a state of intermediate valence occurs, in which f -level is located *very close* to Fermi level, this f -level will be filled only *partially*. For example, experiments show that, on average, f -electron level has only $1/3$ of electrons. At that, a possibility arises that a crystal, regularly or randomly, possesses atoms with *two different types* of f -electrons. If f -electron exists in every third atom, so, on average, one atom holds only $1/3$ of f -electron. Compounds of this type really exist: they include, for example, crystals such as Eu_3S_4 and Sm_3S_4 , very important for electronic technology, as well as well-known ferrimagnetic magnetite (Fe_3O_4).

At high temperatures, crystals Eu_3S_4 or Sm_3S_4 show a rapid exchange by electrons that resembles $\text{Eu}^{2+} \leftrightarrow \text{Eu}^{3+}$ transition and determines a conductivity of metal

type. However, at lower temperatures, these electrons become “frozen” in their centers; therefore the *segregation of valence* occurs, that is, ions Eu^{2+} and Eu^{3+} now become different from each other and occupy in the lattice fixed positions, alternating by a certain way.

This “freezing”-type electronic transition is the *phase transition* that is accompanied by the restructuring of crystal lattice with the *superlattice* arising and essential changing in electrical properties. It is this transition that actually occurs in the magnetite Fe_3O_4 at temperature 119 K. This transition has long been discovered, and it is the first experimentally studied metal-dielectric type of transition. Thus crystals with intermediate value of average valence can be arranged simply as alternation of ions with different but integer-valued valence. Such substances are the *mixed-valence* substances.

However, systems with intermediate valence can be arranged differently. All ions, that is, all centers of crystal lattice, are completely equivalent; at that, the concentration of centers is the same as unit cells, that is, about 10^{22} cm^{-3} . Therefore they are *not impurities* but the principal system, and each cell contains rare earth ions. Intermediate filling of *f*-level, that is, the fractional number of *f*-electrons per one center, remains all times, but it captures the electron from *f*-level, sometime (e.g., 1/3) holds it, and then throws out this electron into conduction band [9]. Then the probability of finding any ion in the state of “*f*-electron” is 1/3 (while the probability of state “without *f*-electron” is 2/3), that is, on average, the chance of *f*-level filling is 1/3.

This situation is symbolically illustrated in Fig. 5.13B: electron moves in the conduction band, then is captivated by a center on closed orbit, almost being localized, sometime it turns again on this orbit, and next jumps back into the conduction band, only to be caught again by some other center, and, possibly, by the same center. Thus, in systems with intermediate valence, all centers are equivalent but each of them shows the *valence fluctuations*, giving on average the fractional filling of *f*-state. From a quantum mechanical point of view, this means that the total electronic wave function Ψ is superposition of wave functions Ψ_f of the *f*-state and the state in conduction band Ψ_e , that is, $\Psi = \alpha\Psi_f + \beta\Psi_e$.

The “weighting factor,” with which Ψ_f -function is included in this sum, determines the probability to find electron in the *f*-orbital, because the average number of *f*-electrons per one center is $n_{ff} = |\alpha|^2$. This process of constant conversion from *f*-state of electron into conduction band and back characterizes the probability of such transition, or the *lifetime of electrons* in the *f*-state. Through uncertainty relation, $\Delta t \cdot \Delta E = h$, the finite lifetime of state means the uncertainty of its energy.

Explanation of heavy-fermion system features. Abnormal behavior of *heat capacity* (Fig. 5.11), as compared with conventional metals, follows from the fact that basic electrical properties of metals in dependence on temperature are determined by the electrons located in the energy range $k_B T$ near Fermi level. Indeed, in conventional metals at temperatures $T = 0$ all states inside Fermi sphere are occupied, while outside of it, when energy $E > E_F$, all states are empty. With increasing temperature, a redistribution of electrons on energy states starts, inasmuch as some

electrons gain energy and move into the empty states above E_F . Accordingly, under Fermi level the holes remain. At that, a rather small number of states of electrons are excited: only in the range of $k_B T \ll E_F$.

However, in systems with heavy fermions the *number of states* in the range $k_B T$ is *much greater* (from two to three orders of magnitude!) than in conventional metals. Therefore factor γ is in many times larger (see Table 5.1). A high value of γ (see Table 5.1), observed in systems with heavy fermions, becomes understandable, as well as a slightly smaller (due to increased width of f -zones) but still quite high values of γ in compounds with intermediate valence.

As for peculiarities of *magnetic properties* (see Fig. 5.12A), it should be noted that heavy fermions below a certain temperature T^* are degenerated, that is, they take their “heavy band” according to Pauli principle, but above temperature T^* they become nondegenerated. Consequently, these fluctuations smooth out various sub-levels and stop the growth of magnetic moment, which results in saturation of magnetic susceptibility $\alpha(T)$. Magnetic properties of heavy-fermion systems also cause anomalies in temperature dependence of electrical resistance (Fig. 5.12B).

Thus heavy fermions arise in such systems that contain the uncompleted $4f$ -shells (or $5f$ -shells), where filling of electronic orbitals is *unstable*; therefore there is a proximity of valence instability [9]. These fermions have a record high value of effective mass, and the heaviest fermions are observed primarily in compounds of cerium and uranium—just in the elements located in the beginning of $4f$ - and $5f$ -periods. Among compounds with heavy fermions, there are magnetics as well as superconductors.

The mechanism of heavy electron appearance *right at the Fermi level* can be rather complicated. Generally, different opportunities should be noted.

One such situation is when f -level itself goes to Fermi level and at low temperatures forms an energy district, where f -electrons are *mixed* with conduction electrons. In this version, heavy fermions are primarily f -electrons themselves that are found near the Fermi level, and they are partially delocalized. In this case, the interaction of f -electrons with other excitations in crystal (phonons, conduction electrons, and maybe others) might play a special role. These types of interactions can result in the narrowing down of energy band and, consequently, the effective mass of electron increases.

In another tested case—electron-phonon interaction influence—effective mass increases due to the “polaron” effect: electrical field of electron deforms crystal lattice in its nearest surroundings, and then electron moves in a crystal, surrounded by “coat” of lattice deformations created by electron itself. Similar interactions with conduction electrons can lead to “electronic polaron” formation: deformed lattice near heavy fermion that additionally increases its mass.

The interaction of f -electrons with each other is another possible mechanism of their effective mass increase. Such interaction between electrons may even result in a complete localization of electrons, that is, their mass becomes “infinite” (Mott’s dielectric). However, if the limit of localization is not exceeded but only draws near, the crystal can remain a metal, but with very a narrow energy band and, correspondingly, with a large effective mass of charge carriers.

5.7 SUPERCONDUCTIVITY IN METALS AND ALLOYS

Superconductivity is one of “cooperative” effects in the ensemble of conduction electrons, but it cannot be explained using simple models within “independent electrons.” Superconductivity is highly interesting in science and is important for technique, in the sense that superconductors have no skin effect; therefore energy losses are significantly reduced. In addition, as found by Josephson, the contact of various superconductors, separated by thin dielectric layer, can produce microwave generation, as well as enabling the creation of a variety of sensory devices.

Superconductivity was discovered about 100 years ago when comparing platinum and mercury electrical resistivity at cryogenic temperatures. The change in resistivity ρ during deep cooling of platinum (Pt) and mercury (Hg), in which at the time of conducting the experiment a superconductivity was discovered, is shown in Fig. 5.16. Platinum is not a superconductor, but in the superconductor mercury below *critical temperature* DC resistance becomes zero: $\rho = 0$ (respectively, conductivity $\sigma = \infty$). In mercury this phenomenon occurs at $T_c \sim 4\text{K}$, but in some others *pure metal* temperature of phase transformation to superconducting phase is higher: for lead $T_c \sim 7\text{K}$ and for niobium $T_c \sim 9\text{K}$. In the *alloy* Nb_3Sn , transition temperature is $T_c \sim 18\text{K}$, while Nb_3Ge reaches the highest transition temperature for superconducting *metallic alloys*: $T_c \sim 23\text{K}$.

Theory of superconductivity in metals (having 3D lattice) was created much later after this discovery. Conduction electrons form the *Cooper pairs* (that belong to bosons), and they can propagate in superconducting metals without any loss of energy [1]. However, according to theory the temperature of superconducting transition cannot exceed 25K (in agreement with experiments conducted on metals and

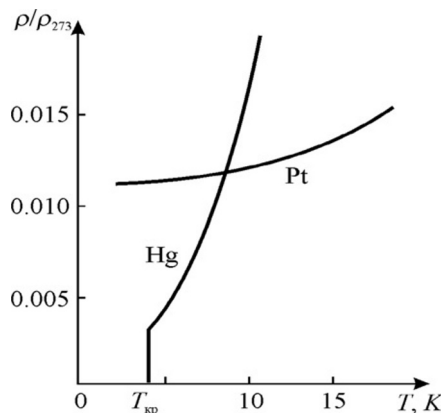


FIG. 5.16

Temperature dependence of resistivity of platinum and mercury; ρ_{273} means resistivity at room temperature.

alloys). This limit follows from the mechanism of electron-phonon interaction in the 3D regular structures.

Indeed, despite many years of research, either in pure metals or in metallic alloys, effort to reach a higher temperature of superconducting phase transition than 25 K was not met with success. For this reason, application of metallic superconductors in power engineering, electronics, and microwave technique necessarily requires devices that could cool to liquid helium temperature (near 4 K). This leads to a very high cost of cryogenic devices. Nevertheless, cryoenergetics with hardware elaborated for helium cooling and superconducting alloys of Nb_3Sn type are widely applied. In microwave cryoelectronics, mainly *niobium* is used (in resonators and millimeter range waveguides). In the devices based on Josephson effect, *lead*, *tin*, and other superconductors cooled by liquid helium are applied.

In contemporary cryoelectronics and some electronic devices, cooling by *liquid nitrogen* is also used (at 77 K, i.e., 100 times less expensive than helium hardware); this has become possible after discovery of *high-temperature superconductivity*.

Zero resistance of superconductors. Temperature of phase transition into superconducting state is *critical temperature* T_c . Most prominent pure metals-superconductors are lead, tin, niobium, and some others. Note that conductors, which are best in normal conditions, such as copper, gold, or silver, cannot turn into superconducting phase: at very low temperatures, they are only cryoconductors.

Most known superconductors are alloys and composite compounds; their total number is up to several hundreds, and growing. In particular, the substances that belong to a family of *high-temperature superconductors* (HTS) consist of three, four, and even five components. In principle, due to *high-pressure* technology, it is possible to transfer into superconducting state even typical dielectrics, such as solid nitrogen and oxygen, but physicists expect the highest T_c in the *solid hydrogen* (however, superconductive “metallic hydrogen” up to now is not reliably prepared).

Magnetic field influence. It was found that superconductivity in *metals* can be destroyed not only by temperature growth, but also under the influence of magnetic field that is also the *critical parameter*, H_c . This effect is shown in Fig. 5.17 in a phase

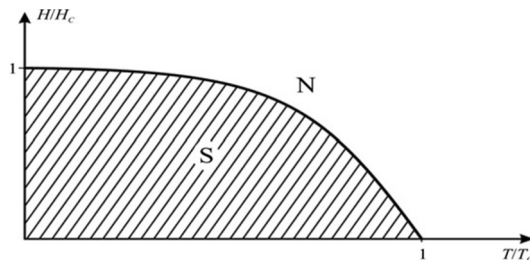


FIG. 5.17

Phase diagram: dependence of superconducting phase S on relative temperature T/T_c and relative magnetic field H/H_c for superconductor of the first type; the border between superconducting S and normal metallic phases N is shown.

diagram of superconducting state S and normal state N . This diagram is drawn using “normalized” temperature (ratio T/T_c) and “normalized” magnetic field (ratio H/H_c). However, such a diagram is characteristic only for superconductors of a *first type*. In the superconductors of a *second type* (that are not considered here), the dependence of T_c on magnetic field is much more complicated [1].

Therefore superconducting state can be realized at very low temperatures and at relatively small magnetic fields. Magnetic field influences the orientation of electron spins: in the *Cooper pairs* (that mainly cause superconductivity) spins of interacting electrons obviously must have opposite orientation.

Meissner effect. When temperature decreases below critical T_c , the superconductor completely forces itself out of the magnetic field: the first-type superconductivity is incompatible with a magnetic field.

While an “ideal conductor” (i.e., not a superconductor) is first exposed to a steady magnetic field, penetrating through it (Fig. 5.18A), and then is cooled down, the magnetic field in it will be same, Fig. 5.18B (in this experiment, the conductor is supposed as diamagnetic). However, the magnetic behavior of a superconductor is quite different from an “ideal conductor” (Fig. 5.18C and D). When passing through phase transition into superconducting state, it will actively exclude any magnetic field presence. Expulsion of magnetic field from the superconductor at phase transition from normal to superconducting state is the *Meissner effect*.

As shown by direct experiments, at temperatures $T < T_c$ under external magnetic field, *regardless of how superconducting state is activated*, inside of superconductor any magnetization is always absent: $B=0$. Otherwise, this result is treated as *zero permeability* of superconductor, that is, $\mu=0$. This fact clearly demonstrates that a superconductor is *quite different* from an “ideal conductor.” Superconductivity can exist when the external magnetic field is less than the value of critical field

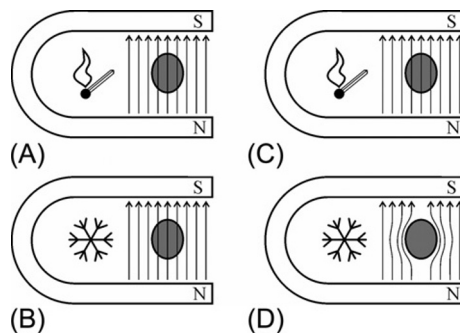


FIG. 5.18

Diagram of Meissner effect. Magnetic field lines are represented by arrows: (A) ideal conductor at elevated temperature; (B) same conductor at extremely low temperature; (C) superconductor above T_c ; and (D) superconductor below T_c .

($H < H_c$), and it is independent of previous history of a sample. Corresponding equilibrium state is thermodynamically stable, and it can be characterized within thermodynamic approaches.

Therefore one can make an important conclusion: superconductivity is such a state of high conduction systems in which two conditions are always fulfilled:

$$\begin{aligned}\rho &= 0 \\ \mu &= 0.\end{aligned}$$

A theoretical explanation of Meissner effect comes from London equations [1]. They show that magnetic field, actually, can slightly penetrate inside a superconductor, but with an exponential decrease over a distance of 20–40 nm. It is described in terms of a special parameter: *London's penetration depth*.

As a rule, at a rather strong magnetic field superconductivity disappears (Fig. 5.17). However, known superconductors can be divided into two classes, according to how *magnetic breakdown* occurs. In type I, superconductivity is abruptly destroyed, when the strength of the magnetic field rises above critical value H_c . However, in type II superconductors, magnetic field, which exceeds *first* critical value H_{c1} , converts the superconductor to a peculiar *mixed state*, when magnetic fluxes can penetrate *locally* in the material, but, as a whole, the superconductor remains as nonresistive up to achievement of *second* critical field H_{c2} (when electrical current in the superconductor becomes too large). Pure metallic superconductors usually belong to type I, while most superconductive alloys belong to type II.

Anomaly of heat capacity at phase transition. The change of energy and entropy in a superconductor can be expressed through magnetic interaction; firstly, because the magnetic field H is able to destroy superconductivity, and, secondly, the surface current in a superconductor creates magnetic moment M that completely compensates the external magnetic field applied to the superconductor. Calculations show that at the point of phase transition into superconducting state specific heat must show a jump (Fig. 5.19), and this expectation is confirmed by numerous experiments. Indeed, a maximum of heat capacity at superconductor transition point exists despite the expected linear increase $C(T)$ in metals (in normal metals the electrons with their half-integer spin obey Fermi-Dirac statistics that cause linear dependence $C = \gamma T$).

However, it is noteworthy that not linear but the *parabolic* temperature dependence of specific heat is seen below phase transition temperature, that is, in superconducting phase: $C \sim T^3$. This fact clearly indicates that statistics of electrons in the superconductor is changed: now it is the Bose statistics that is a characteristic for substances with the *integer spin* (Cooper pairs of electrons are bosons).

Quantization of magnetic flux. If one would take the ring of superconductive material and induce a current in it by the external magnetic field (Faraday effect), this current will flow in the ring *indefinite time*, because any resistance in the ring is absent. To realize this experiment, superconducting ring should be taken at a temperature higher than the transition temperature ($T > T_c$); at that, lines of magnetic field cross the area of a ring (Fig. 5.20A).

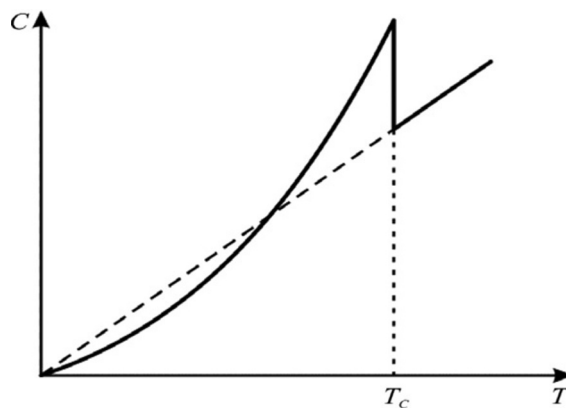


FIG. 5.19

Specific heat capacity temperature anomaly in superconductor phase transition.

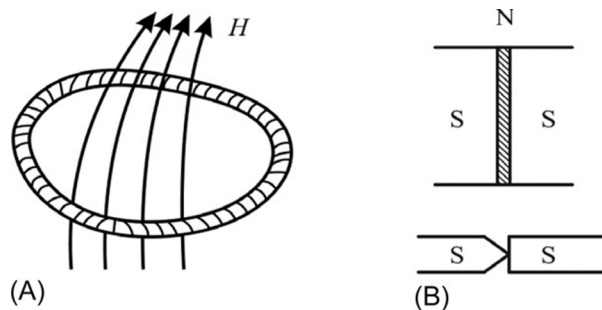


FIG. 5.20

Quantum properties of superconductivity: (A) frozen magnetic flux through superconducting ring; (B) schematic representation of Josephson contacts—weak connection.

Then temperature might be lowered below T_c and the source, creating the magnetic field, is turned off. At the time of switching off the magnetic flux, the decrease starts inducing electromotive force in the ring with a current. This current will prevent the reduction of magnetic flux; therefore after turning off the external magnetic field, the magnetic flux in the ring will remain at the same level, because it is supported by a current in the superconducting ring [2].

If this ring has a resistance R , after turning off the external field the current in the ring (that has inductance L) will disappear with time dependence of $\exp(-t/\tau)$, where $\tau \sim R/L$. Since the superconductor has $R=0$, the time of current existence is $t = \infty$. This means that the magnetic flux is “frozen” due to constant current in the superconducting ring. Moreover, this “frozen” magnetic flux has not any, but a certain, value multiplied by $\Phi_0 = 2.07 \times 10^{-7} \text{ Gs/cm}^2$. This parameter is a fundamental

constant, the *quantum of magnetic flux* that can be expressed through other fundamental constants: $\Phi_0 = hc/2e$. Magnetic flux quantization identically confirms the *quantum nature* of superconductivity.

Josephson effects. These effects also demonstrate the quantum nature of superconductivity. There are two Josephson effects: stationary and nonstationary, and they both belong to a *weak* superconductivity—when two semiconductors are connected through any nonsuperconducting layer. This “weak coupling” can be tunnel junction, thin-film narrowing, and, finally, simple touching of one superconductor to another at a point (Fig. 5.20B).

The *stationary* Josephson effect is a small current, passing through the poor connection, even if this weak link is a dielectric layer. Under created conditions, the weak link does not show electrical resistance when movement of electrons in both superconductors is agreed coherently. The weak connection does not prevent superconducting electrons to be in the same quantum ensemble. In other words, wave function of electrons is able to penetrate the weak connection from one superconducting area to another—this is the *interference*. All electrons in the macroscopic superconductor can be described by a single wave function (such as electrons in individual atom).

The *nonstationary* Josephson effect is the increase of current through a weak connection when voltage V is applied to it. Then, under the influence of voltage, in addition to constant component the *variable current component* appears, described by frequency ν , related to applied voltage by ratio $\nu = 2eV/h$. The frequency of this generation is very high (located in the range of microwaves), and this frequency is a linear function of applied voltage. Nonstationary Josephson effect can be used to generate microwaves in the GHz range [5].

Electron-phonon interaction. It is important that in superconductors the *isotopic effect* is discovered: this means that the temperature of phase transition depends on the *mass* of ions of crystal lattice. Such experimental data clearly indicate the *active part* of lattice oscillations (i.e., ionic cores) to create the superconducting state. According to results of theoretical analysis, *interaction* between electrons and lattice vibrations is the main reason for superconductivity in simple metals and their alloys. Under certain conditions, electron-phonon interaction might have a character of attraction. If this *attraction* is stronger than Coulomb repulsion between electrons, it dominates between charge carriers; as a result, at very low temperature, superconductivity looks as the more ordered and, therefore, more comfortable energy state.

For a simplified analysis of electron-phonon interaction, let us initially assume that in a metal at temperature $T = 0$ no thermally excited phonons exist (it is believed that nothing disturbs a lattice or nothing interacts with it). When an electron moves in crystal lattice with wave vector \mathbf{k}_1 , it can collide with a stationary ion, and due to scattering process this electron will turn into another energy state with wave vector \mathbf{k}'_1 . In such a case, one can say that “electron generates” phonon, which was absent before scattering. Lattice is characterized by invariant translations; therefore the law of impulse conservation has to be implemented:

$$\mathbf{k}_1 = \mathbf{k}'_1 + \mathbf{q}.$$

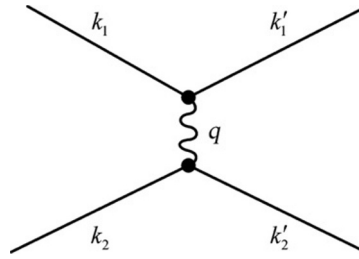


FIG. 5.21

Diagram explaining electron-phonon-electron interaction in superconductors.

Next, this phonon can be absorbed by a second electron with wave vector \mathbf{k}_2 , forcing it to move to state \mathbf{k}'_2 . As the phonon was generated and it disappears, the electron's impulse before and after scattering must be the same:

$$\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}'_1 + \mathbf{k}'_2.$$

It is considered that such scattering corresponds to the *electron-phonon-electron* process, or otherwise, to indirect electron-electron interaction that can be characterized by a diagram shown in Fig. 5.21. When electron goes from state \mathbf{k}_1 to state \mathbf{k}'_1 , the oscillations of electronic density occurs with a frequency

$$\nu = [E(\mathbf{k}_1) - E(\mathbf{k}'_1)]/h,$$

where $E(\mathbf{k}_1)$ and $E(\mathbf{k}'_1)$ are energies of initial and final state of electron, respectively.

Suppose that as a result of such fluctuations of electronic density in one place this density locally becomes increased. Positive ions of lattice will sense this temporary attraction that occurs in this place. They will move to it, and, having relatively larger mass and inertia, will continue their movement even after the compensation of local negative charge is achieved. This, in turn, results in the excess of a positive charge in the same place. Now it becomes the center of attraction for electrons, to where they move toward from nearby regions. As a result, a dynamic picture is created in such a way that it looks like an *attractive interaction* between electrons.

However, it should be noted that attraction by this scheme is only possible if distinctive frequency of such interaction is less than the own frequency of ionic subsystem (last is characterized by Debye frequency ν_D). In order for electron to move to state \mathbf{k}'_1 from its initial state \mathbf{k}_1 , first of all, this state should be free (Pauli principle). This is possible, as is known, only near the Fermi surface (or in the vicinity of Fermi energy) that can be simplistically represented as a sphere of radius k_F in the k -space (Fig. 5.22).

Next it is possible to formulate the rule of interaction of electrons via phonons involving, or, finally, through interaction with them [1]. The electrons, whose energy is different from Fermi energy on a value $h\nu_D$, can attract together (with the remaining electrons continuing to leave). Considerable attraction is peculiar only to those

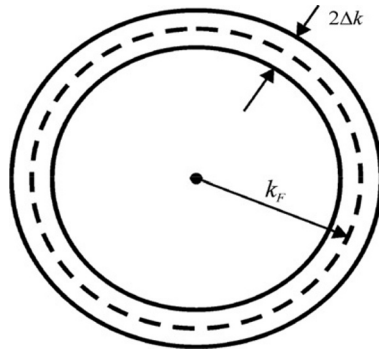


FIG. 5.22

Schematic representation of Fermi surface in metal; interaction of Cooper electrons is possible just near this surface layer.

electrons in which energy states lie in a narrow spherical layer around the Fermi energy; its thickness $2\Delta k$ corresponds to energy $2\hbar\nu_D$ (Fig. 5.22).

High-temperature superconductivity (HTS) is possible not in classic metals but in other crystals—semiconductors and even dielectrics. In these uncommon cases, at low temperatures the conductivity, instead of smooth reduction to zero, can be converted abruptly to endless value (despite relatively small concentration of charge carriers). At that, the energy of electron connection into Cooper pair is carried out not by electron-phonon-electron interaction, but through some other mechanisms (excitonic type) that also can result in superconductivity.

The crystals that have high density of excitonic states appear to be promising materials for elaboration of high-temperature superconductors. For a long time, theoretical predictions showed the possibilities of electron attraction and Cooper pairs rising by means of *excitonic exchange*. In principle, with such predicted mechanisms, superconductivity can be obtained even at 300 K (currently, in 2017, high-temperature superconductivity reaches temperature of about 200K).

When discussing the possibility of superconductivity, the term “exciton” should be interpreted clearly: it means *any polarized excitation in the electronic subsystem* of a crystal, including the variety of vibration modes of spatial or surface type. As classic superconductors are three-dimensional (3D) metals or alloys, while excitons can be extended only in dielectrics, the excitonic superconductor has to be *both metal and dielectric simultaneously*, introducing a system of “crystal in the crystal.” This system might be a complex substance, in which the metallic subsystem allows free movement of electronic pairs, while the dielectric subsystem is the environment for excitons spreading that join electronic pairs. In this case, the dimensionality of a matter should be decreased.

Among other possibilities, excitonic mechanism of electron coupling in Cooper pairs might be possible in the one-dimensional system (1D, needle-like crystal). It is,

for example, the long well-conductive molecule thread with easily polarizable side radicals that can provide attraction of conductive electrons due to excitonic exchange. The presence of excitons makes an appearance of high-temperature superconductivity possible because excitons can compensate Coulomb repulsion of electrons. However, achievement of superconductivity in the 1D crystal is almost impossible due to thermal fluctuations (so-called Peierls prohibition). Nevertheless, in the 1D system phase transition from the *quasimetallic* phase to the *high- ϵ dielectric* phase is possible: quasi-1D highly conductive (above T_c) system at low temperatures turns into “superdielectric” with $\epsilon \approx 2000$ [5].

In the vicinity of phase transition, physical properties of 1D structure are very sensitive to fluctuations. Theoretically, in 1D longitudinally ordered structure a full disordering (with violation of main properties) can occur just in one point. However, in reality, in quasi-1D structures (thin, needle-like, but still macroscopic crystals by their thickness), the situation changes, and stability of system to fluctuations increases due to the interaction between neighboring “threads” of such structure. The degree of “three-dimensionality” is qualitatively assessed by degree of anisotropy of conductivity and permittivity in these crystals. Three-dimensional interaction not only can “extinguish” fluctuations, but also can suppress Peierls transition. Due to this suppression, in some quasi-1D structures superconductivity becomes possible: for example, in the polymer $(\text{SN})_x$ that is, a quasi-1D superconductor, the dielectric phase does not occur; however, temperature of transition is very small ($T = 0.3$ K).

Impact of fluctuations onto the phase transition of metal-dielectric type of crystals is minimal in ordinary 3D structures, in which violation of ordering should occur on certain *surface* inside a crystal. The 2D structures, in terms of resistance to fluctuations, are found in intermediate position, as for destruction of their ordering, that is, “fluctuating break” should be seen on certain *line* (but not in a *point* as for 1D structures). Therefore, in quasi-2D structures, the probability to obtain superconducting state is much greater, than in 1D structures.

Electron formation into the Cooper pairs is promoted by large permittivity (ϵ) that strongly reduces Coulomb repulsion of electrons. As is known, at helium temperatures some paraelectrics and ferroelectrics have very huge permittivity. Indeed, superconducting phase transitions in these *dielectrics* were first discovered in the *doped* strontium titanate (it has $\epsilon \approx 40,000$ at temperature $T \leq 4$ K), as well as in narrow-gap ferroelectric-semiconductor SnTe ($\epsilon \approx 2000$). Although the temperature of superconducting transition in these dielectrics is less than 0.3 K, a *possibility* of superconductivity in these cases seems fundamentally significant.

This opportunity is used to find HTS in the mixed oxides of the *perovskite structure* (which is a typical structure for ferroelectrics). HTS was discovered experimentally only in 1986, although theoretical prediction of this phenomenon was long time before, known as for 2D and 1D nonmetallic structures. Theoretically, the mechanism of electron interaction by excitonic exchange in these structures has no temperature limit.

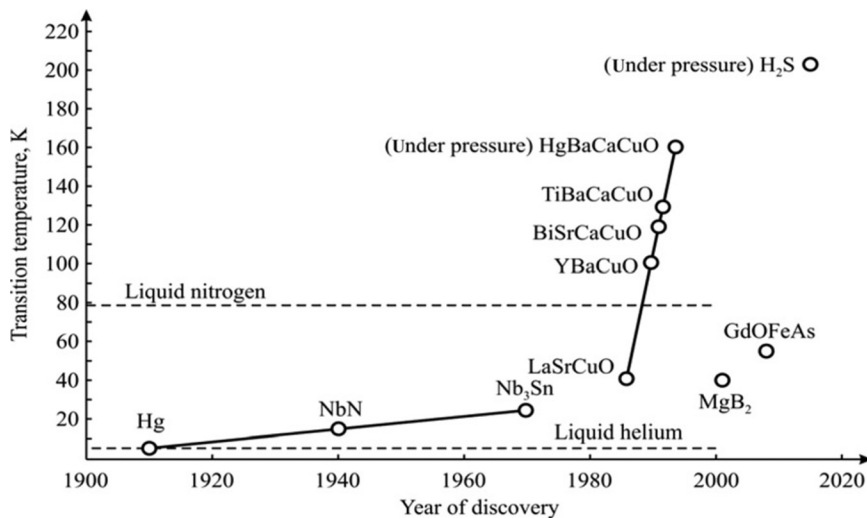


FIG. 5.23

History of superconducting materials research.

A short history of HTS is shown in Fig. 5.23. Among superconductive 2D systems (complex oxides), first several tungstates of A_xWO_3 type were discovered (A is alkali metal). Phase transition into superconductive state in tungstates was observed at temperatures up to 7 K. Then it turned out that in another complex oxides superconducting transition occurs even at temperature near $T = 13$ K, for example, in the compound $LiTi_{2-x}O_4$ and in the *ferroelectric* $BaPb_{1-x}Bi_xO_3$. This solid solution is of interest not only because superconductivity occurs in the material with *low density of charge carriers*, but at relatively high temperature. Similar to most ceramic materials, $BaPb_{1-x}Bi_xO_3$ demonstrates chemical and thermal stability; these ceramics can be made by a standard technology (including thin films). Films of ferroelectric-oxide superconductors are considered as promising for use in various devices of cryogenic (helium) electronics.

Finally, a significant increase of superconducting phase transition temperature is achieved: firstly, up to 40 K in ceramic compound $La_xBa_{1-x}CuO_4$, next superconductivity is discovered *above nitrogen temperature* (liquid nitrogen boils at temperature $T = 77$ K). The mechanism of superconductivity is of a bipolaronic type: polarons bound in Cooper pairs (like electrons in metallic superconductors). The discovery of HTS in the polycrystalline oxides, such as compounds $Y_xSr_{1-x}CuO_4$ ($T_c \approx 35$ K, Nobel Prize for 1987) and $YBa_2Cu_3O_{7-x}$ ($T_c \approx 100$ K), becomes the basis for new components of electronic equipment [5].

In 2001 superconductivity in the fusion MgB_2 (magnesium diboride) was discovered with relatively high transition temperature: $T_c = 40$ K. The crystal structure of this substance consists of boron and magnesium layers (earlier certain compounds of copper and oxygen, so-called cuprates were believed to have HTS properties).

In 2008, several iron-based compounds became known as superconducting at high temperatures. At present, the “record” of HTS is 203 K: in hydrogen sulfide (H_2S) under pressure of 150 gigapascals. Theoretically, there is possibility of superconducting materials development with an operating temperature of 300 K (available experimental information is controversial).

Thanks to the discovery of HTS, it becomes possible to construct high-speed computer memory devices, microwave converters and generators, electronic sensors, and others. Microwave technique from many of HTS mainly uses the composition designated as “1-2-3” with the chemical formula $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. In Fig. 5.23 this compound is designated as Y-Ba-Cu-O that has a transition temperature slightly higher than 100 K. This transition temperature is sufficient for HTS application at a temperature of 77 K, that is, with cooling by liquid nitrogen. This type of cooling costs hundreds of times less than cooling by liquid helium. Moreover, at increased operating temperature of HTS, microwave devices can be applied in the space electronics. The technology of “1-2-3” composition is well developed: Y-Ba-Cu-O is prepared as thin poly- and monocrystalline films, deposited onto dielectric substrates that have low microwave losses, such as MgO , LaAlO_3 , Al_2O_3 (sapphire), and so on.

In high-frequency and microwave technologies, conductors and superconductors are compared by their *surface resistance* R_S , measured in ohms. For ordinary metals the value of R_S is defined by the skin effect, then

$$R_S = (1/2\rho\mu_0\omega)^{1/2},$$

where ρ is resistivity, μ_0 is magnetic constant, and $\omega = 2\pi\nu$ is circular frequency. Thus, R_S in conventional metals *slowly increases* with frequency as $\sqrt{\nu}$.

In superconductors, particularly in HTS, the skin effect is absent, but there is another effect uncomfortable for microwave applications: the *depth of penetration* of electromagnetic field in surface of superconductor. The reason is the presence of not only Cooper pairs of electrons (which do not cause any resistance), but also *ordinary electrons* that make such a resistance in superconductors. It is determined that at high frequencies in superconductors, R_S is *nonzero*, and it rather increases *fast* with frequency: $R_S \sim \nu^2$.

Theoretical calculations are well documented experimentally. Frequency characteristics of the best (at room-temperature) conductor, copper at 77 K, niobium in the normal state (at temperature 300 K), and in the superconducting state (at liquid helium at temperature 5.2 K) are compared in Fig. 5.24. The most important in this figure is the frequency dependence of the Y-Ba-Cu-O film: surface resistance increases with frequency very rapidly as in classic superconductor Nb, so also in the HTS film. Therefore, at millimeter waves (100 GHz and above), superconductors have no advantage over copper.

However, at frequencies lower than 20 GHz, the HTS electrodes have a significant advantage as compared to Cu (even being cooled to liquid nitrogen, 77 K). At that, electrodes made of niobium look better than HTS, but they need very expensive cooling by helium. Superconducting $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ films deposited on dielectric substrates are successfully used instead of the usual superconductive metals in

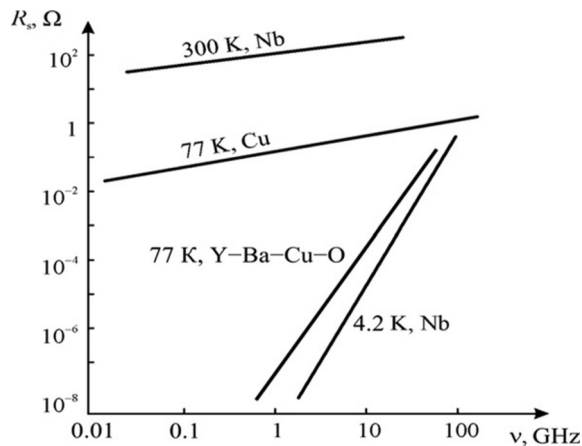


FIG. 5.24

Frequency dependence of surface resistance in superconductors and copper.

microstrip and coplanar microwave devices, thus allowing obtaining record high-quality filters, phase shifters with small losses of energy, and other *passive* microwave devices [5]. There are also some *active* devices based on the HTS films that use Josephson effects.

5.8 SUMMARY

1. Mechanical properties of metals are characterized by great durability, hardness, and other parameters; these properties are dependent on mechanical and thermal processing, determining structural defects and impurities. Annealing of metals results in a significant decrease in their strength and increases their compliance. Assuming metal as a system, in which positive ions are fastened by means of freely mobile electrons, corresponds to basic properties of metals: ductility, plasticity, high values of thermal conductivity and electrical conductivity.
2. Most metals crystallize in one of three main structural types: cubic or hexagonal dense packing and space-centered cubic lattice. In dense packing, each metal ion is located at equal distances from 12 nearest neighbors. Interatomic distances in crystalline structure of metals are characterized by “metallic ionic radius.”
3. The *negative* temperature coefficient of electrical conductivity should be considered as the most characteristic physical feature of metals, that is, electrical conductivity decreases with increasing temperature. At that, temperature dependence of conductivity is close to law $\sigma(T) \sim T^{-1}$. In case

of deep cooling, the $\sigma(T)$ dependence shows saturation, the level of which depends on concentration of static defects. In metals that have ferromagnetic impurities at low temperatures, the $\sigma(T)$ dependence may show even a decrease while cooling.

4. Electromagnetic waves up to optical range are almost entirely reflected from the surface of a metal, so that metals are nontransparent for electrical field and demonstrate metallic luster. Metals have high thermal conductivity λ_e [W/(K m)] caused by high-mobility electrons; at that, thermal conductivity λ_e is proportional to conductivity σ . The uniformity of ratio λ_e/σ for various metals is Wiedemann-Franz law. Electrons in metals are always found in fast movement; they move even at lowest temperatures (near absolute zero). This motion of electrons is chaotic; therefore different electrons move with different velocity. In most metals, electrons move with Fermi velocity: $v_F \sim 10^6$ m/s.
5. In an “ideal metal” with infinite conductivity, electrical field cannot exist. In a real metal only very small electrical field can be applied, as high current will lead to metal melting. In the presence of external electrical field, the current flows through a metal, that is, movement of electrons becomes partially directed; electron *drift* superimposes over their chaotic movement. To calculate, this current one needs to estimate average drift velocity v_{ev} (velocity of chaotic motion is independent of applied field).
6. Direct proportionality of electron drift velocity to strength of electrical field is characterized by the *mobility*. Mobility characterizes the resistance of substance to electron drift in direct electrical field. If inhibitory force is absent, then electrons will move in the electrical field with *acceleration* (such as in vacuum), but not with constant average velocity as they move in crystals. Therefore mobility is degree of electron freedom in crystals.
7. One experimental method of electron concentration n_e measuring in conductors (and semiconductors) is Hall’s effect—the difference of potential *across* investigated sample placed in perpendicular magnetic field, when current flows through the sample.
8. *Magnetoresistance* is the change of material’s electrical resistance in magnetic field. It depends on the sample orientation relative to the magnetic field; that is, magnetic field does not change projection of particle’s velocity on direction of magnetic field, but, due to Lorentz force, bends the trajectory in plane *perpendicular* to magnetic field.
9. Movement of electron in real metal under external electrical field is not continuous but interrupted: as soon as electrical force appears, the electron starts moving with *acceleration* and its velocity gradually increases, but also the force of “friction” increases that is *proportional* to electron’s velocity. After time $\tau = m_e u$ the inhibitory force compensates completely

the acceleration force; therefore electron (or hole) moves with constant velocity. The *time constant*, characterizing installing of steady state of electron movement in metal, is small: $\tau \approx 2 \times 10^{-14}$ s.

10. Parameter τ is average time between two collisions of electron; using its product by Fermi velocity v_F it is possible to estimate average *free path* δ (middle distance between electron collisions): $\tau v_F = \delta$. Charge transfer in metals is carried out by *electrons and holes* located near Fermi surface; they take energy from applied electrical field. As far as electrons (being excited by external field) move to higher levels, they are replaced by other electrons, previously located much deeper under Fermi level.
11. At high and ultrahigh frequencies (10^8 – 10^{11} Hz), electromagnetic field penetrates into conductor (metal) to a *small depth*, and damps the faster the higher field frequency and magnetic permeability of metal. The result is nonuniform distribution of current density in the cross section of conductor—this is the *surface effect* (or skin effect).
12. Heat can be extended in metal rather easily by the same free electrons that determine metal conductivity. *Thermal conductivity* of metals is high as number of electrons per unit volume of metal is large. Therefore electronic thermal conductivity λ_e in metals usually is much higher than heat transport by phonons λ_{ph} in dielectrics that have predominantly lattice-vibration mechanism of thermal conductivity.
13. Despite large electronic conductivity, metals under normal conditions ($T \sim 300$ K) give a rather small electronic contribution C_e to specific heat. However, electronic contribution to heat capacity of metals increases in direct proportion to absolute temperature: $C_e = \xi T$ without any saturation at high temperatures. Besides, it is necessary to note that in metals at very low (cryogenic) temperatures heat transferred by “electronic gas” can surpass phonon contribution to specific heat.
14. In case of two different metal *connection* (and semiconductors also), the *difference of contact potential* occurs between them. If these connections are found at different temperature, a thermoelectric power is caused due to distinction in electronic work functions of two metals; thermoelectromotive power is dependent on free electron concentration in relevant metals.
15. Classic electronic theory of metals is based on Drude hypothesis about *free electronic gas*. This theory enables to explain and describe analytically most experimental data in metals (conductivity, power losses, relationship between electrical conductivity, thermal conductivity, etc.). In addition, some experiments confirmed the hypothesis of electronic gas, such as the curvature of electron trajectory in metal placed in transverse magnetic field, as well as the change of electrical resistance in magnetic field. However, Drude’s theory has contradictions with some experimental data. For

example, this theory cannot explain experimentally observed *paramagnetic* susceptibility in some metals; another discrepancy is the theoretically obtained value of specific heat: experimentally seen specific heat of metals is much less than it follows from Drude's theory (it looks like electronic gas hardly absorbs heat while metal is heated). Quantum mechanics overpasses these contradictions.

16. Unlike classic electronic theory, quantum mechanics show that electronic gas in metals under normal (and higher) temperature is found in the state of *degeneration*. In this state, the energy of electronic gas is almost *independent on temperature*, that is, any alterations in thermal motion of ionic lattice almost do not change energy of free electrons. Therefore, while crystal is heating the energy practically is not consumed by electronic gas, as clearly seen during heat capacity measurements. In a state similar to conventional gases, the electronic gas would have temperature estimated as thousands of degrees.
17. Quantum mechanics explain why for metals normal temperature looks like very "low": because Fermi energy of electrons in metals is large (about 5–10 eV). Electronic gas in metal is always degenerated; therefore energy distribution function at normal temperature varies only in narrow range $k_B T$ near Fermi energy level. From Fermi surface electrons can be easily transferred to the allowed higher energy levels; for this reason, metals are good conductors of electricity.
18. Fermi surface in the space of impulses at 0K separates occupied by electron ground states from the empty states. At nonzero temperature, most electrons, however, are placed *under* Fermi surface, and only some of them shift outside of it (in *energy stripe* $k_B T$). Fermi surfaces in different metals might have quite a different form. In metal energy, increase can be described as the birth of quasiparticles. The increase of electron energy in metal is possible by moving at least one electron under Fermi surface into external allowed level in the impulse space. At that, the appearance of this electron *above* Fermi surface is accompanied by unoccupied state *below* Fermi surface—the hole, which can be interpreted as a kind of quasiparticle that is the antiparticle as to the born electron. Therefore increasing energy of free electrons in a metal always is accompanied by the birth of *two quasiparticles*. Calling hole as antiparticle, there is the possibility of its recombination, when electron will return "in its place" under Fermi surface, and metal again returns closer to its *ground state*, because both quasiparticles—electron and hole—disappear.
19. Behavior of conductors in magnetic field can be determined, taking into account, firstly, that magnetic field does *not change energy* of electron (Lorentz force vector is directed *perpendicular* to velocity of electron), and, secondly, Lorentz force has no effect on electron, which moves along the magnetic field.

20. Mechanisms of electrons *scattering* (collisions) can be divided into three classes: (1) electron collision with static violations of crystal periodicity; (2) electron collision with thermal motion of lattice (phonons); (3) collisions of electrons with each other. At normal (room) temperature, the main cause of electron scattering, and, hence, resistance of metal is not due to crystal lattice defects, but to phonons—thermally excited wave-type motions of ionic lattice. Only at very low temperatures (near absolute zero), when thermal motion in crystal is almost frozen, scattering on defects of structure becomes a more important mechanism of electrical resistance.
21. In recent years, an important class of solids with *anomalous properties* was discovered: they are compounds of rare-earth metals, characterized by incomplete *4f*-shell. Their electronic properties are difficult to explain using existing concepts. These substances are intermediate between magnetic and nonmagnetic materials, as well as between metals and dielectrics, because most of their electrons are found among localized and free states. Studies of these compounds help to understand many properties of metals and magnetics, to extend conceptions of band theory for metallic and dielectric states, as well as to investigate possible types of electronic states in crystals.
22. Some compounds and metals are known, in which electrons have the binary, *ambivalent* nature: keeping largely localized (atomic) nature, they also can show intention to collectivization. Systems with unstable valence (or intermediate valence) belong to compounds of rare-earth metals and actinides, that is, the elements that have incomplete *4f*- or *5f*-shells. These compounds have unique physical properties and anomalous characteristics that explain formation of *heavy fermions*—peculiar electrons that have effective mass 10^2 – 10^3 times greater than mass of free electron.
23. Heavy fermions arise in such a system where the proximity of valence instability exists. Such fermions have large effective mass; the heaviest fermions are observed primarily in compounds of cerium and uranium—the elements located in beginning of *4f*- and *5f*-periods. Among compounds with heavy fermions, there are magnetic materials, disordered in their normal state crystals, as well as superconductors.
24. Superconductivity is one of “cooperative” effects in the ensemble of conduction electrons, and it cannot be explained by simple models of “independent electrons.” Superconductivity is interesting in the sense that superconductors have *no skin effect* and can significantly *reduce energy loss*. In superconductors, two effects were found by Josephson: (1) current that flows indefinitely long time without any voltage applied across a junction that consists of two superconductors, parted by weak link; (2) when external voltage is applied, Josephson junction demonstrates effect of microwave generation. Using these effects a variety of sensitive electronic devices are elaborated.

25. DC resistance of superconductor is zero: $\rho = 0$; respectively, its conductivity is the infinity: $\sigma = \infty$. This phenomenon was first discovered in mercury at temperature $T_c \sim 4\text{ K}$, but in some metals and alloys, phase transition into superconducting state is observed at higher temperatures: for lead $T_c \sim 7\text{ K}$, for niobium $T_c \sim 9\text{ K}$. In alloy Nb_3Sn transition temperature is $T_c \sim 18\text{ K}$, while alloy Nb_3Ge shows highest observed temperature for *metals*: $T_c \sim 23\text{ K}$.
26. The nonstationary Josephson effect arises if voltage V is applied to a weak connection; in addition to constant component of current, the variable component appears that is determined by frequency ν connected with voltage by ratio: $\nu = 2eV/h$. This generation lies in ultrahigh-frequency range, and its frequency is a linear function of applied voltage. Nonstationary Josephson effect is used to generate microwaves in the GHz range.
27. Phase transition temperature of superconductors is their *critical temperature* T_c . For *cryoelectronics*, among most prominent superconducting *metals* are lead and niobium. Best (under normal conditions) metallic conductors, such as copper, gold, and silver, whose resistance at temperature 300K is minimal among other metals, cannot have superconducting state. At low temperatures, they are only *cryoconductors*.
28. Discovery of high-temperature superconductivity (HTS) in polycrystalline oxides, for example, in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ ($T_c \approx 100\text{ K}$), becomes a basis for new components of electronic equipment. At present (year 2017), the “record” of superconductivity appearance is 203 K. Due to HTS the high-speed memory devices for computers, microwave converters and generators, as well as other electronic devices are constructed, cooling by liquid nitrogen (77 K). Theoretically, the possibility exists to develop superconducting materials with an operating temperature 300 K.

REFERENCES

- [1] N.W. Ashcroft, N.D. Mermin, *Solid State Physics*, Holt and Winston, New York, 1976.
- [2] Y.M. Poplavko, *Physics of Metals, Electrical Properties*, Avers, Ukraine, Kiev, 2009.
- [3] R. Waser (Ed.), *Nanoelectronics and Information Technology: Advanced Electronic Materials and Novel Devices*, Wiley-VCH, Weinheim, 2005.
- [4] L.H. Van Vlack, *Materials Science for Engineers*, Addison-Wesley Publishing Co., Massachusetts, 1975.
- [5] Y.M. Poplavko, S.A. Voronov, Y.I. Yakimenko, *Materials Sciences, Part 3, Conductors and Magnetics*, Kiev Polytechnic Institute, Kiev, 2011.
- [6] C. Kittel, *Introduction to Solid State Physics*, John Wiley and Sons, New York, 1976.
- [7] M.I. Kaganov, *Electrons, Phonons, Magnons*, Nauka, Moscow, 1979.
- [8] J.C. Sletter, *Insulators, Semiconductors and Metals*, McGraw-Hill, London, 1967.
- [9] D.I. Khomskii, The problem of intermediate valency, *Sov. Phys. Usp.* 22 (1979) 879–903. <https://doi.org/10.1070/PU1979v022n11ABEH005645>.

Magnetics

CONTENTS

6.1 Basic Definitions	222
6.2 Disordered Magnetics	229
6.3 Ferromagnetism	240
6.4 Antiferromagnetism and Ferrimagnetism	259
6.5 Nanomagnetic Materials	268
6.6 Summary	278
References	285

Magnetism is a special kind of physical interaction from a distance between moving electrical charges (that creates magnetic moment), as well as the interaction between particles and bodies with naturally existing magnetic moments. Magnetic interaction is characterized by the magnetic field.

There is no complete symmetry between magnetic and electrical fields. The sources of electrical field are electrical *charges*, but similar single “magnetic charges” are absent in nature. That is why, the sources of magnetic field are also electrical charges, but only the *moving charges* (even if this movement is hidden in elementary particles). Cyclically moving charges create *magnetic moment*, also called the *magnetic dipole*.

Any material is magnetosensitive in its nature, that is, it interacts with an external magnetic field and has certain magnetic property. In every matter, elementary circular current exists, such as the rotation of electrons around the nucleus (orbital magnetism) and fictitious rotation of electrons around their own axis (spin magnetism). These movements lead to orbital and to spin magnetic moments, both created by the electrons. Magnetic moment of *electronic shells* of atoms determines magnetic properties of any material, because this moment is 1000 times greater than the magnetic moment of the atomic *nucleus*. That is why exactly the peculiarities in *electronic structures* of atoms stipulate differences in magnetic properties of substances [1].

Magnetic materials are widely used in many areas of technologies: electronics, electrical engineering, information, computing and measuring instruments, and others. In recent years, a qualitative “jump” in the development of magnetic materials is seen, and on this basis, new types of electromagnetic and magneto-electronic devices with unique properties are created owing to scientific discoveries in the physics of magnetic materials and advanced technologies.

The current stage of magnetic device development is characterized by the transition from the use of discrete magnetic components to the application of homogeneous magnetic environments, when magnetopolarized charge carriers by their directional movement create magnetic domains. Low-inertia reorientation of magnetic domains is widely applied in magnetic electronic devices. The small size of magnetic domains together with their high mobility can create on their basis various *functional devices* that have large memory (10^9 – 10^{10} bits) and high density of information (10^8 – 10^{10} bits/cm²), as well as characterized by great speed of processing (10^6 – 10^8 bit/s) [2].

Further progress in the creation of materials with new properties is due to the development of *nanotechnologies* for supersmall elements. This progress is associated with the changes in the structure of matter, thus affecting its fundamental properties. Currently, it has become possible to “manage” properties of substances by decrease in their fragmentation (dimensions). At present, most technical implementations of nanoelectronic elements are observed only in the magnetic electronic devices [3]. Based on microelectronics and nanoelectronics, one of the promising areas of functional electronics is the development of *magnetic electronics*, which is qualitatively new stage in the creation of components to build a broad class of logical and storage devices, as well as various information-processing devices.

At present, magnetic materials with large magneto-optical effects are synthesized, and they combine good transparency of the material in visible and near-infrared regions of the spectrum. On this basis, many advanced magneto-optical elements and devices are developed: magneto-optical drives, controlled banners, printers, deflectors, integrated optics elements, various converters, and so on. Owing to a variety of types of magnetic materials, their properties and manufacturing methods promote the creation of new items and devices. Magnetic electronics require the development of electronic equipment through the knowledge of physics of magnetism, features of magnetic interaction in solids, getting control over manufacturing technology materials with different properties, and exact understanding of modern technology and trends.

This chapter focuses on the physical fundamentals of magnetism in solids, processes that determine principles of magnetic electronic devices, as well as operation characteristics, requirements, scope, and prospects of development of magnetic devices.

6.1 BASIC DEFINITIONS

For quantitative description of magnetic phenomena in solids, the *axial vectors* are introduced: magnetic field H , magnetic induction B , magnetic moment M , magnetic flux Φ , and magnetization J .

The *magnetic field* is specified by its direction and strength; it characterizes magnetic effect, which is created by electrical currents or by internal magnetic properties of materials. The symbols B and H are used for two distinct but *closely related* fields. In SI units, the vector H is measured in amperes per meter [A/m], whereas vector B is

measured in Newton per meter per ampere $[N/(m \cdot A)] = \text{tesla [T]}$. Magnetic induction B sometimes is also called as “magnetic field” (it is, most commonly, defined in terms of Lorentz force, which acts on moving electrical charges).

The *magnetization* (magnetic polarization) is also an axial vector field that expresses the density of permanent or induced magnetic dipole moments in a magnetic *material*. It can be compared with electrical polarization, which is the measure of response of a material under an electrical field. Magnetic moment per unit volume is represented by the vector M . Sometimes, during practical investigation of magnetism, the *magnetic flux* Φ *through surface* is also used. This is the surface integral of the normal component of magnetic field B , which passes through the surface; the SI unit of magnetic flux is weber ($\text{Wb} = \text{V} \cdot \text{s}$), whereas the CGS unit for magnetic flux is maxwell [1].

Magnetic *field strength* H , in a more specific description, is defined as a certain distance l from the conductor through which current I flows, and it is given by $H = I/2\pi l$. This ratio determines the field dimension in SI unit $[\text{A/m}]$. According to this definition, magnetic field H is *not dependent* on the magnetic properties of a medium. For example, in the center of a round loop of wire with radius R and circular current I , the magnetic field $H = I/2R$ *regardless of the environment*. The unit of magnetic field H in the CGS system is oersted (abbreviated as $[\text{Oe}]$), which is identical to dyne/maxwell. The oersted is $1000/4\pi \approx 79.6$ $[\text{A/m}]$.

Magnetic *induction* B is the main characteristic of the magnetic field *in the material*: it is the average value of the total intensity of microscopic magnetic fields generated by individual electrons and other elementary particles. In vacuum, magnetic induction B is defined only by an external magnetic field: $B = \mu_0 H$. To make agreement between dimensions of parameters that are used in magnetism, the SI system uses $\mu_0 = 1.25 \cdot 10^{-7} \text{H/m}$ (symbol $[\text{H}] = \text{“henry”}$ is the unit of electrical inductance in the International System of Units). In the Gauss system (GHS), the unit of magnetic induction is $[\text{Gs}]$ and includes a ratio: $[\text{T}] = 10^4 \text{Gs}$.

Magnetic *moment* M is one of the important characteristics of magnetic properties of a *body*. The sources of magnetism are both macroscopic and microscopic electrical currents. Magnetic moment vector is expressed by an analogy with electrical dipole moment, only remembering not electrical dipoles, but *magnetic dipoles* (formed by electrical currents in closed circuits). Magnetic moment has dimension $[\text{A} \cdot \text{m}^2] = [\text{J/T}]$ (joules per tesla), and it is the sum of all elementary moments that are induced in a substance under the influence of magnetic field (or it can be formed spontaneously, as in ferromagnetics).

The *magnetization* J is the density of magnetic moment M , that is, magnetic moment per unit volume of material: $J = M/V$. It corresponds to the macroscopic description of the magnetic state of a body. In the SI system, the dimension of magnetization coincides with the dimension of field strength ($J = \text{A/m} = \text{Wb/m}^2$) and represents such magnetization, when 1m^3 of the material has a magnetic moment of 1A/m^3 (remember that $\text{Wb} = \text{“weber”}$ is the SI unit of magnetic flux; hence flux density is Wb/m^2 , that is, one weber per square meter, which is one tesla).

By formal analogy that is used for dielectrics, in *polar* electrical vectors ($D = \epsilon_0 \epsilon E = \epsilon_0 E + P$, $\epsilon = 1 + \chi$, where D is the electrical induction, E is the electrical

field, P is the electrical polarization, ϵ is the permittivity, and χ is the dielectric susceptibility), the *axial* magnetic vectors are joined by material *tensors of second rank*: permeability μ and magnetic susceptibility α :

$$B = \mu_0 \mu H = \mu_0 H + J,$$

$$J = \mu_0 H,$$

$$\mu = 1 + \alpha.$$

The permeability and the magnetic susceptibility are *relative* values; therefore they are *dimensionless*. In vacuum, relative permeability $\mu = 1$ because in the absence of a substance, magnetic susceptibility is zero: $\alpha = 0$. Similarly, without any substance, relative dielectric permittivity of vacuum $\epsilon = 1$ and dielectric susceptibility $\chi = 0$ [3].

However, the analogy between electrical and magnetic phenomena is purely formal. This follows, for example, from Fig. 6.1, which compares magnetic and electrical dipoles. Electrical dipole is a system of two electrical charges separated in space, with equal magnitude and signs opposite to each other. Conventionally, electrical dipole is indicated by an arrow: it is the *polar vector*. Magnetic dipole is formed by the *movement* of electrical charges, approximately representing circular electrical current: it is the *axial vector*. For both types of dipoles (electrical and magnetic),

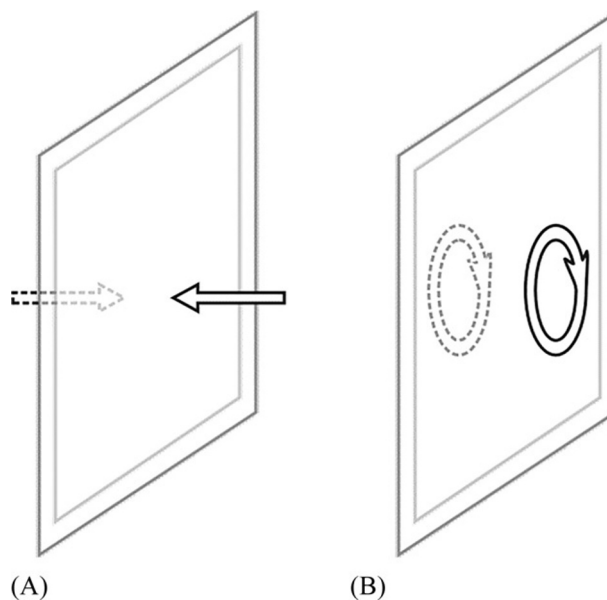


FIG. 6.1

Mirror-like reflection: (A) from electrical dipole; (B) from magnetic dipole.

the *potential* of corresponding fields decreases with distance as $1/r^2$. However, by this factor, they become dissimilar.

Significant differences between electrical and magnetic dipoles can be seen by reflecting them in mirror, as shown in Fig. 6.1. Electrical dipole changes its direction to opposite, whereas reflected magnetic dipole maintains its direction [4]. On the contrary, according to a mental operation called the *inversion in time* (when believed that passage of time is reversed), electrical dipole remains unchanged, whereas magnetic dipole changes its sign to opposite.

In the *macroscopic* examination, magnetic dipole can be represented by electrical current in a closed circuit, whereas in the *microscopic* processes, properties of magnetic matter might also be caused by the internal (hidden) forms of electrical charge movement, which is possible to describe only in *quantum mechanics*.

Microscopic magnetic dipole can be created:

- by changing the orbital moment of the electronic shell of atom (ion or molecule), it results in the diamagnetic component of magnetization;
- by “own rotation” of elementary particles represented by spins of electrons, which leads to the paramagnetic or ferromagnetic components of magnetization;
- owing to the presence of magnetism in some atomic nucleus (it should be noted that nuclear magnetism is weak).

Thus in the atomic scale, electrons can create two types of magnetic moments (microscopic currents): the orbital moment that is due to electron rotation around the atomic residue and the spin moment that is due to natural magnetic momentum of the electron. Practically, magnetism is the characteristic of the orbital and spin magnetic moments of *electrons*. Protons and neutrons also have their own magnetic moments, but nuclear magnetism, compared with electronic magnetism, is very small (around 1000 times weaker) because magnetic moments are related to the mechanical moment and therefore is *inversely proportional to the mass* of particles.

In this way, the smaller the magnetic moment of a particle, the greater is the mass. Therefore the magnetic properties of matter are determined mainly by electrons, as electrons are lighter by nearly three orders in magnitude than the atomic nuclei—proton, which is the lightest. However, in some cases (very rare but important for special studies and applications in physics, chemistry, and biology), nuclear magnetism might have considerable interest. First, only the effect of *nuclear magnetic resonance* has applications in medicine and in solid-state physics; second, the effect of *nuclear demagnetization* is used for deep cooling of matter to achieve experimentally very low temperatures [5].

The energy of magnetic interaction of microscopic particles, although it is smaller than the energy of electrical interaction, is still large enough to affect the structure of matter. As any stable system tends to minimize its energy, internal magnetic moments in substances strive for maximum compensation. For example, in the electronic spectrum of a crystal (see Section 4.6), electrons tend to occupy the lowest possible energy levels; each level can be occupied only by two electrons *with opposite values of spin* (Pauli principle).

For instance, in basic energy state of the helium atom (state $1s^2$), both the spin and the orbital electronic moments are zero; hence the magnetic moment *induced by an external field* can only occur. The same is applicable to hydrogen molecule H_2 . Thus in atoms or molecules with *completely filled electronic shells*, total spin moment and total orbital moment are zero.

Summary spin magnetic moment in the *completely* filled orbitals (2, 6, 10, 14 electrons) is totally compensated. Therefore in most substances, electronic orbitals of atoms and molecules, generally, are entirely filled (self-organized) with the *even* number of electrons (filled *s*-, *p*-, *d*-, and *f*- shells contain 2, 6, 10, and 14 electrons, respectively). Nevertheless, there are some quite uncommon (but very important for practical use) exceptions of stable but only *partially filled d*- and *f*- shells of atoms, in which uncompensated total spin magnetic moment can exist.

The main effects of an external magnetic field that influence matter were discovered in the 19th century by Faraday. First, according to the law of electromagnetic induction, an external magnetic field creates *induced* microscopic electrical current in a substance, whereas the magnetic field is directed *opposite* to the applied field (such reaction of matter to the applied magnetic field always exists). Faraday identified this effect as **diamagnetism**. The prefix “*dia-*” means the opposition to an externally applied field or deviation of magnetic field lines: external magnetic field turns around the diamagnetic, as shown in Fig. 6.2A. That is why diamagnetic *repels with any pole* of a permanent magnet (it is pushed out of the magnetic field but with a small force because this effect usually is very small). Magnetic induction in the diamagnetic becomes *smaller than that in vacuum* [1].

Second, if atoms (or molecules) of a matter have particles with natural nonzero magnetic moments (spin, or orbital, or both), an external magnetic field will orient them *along the field*. The result is the appearance of an additional magnetic moment that is collateral to the external field; Faraday called these materials as **paramagnetics**. The prefix “*para-*” means “consistency” of magnetism in a substance with magnetic field lines; magnetism in the paramagnetic becomes *stronger than that in vacuum*. In Fig. 6.2B, the magnetic field *draws into* a paramagnetic.

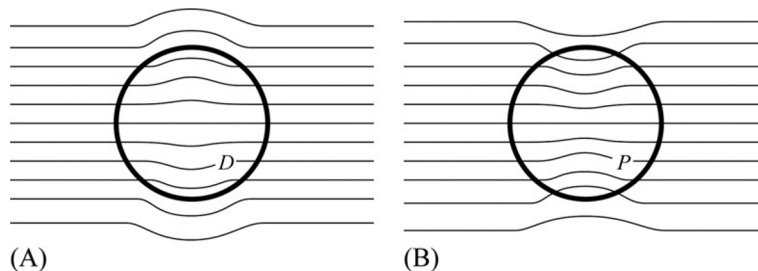


FIG. 6.2

Handling (A) diamagnetic (*D*) and (B) paramagnetic (*P*) in a magnetic field.

The paramagnetic is *attracted by any pole* of a permanent magnet. As the diamagnetism phenomenon exists always, this attraction indicates the preference for paramagnetism over diamagnetism (almost in all cases when both effects take place).

The dependence of magnetization on the magnetic field ($J = \mu_0 \alpha H$) for some typical cases is shown in Fig. 6.3. Magnetic moment induced in matter by an external field can be both positive and negative. Fig. 6.3A shows the comparison of magnetization under an external field for diamagnetic and paramagnetic. In both cases, to obtain a *noticeable effect*, the applied magnetic field has to be large (hundreds of [Oe] = oersted).

Significant magnetic properties, even under a small external magnetic field, can be seen in substances that have a strong *internal magnetic interaction* between particles—carriers of own magnetic moment (atoms, ions, and molecules). Through this interaction, the *involuntary ordering* of internal magnetic moments might be energetically favorable (without action of external magnetic fields). In these cases, a *strong magnetic effect* usually can be seen. The dependence of magnetization, induced by an external magnetic field in the *ferromagnetic*, can be seen in Fig. 6.3B: even if an external magnetic field strength is only 1 Oe, the induced magnetization is thousands of times greater than that in the diamagnetic or paramagnetic substances. Faraday has shown that a ferromagnetic is *attracted to both poles* of a permanent magnet.

Therefore in the diamagnetic, any proper magnetic moments of particles are absent: its magnetization is induced exclusively by the external field. Induced diamagnetic moment disappears very fast on removal of external field—at time of around 10^{-14} s. As to the paramagnetism, it is conditioned by the existence in a material intrinsic (natural) magnetic moments, which are, however, completely disordered if an external magnetic field is absent, as shown in Fig. 6.4A. Magnetization of a paramagnetic (similar to diamagnetic) is also *induced* by an external

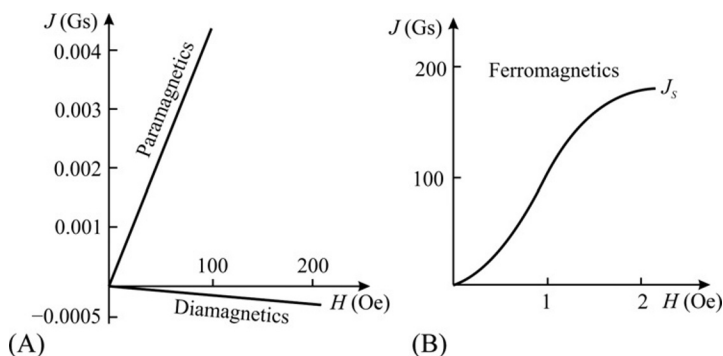


FIG. 6.3

Field dependence of magnetic moment induced in: (A) diamagnetic and paramagnetic, (B) ferromagnetic.

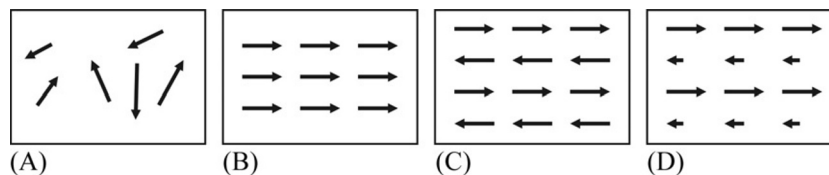


FIG. 6.4

Schemes of magnetic moment ordering in different lattices: (A) paramagnetic; (B) ferromagnetic; (C) antiferromagnetic; and (D) ferrimagnetic.

magnetic field owing to the *orientation* of existing natural magnetic moments. However, the external magnetic field orients only a small part of natural moments and, after switching off the magnetic field, magnetism induced in a paramagnetic disappears, but not so fast as in a diamagnetic (at time of 10^{-9} – 10^{-2} s) [6].

In some solids, their magnetic structures can be characterized by different types of *spontaneous* magnetic ordering. A crystal (or polycrystal), in which natural magnetic moments are oriented in parallel to each other, is a *ferromagnetic* (Fig. 6.4B). Accordingly, *antiferromagnetic* has neighboring atomic magnetic moments oriented in an antiparallel direction, as shown in Fig. 6.4C. Moreover, ferromagnetism and antiferromagnetism can coexist in a single structure; such material is the *ferrimagnetic*, in which compensation of atomic magnetic moments is incomplete, as shown in Fig. 6.4D. Related substances are known as *ferrites*, and they are very important for technical applications. Except for relatively simple collinear ferromagnetics, atomic and electronic structures of antiferromagnetics and ferrimagnetics might have more complicated and even noncollinear magnetic structures (i.e., spiral, triangular, etc.).

Thus magnetic properties of a substance can be divided into weak magnetism (diamagnetism and paramagnetism) and relatively strong magnetism (ferromagnetism, antiferromagnetism, and ferrimagnetism). The magnetization of materials differs significantly from polarization. For comparison, it should be pointed out that in case of electrical polarization in dielectrics, *static* dielectric susceptibility is always positive ($\chi > 0$); that is why, static permittivity of any material surpasses one ($\epsilon > 1$). However, while a matter is magnetized, depending on the nature of magnetism, the value of magnetic susceptibility α can be positive or negative. Hence in a substance, *static* magnetic permeability μ can be both greater than 1 ($\mu > 1$) and less than 1 ($\mu < 1$). The superconductor (in which electrical resistivity is zero, $\rho = 0$) formally is characterized by the value $\mu = 0$ (i.e., it has $\alpha = -1$) being supposedly the “ideal” diamagnetic.

The complexity of the atomic structure of matter, constructed from a wide variety of particles, leads to many forms of magnetic structures. While considering the properties of solids, usually the general term “magnet” is used. The association of magnetic properties of substances with their nonmagnetic properties (electrical,

mechanical, optical, etc.) enables the investigation of magnetic properties as the source of information on the internal microscopic and macroscopic structures of materials.

6.2 DISORDERED MAGNETICS

Magnetic materials can be divided into magnetically ordered and magnetically disordered structures. Magnetic properties of ordered materials usually are strongly pronounced, in particular, this causes widespread use of such materials in electronics and electrical engineering. Magnetically disordered solids, as a rule, show weak magnetism, but sometimes they also demonstrate useful properties for application in instrumental technique and in medicine. As already noted, in the cryogenic technology, weak paramagnetism is used to achieve *very low temperatures*. Similarly, in experimental physics (as in medicine), methods such as *electronic paramagnetic resonance* and *nuclear magnetic resonance* are very important for research and diagnostics, although these methods use weak magnetism [5].

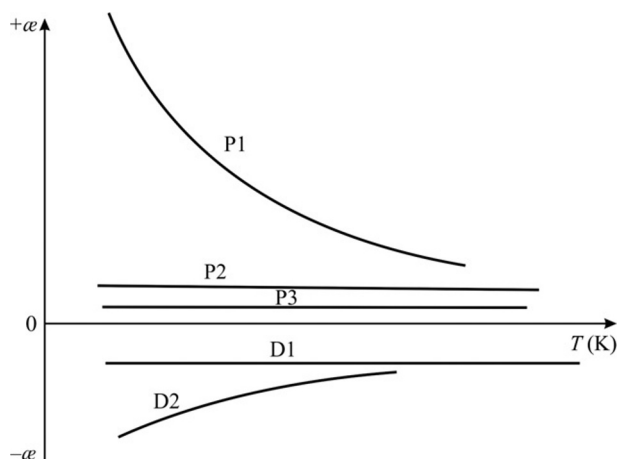
Therefore the terms “weak” and “strong” magnetism are conventional and might be used here only in understanding the engineering of these phenomena. For most calculations related to electromagnetic wave spreading, slowing, or absorption in a material, weak magnetism can be neglected because both diamagnetic and paramagnetic have magnetic permeability $\mu \approx 1$, which is only slightly different from vacuum value $\mu = 1$. At the same time, in strong magnets, the value of μ usually is rather high and can even be very large (sometimes, it reaches thousands).

Diamagnetism. Electrons, which move around the nucleus in their closed orbit, under the influence of an external magnetic field, change its trajectory, so that a new trajectory of their movement becomes helical rotation. Exactly, this phenomenon is related to the diamagnetism in atoms. According to classic representation, the physical nature of diamagnetism lies in the induction of nondamped microscopic currents by a magnetic field owing to the helical rotation of an electron in its closed orbit with variable angular velocity.

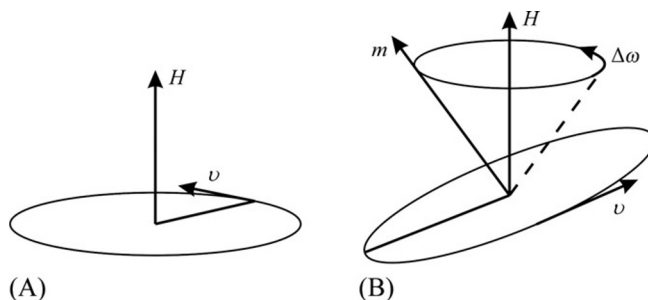
As a rule, diamagnetism represents a very weak response of substances to the applied magnetic field: its contribution to magnetic susceptibility is very small: $\chi = -(10^{-5} - 10^{-6})$. The sign “−” indicates that the induced diamagnetic moment is directed *opposite* to the applied field H . The small magnetic induction B that appears in matter can be compared to the magnetic induction in vacuum: $\mu_D = 0.99999 \dots \approx 1$. As already is shown in Fig. 6.2, any diamagnetic “pushes” out the magnetic field.

Temperature dependences of magnetic susceptibility in various types of weak magnetism are shown in Fig. 6.5.

In any matter, the Larmor diamagnetism is a *common mechanism* that occurs due to the precession of electronic orbitals of atoms, ions, and molecules. In an external magnetic field H , this precession *always occurs* as the manifestation of fundamental

**FIG. 6.5**

Temperature dependence of magnetic susceptibility in case of “weak” magnetism: P1—Curie law for Lanzheven type of paramagnetic; P2—paramagnetism of electronic gas in metals; P3—Van Vlack paramagnetism; D1—Larmor diamagnetism; and D2—diamagnetism in fullerites and nanotubes [3].

**FIG. 6.6**

Larmor precession of electronic orbit in a magnetic field, which leads to diamagnetism: (A) electronic orbit is perpendicular to the magnetic field H , electron moves with velocity v ; (B) electronic orbit is tilted to the field, so that the effect of the magnetic field causes precession of the orbit.

properties of electrons moving in their orbits, as shown in Fig. 6.6. In a magnetic field, the angular velocity ω of an electron decreases by $\Delta\omega$. In an orbital plane, the electron moves in a cone around the magnetic field vector H with constant angular velocity of precession.

Thus, diamagnetism is associated with the orbital movement of electrons and occurs in all atoms and ions (inasmuch as the orbital movement of electrons exists

in any atomic core). Diamagnetism causes a slight deceleration of angular velocity of orbital movement when the atom is placed in a magnetic field. This effect can be explained on the basis of the following general considerations. The movement of an electron in its orbit can be considered as a closed current. In case when a circuit with current is placed in a magnetic field, according to Faraday's law of electromagnetic induction, an additional electromotive force (EMF) arises. As a result, current in a circuit changes, and this modifies the magnetic moment. According to the well-known *Principle of Le Chatelier* in physics, this current should be directed as counteracting to the external field, and this results in counter induction. This means that the induced magnetic moment is directed *against* the applied field that, by definition, is the key feature of diamagnetism. From Le Chatelier principle, it follows that diamagnetism is manifested in materials by repulsion out of the magnetic field.

As the size of the electronic shell of an atom or ion is almost independent of temperature, the diamagnetic susceptibility (that has a negative value), only slightly varies with temperature because of the decrease in material density, as shown in Fig. 6.5, curve D1. In this sense, the diamagnetism induced by the external magnetic field H reminds the *electronic polarization* of dielectrics, which is also explained by a distortion and shift of electronic orbital under an external *electrical* field E . Indeed, in case of electronic polarization, dielectric susceptibility χ_e (as diamagnetic susceptibility) practically is independent of temperature, but the χ_e always has a *positive value* (unlike negative diamagnetic susceptibility α_d). However, it should be noted that electronic dielectric susceptibility in different crystals lies within $\chi_e = 0.8 \dots 4$, that is, 1000 times higher than the diamagnetic susceptibility α_d .

In metals, *in addition* to Larmor diamagnetism, another mechanism of diamagnetism exists (*Landau diamagnetism*) [1]. This diamagnetism is conditioned by conduction of electrons moving under the external magnetic field. By Lorentz force, the magnetic field compels electrons to *move in a spiral*, but not in straight, trajectories. Landau proposed *quantization* of the energy of electrons in metals (when *Landau energy levels* occur). It is necessary to note that this mechanism is also characterized by a very small value of magnetic susceptibility ($\alpha_L \sim -10^{-5}$).

Substances with pronounced diamagnetic properties include the following:

- all matters (atoms and ions) that have no natural magnetic moments;
- organic compounds with nonpolar bonding, in which molecules or radicals have no intrinsic magnetic moment (when the paramagnetic effect in them is less than the diamagnetic effect); hence magnetic susceptibility $\alpha_d \approx -(10^{-5} - 10^{-6})$ and shows significant anisotropy;
- crystalline substances such as certain metals (Zn, Au, Hg, etc.), some metallic alloys, and chemical compounds with prevailing diamagnetism in the ionic cores (ions, similar to atoms of inert gases: Li^+ , Be^{2+} , Cl^- , etc.).

Thus, diamagnetism is *peculiar to all substances*, being a preferred type of magnetism in the materials with completely filled electronic shells (many dielectrics, semiconductors, and certain metals). For example, among materials important for electronics, many *semiconductors* are diamagnetics (in germanium, $\alpha = -8 \cdot 10^{-6}$,

and in silicon, $\alpha = -3 \cdot 10^{-6}$), as well as many *metals* (in copper, $\alpha = -6 \cdot 10^{-6}$; in silver, $\alpha = -22 \cdot 10^{-6}$; and in gold, $\alpha = -30 \cdot 10^{-6}$) and most *dielectrics*.

There are, however, some solids, in which diamagnetism is relatively strong: bismuth, antimony, graphite, and other *semimetals*. For example, in the bismuth, diamagnetic susceptibility is not only increased but also anisotropic: in the main crystallographic directions, in Bi, $\alpha_a = -220 \cdot 10^{-6}$ and $\alpha_c = -310 \cdot 10^{-6}$. It is found that difference $\alpha_a - \alpha_c$ periodically changes in the dependence on magnetic field H (Van Alphen effect).

Increased value of diamagnetic susceptibility ($-\alpha$) is observed also in the *graphite* and other (recently discovered) modifications of carbon and in the *fullerenes* and *carbon nanotubes*. It is noteworthy that in these substances, “ $-\alpha$ ” increases significantly when temperature decreases, as shown in Fig. 6.5, curve D2. The strengthening of diamagnetism in the *semimetals* may indicate a tendency of these materials to have superconducting transition. Indeed, the superconductors absolutely push out the magnetic field (their magnetic susceptibility formally equals $\alpha = -1$). Such behavior of superconductors is caused by the electrical currents flowing in a thin surface layer of superconductor (thickness of this layer is around 10^{-5} cm). This surface current in superconductor shields external magnetic fields; hence in the bulk of superconductor, magnetic field is zero. Except superconductors, there are other cases of “giant” diamagnetism in some materials.

Therefore a relatively weak effect of diamagnetism is inherent *in all matter*, but if a more strong effect—paramagnetism—exists (which is usually characterized by a higher magnetic susceptibility), total magnetic susceptibility turns to the positive value ($\alpha > 0$); hence it is considered that these substances belong to paramagnetics.

Paramagnetism is a property of materials whose structural units (atoms, molecules, ions, and cores) have *natural magnetic moments*. However, without external magnetic field action, these moments are oriented randomly; hence the overall magnetization of a paramagnetic is zero ($J = 0$ if $H = 0$).

When the external magnetic field H is switched on, magnetic moments of atoms in a paramagnetic become *partially oriented* toward the field, and with increase in applied field, magnetization increases, at first—linearly (see Fig. 6.3A). If external magnetic field would be large enough, then most magnetic moments of paramagnetic particles will become already oriented strictly in the direction of field. Therefore, dependence of $J(H)$ becomes nonlinear; as a result, *magnetic saturation* is observed, as shown in Fig. 6.7 [7]. Knowing the value of magnetic moment at saturation and the concentration of paramagnetic particles in a matter, it is possible to determine the magnitude of elementary magnetic moment (e.g., total spin moment is “ $3/2$ ” for Cr^{3+} , “ $5/2$ ” for Fe^{3+} , and “ $7/2$ ” for Gd^{3+}). Paramagnetic susceptibility is *positive* within the values $\alpha = +(10^{-4} - 10^{-1})$. This means that the permeability of paramagnetic is higher than one ($\mu > 1$), unlike diamagnetic in which $\mu < 1$.

Magnetic field, as shown in Fig. 6.7, is large enough to reach almost *complete orientation* of magnetic dipoles (overcoming the impact of disordering by thermal chaotic motion). This is possible because there are *no individual magnetic charges* that would be accelerated in the magnetic field. It should be noted that in similar

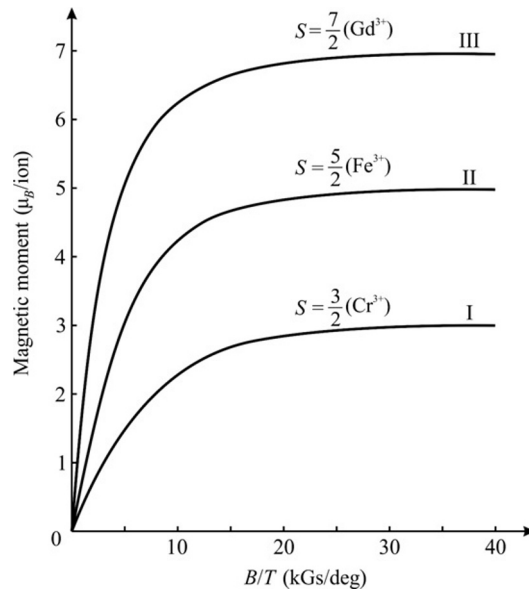


FIG. 6.7

Magnetic moment in paramagnet dependence on magnetic field: I—chromium-potassium alum, II—iron-ammonium alum, and III—gadolinium sulfate [1].

cases of dipole-type dielectrics, usually it is impossible to orient majority of dipoles because under the strong *electrical* field existing electrons are necessarily accelerated and cause the electrical breakdown much ahead of the orientation of most dipoles.

Magnetic moments of atoms or ions that cause paramagnetism are conditioned by the spin moments of electrons (*spin paramagnetism*) or by the movement of electrons in atomic shells (*orbital paramagnetism*). It should be noted that magnetic moments of atomic nuclei also lead to *nuclear paramagnetism*, but usually this effect is negligible (the smaller the magnetic moment of particle, the greater is the particle mass). As a result, total magnetic moments of atoms, ions, and molecules are created mainly by electrons that have a magnetic moment thousands of times greater than that of atomic nuclei.

There are several mechanisms of electronic paramagnetism: temperature dependence of paramagnetic susceptibility, as shown in Fig. 6.5, points to three most important mechanisms. According to the Lenz-Heisenberg-Curie mechanism, when a crystal is cooled, its paramagnetic susceptibility increases according to *Curie law*: $\chi \sim K/T$, where K is the *Curie constant*. In case of Pauli paramagnetism, magnetic susceptibility is practically independent of temperature. This is also seen for the Van Vleck paramagnetism (typical in some molecular compounds): in this case, magnetic susceptibility is small and almost independent of temperature.

Table 6.1 Electronic Construction of *d*-Orbitals in Transient Metals

Element		K (<i>n</i> = 1)	L (<i>n</i> = 2)		M (<i>n</i> = 3)			N (<i>n</i> = 4)	
Symbol	Atomic Number	1s	2s	2p	3s	3p	3d	4s	4p
K	19	2	2	6	2	6		1	
Ca	20	2	2	6	2	6		2	
Sc	21	2	2	6	2	6	1	2	
Ti	22	2	2	6	2	6	2	2	
V	23	2	2	6	2	6	3	2	
Cr	24	2	2	6	2	6	5	1	
Mn	25	2	2	6	2	6	5	2	
Fe	26	2	2	6	2	6	6	2	
Co	27	2	2	6	2	6	7	2	
Ni	28	2	2	6	2	6	8	2	
Cu	29	2	2	6	2	6	10	1	
Zn	30	2	2	6	2	6	10	2	

The Lenzheven-Curie paramagnetism. One reason for the existence of the own magnetic moment in an atom (or ion) might be electronic spins, which are *not compensated* in the noncompletely filled *d*-shells or *f*-shells. For example, in the *transition metals*, listed in Table 6.1, noncompensated atomic magnetic moments are due to some of the *3d*-electrons.

Table 6.1 shows that *3d*-orbital is empty in atoms K and Ca, whereas in atoms Cu and Zn, *3d*-orbital is completely filled (spin magnetic moments of electrons in this case are totally compensated). This means that *atoms* K, Ca, Zn, and Cu are not paramagnetic. In other atoms, listed in Table 6.1, their *3d*-orbital is not completely filled. The exact calculations of *3d*-electron distribution are complicated, but the manner of these electron distribution is expressed by Hund's rules, following which *3d*-electrons are arranged in the *3d*-shell according to their magnetic spins [1].

The conception of *multiplicity* is introduced: it equals to $2S + 1$, where *S* is the total spin angular momentum for all electrons. Applied to electronic shell filling, Hund's rules determine the character of energy level filling by electrons in an atom, under which the ground state must follow such requirements:

- term with maximum multiplicity has the lowest energy (the maximum value of full spin *S* is in accordance with Pauli principle);
- term with the largest value of total orbital angular momentum *L* has the lowest energy (the maximum value of *L* is consistent with the value of *S*);
- full angular momentum *J* (total angular momentum) meets $|L - S|$, if electronic shell is filled less than half, and $|L + S|$, if electronic shell is filled more than half. (When in shell, exactly half of levels are filled; then using the first rule leads to $L = 0$, and hence to equality $J = S$.)

The first Hund's rule is based on Pauli principle and on Coulomb repulsion between electrons. Pauli principle does not allow two electrons to exist in one energy state with the same spins. Thus electrons with the same spin direction should be *separated in space*. However, because of Coulomb interaction, the energy of electrons with the same spin directions is reduced. Thus the average potential energy of parallel spin orientation in might be less than that in the antiparallel spin orientation.

For example, the Mn^{2+} ion may be considered. The $3d$ -shell of this ion has five electrons; hence this shell is filled exactly half. Spins of electrons can be oriented parallel, if electrons occupy *different states*; in the $3d$ -shell, exactly five different states are allowed, which are characterized by the orbital quantum number $m=2, 1, 0, -1, \text{ and } -2$. Each of these states can be occupied by one electron. In this case, it might be expected that the total spin will be equal to: $S=5/2$, and because $\sum m=0$, the only possible value is $L=0$, which is observed experimentally.

Orientation of spins in the first period of transition metal is shown schematically in Fig. 6.8. The limiting number of $3d$ -electrons is 10; hence in the d -shell, up to five electrons may have the same spin orientation (as it is seen in the case of manganese and chromium) before filling these states by electrons with an opposite orientation. Quantum mechanical calculations show that for transition metals, a convergent orientation of electronic spins in the d -shell corresponds to the minimum energy (as the more stable state). In case of chromium, for example, the configuration $3d^5 4s^1$ exists, but not $3d^4 4s^2$. Similarly, copper atom has the electronic configuration $3d^{10} 4s^1$ but not $3d^9 4s^2$ as one might expect.

As magnetic properties of atoms are due, primarily, to spins of electrons, the uncompensated spin orientation, as shown in Fig. 6.8, enables to evaluate the magnetic moment of the atom. For example, single titanium atom has a magnetic moment of two spins (two Bohr magnetons, $2 \mu_B$), whereas single cobalt atom has an own moment of three spins ($3 \mu_B$). Up to five $3d$ -electrons in atoms can be placed with

Z		$1s^2 2s^2 2p^6 3s^2 3p^6$	3d					4s
19	K		□	□	□	□	□	↑
20	Ca		□	□	□	□	□	↑↓
21	Sc		↑	□	□	□	□	↑↓
22	Ti		↑	↑	□	□	□	↑↓
23	V		↑	↑	↑	□	□	↑↓
24	Cr		↑	↑	↑	↑	↑	↑
25	Mn		↑	↑	↑	↑	↑	↑↓
26	Fe		↑↓	↑	↑	↑	↑	↑↓
27	Co		↑↓	↑↓	↑	↑	↑	↑↓
28	Ni		↑↓	↑↓	↑↓	↑	↑	↑↓
29	Cu		↑↓	↑↓	↑↓	↑↓	↑↓	↑
30	Zn		↑↓	↑↓	↑↓	↑↓	↑↓	↑↓

FIG. 6.8

Location spins of electrons in orbitals in transition metals: $3d$ -electrons in atoms can be arranged with the parallel orientation of spins.

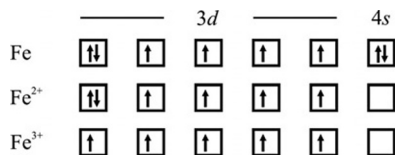


FIG. 6.9

Distribution of 3d spins of electrons in two-valence and three-valence iron ions.

a parallel orientation with spin preservation. The second electron in each state should be oriented antiparallel.

To date, mostly, magnetic moments of *atoms* are considered. The *ions* of 3d-metals, generally, have varying valence, and depending on this, they can have different number of uncompensated spin moments. This fact is significant for magnetic material synthesis for various purposes. A very important example is shown in Fig. 6.9: the distribution of spins in the 3d-shell for two different iron ions: Fe²⁺ and Fe³⁺ compared with the atom of iron (Fe). It is seen that the two-valence iron ion has a total magnetic moment of 4 μ_B , whereas the three-valence iron ion might be characterized by 5 μ_B .

It should be noted that in Fig. 6.9 only simplified models are shown because it does not consider the *spin-orbital interaction*. Considering this interaction (and according to experiments), the magnetic moment of Fe²⁺ is dependent on a given crystal; for the first case, it might have 4.4 μ_B , whereas for the second case, Fe³⁺ can have 6.9 μ_B .

Atoms and ions of the *rare-earth elements* with valence “+3” also might have uncompensated spin moments, but in the 4f-orbital. Location of spins in the 4f-shells for lanthanides was shown previously in Table 5.2. The maximum nonpaired electrons (*seven!*) in the 4f-shell can be seen for gadolinium (Gd), where, instead of 14 possible electrons, only partial filling is observed: 4f⁷.

The *ions* of various rare-earth elements have quite similar chemical properties as their outer electronic shells should be identical: they all have the configuration 5s²5p⁶ (similar to that of the neutral xenon atom). The radius of trivalent *ion*, when transition from one element of this group to the other, gradually reduces from 1.11 Å in cerium to 0.94 Å in ytterbium. This phenomenon is the *lanthanoid compression*. This fact enables to *manage properties* of crystals that contain rare-earth elements by selecting lanthanide ion with the required radius for a given crystal.

Experimentally found values of magnetic moments of rare-earth element ions are shown in Table 6.2. Magnetic properties of rare-earth ions are very appreciable. In lanthanum (La), which is the starting element of rare-earth metal group, the 4f-shell is empty, but in cerium atom, the 4f-shell already has one electron. Further, the number of 4f-electrons consistently increased in each next element of up to ytterbium (Yt), which has 13 electrons in its 4f-shell, and lutetium (Lu) in which 14 electrons completely fill the 4f-shell. It is obvious that the ions La³⁺ and Lu³⁺ are diamagnetics,

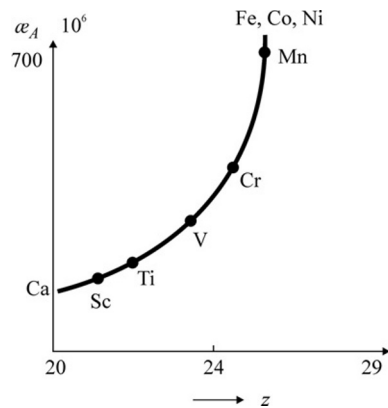
Table 6.2 Experimental Data as to the Number of Bohr Magnetons in Lanthanides

RE Ion	La ³⁺	Ce ³⁺	Pr ³⁺	Nd ³⁺	Pm ³⁺	Sm ³⁺	Eu ³⁺	Gd ³⁺	Tb ³⁺	Dy ³⁺	Ho ³⁺	Er ³⁺	Tm ³⁺	Yb ³⁺	Lu ³⁺
Moment, μ_B	0	2.4	3.5	3.5	–	1.5	3.4	8.0	9.5	10.6	10.4	9.5	7.4	4.5	0

whereas all other *ions* of rare-earth elements (from Ce³⁺ to Yb³⁺) belong to paramagnetic [8].

The difference between the magnetic properties of rare-earth (*4f*) metals and those of transition (*3d*) metals is that the radius of *4f*-shell equals only $\sim 0.3 \text{ \AA}$, and this shell is *hidden* under the outward electronic shells. Therefore, the metals widely used in engineering ferrimagnetic materials (ferrites) that are synthesized with rare-earth elements have the highest electromagnetic *quality factor Q* (i.e., small loss of electromagnetic energy at microwaves). In rare-earth ferrites, there is rather weak connection of deep-seated *active magnetic 4f subsystem* with lattice thermal movement (phonons) that mainly touches external electronic shells. The external electromagnetic field excites exactly the *4f* magnetic subsystem, which being partially screened from phonon losses less energy: under microwaves, rare-earth ferrites have much higher quality factor than ferrites based on transition metals in which the *3d*-shell is not shielded from the thermal movement of ions.

Magnetic susceptibility of paramagnetics can be quite different. If paramagnetism of “electronic gas” in metals (Pauli mechanism) prevails over the diamagnetism of electrons (Landau mechanism), the magnetic susceptibility of metals is $\alpha \sim +10^{-5}$ (in sodium, $\alpha = +16 \cdot 10^{-6}$; in barium, $\alpha = +20 \cdot 10^{-6}$; and so on). But in metals with unfilled *3d*- or *4f*-shells, paramagnetic susceptibility shows increased values: in the range of $\alpha \sim +(10^{-4} - 10^{-3})$. For example, in titanium, $\alpha = +160 \cdot 10^{-6}$; in uranium, $\alpha = +400 \cdot 10^{-6}$; and so on, as shown in Fig. 6.10.

**FIG. 6.10**

Paramagnetic susceptibility in transition metals, *z* is the element number.

Paramagnetism conditioned by d -electrons in transition metals and by f -electrons in lanthanides agrees with the Langevin-Curie mechanism. It should be noted that transition metals, such as those shown in Fig. 6.10, demonstrate a significant increase in paramagnetic susceptibility, if their serial number approaches to the “iron triad” (Fe-Co-Ni). In some chemical compounds, based on d - and f -metals, paramagnetic susceptibility is comparatively high and reaches the value $\alpha \sim 10^{-1}$, for example, the crystal MnCl_2 has susceptibility $\alpha = 14,350 \cdot 10^{-6}$, whereas the crystal CoCl_2 shows $\alpha = 122,000 \cdot 10^{-6}$.

Temperature dependence of magnetic susceptibility in paramagnetic (see Fig. 6.5, curve P1) is well described by the classic Langevin theory: magnetic susceptibility is determined by the formula $\alpha = Nm_a^2/3k_B T$, where N is the concentration of paramagnetic atoms in a substance; T is the temperature, k_B is the Boltzmann constant, and m_a is the magnetic moment of an atom. This formula is obtained by methods of statistical physics for a system of interacting dipoles that are placed in a magnetic field at relatively high temperatures (when $m_a H < k_B T$).

Thus in case of constant field and with temperature increase, the thermal motion grows up and results in magnetic moment disorientation; hence the susceptibility of paramagnetic reduces according to the Curie law: $\alpha \sim K/T$. It is interesting to note that a similar temperature dependence is seen also for *electrical* susceptibility ($\chi \sim K/T$) in the system of *noninteracting* electrical dipoles. It should also be noted that temperature dependence of conductivity in metals also follows the dependence $\sigma \sim K/T$ due to electron scattering by lattice thermal vibrations that, consequently, reduces the mobility of electrons. Thus, the similar temperature dependence of electrical and magnetic parameters (decrease in α , χ , and σ by law K/T) is explained by a growth in the intensity of lattice thermal vibrations (phonons) with increasing temperature. Phonons disorder magnetic and dielectric dipoles and reduce the mobility of electrons [9].

However, Curie law holds true only for relatively weak magnetic fields. In strong magnetic fields, as well as at low temperatures (when $m_a H > k_B T$), magnetization of paramagnetic *nonlinearly* approaches $m_a H$ (this is the “saturation,” when almost all magnetic moments become oriented, as shown in Fig. 6.7). Possible deviations from the Curie law, in particular, deviations from the Curie-Weiss law for α (and χ), which is also seen above phase transition in ferromagnetic (and in ferroelectric: it is caused by the *interaction* between magnetic (or electrical) dipoles).

Paramagnetism may be observed in some chemical compounds whose ions have no magnetic moment in the ground state. This kind of paramagnetism is associated with quantum-mechanical characteristics due to the admixture of excited states with magnetic moment (*Van Vleck paramagnetism*, as shown in Fig. 6.5). In this case, magnetic susceptibility does not depend on temperature, as well as in the case of paramagnetism in electronic gas.

Paramagnetism is widely used in experimental methods such as *electronic paramagnetic resonance* (EPR). This method in solid-state physics enables to determine magnetic moments of individual atoms, ions, and molecules, to promote investigation of complicated structures of molecules, and to perform fine structural analysis of materials used in engineering. Paramagnetic substances are also applied in *cryotechnology* to achieve extremely low temperatures (paramagnetic cooling).

The Pauli paramagnetism in metals. As an electron has its own magnetic moment that is approximately equal to that of Bohr magneton, it would be expected that *conduction electrons* in metals would make great contribution to the paramagnetic properties of metals, described by the Curie law: $\alpha = N\mu_B^2/3k_B T$, where N is the concentration of electrons (in metals, it equals $N \sim 10^{22} \text{ cm}^{-3}$). However, experimental studies show that magnetic susceptibility of normal (nonferromagnetic) metals does not depend on temperature, and its value can be estimated as only around 10^{-2} compared to a value defined by the Lenz-Heve mechanism. As a result, the paramagnetism of conduction electrons is quite small that in many metals the Landau diamagnetism present within them dominates.

As it was shown by Pauli using the quantum theory, the weakness of paramagnetism of free electrons in metals can be explained by Fermi-Dirac statistics. Wave functions of conduction electrons are quite different from electrons located in the atomic shells (where any level of valence electrons has two spin states). For most electrons in a metal, the probability of an event that under the influence of an external field their spins can change their direction is zero because *most energy states are already occupied* by the electrons with opposite spins. Indeed, in the valence band of metal, all “deep” levels (which are located significantly below the Fermi level) are completely filled with electrons with opposite spins; hence these electrons cannot orient their spin moments according to the applied magnetic field. Only for a *small fraction* of electrons, located *near the Fermi level* (whose energies are in the range of $k_B T$), their spins are able to follow the direction of the applied magnetic field. However, Fermi energy is much greater than the heat energy: $E_F \gg k_B T$. Thus only a little part of the total quantity of conduction electrons (in proportion to $k_B T$) contributes to paramagnetic susceptibility.

For this reason, it would be expected that paramagnetic susceptibility α_{P2} (Fig. 6.5) must increase along with temperature increase in proportion to T . However, the opposite effect also works: owing to temperature fluctuations of the crystal lattice (that intensity is also proportional to $k_B T$), the contribution of “free” electrons to paramagnetism α_{P2} should decrease with increasing temperature as $1/T$. As a result, Pauli paramagnetism shows temperature constancy of α_{P2} , as shown in Fig. 6.6. Free electrons in a metal usually behave as in diamagnetic, hence it is possible a paramagnetic contributes to magnetic susceptibility; typically, the paramagnetism of free electrons is larger than their diamagnetism; hence the total contribution of free electrons to magnetic susceptibility usually has a paramagnetic nature.

Thus paramagnetism of conduction electrons in most metals makes contribution to paramagnetic susceptibility, which is not subjected, with the Curie law being practically independent of temperature.

The spin of electron. If a primary physical cause of diamagnetism is the *orbital* motion of electrons in atoms and ions, the paramagnetism is conditioned by the *spin moments* of particles. The value of spin is denoted by the letter s , particle with spin $1/2$, and electrical charge e has a magnetic moment $\mu_B = e\hbar/2mc$. This value (called as Bohr magneton) equals $\sim 10^{-20} \text{ erg/Gs}$. It should be noted that electronic magnetic moment is an unusual vector because it can be oriented in space only by two ways: either on the field or against it.

The ratio of magnetic moment of particle to its mechanical moment is the constant “ γ ” that is *magnetomechanical ratio* (or gyromagnetic ratio). Its unit in SI is “radian per second per tesla”: [rad/(s·T)]. However, very often, another term *gyromagnetic ratio* is used for different but closely related quantity, namely, the g -factor, which unlike γ is dimensionless [9].

Magnetic moment of an atom is expressed by the formula $\mu_{at} = \gamma \hbar J = g \mu_B J$, where $\hbar J$ is the total angular momentum, which is the sum of the orbital moment $\hbar L$ and spin moment $\hbar S$. Bohr magneton is determined as $\mu_B = e \hbar / 2m$, which is very close to the spin of free electron. The value of g -factor for an electron is defined as $g = -\gamma \hbar / \mu_B$, and it is also called the *spectroscopic splitting factor*. For electrons, $g = 2.0023$, but usually the value $g = 2$ is used.

The nuclear magnetism. For better understanding, the nature of particle *interactions*, which examines not only electrical but also magnetic properties of materials, is necessary. According to experiments, the “classic” size of nucleus is around 10^{-13} cm, which is negligibly small compared with the size of an atom (10^{-8} cm). As the mass of cores in four orders of magnitude is greater than the mass of electron, it might be considered (while electronic processes are studied) that atomic core is “infinitely heavy.” This approach is so-called *adiabatic hypothesis*, when the condensed matter theory is applied to justify electronic spectrum. Electrical fields in atomic nuclei are very large, and it is determined by the number of protons in the nucleus. However, nuclear magnetism is 1000 times weaker than electronic magnetism; hence in technical applications, as a rule, magnetism of cores can be ignored.

It is necessary to mention that, in general, magnetic interactions are much weaker than electrical interactions. Actually, the energy of magnetic interaction in an *atom* is appreciated as $U_{\text{mag}} \approx \mu_B^2 / a^3$, where μ_B is the Bohr magneton and a is the average distance between electrons. Energy of electrostatic interaction between two electrons under the same conditions equals $U_{\text{elec}} = e^2 / a$. The ratio of these two energies is

$$U_{\text{mag}} / U_{\text{elec}} \approx \alpha_Z^2$$

where $\alpha_Z = e^2 / \hbar c \approx 1/137$ is the *Sommerfeld constant* (fine structure constant), characterizing the strength of electromagnetic interaction between charged elementary particles. Thus the *magnetic interaction of electrons is much weaker than their electrostatic interaction*. In physics of magnetism, it is important because small “fine structure constant” results in a small value of diamagnetic susceptibility. It can be shown that this susceptibility is estimated as $\alpha_{\text{dia}} \approx \alpha_Z^2 \approx (1/137)^2 \approx 5 \cdot 10^{-5}$, which is well consistent with experimental data [8].

6.3 FERROMAGNETISM

Magnetic crystals and polycrystals with high ordering of spin and orbital magnetic moments demonstrate the so-called strong magnetism. In this case, permeability is large and corresponding materials can be the sources of strong magnetic fields that are widely used in engineering.

When considering paramagnetics, it was shown that some atoms, whose electronic shells are not completely filled, have their own magnetic moments and behave as small permanent magnets. The degree of magnetization of such crystal is determined by the total magnetic moment, which is the vector sum of the magnetic moments of atoms.

Natural magnetic moment involves atoms and ions of *transient groups* of Mendeleev's periodic table because they are characterized by unfilled inner electronic shells that are available to hold the spin of unpaired electrons. An example (Table 6.1) is the iron atom, in which 26 electrons move around the nucleus; 18 of them fill the inner orbitals (as well as in the argon atom). However, in the 3*d*-shell of the iron atom, only 6 of the possible 10 electronic states are occupied; hence the 3*d*-shell in iron is not filled completely, as there are four empty states (see Fig. 6.9). Moreover, four magnetic moments of electrons in the 3*d*-shell of Fe atom are *self-ordered*, thus making a system with uncompensated magnetic spins. Such feature of the 3*d*-shell, which determines big *intrinsic magnetic moment* of an atom, is peculiar to several elements of the iron group.

If a crystal is formed from atoms that have natural magnetic moments (such as iron), different ways of magnetic moment orientation may be realized. The simplest types of regulation in two-dimensional case are shown in Fig. 6.4. The tip of the arrow shows the north pole of a magnet linked to an atom. If magnetic moments are oriented randomly, as shown in Fig. 6.4A, then the total magnetic moment of the crystal is zero (this corresponds to paramagnetic). When one applies a magnetic field to such a crystal, the *forced ordering* of magnetic moments occurs with their overwhelming focusing according to a field that creates deposit in total magnetic moment (paramagnetism). In Fig. 6.4B, the ordered structures are shown very simplistic—only as a comparison with disordered structures.

Different ordered structures are shown in a more detailed way in Fig. 6.11. In the simplest *ferromagnetic* structure (Fig. 6.11A), all magnetic moments of atoms are directed equally. Examples of such ferromagnetics are metals: Fe, Ni, Co, Gd, and Dy. These strictly magnetically ordered metal crystals can behave like permanent magnets (if they have a single-domain structure).

Simplest-ordered *antiferromagnetic* structures might also be collinear, but magnetic moments in them are directed oppositely; hence they are totally self-compensated, as shown in Fig. 6.11B. Axis, at which all these moments are directed, is the *antiferromagnetic* axis. Typical representatives of crystals with antiferromagnetic structure are some oxides of transition metals (Mn, Ni, Co, and Fe) and many of their fluorides, chlorides, sulfides, selenides, and others [8].

Crystallographically, all magnets that have a structure with a similar direction of their magnetic moments might be presented as *magnetic sublattices*. In an illustrated case, shown in Fig. 6.11B, some of the magnetic moments of atoms are directed “up,” thus forming one sublattice, whereas atoms with a direction opposite of their magnetic moments form another sublattice. These two sublattices consist of atoms that are located in the equivalent positions (two *equivalent* magnetic sublattices).

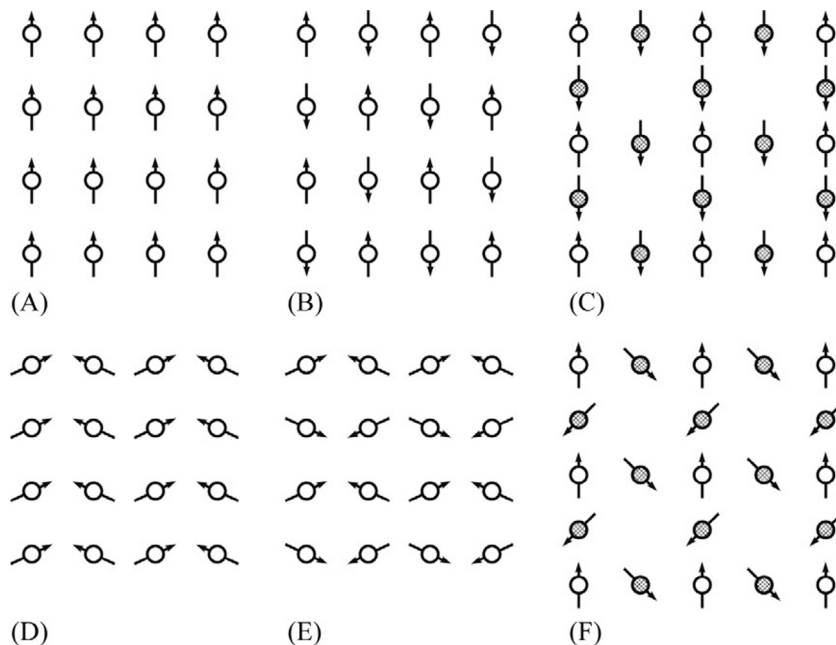


FIG. 6.11

Different types of ordered magnetic structures: (A) ferromagnetic structure; (B) antiferromagnetic structure; (C) ferrimagnetic structure; (D) weak ferromagnetic structure; (E) weak antiferromagnetic structure; (F) strongly noncollinear ferrimagnetic structure.

In general, a magnetic structure may contain several sublattices, formed by atoms that are crystallographically located in *nonequivalent positions*. Magnetic sublattice is a set of all atomic magnetic moments that can be obtained using parallel translations at distances that are divisible to the period of unit cell. In the collinear *ferrimagnetic* structure (Fig. 6.11C) the neighboring atoms also show antiparallel orientation, but the *total* magnetic moment of elementary cell of a crystal is *different from zero*. Therefore this structure has spontaneous magnetization, as magnetic moments of ions that belong to different sublattices are noncompensated.

Partial compensation of magnetic energy may be conditioned by several ways. First, elementary magnetic cell may account different numbers of ions belonging to two sublattices (magnetic moments of ions in this case might be the same). Second, magnetic moments of ions of two different sublattices may have different size. Most often, both causes are observed, as shown in Fig. 6.11C. Ferrimagnetism usually is called as *noncompensated antiferromagnetism*, which better reflects the nature of this phenomenon.

The types of magnetic structures that belong to *collinear* magnetic structures are shown in Fig. 6.11A–C. There are also different types of *noncollinear* magnetic

structures, and some of them are shown in Fig. 6.11D–F. *Weakly* noncollinear magnetic structure (Fig. 6.11D) is inherent to the weak ferrimagnetics, and they are characterized by small resultant magnetic moment (in Fig. 6.11D it is directed upward). This causes a slight slanting of antiferromagnetic ordering of magnetic moment sublattices. Weakly noncollinear magnetic structure is seen in Fe_2O_3 (hematite); in the crystals FeBO_3 and FeF_3 ; in the carbonates MnCO_3 , CoCO_3 , and NiCO_3 ; in the orthoferrites RFeO_3 , as well as in the orthochromites RCrO_3 (R is a rare-earth element).

There are also such weakly noncollinear antiferromagnetic structures (Fig. 6.11E) that have no resultant moment. Triangular (corner) structures shown in Fig. 6.11F belong to *strongly* noncollinear magnetic structures. In this case, magnetic lattice is formed by blackened atoms, divided into two sublattices, whose magnetic moments are directed at an angle to each other; as a result, magnetic moment can be created, and it is antiparallel to the moment of the third sublattice. All these are very special cases of ferromagnetic structures. There are also more complicated cases of “screw” and “helical” magnetic ordering, which is not considered here.

The physical nature of ferromagnetism. It is necessary to consider why in some materials (ferromagnetics) natural magnetic moments of individual atoms become spontaneously ordered, whereas in other materials (paramagnetics) no ordering is observed.

When a permanent magnet is placed in a constant magnetic field, then its magnetic moment tends to take a position, coincident with the direction of the applied field. In the majority of crystals, which contain *d*- and *f*-atoms, each structural unit has its own magnetic moment that creates around itself a magnetic field. If this field would be large enough, it can force magnetic moments of the nearest neighboring ions to be oriented in parallel. This happens in case when the energy of interaction of magnetic moments of neighboring ions is larger than the energy of thermal fluctuations ($k_B T$) in crystal lattice. It is determined that two types of interaction between magnetic moments of neighboring ions might exist: *dipole* interaction and *exchange* interaction. Exchange interaction is a purely quantum effect, and usually, it is stronger than dipole type of interaction.

The main ferromagnetics are listed in Table 6.3. In most of them, carriers of ferromagnetism are uncompensated ion spins *associated with orbital moments of electrons* belonging to crystal lattice. As known, electronic magnetism is manifested as spin with orbital moments. Magnetization of a ferromagnetic summarizes magnetic moment M consisting of ordered magnetic moments of electrons and appropriate mechanical moment P . Ratio M/P equals $-q\mu/2m$, if magnetization is caused by the orbital magnetic moments of atoms, but equals $-q\mu/m$, if magnetization is caused only by spin magnetic moments [4].

There are some important experiments related to these assumptions:

1. Magnetomechanical effect (*mechanical moment* arising at magnetization) was studied by Einstein and de Haas. The iron rod was hung on elastic string inside a solenoid; when magnetized, the rod turns and twists the string. If the direction of

Table 6.3 Curie Temperature and Magnetic Saturation Induction of Ferromagnetics

Matter	T_C (K)	$4\pi B_S$ (Gs)
Fe	1043	21,580
Co	1604	17,900
Ni	631	6084
Gd	293	–
Dy	87	–
CrTe	339	3100
FeCo	1243	24,000
MnBi	633	7800
NiMn	733	9000
EuO	97	–
EuH _{1.86}	24	–
MnAs	318	8400
MnB	533	1850
GdFe ₂	803	5000

the magnetic field changes, the direction of rod rotation also changes. From this experiment, the value of gyromagnetic ratio is determined as $M/P = -q\mu/m$, which implies that this effect is caused by the *spins* of electrons.

2. In the reciprocal experiment, the magnetization of iron rod occurs in case of its rapid rotation. This means electron aspiration (representing the so-called whipping tops with angular momentum) to be oriented in the direction of the axis of rod rotation. Along this experiment, mechanical and magnetic moments of electrons were oriented. This also confirms the *spin* model of magnetization.
3. In another experiment, a previously magnetized rod was subjected to rapid heating above the Curie point. As a result, previously oriented “whipping tops” acquire a random direction; hence the demagnetization stimulated rotational momentum of the rod that can be directly measured in the experiment. In this case, also the gyromagnetic ratio indicates that ferromagnetism is due to the spin momentum of electrons.

However, convincing calculations show that *only spin interaction cannot provide their parallel orientation*, which is the main characteristic of ferromagnetic at temperatures below the Curie point. Theory is obliged to assume (F.R. Weiss) that stable orientation of spins can be caused by the *molecular field*, which is nonmagnetic by its nature. It was first shown by Y. Frenkel that the forces that compel orientation of magnetic moments have an electrostatic nature. The spontaneous orientation arises as a result of the *exchange interaction* of spins and orbital moments of electrons in a crystal lattice.

Exchange interaction is repeatedly considered in quantum mechanics, for example, to describe constitution of a hydrogen molecule. In case of small particles that have magnetic moment (such as electrons), their arrangement in the magnetic field is determined by a fact that projection of spin vector on the magnetic field direction can take only two values: $\pm(1/2)\mu_B$. For two-electron system in the H_2 (as example), it cannot be specified which of the two electrons has a definite state. However, following Pauli principle, two electrons cannot be located on a single energy level with the same spin quantum number. In quantum mechanics, this is considered by introducing the *antisymmetric wave function*, that is, two electrons that interchange their wave function must change their sign.

Exchange interaction has an electrostatic nature; however, it is not a simple Coulomb-type, but a quantum interaction. During mechanism of exchange interaction, the direction of electronic spins of neighboring atoms is coordinated. Such interaction is titled as “exchange” because in the process of interaction between neighboring electrons, magnetic atoms appear as if their *places are changing*. The result of exchange interaction is the establishment of electronic spin moment orientation in parallel to each other; hence *spontaneous magnetization* arises without any external field.

As both spin and orbital moments of electrons are interrelated, it can be argued that spontaneous magnetization is created by the ordering of magnetic moments of atoms. While heating to the Curie temperature lattice, thermal motion destroys orderly setting of atoms, established by exchange interaction. It follows that the greater the exchange interaction in ferromagnetic, the higher should be its Curie temperature at which magnetic ordering becomes destroyed.

In the *exchange integral*, both positive and negative members are included; hence it might have both positive and negative signs. This sign identifies what kind of spin orientation of electrons is involved in the bonding exchange and is energetically more favorable: parallel (corresponding to ferromagnetism) or antiparallel (corresponding to antiferromagnetism). Thus, exchange interaction characterizes the difference in Coulomb energy between parallel and antiparallel orientation of spins. For ferromagnetics and antiferromagnetics, exchange integral has an opposite sign.

As exchange interaction occurs only in case of *overlapping orbitals*, it follows that this interaction has a *short-range nature*: between adjacent orbitals. Conversely, spin-type interaction (between own magnetic dipoles in the lattice of magnetic ions) is called as *long-range* dipole-dipole interaction. Thus the main magnetic interactions are exchange interactions (short range) and dipole-dipole interactions (long range).

Results of exchange integral calculation in dependence on the ratio of lattice constant a and radius r of the $3d$ -shell for different metals of the iron group are shown in Fig. 6.12. It can be seen that only for ferromagnetic metals—iron, cobalt, and nickel—exchange integral is positive, that is, parallel location of spins for neighboring atoms appears energetically favorable [9].

The value of exchange integral correlates with the Curie temperature: that is, the greater the exchange energy, the higher the ordered structure of spins can resist to action of thermal phonons. Indeed, the greatest value of exchange integral is

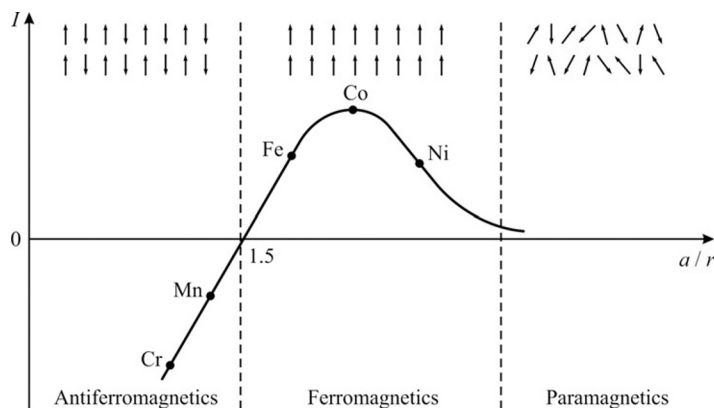


FIG. 6.12

Calculated data for exchange integral for different metals of the iron group in ferromagnetics.

observed in the cobalt with a Curie point of around 1400 K. Exchange integral is smaller in iron ($T_C = 1040$ K), whereas the lowest in nickel ($T_C = 509$ K). The density of electronic states (partially filled orbitals) in the ferromagnetic must be big enough so that kinetic energy cannot exceed exchange energy.

Interestingly, *manganese* (and other representatives of $3d$ -metals, in which the ratio of $a/r < 1.5$) is not ferromagnetic, but the value of exchange integral in Mn is very close to Fe, as shown in Fig. 6.12. Therefore, if lattice constant of manganese would be slightly increased so that the ratio of a/d will be > 1.5 , it would be expected that manganese will become ferromagnetic. Experiment confirms this expectation: ferromagnetism in Mn occurs after its doping by a small amount of nitrogen, which causes the increase in manganese lattice parameter. Similarly, many manganese-based alloys are also ferromagnetics despite not having components that are ferromagnetics in pure crystal. For example, the alloy Mn-Si-Al is very important for application, as well as the compounds MnSb, MnBi, and some others that contain manganese atoms at distances larger than those of pure manganese atoms.

Apparently, for ferromagnetism emergence, it is important to have certain “optimum” in the atomic distance in the crystal lattice. When atoms approach very close to each other (Ti and Cr), then significant dispersal appears in electronic energy band with a rapid increase in kinetic energy, and ferromagnetism is absent. The point is that atoms are located very far from each other, and exchange interaction becomes insufficient for ferromagnetism. In the iron group of metals, only spin interaction (i.e., dipole-dipole attraction) is not large enough for ferromagnetism formation.

Thus the presence of unfinished internal electronic shells in some atoms, as well as the positive sign of exchange integral (which results in parallel orientation of spins), is the necessary and the sufficient conditions when ferromagnetism exists.

Temperature characteristics of ferromagnetics. It needs to be recalled that magnetization J (density of magnetic moment M in a sample) is defined as the total

magnetic moment per unit volume, induced by an external field H (in which measurement is performed). Magnetic “response” of crystal to an applied field H is characterized by magnetic susceptibility α because $J = \mu_0 \alpha H$. However, in the ferromagnetic materials, $\alpha \gg 1$, and therefore magnetic susceptibility practically equals to permeability that follows from ratio $B = \mu_0 \mu H$, so that in ferromagnetics, $\alpha \approx \mu$ and $B \approx J$.

Permeability temperature dependence. While cooling from high temperatures (i.e., cooling from the disordered paramagnetic phase), permeability (and magnetic susceptibility) of ferromagnetic increases and reaches the maximum at the Curie temperature T_C , as shown in Fig. 6.13. In the paramagnetic phase, above phase transition point, the Curie–Weiss law is fair: $\alpha \approx \mu = C/(T - \theta)$, where C is the Curie–Weiss constant and θ is the Curie–Weiss temperature [3].

Once a crystal is ferromagnetic, then spontaneous *internal magnetic field* H_{sp} appears; hence it is measured in a *small* external magnetic field $\alpha \approx \mu$ below its sharp maximum and rapidly decreases with temperature lowering due to *saturation* process occurring in H_{sp} . (However, in *strong* measuring magnetic field $\mu \approx \alpha$ continues its smooth increase, hence the sharp maximum $\mu(T_C)$ is seen only in a small magnetic field.) As it follows from temperature dependence of the *inverse* magnetic susceptibility, near phase transition, the increase in $\alpha(T)$ becomes a little slower, and therefore $\theta \neq T_C$.

Temperature dependence of spontaneous magnetization. Magnetization that arises below the Curie point is *spontaneous*: $J_{sp} \approx B_{sp}$. Temperature dependence of spontaneous magnetization in iron, nickel, and cobalt is shown in Fig. 6.14. On

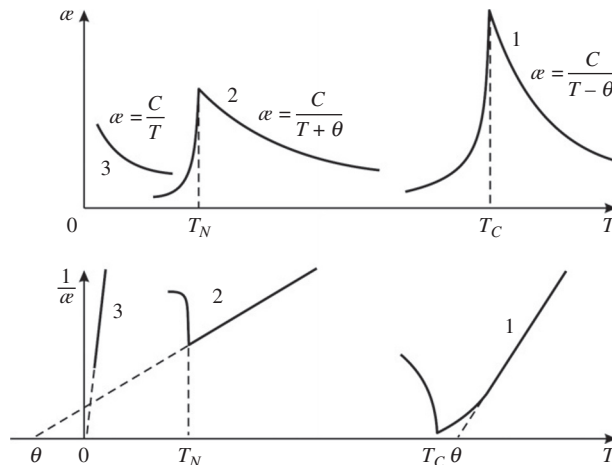


FIG. 6.13

Temperature dependence of magnetic susceptibility and inverse susceptibility for various magnetic materials (θ is the Curie–Weiss temperature, T_C is the Curie temperature, T_N is the Neel temperature): 1—ferromagnetic, 2—antiferromagnetic; 3—paramagnetic.

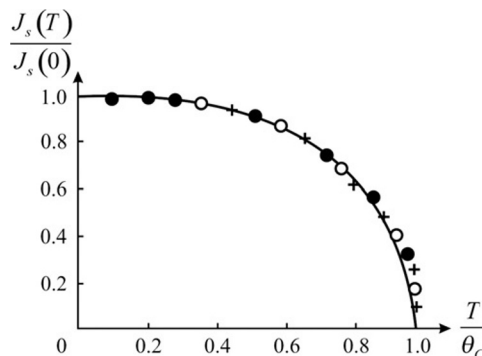


FIG. 6.14

Temperature dependence of spontaneous magnetization for different magnetic materials: ●—iron, ○—nickel, and +—cobalt.

the vertical axis, the relative value of magnetization is designated. Dependence of J_{sp} on T/θ is depicted by the *same curve* for these three ferromagnetics. As temperature increases, magnetization decreases, and the Curie point (as well as above it) becomes zero [1].

The temperature at which phase transition occurs from ferromagnetic ordered state into paramagnetic disordered state is *ferromagnetic Curie point*, T_C . Above this temperature, a substance ceases to be ferromagnetic and behaves similar to many other paramagnetic solids. Afterwards, when cooling to temperature below critical, spontaneous magnetization occurs again, and dependence $J_{sp}(T)$ is restored. In other words, spontaneous magnetization of the material decreases with increasing temperature and vanishes at the critical point.

The value of saturation in $J_{sp}(T)$ curve depends on the fundamental properties of ferromagnetic. As this value corresponds to magnetization inside domain, it does not depend on the method of preparation of the ferromagnetic sample. This feature of spontaneous magnetization temperature dependence is explained by F.R. Weiss: in ferromagnetic, an *internal* (molecular) field exists, which orients all elementary magnets along one direction.

This field is directly proportional to existing magnetization. Thermal fluctuations seek to destroy orientation of elementary magnets, and the more intense, the higher is the temperature. The violation in ordering means less spontaneous magnetization, but, for its turn, it decreases the field that organizes magnetic dipoles. Thus there is a kind of “positive feedback”: the aspiration for magnetization to zero as temperature increases is progressively increasing with decreasing magnetization. On the contrary, when temperature decreases, magnetization gradually increases.

To explain internal (or molecular) Weiss’s field existence, it is insufficient to consider only magnetic forces, acting between elementary dipoles. Calculations show that magnetic forces between the spins cannot play a vital role for internal forces; according to Weiss’s theory, these forces have around three orders of

magnitude smaller than it is necessary to overcome the action of heat disordering. That is why as it is considered before, the electrical forces also act between electrons that stipulate for their exchange interaction.

Heat capacity of ferromagnetic. Temperature behavior of spontaneous magnetization resembles a melting process. When solid melts, the crystalline ordering of atoms suddenly disappears, and the solid turns into a liquid (disordered) state. The intensity of thermal vibrations of atoms becomes large enough to overcome the forces that seek to maintain atoms in the ordered state. The process of overcoming the strength of bonds between atoms at melting temperature results in the *large heat capacity anomalies* in melting point [5].

The analogy between heat capacity temperature dependence in the point of spontaneous magnetization disappearance and critical temperature of crystal melting is confirmed in ferromagnetic. Heat capacity of ferromagnetic shows a similar behavior, namely, the sharp maximum at critical temperature. Still heat capacity maximum in T_C is not “infinite” because temperature does not remain constant when heat is admitted.

Dependence of specific heat on temperature in a typical ferromagnetic is shown in Fig. 6.15 in comparison with the heat capacity of a nonferromagnetic metal. In a nonmagnetic metal, as shown in Fig. 6.15A, lattice heat capacity (curve 1) dominates, whereas electronic contribution to heat capacity is small and increases linearly with temperature (curve 2). In the ferromagnetic, a sharp maximum of heat capacity is observed at the Curie temperature caused by excess energy, necessary for disordering magnetic moments, as shown in Fig. 6.15B.

Moreover, in heat capacity behavior of ferromagnetic, another significant property is seen—pronounced deviation of $C(T)$ dependence, which is quite different from the smooth curve with saturation at high temperatures, observed in nonmagnetic metals. It means that inherent to ferromagnetic *destruction* of spin, ordering

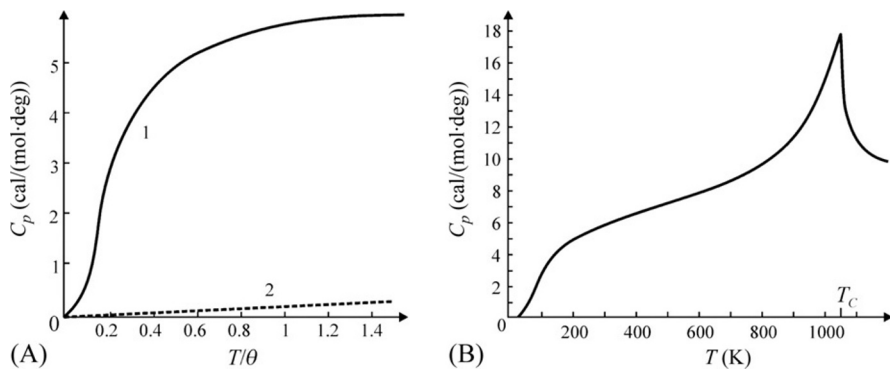


FIG. 6.15

Temperature dependence of heat capacity for lattice (1) and electronic (2) contributions in metal: (A) nonmagnetic metal; (B) ferromagnetic metal (iron).

adds some energy throughout the temperature range. Therefore heat capacity of ferromagnetic, as shown in Fig. 6.15B, is significantly increased in comparison with usual metals. A particular noticeable effect of increased heat capacity is seen in the range of fast decrease of spontaneous magnetization with temperature.

Thus peak-type anomaly of specific heat in the critical point is caused by disordering of elementary magnet orientation, which occurs at a narrow temperature interval. The energy of magnetization approximately equals thermal energy required to destroy spontaneous magnetization. That is why in the case of Fe, Co, and Ni, in which Curie temperature equals $\sim 1000\text{K}$, magnetization energy is of around 0.1 eV per atom, or 2000 kcal/mol . When energy is applied to the lattice of ferromagnetic while temperature increases from absolute zero, the heat capacity of a ferromagnetic crystal looks bigger than the heat capacity of a nonferromagnetic crystal. The point is that contribution to specific heat of magnetic materials is made not only by phonons but also by magnons (see Section 4.3). Therefore heat capacity of ferromagnetic metals in a wide temperature range substantially surpasses the heat capacity of conventional metals. Effect of increased heat capacity is especially noticeable at temperatures just below the Curie point because at this temperature the magnetization decreases faster.

Domain structure of ferromagnetic. Experiments show that magnetic moment of bulk ferromagnetic materials at temperatures below the Curie point is much lower than its theoretical prediction for a case, when all magnetic moments would be directed equally. This is due to the formation of *domains* in the structure of a ferromagnetic.

Domain is a region in a magnet, in which all magnetic moments of atoms are directed equally; hence in each domain, its magnetization reaches saturation, that is, takes a maximal possible value at *given temperature*. However, in different domains of crystal (or polycrystal), vectors of magnetization are not parallel to each other. Thus total magnetization of a ferromagnetic sample appears much lower than in the case of complete ordering of atomic magnetic moment orientation [4].

Simplified examples of domain structure are shown in Fig. 6.16. These structures are formed because they *reduce external magnetic energy* of a pattern in the process of domain formation. Suppose that ferromagnetic crystal totally consists of one domain, then, under the influence of exchange forces, electronic spins of all atoms are lined up, parallel to each other. Consequently, the crystal *creates a magnetic field in the surrounding*. However, this situation is not sustainable because it corresponds to the maximum energy of magnetic interaction.

More stable is such configuration of domains, at which the magnetic field of neighboring areas is partially compensated, that is, their magnetization is directed opposite to each other (Fig. 6.16A in the top shows two neighboring domains). During further crystal division on domains, the energy of the magnetized crystal becomes more reduced, but to a certain limit. The fact is that there are walls between domains, which lead to some *stresses* in a crystal. Such transitional layer between domains is the “Bloch wall,” as shown in Fig. 6.17. This transitional layer wall separates two domains that are magnetized in different directions.

The conception of Bloch walls is due to the fact that change in spin direction when transition from one domain to another (having different directions of

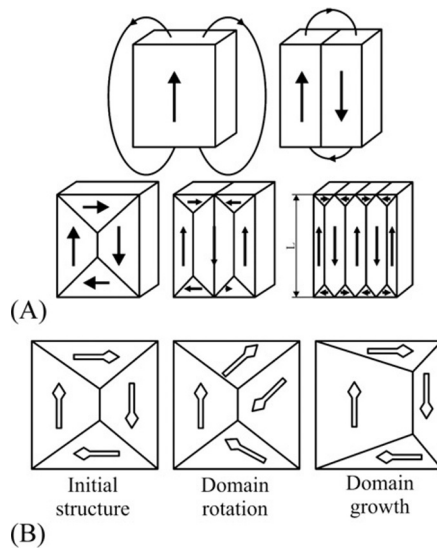


FIG. 6.16

Examples of ferromagnetic domain structure (A) and domain structure change; (B) domain walls shift and domain growth under magnetic field influence.

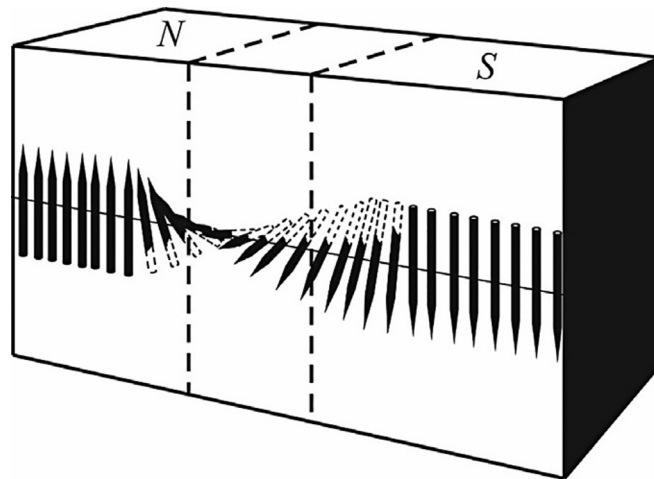


FIG. 6.17

Structure of transition layer between domains (Bloch wall).

magnetization) cannot occur abruptly at one atomic plane. This changing takes place only *gradually* and captures many atomic planes. The exchange energy would be lower if these directions will be distributed between many spins than the change that occurs abruptly. In iron, thickness of the transition layer—domain wall—is around 300 lattice constant ($\sim 1000 \text{ \AA}$).

Thus without an external magnetic field, a ferromagnetic crystal usually is composed of many small individual plots, magnetized to saturation—domains. Domains are separated by the layers—domain walls—in which spins change their orientation inherent in one domain to orientation inherent in the neighboring domain. The resulting size of domains depends on many factors, but usually their size is no more than several micrometers.

Magnetization curve of ferromagnetic. Spontaneous magnetization is an anisotropic property, and therefore it is turned, first, to the direction of “easy magnetization.” Without external field action, all domains are oriented relatively to each other by such a way, at which the total magnetic moment of the ferromagnetic would be zero, as it meets the minimum free energy of the system. When external field H increases, ferromagnetic becomes magnetized, gaining nonzero magnetic moment. The following physical phenomena are observed in a ferromagnetic when magnetization can be divided into three stages.

1. Process of *domain boundaries displacement*. Let us put the crystal, as shown in Fig. 6.16B, into an external magnetic field H . Magnetic vector orientation in different domains relative to H initially is not similar. If the field H increases, the growth of the most favorably oriented domain is energetically more advantageous as compared with other domains. This increase is due to a shift in domain walls. Hence the first step in the magnetization process is the *displacement process*. The shift in domain walls takes place until all favorably oriented domains would extend to the entire crystal

Magnetization curve B of a ferromagnetic crystal is shown in Fig. 6.18. The process of domain wall displacement corresponds to section a on this curve. At small values of H , magnetization B increases slowly, but in case of a stronger magnetic field, this process occurs abruptly, thus causing the *Barkhausen effect*—fast jumping domain walls that are accompanied by a noise.

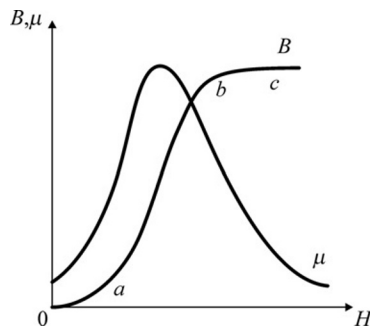


FIG. 6.18

Magnetic induction B and magnetic permeability μ dependence on the magnetic field.

2. Next is the process of *domain rotation*. With further increase in field H , spontaneous magnetization changes its direction toward the field. Here, the magnetization process runs much slower than that in the first stage, and it ends when vector B_{sp} acquires its location along the vector H ; hence magnetization reaches *technological saturation*, as shown in Fig. 6.18, section *b*.
3. The *paraprocess*. After reaching technological saturation, the increase in magnetization with field H slows down, but not terminated. The reason is that at given temperature (over absolute zero), the spins of not all electrons are spontaneously magnetized and oriented in parallel to each other: thermal motion of atoms *partially disorients* spontaneous orientation of spins. However, creating a strong magnetic field can cause *more complete orientation* of all spins. Magnetization, corresponding to the paraprocess, is shown in Fig. 6.18 in section *c*.

As it follows from the ratio $B = \mu_0 \mu H$, magnetic permeability depends on the *rate* of magnetization change in magnetic field: $\mu \sim dB/dH$. In the region of sharp increase in magnetic induction, permeability reaches its maximal value, as shown in Fig. 6.18. Next, when the rate of $B(H)$ dependence slows down, permeability $\mu(H)$ decreases. In different ferromagnetic materials, the initial value of permeability is $\mu = 10^2 - 10^3$, but in its maximum permeability, it reaches up to values of $\mu = 10^4 - 10^6$.

Magnetic hysteresis. A complete cycle of magnetization is shown in Fig. 6.19. During magnetic field increase, magnetic moment M first increases to its maximum (i.e., to the *spontaneous magnetization* M_s), but when magnetic field decreases, magnetization *remains behind*; hence when magnetic field becomes $H=0$, magnetic moment does not disappear, but its value gains to the *residual value* M_r .

The phenomenon of magnetization backlog when magnetic field is changing is the magnetic hysteresis. To destroy residual magnetism, it is necessary to apply the counter field H_c that can reverse magnetization in ferromagnetic. This field is the

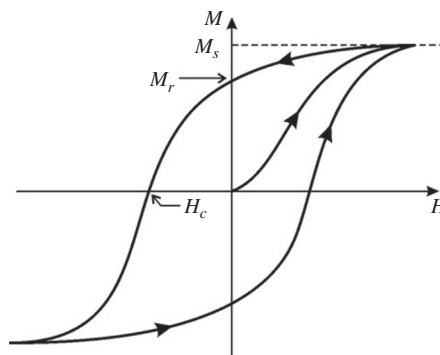


FIG. 6.19

Magnetic hysteresis.

coercive field (retentive force) [9]. As shown in Fig. 6.19, the closed curve, describing the reversal cycle, is the hysteresis loop. The area of this loop is proportional to the energy that an external field spends to reverse polarization in the unit volume of ferromagnetic. During the reversal process, this work turns into heat and characterizes the hysteresis losses. Therefore, in case of repeated reversal magnetization, ferromagnetic becomes heated, and the more intense, the bigger is the area of hysteresis loop. Heating is a result of the internal friction that occurs at continuous reorientations of magnetic domains. At higher frequencies, ferromagnetic is heated additionally: due to Foucault currents arising in a ferromagnetic that usually is a good conductor.

Depending on the shape and area of hysteresis loop, ferromagnetic materials are divided into “soft” materials (small coercivity) and the “hard” (high coercivity) materials. Different application of magnetic materials requires different types of magnetization curve. Materials used in the electrical transformers and electrical machines should show a *quick response* to magnetic field because they have to change their magnetization many times per second. This might result in a partial loss of efficiency and material heating, especially, if the ferromagnetic is rather “hard” (with increased coercive field).

That is why many applications require ferromagnetics with a very low coercive force that reduces the area of hysteresis loop—*magnetically soft materials* (with small coercive field and large permeability). The value of magnetic permeability of the best iron-nickel alloys (permalloys) reaches 10^5 with high induction of saturation: $B_s \sim 1$ T (tesla) while their coercive force is only 0.3 A/m. Hysteresis loop in the permalloy is quite narrow that its reversal losses are around 500 times smaller than those in iron.

Permanent magnets are used to create large permanent magnetic fields; they must have increased coercive force that corresponds to a very wide hysteresis loop. They do not need any reverse magnetization—on the contrary, they must consistently hold the maximally magnetized state: these are the *magnetically hard materials*. They also need high values of saturation in magnetization. The example is the alloys of Al-Ni-Fe type, whose coercive force reaches to around 10^5 A/m, whereas saturation of induction is near 1.5 T. In alloys with cerium, samarium, and yttrium, the coercive field of permanent magnets can reach a value $\sim 10^6$ A/m. In alloys with rare-earth metals, a very large coercive fields is achieved, for example, $H_c = 2 \cdot 10^6$ A/m in the SmCo alloy [2].

Anisotropy of magnetic properties. Magnetic and, in particular, ferromagnetic phenomena in a *single crystal* are anisotropic, although in conventional polycrystalline materials, this phenomenon is imperceptible. The anisotropy of magnetization is caused by different forces of spin-orbital interaction in a structure, which are found in a ferromagnetic single crystal.

Owing to the features of spin-orbital interaction in electrons, along a *peculiar axis*, magnetization occurs most easily and magnetic saturation can be achieved at much lower values of an external magnetic field. These axes are the *easy-magnetization directions*. In iron crystal, for example, this direction is [100] type,

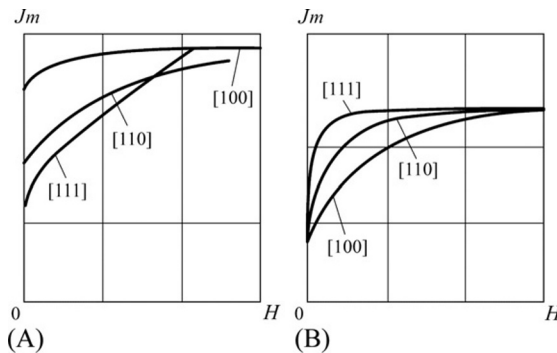


FIG. 6.20

Anisotropy of magnetization in crystals of iron (A) and nickel (B).

as shown in Fig. 6.20A, whereas in directions [110] and [111], magnetization of iron is more difficult, and magnetic saturation is achieved at much higher fields.

In the nickel crystal, by contrast, the direction of easy magnetization is axis of [111] type, whereas the hardest directions of magnetization are [100] type, as shown in Fig. 6.20B. Essential differences in magnetic anisotropy in Fe and Ni crystals result in the *frustration* (uncertainty) of easy magnetization selection in the Fe-Ni alloy (permalloy) that leads to magnetically soft properties.

Magnetostriction. The magnetization of a ferromagnetic sample by all means is accompanied by the change in size and shape. This phenomenon is called the *magnetostriction*. The cause for this effect (which is now widely used in engineering) is large *spin-orbital coupling* in ferromagnetic materials. Fig. 6.21 schematically shows longitudinal deformation (expansion) of ferromagnetic in the magnetic field, accompanied by its transverse deformation (compression). The sample of polycrystalline ferromagnetic with length l , placed in the magnetic field, can either lengthen or shorten on a value Δl induced by magnetic field relative deformation $x = \Delta l/l$ is proportional usually to the *square* of an applied magnetic field: $x \sim H^2$. Nickel

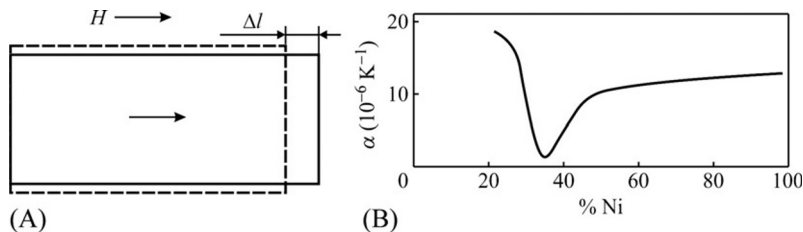


FIG. 6.21

Mechanical and thermal effects in ferromagnetics: (A) magnetostriction; (B) invar effect (coefficient of thermal expansion of nickel/iron alloys).

sample is characterized by shortening in the direction of the applied field ($x \sim -4 \cdot 10^{-4}$); conversely, iron and steel samples in a weak magnetic field become slightly elongated, but in a strong field, they shorten; cobalt sample in a weak field shortens, but in a strong field, it elongates [3].

There are some special ferromagnetic alloys, in which magnetostriction is very large. This effect is used in the magnetostrictive vibrators and produces ultrasonic oscillations with high frequency (up to several megahertz). These vibrators are used, for example, in ultrasonic processing of solids and cleaning them from dirt, in the sonars that are designed to measure the depth of water, and many other facilities and appliances including household.

Thus any process of ferromagnetic magnetization is accompanied by the magnetostriction; it becomes apparent in the orientation of magnetic moments of atoms under magnetic field influence. This process resembles the magnetization of a paramagnetic, hence it is the paraprocess. Magnetostriction is particularly strong near the Curie point where it reaches the maximum value. In the ferromagnetics of hexagonal structure, for example, in the rare-earth metal gadolinium (Gd), the paraprocess occurs; hence the magnetostriction is anisotropic.

The *anisotropic magnetostriction* accompanies the magnetization processes in the *weak* magnetic fields (whereas paraprocess is seen in strong fields only). The components x_{ij} of strain tensor are different in their size and sign. The characteristic feature of anisotropic magnetostriction gives rise to the *change in shape* of the studied sample with a very small change in its volume. In recent theories, two mechanisms of anisotropic magnetostriction are considered: magneto-dipole and one-ion.

The *magneto-dipole* mechanism supposes the interaction of atomic magnetic moments; these moments resemble magnetic dipoles (elementary magnets). Anisotropic magnetostriction in the *3d*-metals (Fe, Ni, and their alloys) and in some ferrites can be described only by the magneto-dipole mechanism. However, this mechanism makes small contribution to the anisotropy of magnetostriction.

The *one-ion* mechanism is more suitable to describe the phenomenon of anisotropic magnetostriction. According to the quantum theory, in this case, orbital electronic cloud of ions acquires *nonspherical* (anisotropic) configuration. When such ellipsoid-like atoms rotates in the magnetic field, magnetostriction might be very large and anisotropic; this is peculiar in some rare-earth metals.

Giant magnetostriction. If the magnet has large magnetostriction, it can be used in various sensors and actuators. However, almost all ferromagnetic materials have only small magnetostriction with deformation of around 0.001%; hence their practical application is difficult. By contrast, the *giant* magnetostriction is 100-fold greater than the usual magnetostriction; it is first found in Terfenol-D (Tb-Dy-Fe alloy) and later in the rare earth ferromagnetic alloy TbCo₂-DyCo₂.

Magnetoelastic effect. According to the Le Chatelier principle (when a system is disturbed in its equilibrium, this system will adjust itself in such a way that any change should be minimal) such mechanical deformation of ferromagnetic, which

causes the change in shape or size, obviously leads to magnetization. This change in magnetic properties of the ferromagnetic in case of its *deformation* is observed experimentally: this is the *magnetoelastic effect*. Some ferromagnetic materials are quite sensitive to external influences that this property is used for strain and tension measurements.

Thermal expansion and invar effect. It is known that thermal expansion of solids at their heating is caused by the vibration of atoms or ions near their equilibrium positions in *anharmonic lattice*. In the weak magnets (diamagnetic and paramagnetic), this anharmonicity is the only reason of their size change during heating. As a result, these substances mostly show only expansion with temperature increase.

However, in ferromagnetics, during their thermal deformation, another very important (and unique) phenomenon is observed: compression while heating in a certain temperature range. The point is that thermal deformation is essentially connected with *spontaneous magnetization*. Conditioned by exchange interaction, magnetostriction depends not only on the external magnetic field but also depends on the *internal magnetization* that in the ferromagnetic is changed with temperature (without any external field). This appears as “thermally induced magnetostriction,” which sometimes is called *spontaneous thermostriction* (as it occurs when an external magnetic field is not applied). This effect is particularly large in the vicinity of the Curie point, that is, when phase transition to magnetically ordered phase occurs.

The effect of spontaneous magnetostriction (with coefficient α_f) affects total thermal expansion coefficient of ferromagnetics because it *compensates* the usual (anharmonic) lattice effect having a positive coefficient α_a . Thermostriction has a sign opposite to that of the regular thermal expansion coefficient; thus the resulting thermal expansion coefficient in ferromagnetic materials can be both positive and negative and even *might be close to zero* in a certain temperature range, as shown in Fig. 6.21B. The group of ferromagnetic materials, in which total thermal expansion coefficient is practically zero ($\alpha = \alpha_a + \alpha_f \approx 0$), assumes the term *invar alloys*. The phenomenon of thermal expansion coefficient *compensation* by spontaneous magnetostriction is the *invar effect*. The term “invar” comes from the word *invariable*, which reflects their ability not to expand and not shrink when the temperature changes. Invar materials are used when high-dimensional stability is required in the precision instruments.

Invar-type metallic alloys (which practically do not change their size while heating or cooling) have long been used in industry. At present, there are many alloys such as “invars,” but always the nature of small coefficient of thermal expansion is magnetic. For example, note that Invar H-36 is the alloy of iron and nickel (36%); Kovar is the alloy of iron, nickel (29%), and cobalt (17%); and others. The α -values in them may be dependent on combinations of components. In the gadolinium crystal, the invar effect is anisotropic, that is, it is diverse in different axes of Gd hexagonal crystal; this opens additional opportunities for technical applications of gadolinium.

Thus in the ferromagnetic alloys, the coefficient of thermal expansion is “manageable,” including, if necessary, $\alpha \approx 0$. Such alloys are widely used in technologies: both in electronics and instrumentation, as well as in the aviation and constructions.

The magnetocaloric effect consists in changing of the material’s temperature (cooling; interesting, of course) during magnetic adiabatic demagnetization or magnetization.

There is thermodynamic explanation of the magnetocaloric effect. In the adiabatic condition (when there is no heat energy exchange with the environment), the magnet does not absorb or return heat ($dQ=0$), and therefore the entropy S is constant: $dS=dQ/T=0$. Therefore under this condition ($dS=0$) and constant pressure ($p=const$), the entropy is considered as a function of temperature T and external magnetic field H : $S=f(T, H)$. Temperature change in ferromagnetics, that is, its cooling ($\partial T < 0$) or heating ($\partial T > 0$), depends on the sign of derivative and on the change in external magnetic field: when $\Delta H > 0$, magnetization occurs, and if $\Delta H < 0$, the demagnetization occurs.

Magnetocaloric effect also occurs in the *paramagnetic*, and it is caused by an increase (or decrease) in the amount of equally oriented atomic magnetic moments (spin or orbital) when the magnetic field is switched on (or off). This effect is studied and applied for a long time. The effect of *adiabatic demagnetization* of a paramagnetic is used to achieve extremely low temperatures. Specific heat at low temperatures is very small ($C_{p,H} \sim T^3$); therefore, the method of paramagnetic cooling is very effective, if initial temperature is sufficiently low.

At normal temperature, another magnetocaloric effect can be applied—in the vicinity of *ferromagnetic phase transition* (in gadolinium, this transition occurs at a temperature of 260 K). In ferromagnetic on a stage of *paraprocess*, a strong magnetic field can orient the magnetic moments that were not yet oriented because of thermal motion. Classic ferromagnetics (Fe, Co, Ni, Gd, and their alloys) are characterized by negative derivative ($\partial M/\partial T < 0$); hence if one increases magnetic field, heating will be observed ($\partial T > 0$), but when the field is turned off, *magnetic cooling* causes $\partial T < 0$ (as $\Delta H < 0$). Predefined by the paraprocess, the magnetocaloric effect can show rather high values in the vicinity of the Curie point.

Therefore in the ferromagnetic with paraprocess participation, not only positive but also *negative* magnetocaloric effect occurs. This can be explained by an example of ferromagnetic compounds of rare-earth metals with iron, where magnetic atomic structure is represented by two sublattices, in which magnetic moments are oriented in antiparallel: sublattice of iron ions (labeled M_1) and sublattice of rare-earth ions (M_2). At the temperature of compensation T_{com} , the magnetization M_1 of iron sublattice is equal to the magnetization M_2 of rare-earth ions. If $T < T_{com}$, then $M_2 > M_1$, whereas if $T > T_{com}$, then $M_2 < M_1$. It means that the total magnetocaloric effect is negative.

Magnetic refrigerator can work at room temperature; the low T_C of magnetic materials (such as gadolinium) or various alloys of rare-earth elements can be applied. The operating temperature range is sufficient for the application of the

magnetocaloric effect in devices such as home refrigerators, air conditioners, and devices for cooling products or electronic equipment. Recently, the *giant magnetocaloric effect* was discovered in the intermetallic compounds based on rare-earth elements in the system silicide-germanide $\text{Gd}_5(\text{Ge-Si})_4$. This kind of materials provides promising application of magnetocaloric cooling.

6.4 ANTIFERROMAGNETISM AND FERRIMAGNETISM

Antiferromagnetic interaction. In case of the negative sign of exchange integral, the antiparallel orientation of spins in the lattice sites of the crystal is more profitable, as shown in Fig. 6.12. Spins are ordered, but no spontaneous magnetization occurs because magnetic moments of neighboring spins compensate each other, as shown in Fig. 6.4C. Such crystal has two magnetically opposite sublattices that are interpenetrated. The well-known antiferromagnetics are listed in Table 6.4.

The structure of antiparallel arrangement of spins is formed spontaneously at temperature below the *Neel temperature* (T_N) in competition with chaotically disordered thermal motion. When an antiferromagnetic is heated above the Neel point ($T > T_N$), uncompensated spins that partially fill *d*- or *f*-shells form something *similar to a paramagnetic* system that is characterized, however, by a *special* temperature dependence of magnetic susceptibility: $\chi = C(T + \theta)$, where C is the Curie-Weiss constant, θ is the characteristic temperature, which in contrast to the paramagnetic phase of ferromagnetics is located in the *negative* range of Kelvin temperature scale, as shown in Fig. 6.13, curve 2.

As examples of antiferromagnetics, some *d*- and *f*-metals are to be mentioned: Cr with Neel temperature $T_N = 311$ K, Mn with $T_N = 100$ K, and numerous other compounds. Antiferromagnetics are also many oxides of *d*- and *f*-metals: MnO with $T_N = 122$ K, FeO with $T_N = 198$ K, NiO with $T_N = 650$ K—this is the highest Neel temperature [9].

Temperature dependence of magnetic susceptibility of an antiferromagnetic indicates a *sharp anisotropy* in their magnetic properties at temperatures below phase transition, as shown in Fig. 6.22.

Table 6.4 Neel Temperature of Some Antiferromagnetics

Crystal	T_N (K)	Crystal	T_N (K)
MnO	122	KCoF ₃	125
FeO	198	MnF ₂	67.34
CoO	291	FeF ₂	78.4
NiO	650	CoF ₂	37.7
RbMnF ₃	54.5	MnCl ₂	2
KFeF ₃	115	VS	1040
KMnF ₃	88.3	Cr	311

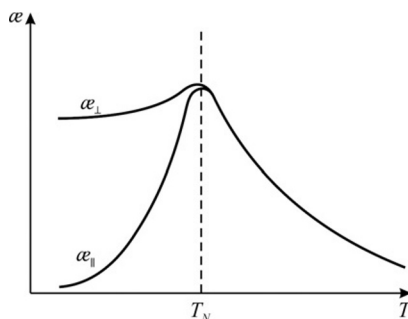


FIG. 6.22

Typical temperature dependence of susceptibility of an antiferromagnetic.

It should be noted that antiparallel spontaneous orientation of electronic spins in the closely located ions strongly *reduces electrical conductivity*—an antiferromagnetic below the Neel temperature is converted from conductors to semiconductors (or dielectric). In the disordered (paramagnetic) phase, an antiferromagnetic does not have any band gap in their electronic spectrum (as metal). However, as temperature decreases and transition to antiferromagnetic phase occurs (at Neel point), in the electronic spectrum of most antiferromagnetic compounds, the *energy gap opens*. Therefore, Neel phase transition in antiferromagnetics might be simultaneously the “dielectric-to-metal” phase transition. Electrical conductivity in the antiferromagnetic phase is 1000 times lower than that in magnetically disordered (conducting) phase.

However, the permeability of antiferromagnetics is small ($\mu \approx 1$), which is obviously insufficient for their technical application as magnetic materials. The smallness of permeability is a consequence of the fact that at low temperatures (in the antiferromagnetic phase), atomic magnetic moments of sublattices *totally compensate* each other; thus the resulting magnetic moment is zero.

When temperature increases and antiparallel orientation of spins become disordered, the value of magnetic susceptibility α increases and reaches maximum at the Neel point, as shown in Fig. 6.22, whereas disordering of spins occurs in a manner similar to that in a paramagnetic. Simultaneously, movement of valence electrons (which in antiferromagnetic phase are constrained by strongly ordered opposite spins) becomes free; hence with a transition into disordered (paramagnetic) phase, the crystal turns into conductor.

Ferrimagnetism. In addition to totally magnetically compensated antiferromagnetics, there are many crystals and polycrystals in which magnetic moments of the sublattices, although being directed opposite to each other, have significant difference in their magnetization, as shown in Fig. 6.4D. These materials have rather complicated structures with varying kinds of atoms that form them and with variable number of uncompensated electrons in the *d*- shells (or *f*-shells). These magnets show properties *similar to those of ferromagnetic materials* because they hold spontaneous

Table 6.5 Curie Temperature T_C and Magnetic Saturation Induction B_S at 4 K in Some Ferrimagnetics

Crystal	T_C (K)	$4\pi B_S$ (Gs)
Fe_3O_4 (magnetite)	858	6400
CoFe_2O_4	793	6000
MgFe_2O_4	713	1800
CuFe_2O_4	728	2000
MnFe_2O_4	573	7000
$\text{Y}_3\text{Fe}_5\text{O}_{12}$	560	2470

magnetization, and the total magnetic moment in their lattice is nonzero [3]. These *ferrimagnetics* are very important for application of substances; some of them are listed in Table 6.5.

Therefore magnetic moments of ferrimagnetics are directed in antiparallel orientation, but they are noncompensated. Electronic interaction in such lattices is known as the *indirect exchange interaction*, at which there is no direct overlap of magnetic ion wave functions. However, the overlap of wave functions of diamagnetic anions (e.g., O^{2-}) with the wave functions of magnetic cations (e.g., Fe^{3+}) enables the exchange interaction *through the virtually excited state*, as shown in Fig. 6.23.

The $2p$ -shell of the oxygen ion in its main state is completely filled, as shown in Fig. 6.23A, and despite the overlap with iron ion wave functions (p -orbitals of O^{2-} and two d -orbitals of Fe^{3+}), any exchange interaction between them is absent. However, in the *excited state*, as shown in Fig. 6.23B, one of the p -electrons of oxygen transfers to the $3d$ -shell of the iron ion. In compliance with Hund's rules, this electron has to move, whose spin is antiparallel to the spins of electrons in the half-filled shell of the Fe^{3+} ion. Leaving $2p$ -shell, the electron, because of negative exchange interaction, orients spins of neighboring iron ions, as shown in Fig. 6.23B.

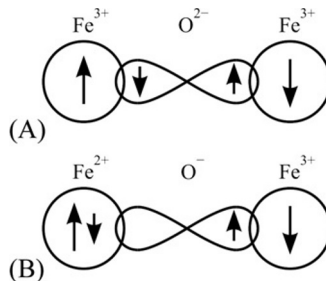
**FIG. 6.23**

Diagram illustrating indirect exchange interaction in the system $\text{Fe}^{3+} - \text{O}^{2-} - \text{Fe}^{3+}$: (A) basic state; (B) excited state.

The intensity of indirect interaction increases with the extension of overlapping of electronic shells, that is, with the *strengthening of covalent bond*. Because covalent bonds are noncentral, the indirect exchange interaction reaches maximum when three interacting ions are not collinear. Therefore, the compensation of magnetic moments is not complete in the complex structure of antiferromagnetic.

Permeability of ferrimagnetics, being less than the values of permeability in typical ferromagnetics, nevertheless, is rather big for successful technical applications: $\mu \sim 10^2 - 10^3$. With regard to other physical properties (hysteresis, nonlinearity, and domain structure), the ferrimagnetic is close to ferromagnetic, but its magnetization decreases with increase in temperature nonmonotonically (as in the case of ferromagnetics), sometimes passing through a zero before reaching the final Curie temperature, as shown in Fig. 6.24.

Several different sublattices that exist in a ferrimagnetic make temperature dependence of spontaneous magnetization rather complicated compared with that in a conventional ferromagnetism, as shown in Fig. 6.14. This is because temperature dependence of spontaneous magnetization may be different for various sublattices of a ferrimagnetic.

It is necessary to recall that most of the ferromagnetics are metals (with high conductance), and therefore they cannot be used at increased frequencies owing to high losses (conditioned by Foucault currents). Therefore even for electrotechnical applications (at a frequency of 50 or 60 Hz), and especially for mobile (transport) electrical engineering (a frequency of 400 Hz), the iron, permalloy, or any ferromagnetic metal should be divided into the separate plates (or even into a thin foil) with the electrical insulating layers between the plates (or foil).

One of the possible ways of using ferromagnetic metal at radio frequencies is to reduce losses from Foucault current by using micron-sized ferromagnetic particles pressed together with polymer (*magneto-dielectric composites*). Nevertheless, this technology cannot prevent losses from Foucault currents in the microwave range (where magnetic materials are widely used, particularly in information and computing technique).

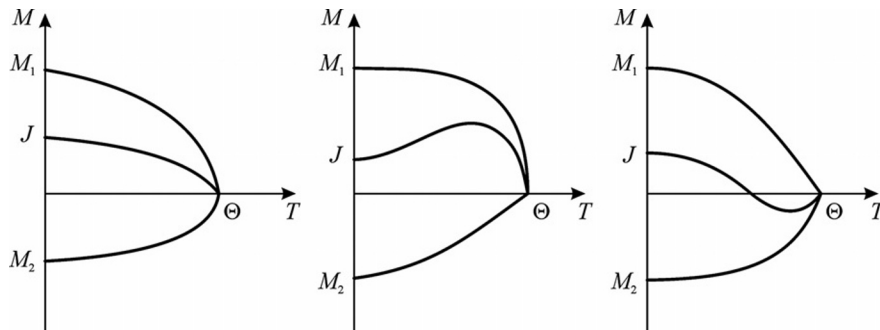


FIG. 6.24

Different possibilities of ferrimagnetics magnetization.

However, insulating properties that are extremely necessary are achieved in the ferrimagnetics by the combination of depressed conductivity with rather strong magnetism in the *elementary crystal cell*. Therefore, the main advantage of ferrimagnetics is rather high permeability with significant manifestation of magnetic properties, combined with high adequate electrical resistance. This is especially important in microwave technology because of the provided small loss of electromagnetic energy [5].

In technologies, ferrimagnetics usually are called *ferrites*, and the most important representative among them is the *magnetite* $\text{Fe}_3\text{O}_4 = \text{FeO} \cdot \text{Fe}_2\text{O}_3$. Its unit cell is ferrospinels that has a cubic lattice formed of eight molecules of $\text{FeO} \cdot \text{Fe}_2\text{O}_3$. In this mineral, the negative oxygen ions form face-centered lattice, in which the compound Fe_3O_4 has one divalent (Fe^{2+}) and two trivalent (Fe^{3+}) iron ions. As shown in Fig. 6.9, the Fe atom and Fe^{2+} and Fe^{3+} ions have different number of uncompensated spins. One sublattice of ferrite contains one half of trivalent iron ions, and another sublattice—the second half of trivalent iron ions and divalent iron ions (or metal that replaces iron). Magnetic moments of sublattices are antiparallel. Therefore magnetic moments of trivalent iron ions are compensated, but spontaneous magnetization is caused by magnetic moments of divalent iron ions (or any other metal that replaces iron).

In various ferrites with a structure of magnetite, the divalent iron ions Fe^{2+} can be substituted by the divalent ions of other metals such as Mg^{2+} , Ni^{2+} , Co^{2+} , Mn^{2+} , and Cu^{2+} . The general formula of ferrites with spinel-type structure is $\text{MeO} \cdot \text{Fe}_2\text{O}_3$, where Me is the divalent metallic ion. Only the divalent metal ions cause spontaneous magnetization of many ferrimagnetics. Some ferrites (manganese and nickel ferrites) have rather high permeability: up to several thousands. In other ferrites, usually $\mu \sim 100$ (however, ferrites based on zinc and cadmium are not ferrimagnetics).

Ferrites based on rare-earth elements. Magnetic materials based on the rare-earth elements have gained considerable scientific and technological interest. It is necessary to remember that rare-earth elements (or lanthanides) are the elements of the third group with numbers 57–71 (La, Ce, Nd, Sm, etc.). The elements scandium Sc and yttrium Y are attributed to this group owing to similar properties. Alloys and compounds of these elements have pronounced magnetic properties. Their difference from magnets of iron type (*3d*-metals) lies in the fact that the magnetic moment of lanthanides (*4f*-metals) is determined mainly by the *spin properties of electrons*, whereas the importance of orbital moment is smaller (nevertheless, orbital moment also has some influences on their magnetic properties).

From 14 rare-earth elements, *ferromagnetism* is observed only in six of them (gadolinium, erbium, dysprosium, holmium, thulium, and terbium) but in most of them (except gadolinium) at a certain temperature, the ferromagnetism gets converted into antiferromagnetism. In the gadolinium, similar to iron, the ferromagnetic state occurs *directly* from the paramagnetic state (at a temperature of 290 K). Five rare-earth elements (cerium, praseodymium, samarium, europium, and promethium) are antiferromagnetics. Magnetic moments of lanthanum and lutetium are zero (they are diamagnetics). Thus magnetic properties of rare-earth metals are quite different and complicated.

Permanent magnets based on rare-earth alloys. These ferrites have very large magnetic anisotropy (from two to three orders of magnitude greater than that in the iron group) that allows the manufacture of permanent magnets. Of particular interest are the intermetallic compounds of RCo type, where R is the rare-earth metal. For example, coercive force of samarium-cobalt alloy (SmCo) is $\sim 20,000$ A/m (coercive force of conventional ferromagnetics is less in order of magnitude). The RCo type alloys are widely used in engineering as permanent magnets: very high coercivity with a large magnetic induction can essentially reduce the weight and size of magnetic systems.

In recent years, some alloys based on NdFe have found important applications; their parameters exceed the parameters of SmCo magnets. However, as a record of the properties of hard magnets, a single crystal of terbium-cobalt alloy is considered. At low temperatures, the coercivity of the TbCo compound surpasses the NdFe and SmCo magnets by seven times.

Rare-earth microwave ferrites. In the microwave range, the properties of ferrites that make them possible to create the *nonreciprocal* devices are used, i.e., the devices, having different specifications for different directions of energy spreading (valves and circulators), as well as microwave devices with fast controlled parameters (phase shifters and switches).

The microwave technique also widely uses the phenomenon of *microwave magnetic resonance*. Atoms in a magnet appear similar to mechanical “whippings” (gyroscopes). Magnetic moment of such whipping is directed along the axis of rotation. If an external magnetic field is applied directed at some angle to the axis of whipping rotation, this axis will rotate around the direction of the applied field. This phenomenon is called *precession*. The frequency of precession depends on the material and on the field strength. If damping of such oscillation would be absent, this precession will exist indefinitely and oscillator will be lossless. However, owing to losses (energy dissipation on phonons and defects), this precession decreases, and the direction of magnetic moment is set along the direction of the magnetic field. If constant magnetic field and alternating field of some frequency are *simultaneously* applied to the magnetic crystal, it can increase precession angle. This angle reaches the maximum value, when frequency of the external field coincides with the frequency of precession. This phenomenon is the *gyromagnetic* (or ferromagnetic) *resonance*.

Gyromagnetic resonance finds technical application because, at the resonant frequency, the energy loss in a magnet is maximal, which provides maximum selective absorption. The higher the quality of the magnetic crystal, the bigger is the absorption of energy and the narrower is the band of magnetic resonance. The best results are observed in case of the yttrium-iron garnet. The microwave filters are built based on the magnetic resonance in yttrium-iron garnet (and in some other ferrites, the magnetic resonance close to it). The quality factor of such filters reaches 10,000. The rare-earth ferrites are used also as the terminators of microwave power. That is the reason why properties of single crystals of ferrites are of particular interest in the microwave technology.

The monocrystalline ferrites are specially grown crystals that have remarkable highly ordered crystalline structure. The feature of single-crystal ferrites is their increased resistivity and *high optical transparency*. These properties enable to apply them not only in microwave devices but also in magneto-optical devices. Properties of ferrites depend on their crystalline structure. They might have cubic symmetry as *ferrogarnets* with the general structural formula $3\text{Me}_2\text{O}_3 \cdot 5\text{Fe}_2\text{O}_3$ (where Me^{3+} is a rare-earth element), and they might also have a rhombic symmetry such as *orthoferrites*, with the general structural formula MeFeO_3 (where Me is a rare-earth element or yttrium).

The axis of easy magnetization in ferrites might be different in various crystals. In crystals with cubic symmetry, the axis of easy magnetization is [111]. The cubic cell has four diagonals; hence such crystals have four axes of easy magnetization. In the crystals with rhombic structure, the easy magnetization axis coincides with axis [001]. In these crystals, the axis of easy magnetization is only one; hence they are called the magneto-monoaxial.

Monocrystalline ferrites are light transparent materials, and this is their important feature when they are used in optical spectrum. The value of absorption coefficient is relatively small. For example, an orthoferrite plate with thickness of 1 mm in the wavelength range 1.5–5 μm transmits 95% of light, whereas a 30- μm -thick plate transmits 50% of red light (with a wavelength of 0.6 μm). Such good properties are peculiar to high-quality crystals only. If the raw material for ferrites is not very clean or the ferrite plate is not polished properly, optical transparency will be much less. Magneto-optical effect in ferrites has important application in the optoelectronics and instrumentation.

Giant magnetostriction. Ferrimagnetic materials based on rare-earth elements might have a very large coefficient of magnetostriction that makes them promising for use in the area of actuators. The main point of magnetostriction effect is the change in sample shape and size when it is placed in the magnetic field. Previously, magnetostriction is considered as a very small effect (in ordinary ferromagnetics, the possible strain is only 0.003%). However, in the rare-earth metals (terbium Tb, dysprosium Dy, and some alloys), the *giant* magnetostriction effect is discovered when the strain is higher in two orders of magnitude: 0.5% for the alloy TbDyZn. Another alloy, namely, terbium-iron (especially the TbFe single crystal) is the best magnetostrictive material in modern engineering.

Application of magnetostrictors, based on rare-earth materials, makes it possible to create *power actuators* (need, for instance, in the adaptive optics for large reflecting telescope). They can also be used in radio engineering and telecommunications as enormous power sound sources, superpower ultrasonic transducers, magnetostrictive highly microshift mechanisms, and screw devices to develop ultrasensitive audio receivers [5].

Magnetic semiconductors and dielectrics. Magnetic materials, depending on the type of chemical bond, are divided as magnetic metals, magnetic dielectrics, and magnetic semiconductors. Previous sections dealt mostly with the magnetic metals and alloys that are characterized by the special type of particle bonding: ionic lattice

crowded by electronic gas. In the magnetic semiconductors and magnetic dielectrics, their chemical bonds are mixed (ionic-covalent) and depend on anion and cation affinity to electron. During chemical bond formation in magnetic semiconductors and magnetic dielectrics, a significant role is played by spin magnetic moments of electrons in unfilled *d*- or *f*-shells of ions.

Magnetic semiconductors and magnetic dielectrics are predominantly compounds whose components are transition metals and rare-earth elements that show ferro-, antiferro-, or ferrimagnetic ordering of their lattice. This arrangement significantly affects optical and electrical properties of a material. Magnetic controlling by *optical properties (magneto-optics)*, peculiar exactly to dielectrics and wide-band semiconductors, is needed in high optical transparency.

It is evident that for magnetic controlling properties by semiconductors, a strong relationship between unfilled *3d*- and *4f*-shell magnetic ions with free charge carriers is necessary. Owing to spin ordering in a lattice, its magnetic ions affect the movement of free charge carriers in a crystal, and these carriers, in turn, can affect the magnetic ordering in a lattice.

It is well known that to change magnetic moment orientation in local parts of magnetic substance, an external magnetic field should be applied. This is the basis of traditional magnetic memory, which is widely used in computers. The necessity to increase the *density* of memory cells in devices strikes against a problem of *smallness* of managing magnetic fields. Therefore a possibility of magnetic material *local reversing* by a beam of *spin-polarized electrons* (during the time of their passing through this local area) looks as very important. Some magnetic semiconductors can introduce the spin-polarized electronic current between *p*- and *n*-type areas. In case when the spins of charge carriers are *preliminary ordered* by an external field, this current creates spin-ordering in the adjacent semiconductor, which can be preserved awhile even at room temperature. Significance of this method is the possibility to control spin orientation by the *electrical field* instead of the magnetic field.

Magneto-optical phenomena. Telecommunications, instrumentation, electronics, and computing—all these technologies now use optical frequency range of electromagnetic waves. In optical devices, mainly the transparent medium should be used: magnetic dielectrics and wide-gap semiconductors (metals strongly reflect electromagnetic waves but magnetically tunable reflection also can be used). Physical phenomena, applied in the magneto-optics, are highly varied. They are based on the dependence of optical properties on the direction of light propagation (anisotropy) and light beam controlling using magnetic semiconductors and dielectrics.

Next, the optical phenomena that are caused by the influence of a magnetic field on a light-transparent magnet will only be considered. This area of science and technology (called the *magneto-optics*) studies and uses the change in the optical properties of a matter under the influence of a magnetic field. Magnetic materials, which are utilized in functional magneto-optical devices, can be divided into two groups.

The *first group* includes materials with relatively low optical absorption, which are applied for spatial-temporal modulation of light in the amplitude or phase. To elaborate magneto-optical functional devices, the essential importance for material

selection is their optical absorption. Among various magnetic materials, relatively small absorption in the visible and near-infrared parts of the spectrum might be expected only in nonmetallic materials. These are the following ferrimagnetics:

- ferrite-garnets of the general formula $R_3Fe_5O_{12}$;
- orthoferrites of the general formula $RFeO_3$ (R—rare-earth ion);
- ferrites with a spinel structure, for example, $CdCr_2Se_4$ and $CuCr_2Te_3I$.

To choose the most transparent magneto-optical materials, one should be guided by the fact that the intensity of absorption is caused by electro-dipole transitions in the $3d$ -ions (usually Fe^{3+}); this absorption can be reduced by the decrease in nonequivalent positions of iron ions in the structure of a magnetite.

The *second group* of magneto-optical materials includes thin magnetic films, based on intermetallic compounds with rather high absorption coefficient in the visible and infrared ranges. Magnetic *amorphous film* designed for magneto-optical devices and film composition follows the general formula R-Me-Z, where R is the rare-earth ion, Me is the transition metal (Mn, Ni, Fe, and Co), and Z is the non-magnetic metal (Mo, Cu, and Au).

The fundamental cause of the magneto-optical effect is the *splitting of energy levels* of atoms in the magnetic field. During isolated atoms study, this splitting was found, which was termed as the *Zeeman effect*. However, in crystals, the magneto-optical effects are also a result of the Zeeman effect. Magneto-optical effects, first, change light polarization characteristics, and, second, can control the distribution of polarized light in the dispersion medium.

In addition to ordinary *optical anisotropy* that occurs in medium under the influence of an electrical field or mechanical strain, *circular anisotropy* occurs in the magnetic field, which is caused by nonequivalence of polarization rotation in a plane perpendicular to the field. This important fact is the result that the *magnetic field is axial*. Because of absorption, the left-hand and the right-hand polarized light become different; hence amplitudes of output components can be various. This is the *magnetic circular dichroism*. Its existence leads to the fact that after light passage through the medium, the linearly polarized light turns into the elliptically polarized light.

In the absorbent medium, the magnetic *linear dichroism* also appears, and this is a difference between absorption coefficients of two linearly polarized waves while they pass through the magnetized medium. The presence of dichroism leads to the rotation of orientation angle of ellipse during propagation.

When light spreads perpendicular to the direction of magnetization J , then in the magnet, a *linear double refraction* is observed, called the Cotton-Mouton effect. It is caused by a difference in refractive index of two linearly polarized components of light waves, polarized parallel and perpendicular to J . Appeared phase change gives rise to elliptically polarized light at the exit of a medium. The Cotton-Mouton effect, in contrast to Faraday effect, is even: its value is proportional to the *square* of magnetization.

Along with magnetic optical effects that occur during the passage of light through a magnetized medium, there are some effects that are due to *light reflection* from the

surface of the magnetic sample. Therefore optical anisotropy that is acquired in the magnetic field can be detected not only in transmission characteristics but also by light reflection from the magnetic surface. When magnetization of the active medium changes the *polarization* of reflected light, the nature and extent of this effect depends on the relative position of sample surface on incident light polarization and on magnetization vector. This effect is observed mainly in the ferromagnetics; it is the *magneto-optical Kerr effect*. Depending on the relative orientation of magnetization J , on the direction of light propagation k , and on the normal n to surface, there are three types of Kerr effects: polar, equatorial, and meridional.

The *polar effect* is the rotation of polarization plane with the appearance of ellipticity during the reflection of linearly polarized light from the surface of the magnetic material when magnetization is parallel to normal: $J \parallel n$.

The *equatorial effect* is observed in the absorption of a magnetic material, and it involves change in intensity and phase shift of linearly polarized light, reflected from the magnetized medium, when magnetization is perpendicular to J and to the plane of incidence: $J \perp n$.

The *meridional Kerr effect* is the rotation of polarization plane and appearance of ellipticity as a result of linear polarized light reflection from magnetic surface, when magnetization J is perpendicular to normal n and located in the plane of light incidence.

To control light *transmission*, magnetization of the working medium should be changed by an external magnetic field. To do this in transparent ferromagnetic, first, the displacement of domain walls should be used, and, second, the rotation of magnetization vector in a magnetic field is necessary. In magneto-optical modulators that use the first process, optically transparent active medium often is applied: ferrite-garnets $R_3Fe_5O_{12}$ and orthoferrites $RFeO_3$ (where R is the rare-earth ion). In orthoferrites, abnormally large Faraday rotation is observed, although the saturation of magnetization in orthoferrites is significantly less than that in ferrite-garnets.

Faraday's rotation is proportional to sample thickness, and it can be observed only when light extends *along the optical axis* of orthoferrite. The magneto-optical quality Q in $Nd_{0.8}Pr_{0.2}FeO_3$ reaches 14 degrees/dB, exceeding the value of merit for all known magnets. In $YFeO_3$, the magneto-optical Q factor is lower. The main feature of orthoferrites is the high mobility of domain walls that make them promising to create high-speed magneto-optical devices.

6.5 NANOMAGNETIC MATERIALS

Apparently, the possibilities of volumetric materials used by engineers already reached their maximum. It is believed that it is barely possible to get any significant improvement in their performance only through a more thorough technology or by changing in components. Therefore it might be assumed that a subsequent creation of materials with *new properties* should be associated with fundamental changes in the *structure* of substances, thus affecting properties that are necessary for contemporary

applications. It is considered that one of most promising and relatively new research areas in the field of material science is the creation of materials that are condensed from *very small* crystals, clusters, and fragments that consist of around 10^2 – 10^5 atoms.

The main reason for the difference between nanomaterials and conventional materials is that the *ratio of surface to volume* in nanomaterials is rather big. The smaller the size of nanocluster, the greater is the influence of its surface properties compared to its bulk properties. In a certain sense, the nanostructure transforms properties of the crystal surface into the volumetric properties of condensed nanomaterial. In other words, the properties of nanoformed substance depend on the ratio of the number of atoms located on the surface of nanocluster to the number of atoms located in its volume, and this ratio might be quite different. Therefore, by controlling the size and the shape of clusters, the properties of a nanomaterial can be purposefully changed.

Nanostructurization of *magnetic materials* enables to operate in a wide range of their characteristics. Nanotechnology can be used, primarily, to create a material with adjusted type of magnetization curve: both for extremely magnetically soft materials and for extremely magnetically hard materials. Fundamental magnetic properties of a matter in their nanostate vary considerably owing to correlation in the interaction of spin and orbital moments in lattice cells, located on the surface of the nanoparticle. The properties of ferromagnetics and ferrimagnetics in their nanostate can be changed, especially strongly. In the magnets formed from nanoclusters, the nature of short-range ordering becomes different; that is why new properties appear (sometimes, very important for technical application).

Some examples of atomic magnetic moment dependence on the size of nanoparticles in the main ferromagnetics are shown in Fig. 6.25. Magnetic moment of atoms in *bulk* ferromagnetic usually is less than the number of uncompensated spins in atoms (which are $2 \mu_B$ for nickel, $3 \mu_B$ for cobalt, and $4 \mu_B$ for iron). Owing to the *spin-orbital interaction*, the effective magnetic moment of atoms in bulk ferromagnetic is much smaller ($0.6 \mu_B$ for nickel, $1.8 \mu_B$ for cobalt, and $2.2 \mu_B$ for iron). Changing the number of atoms in the nanoparticles leads to a significant *increase* in effective magnetic moment with decrease in particle size (finally, it becomes very close to the magnetic moment of a single atom). This effect significantly increases the permeability of magnetic nanocomposites and causes a number of other effects.

Thus the effect of nanostructuring on ferromagnetic characteristics is very significant, particularly, the effect of nanoparticle size that is used in the composed form of nanoparticle magnetic materials. Therefore, in modern materials technology, it is possible to “design” their properties particularized for various fields of technology by changing the size of grain structure.

Soft magnetic nanomaterials. The change in orientation of magnetic domains or clusters under an externally applied magnetic field influence might be possible even under small magnetic fields: these are *soft magnetic* materials. For example, very soft magnetics are the films of amorphous alloys, made with the compound $\text{Fe}_{70}\text{Si}_{13}\text{Nb}_3\text{Cu}_9$ and obtained by rapid cooling of melt that is poured on the cold copper

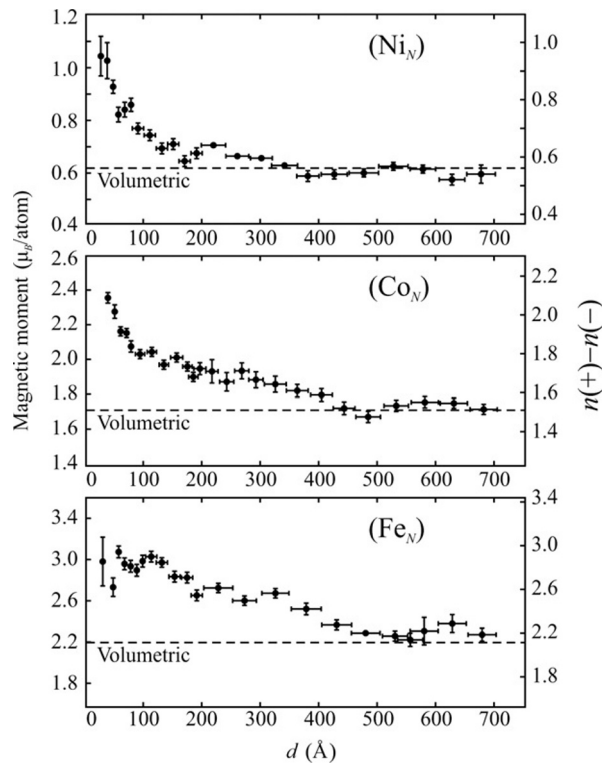


FIG. 6.25

Magnetic moment per atom for nanostate ferromagnetics Ni_N , Co_N , and Fe_N depending on the particle size in the Angstrom unit scale ($1\text{Å}=0.1\text{ nm}$).

rotating drum: the prepared alloy consists of disordered 10-nm nanoparticles. Having a large saturated induction (1.2T), this alloy shows small coercive force (0.5 A/m).

One reason for the increase in magnetic softness and therefore the appearance of huge magnetic permeability ($\mu \sim 5 \cdot 10^5$) is the light orientation of spins. The point is that the concentration of structural defects in the nanoparticles is small (in them, any defects easily diffuse on the surface); as a result, the orientation of spins in the external magnetic field becomes much easier. Nanomagnetic cluster can have its magnetic moment orientation such as that shown in models given in Fig. 6.26 [10].

As shown by an experimental study with magnetically soft materials (nanosized powders of amorphous alloy of the compound $\text{Fe}_{70}\text{Ni}_{10}\text{CO}_2$ with a grain size of 10–15 nm), the hysteresis loop coercivity is practically absent, as shown in Fig. 6.27B, because each of the nanoscale grain constitutes only one domain. (In common ferromagnetics, hysteresis is caused by the orientation of big domains.)

Magnetic materials of this type, showing no hysteresis, are the *superparamagnetics*. This term means that at temperatures below the Curie point and in wide

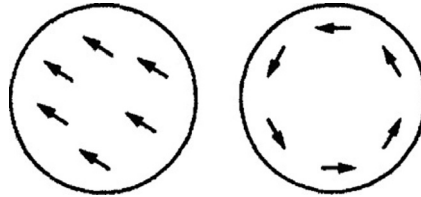


FIG. 6.26

Different orientation of magnetic moments in ferromagnetic nanoparticles.

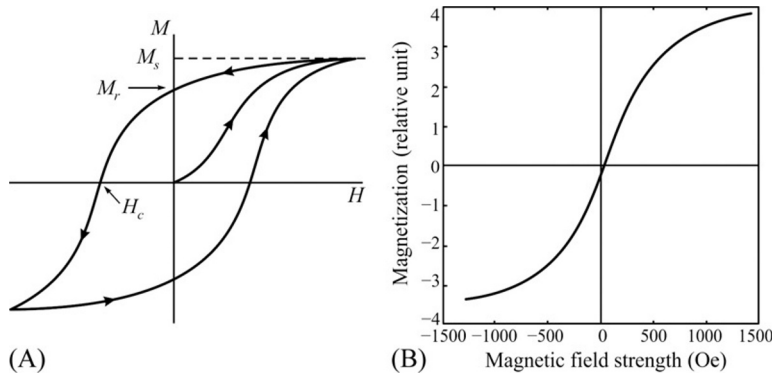


FIG. 6.27

Typical magnetic hysteresis (A) magnetization M and coercive field H_c ; (B) nonhysteresis magnetization curve for the nanomaterial Ni-Fe-Co.

temperature range, *superparamagnetics* are similar to paramagnetics. In typical ferromagnetic (or in ferrite), their spontaneous magnetization, occurring below the Curie point, is accompanied by the internal force that strongly maintains spins in their oriented state; hence to change their direction, one needs to overcome the coercive field (H_c as shown in Fig. 6.27A). In this case, in a ferromagnetic, a sufficiently large energy of anisotropy exists, which makes magnetic moment to choose one or the other direction. However, in the nanoparticles, owing to significant violations in the structural bonds of atoms, energy of anisotropy is practically absent; hence the direction of orientation of electronic spins can be easily changed even under a very weak magnetic field. Therefore during the reversal of magnetization, there is no hysteresis.

The more pronounced the magnetic softening, the smaller is the nanoparticle size. However, unlike conventional paramagnetic (whose susceptibility at low temperatures demonstrates the Curie law: $\alpha \sim K/T$), in case of superparamagnetics, the limitative temperature exists, below which the possibility of “soft” (noncoercive) orientation of magnetic moments is limited. The reason of this restriction is that orientation of magnetic moments in nanoparticles is supported by thermal motion in a

lattice, which at low temperatures becomes insufficient. The temperature, at which this movement is blocked, depends on the size of nanoparticles.

Magnetically hard nanomaterials. Nanotechnology enables to also control the *coercive field* that is very important to achieve a great H_c . With traditional (bulk) materials, some powerful permanent magnets are made from the alloys of neodymium, iron, and boron. Among them, a very big residual induction is reached (1.3 T) with a coercive force value of 10^6 A/m, which is more than a million times higher than that in magnetically soft alloys. However, nanotechnology can significantly improve even these data. Some results of grain size have an influence on the properties of the alloy $\text{Nd}_2\text{Fe}_{14}\text{B}$ as shown in Fig. 6.28A. From this figure, it follows that residual magnetization *increases* significantly if the grain size becomes smaller than 40 nm (the coercive field becomes three times higher). Another approach on how to change parameters of magnetization curve for this material is to create a mixture of nanoscale particles of the magnetically hard compound $\text{Nd}_2\text{Fe}_{14}\text{B}$ and the magnetically soft α -phase of iron.

The study of soft iron particles influence on magnetically hard matter confirms that coercive field can be further increased. This is due to the exchange interaction between hard and soft nanoparticles, which turns magnetization of soft phase particles exactly in the direction of hard particle magnetization. By size reducing of nanoparticles, the granular magnetic material can be significantly improved [2].

Nanomagnetic films for computer memory devices. The study of magnetic materials, mainly the films produced by nanotechnology, aims to increase the capacity of magnetic information drives—such as the hard drive of a computer. The unit of storage information is *bit*; to reach the density of 10 Gb (10^{10} bits) per square inch, the single bit should have a length of $\sim 1 \mu\text{m}$ and width of $\sim 70 \text{ nm}$. The thickness of the

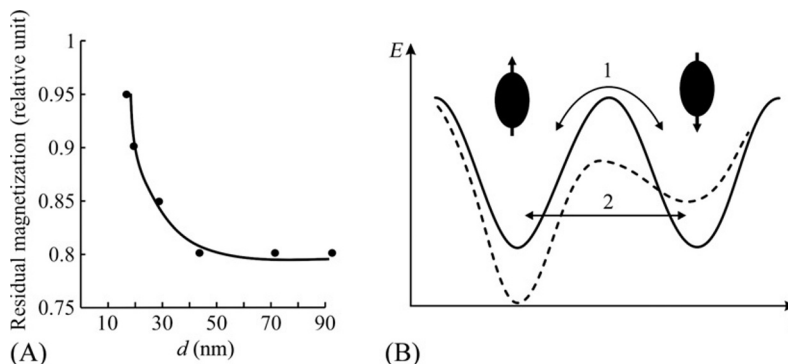


FIG. 6.28

Residual magnetization M_r dependence on particle size (A); scheme of double potential well, which shows energy dependence on magnetic moment orientation in external magnetic field absence (*solid line*) and presence (*dotted line*): 1—thermally activated switching, 2—tunneling (B).

magnetic layer in this case should be around 30 nm. Magnetic storage medium such as hard drives can be realized by using small crystals of chromium-cobalt alloy. Problem of bit size diminution < 10 nm is the “self-erasing” of memory because magnetization vector of microarea can change its direction under the influence of thermal fluctuations. One solution to this problem requires the use of nanoscale grains with high values of saturated magnetization.

Another example: using nanotechnology, magnetic FePt nanograins that have higher magnetization than their analogs are obtained. Particles of FePt are formed by heating the solution of platinum acetylacetonate and carbonyl iron with addition of a reducing agent. After spraying this solution on the substrate, it evaporates leaving passivating particles. The thin film, obtained as a result of these operations, consists of a hard carbon layer containing FePt particles with a size of around 3 nm. This kind of magnetic nanoparticles enables to achieve a density of 150 Gb per square inch, which is 10 times greater than that of most existing commercially available medium.

When the size of magnetic nanoparticles is very small, magnetic vectors of atoms are oriented by a magnetic field equally within a grain, avoiding difficulties that arise in other cases (when the adjacent domain walls exist with different directions of magnetization). The reason is a peculiarity of nanoclusters: the reducing defects concentrate within cluster (defects easily diffuse to the surface).

Typically, a magnetic medium uses elongated magnetic grains. Dynamic properties of such magnetic particles can be described by a model, which assumes that without applying magnetic field the ellipsoidal grains have only two possible directions of their magnetic moment: “up” or “down,” relative to the long axis of magnetic elongated particles, such as those shown in Fig. 6.28B.

Magnetic energy dependence on magnetic moment orientation is characterized by symmetric potential well with two minima, divided by a potential barrier. Under the influence of thermal fluctuations, the elongated particle can change orientation of its magnetic vector. This particle can also (but far less likely) change its magnetic orientation by means of quantum-mechanical *tunneling*. This occurs when heat energy $k_B T$ is much less than the height of barrier. Tunneling is a purely quantum-mechanical effect, which arises from the fact that there is a probability of magnetic state changing from the direction “up” to the direction “down.” In the external magnetic field, the potential, which divides minima, changes, as shown in Fig. 6.28B, by a dotted line; in case when the magnetic field is equal to coercive force, one of the levels becomes unstable.

This model explains many magnetic properties of small magnetic particles, for example, the shape of hysteresis loop. However, this model has a limitation with coercive field value because it allows only one way for reorientation. Magnetic energy of a particle is assumed as a function of *collective orientation* of spins in magnetic atoms and a function of an external magnetic field. This model considers only simple (linear) dependence of magnetic energy of particle on its size. However, when the size of particle becomes approximately 6 nm, most atoms are located on the particle surface. This means that they can have magnetic properties that significantly differ from those of larger particles.

It is shown that replacement of magnetically soft α -iron in a variety of chemicals substances leads to change in coercive force up to 50%, thus indicating the importance of particle's surface in the formation of magnetic properties in a grain. Thus, the dynamic behavior of very small magnetic particles is more complicated than it follows from considered model.

Nanomagnetic particles in the interstices. An interesting research in the field of nanomagnetism is to create porous material crowded by magnetic nanoparticles. In nature, there are substances with molecular cavities filled with nanoscale magnetic particles. One example might be ferritin—a biological molecule that contains 25% iron by weight, whereas its shape is a symmetric protein shell. It has a hollow sphere with inner diameter of 7.5 nm and outer diameter of 12.5 nm. In biological systems, this molecule plays a part of “repository” of iron Fe^{3+} in an organism. One quarter of iron in human organism is contained in molecules of ferritin and 70% in hemoglobin.

Under normal conditions, the cavity of ferritin is filled by quasicrystalline iron oxide: $5\text{Fe}_2\text{O}_3 \cdot 9\text{H}_2\text{O}$. This iron oxide solution can enter from outside into a cavity, in which the number of iron atoms can vary from a few to several thousand. Magnetic properties of this molecule depend on the number and type of particles in a cavity. It can be both ferromagnetic and antiferromagnetic.

In ferritin at low temperature, the quantum tunneling can be observed. Even when magnetic field is absent, the magnetization demonstrates tunneling between two minimal positions. Resonant frequency of this tunneling depends on the total magnetic moment; frequency dependence on the number of iron atoms in the ferritin molecule is shown in Fig. 6.29B. It is seen that resonant frequency decreases with the increase in atom quantity in a cluster. Under external magnetic field, this resonance disappears because the symmetry of double-well potential becomes broken.

The magnetoelectronics (spintronics). In recent years, a new science and technology has developed—magnetoelectronics, or as it is now called the *spintronics* that deals with the study and application of some effects and devices that use electronic spin. Spintronics studies magnetic and magneto-optical interactions in metallic and

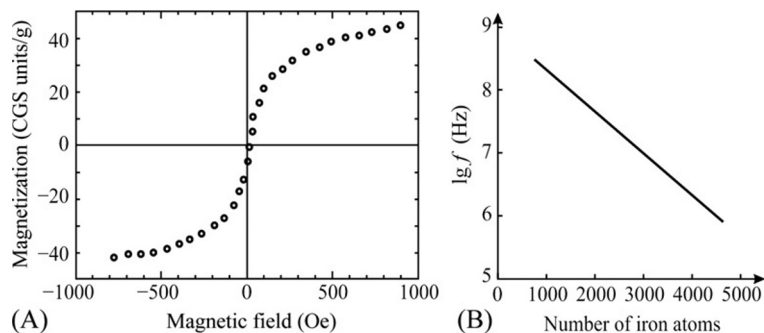


FIG. 6.29

Magnetic properties of ferritin: (A) magnetization curve; (B) resonant frequency dependence on number of iron atoms in the cavity of molecule.

semiconductor structures, as well as quantum phenomena in the magnetic structures of nanometer size [5].

Thus spintronics is a new research direction of nanoelectronics, in which, in addition to electron's charge, its *spin is used* for information processing. In the modern electronics, many devices already exist based on spin phenomena. This is, for example, the capping (manufactured by IBM) that reads information from magnetic disks, and new type of magnetic memory—magnetic random access memory (MRAM). These devices operate using giant magnetoresistance effect (GMR).

Giant and colossal magnetoresistance. The magnetoresistance effect is caused by electrical conductivity change in the magnetic field. In metals and semiconductors, this phenomenon is long time known: electrical resistance is caused by scattering of electrons during their collision with lattice. Under magnetic field, conduction electrons should move on helical trajectories. The elongation of trajectory increases the number of collisions and hence resistance increases. However, in *ordinary* metals, the effect of magnetoresistance is small: *increase* in resistance occurs only on part of percent (that is why this effect in ordinary metals is rarely used in practice). Nevertheless, this *positive* magnetoresistance effect is more noticeable in strong magnetic fields and at low temperatures, when electron trajectory is characterized by much bigger free path.

In *ferromagnetic* materials, magnetoresistance effect is *negative* and reaches already several percent. The point is that in the absence of an external magnetic field, a ferromagnetic divides on magnetic domains, in which magnetic moments have different orientation; domain boundaries lead to additional scattering of conductive electrons. Under external magnetic field, the influence of domain boundaries disappears; hence the entire sample becomes close to a single domain that is completely magnetized, and its resistance *decreases*. It is noteworthy also that electrical resistance of magnetic materials depends on the angle between magnetic field and current. This phenomenon is *anisotropic magnetoresistance*. This effect, despite relatively small size, is used in some devices to measure magnetic field in automation systems and in alarm information devices.

Recently, the application of magnetoresistance effect becomes wider owing to *GMR* discovery. It is seen in materials created artificially by the deposition on a substrate some alternating ferromagnetic and nonferromagnetic layers of nanometer thickness. The scheme of such layered structure with magnetization vector direction in layers is shown in Fig. 6.30A.

The GMR effect was first observed in the films with alternating layers of iron and chromium, but later many other combinations of layers are discovered. In films composed of cobalt and copper layers, magnetoresistance is much larger than that in the Fe-Cr films. The simplest device may consist of two ferromagnetic layers placed parallel to each other, in which electrical resistance depends on the relative orientation of spins in magnetic layers. If magnetic moments in ferromagnetic layers are found as parallel, the device has *smaller* resistance; if magnetic moments are found as antiparallel, the resistance *increases* greatly. Electrical current can flow both perpendicular and parallel to segments. In both cases, the change in resistance is sufficiently big (~40%).

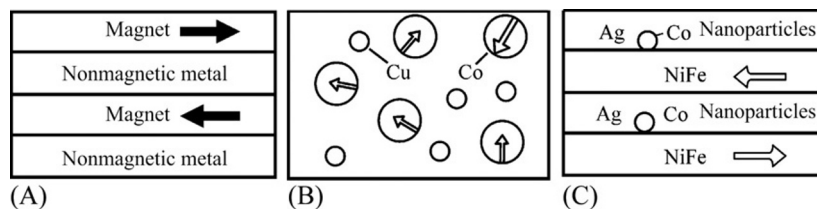


FIG. 6.30

Three structures that show giant magnetoresistance: (A) layers of nonmagnetic material with ferromagnetic layers, magnetized in opposite directions; (B) randomly oriented ferromagnetic cobalt nanoparticles (*large circles*) in nonmagnetic copper matrix (*small circles*); and (C) mixed system consisting of silver layers with cobalt nanoparticles and magnetic layers of alloy Ni-Fe.

The effect of constant magnetic field on the resistance of multilayer iron–chromium system is shown in Fig. 6.31A. The degree of resistance change depends on the thickness of iron layers and reaches the maximum at thickness of 7 nm, as shown in Fig. 6.31B.

This effect occurs from electron scattering dependence on the direction of their spin relatively magnetization vector. The electrons whose spins are directed opposite to that of magnetization B scatter more intense than electrons whose spins are directed equally with B . Application of constant magnetic field along layers orients magnetization vector of all layers in one direction. Conduction electrons whose spins are directed opposite to magnetization are scattered on metal–ferromagnetic boundary more strongly than electrons whose spins are oriented in the direction of magnetization. As both channels operate in parallel, the channel with less resistance determines the impedance of a film.

The effect of magnetoresistance in layered materials is used in sensitive detectors of magnetic field, and this effect is the basis for creation of new highly sensitive

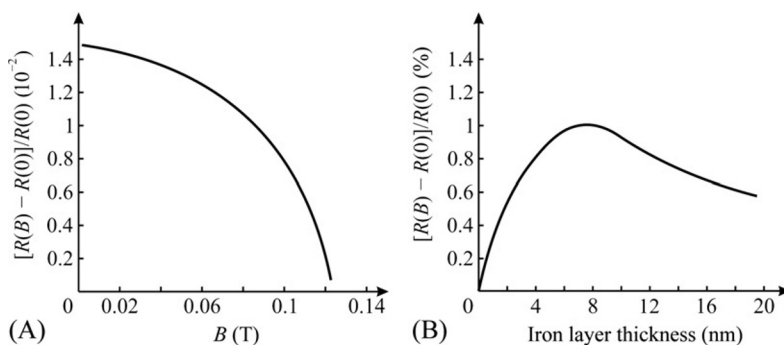


FIG. 6.31

Magnetoresistance in Fe-Cr multilayered structure: (A) magnetic field applied parallel to surface layers; (B) dependence on magnetic layer thickness.

magnetic head (capping for disk) that reads information. Until this effect discovery in magnetic storage devices, the induction coil was used for operation with magnetic small cells: such as in recording mode and for information reading. Giant magnetoresistive reading head is much more sensitive than the induction one.

Compound materials, consisting of single-domain ferromagnetic nanoparticles with randomly oriented magnetization, being placed in the nonmagnetic matrix also show GMR. A scheme of such system is shown in Fig. 6.30B. Unlike layered structures, magnetoresistance in this system is isotropic. Magnetization vectors of ferromagnetic nanoparticles are oriented in the magnetic field, which *reduces* electrical resistance. The influence of magnetic field on resistance increases with increase in field strength and with decrease in size of magnetic particles. Typical measurement results with film consisting cobalt nanoparticles in the copper matrix are shown in Fig. 6.32A. A hybrid system, composed of metallic nanoparticles in matrix, placed between two ferromagnetic layers (Fig. 6.30B) demonstrates similar properties.

The GMR effect is used in the heads of hard disks. On the basis of such magnetic structures, many sensors, switches, and nonreciprocal devices are elaborated. Low cost and low-power consumption promote high competitive ability of these devices. Magnetic storage devices based on GMR devices can compete with conventional semiconductor storage devices by the integration density, speed, and cost.

Some materials have even much higher magnetoresistance effect, as shown in Fig. 6.31, and this phenomenon is called the *colossal magnetoresistance*. Corresponding materials also have many opportunities for use, for example, in magnetic recording heads or sensitive magnetometer elements. These materials have perovskite structure, such as LaMnO_3 , where manganese, similar to lanthanum, has a valence of “+3.” If La^{3+} ions are partially replaced by bivalent ions such as Ca, Ba, Sr, Pb, or Cd, then, in accordance to electroneutrality law, some manganese ions should change their state from Mn^{3+} to Mn^{4+} . The result is a system with *mixed*

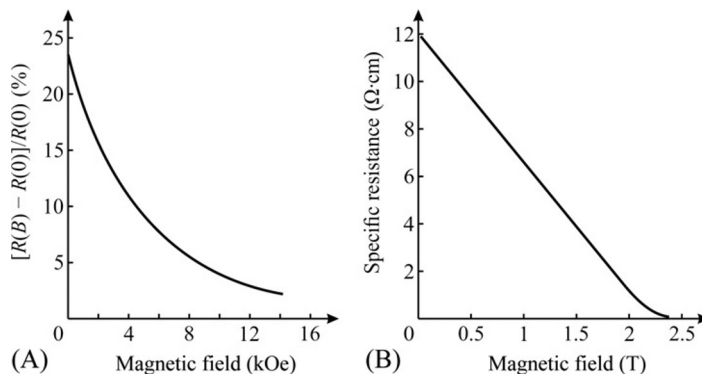


FIG. 6.32

Dependence of magnetoresistance change on applied magnetic field: (A) for thin film of cobalt nanoparticles in copper matrix; (B) for La-Ca-Mn-O near the Curie point (250 K).

valences of $\text{Mn}^{3+}/\text{Mn}^{4+}$, in which large number of mobile charge carriers exists. It is found that this system shows very large magnetoresistance. For example, resistance of the $\text{La}_{0.67}\text{C}_{0.33}\text{MnO}_x$ system at magnetic inductance of 6 T can be changed in hundreds of times. The dependence of the resistivity of thin film made up of this material on the applied magnetic field is shown in Fig. 6.32B.

Tunneling magnetoresistance is the effect of spin-dependent electron tunneling through the nanometer layer of dielectric or semiconductor, located between two ferromagnets. This structure has a construction similar to that as shown in Fig. 6.30B, but instead of the copper layer, a dielectric layer (Al_2O_3) or semiconductor layer is used. As in the case of giant magnetoresistance, electrons show tunneling, thus creating current from one ferromagnetic to another, if they have parallel magnetization.

Because the magnetization of ferromagnetics is antiparallel, probability of tunneling greatly reduces, and hence current through a structure decreases sharply owing to significant increase in resistance. At room temperature, the change in resistance is around 30%, which enables to apply this effect in devices. As in the case of giant magnetoresistance, soft and hard ferromagnetics are used. The state of magnetization is stored up to new magnetic switching; that is why, the switching can be used as a transfer of bit of information in electronic memory. New computer memories are developed based on the tunnel magnetoresistance effect. Such devices use very small currents; hence they have low power consumption. The imperfection of these devices is that currents are directed perpendicular to layers. As a result, decrease in the area of layers increases electrical resistance of device.

Development of spintronics significantly increases the operation speed and the density of processed information.

6.6 SUMMARY

1. In addition to electricity, magnetism is the manifestation of electromagnetic interaction. This interaction becomes apparent as moving electrical charges *influence* each other at a *distance* from the magnetic field. Microscopic sources of electrical field are electrical charges (electrons or protons). Microscopic sources of magnetic field are orbital and spin magnetic moments of elementary particles, atoms, and molecules. At the macroscopic scale, magnetic field is created by electrical current or permanent magnets.
2. Classic statistical physics proclaims that electronic systems cannot have thermodynamically stable magnetic moment, but this assertion contradicts experiments. Quantum mechanics, which explains the stability of atom, account for magnetism in atoms and in macroscopic bodies. In atoms and molecules, magnetism is caused by the following:
 - spin magnetic moments of electrons (*spin* magnetism);
 - moving electrons in shells of ions and atoms (*orbital* magnetism);
 - spin magnetism of some nucleons (*nuclear* magnetism).

Nuclear magnetism is very small compared to electronic magnetism, but it is still used in instrumentation as a method to study matter by the *nuclear magnetic resonance* method and to obtain *very deep cooling* by the nuclear demagnetization method.

3. All substances, but in varying degrees, respond to external magnetic fields; hence they can be characterized by a certain *magnetic susceptibility*. However, usually only those substances are called as magnetics, in which ions or atoms without any external magnetic field influence have *unpaired* electronic spins. Existence of noncompensated spin magnetic moments in some atoms or ions is caused by *partially filled 3d- or 4f-shells*.
4. Interaction of orbital and spin magnetic moments in atoms with many electrons follows from the law of space quantization; the resulting magnetic moment M_j is determined by the total angular quantum number j and equals to $M_j = g_j [(j + 1)\mu_B]^{1/2}$, where g_j is the Lande factor (factor of magnetic splitting) and μ_B is the Bohr magneton—unit of magnetic moment, caused by electronic spin: $\mu_B = e\hbar^2/2m_e c$. Factor of magnetic splitting, or g -factor, describes important magneto-mechanical ratio that shows how orbital and spin moments of individual electrons are put together. In case of purely orbital moment $g = 1$, whereas for pure spin moment $g = 2$.
5. Magnetic moment of atom in a crystal may differ significantly from the magnetic moment of the same atom that is in a free state because of spin-to-spin or spin-to-orbital interactions. These differences are particularly large for third group of d -metals (iron group), in which $3d$ -shell has larger radius. Interaction of magnetic electron in shells with their surroundings not only affects the size of atomic magnetic moment but also causes the exchange relationships between all magnetic atoms of a crystal. However, magnetic moment of the *rare-earth element* crystals is determined by the deep-laid $4f$ -electrons; hence it is approximately equal to the magnetic moment of a free atom.
6. The energy of magnetic interactions between atoms in crystals can be estimated through the *ferromagnetic Curie point* or the *antiferromagnetic Neel point*, above which magnetic ordering becomes thermally destroyed and the crystal turns into a disordered paramagnetic state. Typically, these temperatures do not exceed 1000 K, and therefore magnetic interactions are much smaller than electrical interactions in crystals.
7. According to the nature of interaction with magnetic field and the nature of internal structure, all magnetic materials can be divided into several types.

The disordered magnetic materials:

 - *diamagnetics*, magnetically neutral materials (in which atoms and molecules have no intrinsic magnetic moment); magnetic behavior of the material is determined by law of Faraday electromagnetic induction, whereby

molecular currents in matter vary to compensate change in magnetic flux through matter;

- *paramagnetics*, in which particles have their own nonzero magnetic moments that can be oriented along the applied magnetic field.

The substances with magnetic ordering of different types:

- *ferromagnetics*, in which owing to exchange interaction, the parallel orientation of magnetic moments of atoms or molecules exists, which is energetically favorable in macroscopic areas (domains);
 - *antiferromagnetics*, in which exchange interaction is such that two or more sublattices of crystal are oriented antiparallel, which in sum shows zero magnetization;
 - *ferrimagnetics*, which in contrast to antiferromagnetics cannot reach full compensation of magnetic moments of sublattices; hence these materials, generally, have nonzero spontaneous magnetization.
 - substances with *special magnetic ordering*: spin glasses, superparamagnetic ensembles of particles, molecular magnets and clusters, plasma, elementary particles (in solid-state physics, magnetic properties of plasma and elementary particles are not considered).
8. For examples of magnetically *disordered* structures, the electronic and nuclear paramagnetism in crystals can be considered. Magnetic *ordering* of different degrees is seen in ferro-, antiferro-, and ferrimagnetics including nanoparticles. At sufficiently high temperatures (when thermal motion in crystals obstructs magnetically ordered structure), any substance becomes either diamagnetic or paramagnetic.
9. *Diamagnetism* is explained by the precession of electronic orbits in atoms, ions, and molecules, and therefore it is inherent to *all substances*. Diamagnetic susceptibility of materials is *negative* and *small* ($\alpha \sim -(10^{-6}-10^{-5})$) and almost *independent of temperature*. Most *elements* listed in the Periodic System of Mendeleev have nonzero magnetic moment (e.g., *atoms* of sodium and chlorine are paramagnetics). However, most *crystals* consist of ions or molecules. That is why the molecular crystal Cl_2 and the ionic crystal NaCl are diamagnetics. One reason for molecule and crystal formation is that their energy *reduces* through self-organization of “completely magnetically compensated” electronic shells with zero magnetic moment. Thus in nature, there are much more diamagnetics than it would be expected from consideration of electronic shells of atoms in the Mendeleev Periodic System.
10. *Paramagnetic* susceptibility is *positive* ($\alpha \sim +(10^{-6}-10^{-1})$); typically, it characterizes atoms and ions that have internal permanent magnetic moment. This susceptibility depends strongly on temperature, usually, by Curie law: $\alpha \sim K/T$. However, for most metals, their paramagnetic susceptibility is originated by “free electronic gas,” equals approximately $\alpha \sim +(10^{-6}-10^{-5})$ and does not depend on temperature that is explained by Pauli mechanism of paramagnetism.

11. In most *dielectrics* that practically have no conductive electrons, the magnetic moments of electronic shells of ions, atoms, or molecules are totally compensated (magnetic moments, sometimes, might have some nucleus, but nuclear paramagnetism is extremely small). Therefore dielectrics usually are the diamagnetics. Only some dielectrics whose atoms or ions have uncompensated spins in the $3d$ - or $4f$ -electronic shells show Langevin-Curie paramagnetism.
12. In most *metals*, diamagnetism or paramagnetism dominates. Paramagnetism of metals might have spin nature, inherent to conduction electrons (Pauli paramagnetism). Owing to large Fermi energy of electrons in metals, the paramagnetism of conduction electrons practically is independent on temperature. Therefore for example, in alkali and alkaline earth metals, in which electronic shells of *ions* have no magnetic moment, the paramagnetism is solely due to conduction electrons and characterized by positive (paramagnetic) susceptibility that is independent of temperature.
13. Electrons in metals can also show diamagnetic effect (Landau's diamagnetism), as the movement of electrons in magnetic field is quantized: when external magnetic field is absent ($H=0$), electrons in metal have no discrete stationary states, but these states occur at $H \neq 0$. The fact that under the influence of Lorentz force, moving electron revolves around H with cyclotron frequency $\omega_c = eH/m_e c$; this complicated movement can be represented by oscillator whose frequency can take only discrete values. This effect, called *Landau's diamagnetism*, often makes small negative contribution to magnetic susceptibility of metals.
14. Most *nondoped semiconductors* are diamagnetics. Paramagnetic susceptibility of *doped semiconductors* may be caused by conduction electrons; in the simplest case, such magnetic susceptibility exponentially depends on temperature: $\chi = AT^{1/2} \exp(-\Delta E/2k_B T)$, where A is the constant of a given substance and ΔE is the band gap of semiconductor. However, semiconductor structure specifics can greatly alter this general relationship. For example, at low temperatures, semiconductors, typically, are diamagnetics, but at high temperatures, they can manifest paramagnetism of electronic gas, which exceeds diamagnetic contribution to susceptibility.
15. Magnetic crystals and polycrystals with ordered spin and orbital magnetic moments show "*strong*" magnetism—in sense that their permeability can be large, and they can be a source of strong magnetic fields, which is widely used in engineering. The degree of magnetization of these crystals is determined by total magnetic moment, which is the vector sum of atomic magnetic moments. Proper magnetic moment has atoms of transition group of periodic table of elements because they are characterized by noncompletely filled inner electronic shells, which are available to hold unpaired electron spin.
16. In simple *ferromagnetic* structures, all magnetic moments of atoms are directed equally. Examples of such metals are the ferromagnetics Fe, Ni, Co, Gd, and Dy.

These magnetically ordered metals behave similar to permanent magnets (in case of single-domain structure). The simplest ordered *antiferromagnetic* structures are also collinear, but their magnetic moments are directed oppositely, and they are totally self-compensated. The axis, on which all these moments are directed, is called antiferromagnetic axis. In collinear *ferrimagnetic* structure, the neighboring atoms also show antiparallel orientation of magnetic moments, but the total magnetic moment of elementary cell of crystal is nonzero. Hence these structures have *spontaneous magnetization*, inasmuch as magnetic moments of ions in different sublattices are noncompensated.

17. The carriers of ferromagnetism are uncompensated electronic spins, associated with electronic orbital moments in ions. In both cases, the electrons have both spin and orbital moments. However, calculations show that only spin-type magnetic interaction are *not able to provide parallel orientation of spins*, which is the main characteristic of ferromagnetic at temperatures below the Curie point. It is assumed (by Weiss) that stable orientation of spins can be caused by a *molecular field* that has nonmagnetic nature. Forces that coordinate magnetic moment orientation of ions have *electrostatic nature*. They occur as a result of spin and orbital *exchange interaction* of electrons.
18. While cooling from high temperatures (i.e., cooling from disordered paramagnetic phase), permeability of a ferromagnetic increases and reaches the maximum at the Curie temperature T_C . In paramagnetic phase, above the phase transition point, *Curie-Weiss law* can be implemented: $\chi \approx \mu = C(T - \theta)$, where C is the Curie-Weiss constant and θ is the Curie-Weiss temperature (the latter is slightly different from the phase transition temperature T_C).
19. Sharp maximum of *heat capacity* is observed at the Curie temperature of ferromagnetic; it is caused by excess energy necessary for magnetic moments disordering. Moreover, in behavior of heat capacity of ferromagnetic, another significant anomaly is seen: pronounced increase in heat capacity in the ferromagnetic phase (differing from the smooth curve of saturation, observed in nonmagnetic metals). Thus, the spin ordering is inherent to ferromagnetic, and for its destruction, it is necessary to add energy throughout the temperature range.
20. It is seen that the magnetic moment of bulk ferromagnetic materials at temperatures below the Curie point is much lower than its theoretical determination that can be defined for the case, when all atomic moments are directed equally. This is due to the formation of *domains*: regions, in which all magnetic moments of atoms are directed equally; hence in each domain, magnetization corresponds to saturation, that is, it takes the maximum value. However, in different domains of magnetic crystals (or polycrystal), vectors of magnetization are not parallel to each other. Thus total magnetization of ferromagnetic sample is lower than in case of complete ordering of atomic magnetic moments. Therefore without external field, ferromagnetic crystal is

composed of many small individual plots, magnetized to saturation—domains. Domains are separated by layers—domain walls—in which spins gradually change orientation, inherent in one domain, to orientation, inherent in the neighboring domain.

21. At reversal operation (changing the direction of external field H), magnetic moment M first increases to its maximum—to *spontaneous* magnetization M_s . With decrease in external field, magnetization remains behind; hence if magnetic field again becomes zero ($H=0$), the induction is not zero, but its value gains to *residual* value M_r . Phenomenon of magnetization lateness while magnetic field changes is the *magnetic hysteresis*. For residual magnetism disappearance, it is necessary to apply the counter field H_c that can reverse magnetization of ferromagnetic. This field is the *coercive field* (retentive force). Depending on the shape and area of hysteresis loop, ferromagnetic materials are divided into the “soft” (low coercivity) and the “hard” (high coercivity).
22. Magnetization of ferromagnetic materials is accompanied by changes in the size and shape of magnetic sample. This phenomenon is called *magnetostriction*. The reason for this effect (which is widely used in engineering) is large spin-orbital coupling in ferromagnetic materials. Change in magnetic properties in case of ferromagnetic deformation is observed experimentally, and it is called the *magnetoelastic effect*. Some ferromagnetic materials are quite sensitive to internal stresses that this property is used for strain and tension measurement.
23. Magnetization makes essential influence on ferromagnetic deformation—the *magnetostriction*. Conditioned by exchange interaction, it depends not only on the applied magnetic field but also on temperature change (without any external field). The thermally induced magnetostriction (sometimes called as *thermostriiction*) is the spontaneous effect (as it occurs when an external field is not applied), and it is the greatest in the vicinity of the Curie point, that is, when transition to magnetically ordered phase occurs. Some ferromagnetic materials assume the name *invar alloys*: in them, *negative* (ferromagnetic) deposit to thermal expansion (α_f) compensates typical for all crystals *positive* (anharmonic) thermal expansion coefficient (α_a); hence total coefficient can be practically zero ($\alpha = \alpha_a + \alpha_f \approx 0$).
24. The *magnetocaloric effect* involves change in material temperature during magnetic *adiabatic* magnetization or demagnetization. Under adiabatic condition (when there is no heat energy exchange with the environment), a magnet does not absorb or return heat ($dQ=0$), and therefore its entropy S does not change: $dS = dQ/T = 0$. Recently, the *giant magnetocaloric effect* was discovered in intermetallic compounds based on rare-earth elements, for instance, in the silicide-germanide system: $\text{Gd}_5(\text{Ge-Si})_4$. This effect provides application of magnetocaloric cooling. Another effect of adiabatic demagnetization of *paramagnetics* is used for achieving extremely low temperatures.

25. The *antiferromagnetic interaction* occurs in case of negative sign of exchange integral; hence antiparallel orientation of spins in lattice cells of crystal is energetically more profitable. Spin locations are ordered, but no spontaneous magnetization occurs because neighboring moments are directed antiparallel and cancel each other. In such a crystal, two (or more) magnetically opposite sublattices are interpenetrated.
26. The structure with antiparallel arrangement of spins is formed lower than the temperature called the *Neel temperature* (T_N), when spin interaction surpasses chaotic thermal motion. If the crystal is heated above this temperature, the uncompensated spins form a kind of paramagnetic system that is characterized by a very special temperature dependence of magnetic susceptibility: $\alpha = C(T + \theta)$, where C is the Curie-Weiss constant, θ is the characteristic temperature, which in contrast to paramagnetic phase of the ferromagnetic is located in the *negative part* of Kelvin temperature scale.
27. In addition to totally magnetically compensated antiferromagnetics, there are many crystals and polycrystals in which magnetic moments of sublattices (although being directed opposite to each other) have significant difference in their size—the *ferrimagnetics*. They have complicated structures with varying nature of atom location that forms some uncompensated electrons in $3d$ - or $4f$ -shells. Ferrimagnetics have properties similar to those of ferromagnetic materials because they have spontaneous magnetization owing to total magnetic moment of sublattices is nonzero.
28. The *nanophysics* represents scientific direction in the field of material sciences that at present is one of the most promising fields. It dedicates the creation and study of structures and properties of materials, condensed in the form of very small crystals, clusters, and fragments that have around 10^3 – 10^5 atoms. The main reason for differences between nanomaterials and customary materials is that in these substances the *ratio of surface to volume is rather big*. The smaller the size of nanocluster, the more the surface properties dominate over bulk properties. In some sense, the nanostructures transform properties of crystal surface into volumetric properties.
29. The *nanotechnology* is a scientific and technical direction for creation of materials, devices, and functional structures of nanometer size. Only because of small size of units (particles, granules, and phases), the nanomaterials exhibit unique mechanical, optical, electrical, and magnetic properties. The nanostructured *magnetic materials* can operate with a wide range of characteristics. Moreover, nanotechnology can be used to create materials with the prescribed type of magnetization curve—both for record magnetically soft materials and for extremely magnetically hard materials.
30. Such magnetic materials that show great ability to magnetization and, at the same time, the lack of hysteresis are *superparamagnetics*. The essence of this term is that for temperatures below the Curie point in a wide temperature range, they exist as if they were in the paramagnetic phase.

31. The *magnetoresistance* is the effect of electrical conductivity change in solids while placing them in the magnetic field. Multilayered structures composed of layers of nonmagnetic material alternating between oppositely magnetized ferromagnetic materials show significant change in their resistance when putting them in the magnetic field. This phenomenon is called as the effect of *giant magnetoresistance* (GMR). This effect can be either *longitudinal*, when electrical current flows in plane of layers, or *transversal*, if current is perpendicular to layers.
32. Recently, new scientific and technology direction is under developing—magnetolectronics, or as it is now called—*spintronics*, which deals with the study and practical application of such effects in devices that use electronic spin. Spintronics studies magnetic and magneto-optical interactions in metallic and semiconductor structures, as well as quantum phenomena in magnetic structures of nanometer size.
33. Electrical current, passing through magnetic crossing, under certain conditions can be accompanied by the transfer of polarized spins, thus leading, in particular, to switching of magnetization in layered nanoscale structures. Theoretical examination of this phenomenon is based on the conception of exchange interaction between electrons—carriers—and magnetic lattices in ferro-, ferri-, or antiferromagnetics.
34. Magnetic materials are widely used in electrical engineering and electronics, but in recent years of rapid development of information technology, they acquire even greater importance. Magnetics are the main working bodies in magnetoacoustic, microwave technologies, magneto-optics, and magnetolectronics (spintronics). Improvements in corresponding working elements are based on the main principle of magnetism; this supported the rapid growth of microelectronic and, particularly, nanoelectronic technologies.

REFERENCES

- [1] S.V. Vonsovsky, *Magnetism*, Nauka, Moscow, 1971.
- [2] R. Waser (Ed.), *Nanoelectronics and Information Technology: Advanced Electronic Materials and Novel Devices*, Wiley-VCH, Weinheim, 2005.
- [3] Y.M. Poplavko, S.A. Voronov, Y.I. Yakimenko, *Materials Sciences Part 3, Conductors and Magnetics*, Kiev Polytechnic Institute, Kiev, 2011.
- [4] M.I. Kaganov, V.M. Tsukernik, *Nature of Magnetism*, Nauka, Moscow, 1982.
- [5] Y.M. Poplavko, A.V. Borisov, *Magnetolectronics*, Kiev Polytechnic Institute, Kiev, 2016.
- [6] N.W. Ashcroft, N.D. Mermin, *Solid State Physics*, Holt and Winston, New York, 1976.
- [7] C. Kittel, *Introduction to Solid State Physics*, John Wiley and Sons, New York, 1976.
- [8] M.I. Kaganov, *Electrons, Phonons, Magnons*, Nauka, Moscow, 1979.
- [9] L.H. Van Vlack, *Elements of Materials Science and Engineering*, sixth ed., Addison-Wesley Publishing Co., Reading, MA, 1989.
- [10] H.S. Nalva (Ed.), *Nanostructured Materials and Nanotechnology*, Academic Press, New York, 2002.

Dielectrics

CONTENTS

7.1 Main Features of Dielectrics	287
7.2 Macroscopic Description of Polarization	294
7.3 Different Mechanisms of Polarization	301
7.4 Optical and Far-Infrared Polarizations	311
7.5 Thermally Activated Polarizations	319
7.6 Clausius-Mosotti-Lorentz Equation	331
7.7 Dynamics of Electrical Polarization	336
7.8 Dielectric Losses and Dielectric Spectroscopy	349
7.9 Electroconductivity in Dielectrics	360
7.10 Electrical Breakdown	383
7.11 Summary	399
References	408

Dielectrics that include a majority of crystals, polycrystals, many amorphous materials, many liquids, and all gases might have quite different structures. The main physical property that unites such different substances is their special behavior in the electrical field; namely, in dielectrics the local displacement of electrical charge occurs—*polarization*—and, at the same time, DC charge transference through dielectric (electrical conductivity) is practically absent. Properties of gases, liquids, crystals, and amorphous substances are so different that these materials are studied in different fields of physics; therefore, it is not so easy to establish any common and consistent approach in the physics of dielectrics. For this reason, it is necessary to identify important features that characterize the substance as a dielectric.

7.1 MAIN FEATURES OF DIELECTRICS

It is known that physical properties of any substance can be divided into three main classes: mechanical, thermal, and electrical (magnetic properties usually do not play significant part in dielectrics, while their optical properties are determined by electrical properties). The *mechanical* properties, reflecting internal bonds between atoms, which include the elasticity, strength, and hardness, were described earlier in [Chapter 2](#). The *thermal* properties, conditioned by the internal energy of

molecular, atomic, and electronic structures, are characterized by thermal expansion, heat capacity, and thermal conductivity (described in Chapter 3).

The *electrical* properties, caused by forced movement of electrical charges, are electrical polarization, electrical conductivity, absorption of electrical energy, and electrical breakdown. This chapter is dedicated exclusively to the description of the electrical properties of dielectrics.

Therefore, in physics of dielectrics, their electrical properties are positioned at the first place. Mechanical and thermal properties in different classes of dielectrics are considered only if they relate to electrical properties or to technical applications of given dielectric. Regarding electrical properties of dielectrics, in addition to *polarization* and the *electrical conduction*, the *dielectric losses* (absorption of electrical energy) deserve an in-depth study. Next important points are physics of *electrical aging* (i.e., electrical property variation over time in the increased external electric field), as well as the nature of the *electrical strength* (i.e., insulating property preservation in strong electrical field), and the *electrical breakdown*, when insulator turns into a conductive material.

It should be noted that besides the proper electrical properties listed, in contemporary physics and technology of dielectrics several phenomena are emphasized that concern the *interdependence* of electrical, mechanical, and thermal properties of a dielectric [1]. Typically, such relationship occurs in the noncentrally symmetrical crystals and textures, as well as in the liquid crystals. For example, strong interdependence of electrical and mechanical properties is most evident as *piezoelectricity*, while the tie of electrical and thermal properties is manifested as *pyroelectricity*. Electro-optical phenomena and photoelectric effects characterize the relationship of electrical and optical properties of dielectrics; interdependence of electrical and magnetic properties in certain crystals is associated with some electromagnetic phenomena, and so on. Thus, physical phenomena in dielectrics might be quite complicated and different [2].

It is possible to classify various electrical, thermal, and mechanical properties of dielectrics as *reversible* and *irreversible* [3]. In the case of reversible properties, the initial state of a matter can be almost completely restored after the removal of external electric field. Electrical polarization could be mentioned as an example of reversible electrical properties, while an example of thermal reversible properties is heat capacity; in case of mechanical properties, an example of reversible property is elastic deformation.

In case of irreversible phenomena, the consequence of impact remains even after eliminating of cause. Considering electrical properties of dielectrics, the irreversible properties are dielectric losses, electrical aging, and electrical breakdown. The irreversible thermal phenomena are melting or evaporation after material heating; the mechanical irreversible phenomena include destruction of material under stretching or compression.

In connection with the investigation of dielectric properties, it is also useful to distinguish the *transport phenomena* in dielectrics, which occur when the charge, the energy, or some part of material is transferred. For example, electrical

conductivity is electrical charge transfer, while thermal conductivity is thermal energy transportation. Plastic deformation is another example of irreversible mechanical transfer of substance.

Among the wide variety of different properties of dielectrics, it is necessary to highlight their main property that determines a special behavior of dielectrics under both electrical field and other external influences. This property is *electrical polarization*. For better understanding of this phenomenon, electrical polarization should be compared with electrical conductivity, because both arise simultaneously when external electrical field is applied.

From *macroscopic* representations of these phenomena, it should be noted that during charge transfer electrical current varies *in phase* with applied electrical field, and this current exists all time as long as electrical field is applied (for both alternating and direct (constant) voltage). At the same time, the charge separation (polarization) results in a bias current *differing in phase by $\pi/2$* from applied sinusoidal electrical voltage. In the event that voltage is switched, the bias current exists only at the time when electrical field is changing, and it is absent in the case of direct voltage [4].

Continuing to discuss the differences of these two phenomena, it should be first noted that conductivity relates to transport phenomena, while polarization (with rare exceptions) is reversible property. It is also important to note that the internal structure of dielectrics is insignificant for free charge carrier movement. However, when external electrical field is applied to dielectric, a small displacement of bounded electric charges prevails and *symmetry* of dielectric structure can be changed (due to electrical polarization). At last, *all particles* of dielectric participate in the polarization, changing their mutual position (usually they are shifted by a very short distance), while manifestation of electrical conductivity is determined by only *a few particles* (that are relatively free, and, therefore, capable of moving through the dielectric).

The *electrical polarization* (i.e., charge separation in the electrical field) is the main property of dielectrics. To describe polarization, the vector \mathbf{P} is introduced, which is numerically equal to electrical moment per unit volume of dielectric, and directed in accordance with applied field \mathbf{E} direction and *symmetry* of dielectric. In isotropic dielectric, vectors \mathbf{P} and \mathbf{E} are collinear, but in anisotropic dielectric, directions of these vectors may be different.

Macroscopic electrical field, arising in dielectric between a pair of same electrodes is smaller than that in vacuum, because it is determined only by the “free” part of electrical charge located on the electrodes. This charge is partly compensated by the polarization, which binds the part of total electric charge located on electrodes. For this reason, the electric induction vector \mathbf{D} needs to be introduced, which characterizes the *total charge* on the electrodes (similar to vector \mathbf{E}_{vac} describing electrical field in vacuum capacitor). Vectors \mathbf{P} , \mathbf{D} , and \mathbf{E} are constrained by simple relation, $\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}$, which can be obtained in macroscopic electrodynamics from Maxwell’s equations. In case of dielectrics, these equations are simplified by the assumption that in dielectrics relative magnetic permeability may be considered $\mu = 1$ as in vacuum [5].

The polarized state of a dielectric, when it is induced by external electrical field, can be described by several mechanisms of elastic, thermal, and space-charged polarizations, in which locally bound electrons, ions, or dipoles shift in the electrical field.

The *quasielastic* (deformation type)-induced polarization is the least-inertial polarization; generally, it is slightly affected by temperature; so dielectrics with this polarization are possible to use up to very high frequencies. Electronic quasielastic polarization (deformation of electronic shells of atoms, molecules, or ions in the electrical field) is a common mechanism for *all dielectrics*. Ionic quasielastic polarization belongs to such dielectrics and semiconductors in which the ionic character of crystal lattice is markedly pronounced. Quasielastic dipole polarization is observed only in case when external electrical field induces reversible change in the direction of spontaneously oriented dipoles.

The *thermally supported* (relaxation) induced polarizations of solid dielectrics are mainly caused by the structural defects, and this polarization is essential, if electrons, dipoles, or ions are weakly bounded in the structure of dielectric. Remaining localized in the nanoscale areas, these charged particles, under the influence of thermal motion, make thermally activated jumps, moving at a distance of the order of atomic dimensions. Electrical field influences the direction of this thermal hopping, which becomes asymmetric and generates electrical moment. The inertia of thermally activated polarization is much larger in comparison with the quasielastic polarization.

The *space-charge* (migration) induced polarization is the additional mechanism of polarization, which is observed in solids with heterogeneous structure. The causes of this polarization may be a presence of layers or microregions with different conductivity (e.g., the existence of semiconductor inclusions in technical dielectrics). Conditioned by the space-charge migration polarization is the lowest-frequency (highest inertia) mechanism.

The *relative permittivity* is one of the most important macroscopic characteristics; static relative permittivity is always positive: $\epsilon > 1$. Because of anisotropy, the value of permittivity may be different in various directions. This parameter depends on many conditions. As different mechanisms of polarization have different sluggishness, permittivity is dependent on frequency. In this regard, it is possible to distinguish between dynamic and static dielectric constant as well as the difference between permittivity of mechanically clamped dielectric and mechanically free dielectric (in case of piezoelectric). In polar dielectrics (pyroelectrics), there is also the difference between isothermal and adiabatic permittivity.

Thermodynamic consideration of dielectric polarization phenomenon allows calculation of the free energy of polarization, as well as change in total energy, the entropy, and other thermodynamic functions at polarization. These methods can be used to calculate the energy function for polar dielectrics, which is of interest when permittivity is not a constant, but depends on temperature and electrical field intensity.

The *electroconductivity* (i.e., charge transfer in external electrical field) increases energy of losses in the dielectric, and, moreover, can lead to electrical breakdown.

Various types of moving charged particles may exist in dielectrics. Movement of charged particles (mainly electrons and ions) over rather long distances results in conductivity. In case of external field absence, free charge carriers (being generated by thermal fluctuations) move chaotically. However, when electrical field (i.e., gradient of electric potential) is applied, or at the presence of temperature gradient on the random motion of free charges, the directional transfer is superimposed, leading to electrical current [6].

Depending on the physical nature of charge carriers, the conductivity in dielectrics might have electronic, ionic, polaronic, or molar-ionic character. The mechanisms of charge transport in electrical field can be divided into *drift* (electrons and holes), *hopping* (for polarons and ions), and *diffusion* (available for electrons, polarons, and ions). In case of ionic conductivity not only electric charge is transferred, but also a part of dielectric substance: negative charge carriers (anions) are settled and discharged at the anode, while carriers of positive charge (cations) establish their residence and discharged at the cathode. Therefore, this is a case of not only the charge but also the mass transfer occurs by electrical field.

The value of conductivity (σ) depends both on the concentration and on the mobility of charge carriers. In dielectrics with increased polarizability, on the one hand, charge carrier mobility reduces because charge carrier motion is braking due to the fact that in solid dielectrics they are surrounded by polarized nanoregions, while in liquids and gases free carriers may be joined to electrically neutral molecules forming charged complexes. On the other hand, increased polarizability (i.e., high permittivity) weakens the strength of Coulomb interaction between charged particles, thereby increasing probability of charge carrier generation that results in an increase in their concentration that, in turn, increases conductivity.

Electrical conductivity of dielectrics depends on many factors: $\sigma(T, E, \omega)$. With increasing temperature, dielectric conductivity rises exponentially; the same effect can be obtained in the strong electrical fields, because new charge carriers can be excited by field. At relatively low frequencies (10^{-2} – 10^6 Hz), a significant contribution to conductivity is provided by the delay of polarization, resulting in the increase in effective conductivity with frequency rise. However, at further increase in frequency (typically, in megahertz range), conductivity reduces owing to charge carrier inertia.

Two most important electrical properties of dielectrics—polarization and conductivity—are largely *interdependent*. For example, the larger the *ionic* conductivity in liquid dielectrics, the bigger is the polarizability (i.e., greater permittivity), because, according to Coulomb law, large permittivity weakens the forces of charge attraction, increasing the probability of dissociation of molecules into ions. Therefore, polar liquids, such as water and alcohols, can be attributed to the ionic semiconductors rather than insulators.

A quite different relationship of conductivity and polarization is observed in dielectric crystals that have predominantly *electronic* conductivity: higher polarizability and lower conductivity. Low concentration of free charge carriers in dielectrics leads to the fact that in dielectrics the *electrostatic field can exist* for relatively

long time. Besides, polarization gives rise to the fact that conduction electrons that appear in dielectric as a result of various reasons (but in small quantities) turn into the state of low mobility, because they polarize by its electrical field the local areas in dielectric and can move only with those areas (polarons). The mobility of charge carriers in this case is reduced by tens and hundreds of times. Consequently, even the availability of free electrons (generated in dielectric by activation of impurities) cannot provide appreciable conductivity—for the very reason that local areas of polarization hamper charge carrier movement [7].

In particular, phenomenon of polarization is possible because dielectric has a few of free electrons and they have low mobility so that they cannot shield electrical field leading to polarization. On the contrary, in conductors the electrical field is screened by the free charge carriers so that polarization in these conditions is usually impossible: in metals, for example, screening radius is almost equal to the interatomic distance.

Thus, in dielectrics, a relatively stable state exists with extremely low electronic conductivity. Nevertheless, this stability can be compromised by the heating of dielectric to high temperatures or by its external irradiation of high-intensity flows. In both cases, charge carriers are generated in very high concentrations, so dielectric-insulator can be transformed into a rather conductive material.

The sustainable nonconductive state of dielectric may be impaired by a very large electrical field, when the velocity of electrons, moving under the influence of electrical field, becomes so high that polarization cannot capture them into the state of polarons. Excited by the field fast electrons cause the *impact ionization*, wherein concentration of charge carriers in dielectric grows rapidly that ultimately leads to the breakdown—insulator turns into conductor.

It should be noted that in certain dielectrics high electrical resistance of nonconductive state may be broken even in a weak electrical field and without excessive heating or irradiation. In such substances, even a relatively small change in external conditions—temperature, pressure, magnetic, or electrical fields—results in the spasmodic (5 or even 10 orders of magnitude) increase in conductivity with the conversion of insulator to conductor. This transformation, in contrast to electrical breakdown, is *reversible*. Such exceptional phenomena are observed experimentally, for example, in vanadium oxides and in other transition metals oxides, as well as in quasi-1D conductive polymeric compounds (such as TTF–TCNQ). This phenomenon is known as phase transitions of “insulator-metal” type [8].

It should be noted also that a sharp (stepwise) increase in electronic conductivity in dielectrics and wide-gap semiconductors sometimes is observed under increased electrical field owing to “currents limited by space charge.” In contrast to breakdown, these phenomena are also reversible.

Under normal conditions, conductivity of dielectrics usually is less than 10^{-10} S/m; for this reason, the electrostatic field can exist in dielectric for a long time. At alternating electrical field, reactive current usually greatly exceeds active current. In addition, dielectrics are the only possible medium for electromagnetic field propagation over long distances.

The *dielectric losses* (absorption of energy) represent a portion of electrical energy that in the *alternating* electric field is converted into heat. Quantitative description of dielectric losses includes several parameters: dielectric loss tangent ($\tan\delta$), the loss factor ($\epsilon'' = \epsilon' \cdot \tan\delta$), and the specific power of losses (p). The frequency dependence of dielectric absorption is convenient to describe by a complex permittivity: $\epsilon^* = \epsilon' - i\epsilon''$. Frequency and temperature dependence of dielectric losses are determined by the peculiarities of physical mechanisms that describe dispersion of electrical energy in dielectric. The most important of these mechanisms are: (1) delay of migration mechanisms of polarization (below frequencies 10^3 Hz), (2) delay of thermally activated polarization in the frequency range of 10^3 – 10^8 Hz, and (3) damping of quasielastic polarization (at THz frequencies).

The delay of polarization results in the $\epsilon'(\omega)$ decrease with frequency increase, accompanied by the maximum of loss factor $\epsilon''(\omega)$. In case of thermally induced polarization, the dependence of $\epsilon^*(\omega)$ is described by Debye relaxation equation while for quasielastic polarization $\epsilon^*(\omega)$ dependence is described by resonant Drude-Lorentz equation. Dispersion of dielectric constant is *interdependent* change with frequency of real (ϵ') and imaginary (ϵ'') parts of complex permittivity (ϵ^*). The main properties of dispersion that should be satisfied by any dispersion equation $\epsilon^*(\omega)$ are the Kramers-Kronig relations. In dielectrics, in a wide range of frequencies and in different crystallographic directions, several areas of $\epsilon^*(\omega)$ dispersion are usually observed that form the *dielectric spectrum*.

The *relaxation dispersion* is characterized by a gradual decrease in $\epsilon'(\omega)$ with frequency increase over the entire range of dispersion ($\partial\epsilon'/\partial\omega < 0$). Blurred relaxation spectra can be described by different empirical formulas, as well as by a model with normally distributed relaxation oscillators. The *resonant dispersion* in all frequency range is characterized by $\partial\epsilon'/\partial\omega < 0$ (with the exclusion of narrow range of resonance). The expanded resonance spectrum may be due to increased damping of oscillator describing dispersion, as well as by distribution of many oscillators in the frequency range.

The electrical breakdown, typically, is the consequence of increased conductivity in the strong electrical fields. Upon reaching a certain threshold of electrical field (E_{br}) the current, passing through the insulator, shows a dramatic increase, then, electrical discharge occurs, which results in the mechanical destruction of dielectric. The main physical mechanism of breakdown is the percussive ionization of dielectric by acceleration in field electrons, which results in a sharp increase in charge carrier concentration with electronic avalanche.

Electronic breakdown is characterized by a very rapid evolution, when dielectric loses its electrical strength. The channels of electronic breakdown in crystals are straight and oriented according to crystallographic axes. The value of E_{br} in crystals exhibits anisotropy. As a rule, electronic avalanche initiates a streamer that spreads with *photoionization* processes. In case of small thickness of dielectric, the electronic breakdown becomes multiavalanches.

Electrothermal breakdown occurs when dielectric loses its thermal stability, owing to the heat generation in strong electrical field (due to conductivity or

dielectric losses). At that, the heating occurs so intensively that it cannot be compensated by the process of heat removal.

Electrochemical breakdown is conditioned by chemical reactions that stimulate the increase of current in the dielectric. This results in irreversible changes in insulating properties and significantly reduces dielectric strength and reliability. This breakdown is caused by the *electrical aging* (degradation) of dielectrics. In the *inorganic* dielectrics, the aging process occurs mainly in *direct* electrical field, and it is conditioned by the formation of structural defects, contributing to electric current increase. In *polymers*, the aging occurs mainly in *alternating* electrical fields owing to occurrence of partial discharges in the gas inclusions that results in the erosion of polymer surface or in the dendrite occurrence in the volume of polymeric insulation.

7.2 MACROSCOPIC DESCRIPTION OF POLARIZATION

Electrical polarization is the most important property of dielectrics. In this chapter, only polarization *induced by external electrical field* is discussed. However, in the noncentrosymmetric dielectrics polarization can be induced by *mechanical action*, while in polar dielectrics by *temperature change* (these two kinds of polarization will be discussed later in [Chapter 9](#)).

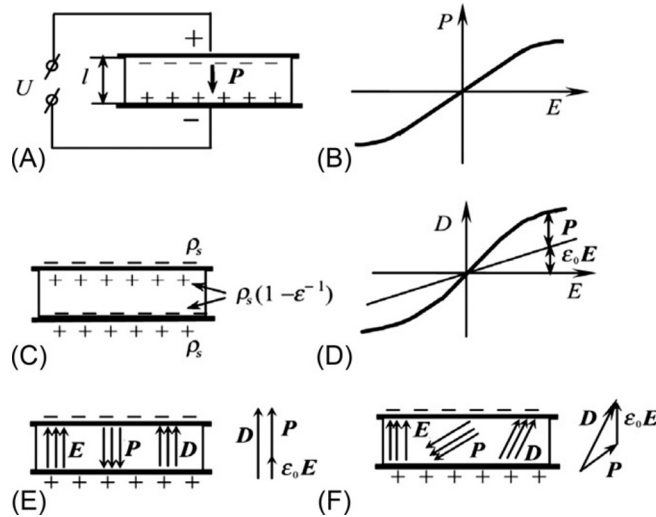
Electrical charges in dielectric structures are bound very tightly; therefore, concentration of free charge carriers that are involved in electrical conductivity is usually very small. In this connection, in further consideration of dielectric polarization, it is assumed for simplicity a complete absence of electrical conductivity: $\sigma = 0$.

Electrically induced polarization originates under the influence of external electrical field: electrically charged particles (from which dielectric is formed) are displaced from their equilibrium position, creating *induced* electrical moment: $M = \sum_{i=1}^N q_i x_i$, where N is number of charged particles in dielectric, q_i is electrical charge of i th particle, and x_i is displacement of this charge from its equilibrium position under influence of electrical field. The unit of measurement of electrical moment is “Coulomb multiplied by meter” $[M] = C \cdot m$.

Electrical field, induction, and polarization. The bulk density of electrical moment is *polarization*: $\mathbf{P} = \mathbf{M}/V$, where V is volume of polarized dielectric. The unit of measure of polarization is $[P] = C/m^2$ that corresponds to another definition of P as a surface density of bound charges near electrodes of polarized dielectric capacitor ([Fig. 7.1A](#)). Therefore, polarization quantifies the magnitude of electrical moment in dielectric and depends on the value of electrical field, as well as on structural characteristics (chemical composition) of a dielectric. Obviously, the higher the electric field strength, the larger the polarization. Unit of electrical field intensity is $[E] = V/m$.

In general, the $P(E)$ dependence may be more complicated ([Fig. 7.1B](#)), but in majority of dielectrics, if electrical field is not very large, the relationship between P and E can be regarded as linear:

$$P = \epsilon_0 \chi E, \quad (7.1)$$


FIG. 7.1

Macroscopic description of electrical polarization: (A) bound charges near electrodes of polarized dielectric capacitor; (B) polarization dependence on electric field intensity; (C) dielectric with electrodes is represented as electrical capacitor; (D) dependence of electrical induction on electric field intensity in strong electric fields; (E) vectors \mathbf{E} , \mathbf{P} and \mathbf{D} in an electrical capacitor containing isotropic dielectric; (F) vectors \mathbf{E} , \mathbf{P} , and \mathbf{D} in an electrical capacitor containing anisotropic dielectric.

where χ is *dielectric susceptibility* (dimensionless parameter). In vacuum susceptibility is absent ($\chi = 0$), while in most of dielectrics it is found as $\chi = 0.5\text{--}10$. However, in some crystals (which, by the way, are of largest interest in electronics), $\chi = 100\text{--}10,000$ and even more. At that, parameter $\epsilon_0 = 7.854 \cdot 10^{-12} \text{ F/m}$ is *electrical constant* that in SI agrees dimensions of P and E (here [F] is farad, the unit of capacitance, [F] = C/V).

Except polarization, there exists another important parameter to describe the *electrically induced polarized state* in dielectric, namely, *electrical induction*:

$$D = \epsilon_0 E + P. \quad (7.2)$$

Induction is defined by the same unit as polarization: $[D] = \text{C/m}^2$ and it is also characterized by a surface density of electrical charge on the electrodes (Fig. 7.1D). If dielectric with electrodes is represented as electrical capacitor (Fig. 7.1A), the electrical induction characterizes *total charge* on the electrodes of a capacitor: $D = \rho_s$, while polarization P characterizes only a *portion of total charge* that is related to the opposite sign charges, close-fitting to the surface of polarized dielectric. Parameter ϵ is introduced as a proportionality factor between induction and electrical field:

$$D = \epsilon \epsilon_0 E. \quad (7.3)$$

Dimensionless parameter ϵ is the *relative permittivity* (*dielectric constant*) that is related to dielectric susceptibility χ by simple equation $\epsilon = 1 + \chi$.

Dielectric constant is usually determined from electrical capacitance of capacitor. In case of *planar* capacitor C with surface area S and thickness h , permittivity is $\epsilon = \frac{Ch}{\epsilon_0 S}$. Such capacitors are most often used in industry and for investigation of dielectrics. For *cylindrical* capacitor (C) that has outer and inner diameters D and d , respectively, and axial length l , the permittivity can be computed as $\epsilon = \frac{C \ln(D/d)}{2\pi\epsilon_0 l}$. Dielectric components of *cylindrical* form are, for example, coaxial cables, piezoelectric elements, etc. If studied capacitor C has a *spherical* form (such as piezoelectric elements for sonar or spherical pyroelectric element), it is characterized by values of outer and inner radii of spheres r_1 and r_2 , and its permittivity can be calculated as $\epsilon = \frac{C(r_2 - r_1)}{4\pi\epsilon_0 r_1 r_2}$.

Dielectric nonlinearity. One type of possible nonlinear dependences of polarization on a field, $P(E)$, is shown in Fig. 7.1B; it is associated with dielectric constant dependence on electrical field, because $\epsilon = 1 + P/\epsilon_0 E$. In principle, the $\epsilon(E)$ changing should be observed in all dielectrics. However, in most of them this nonlinearity can be considered only when the dielectric is subjected to extremely large dielectric electric field, 10^7 – 10^{10} V/m, while the dielectric strength in solid dielectrics usually equals $E_{br} = 10^8$ – 10^9 V/m.

Thus in majority of dielectrics, electrical breakdown occurs earlier than they might show noticeable nonlinearity. However, in some dielectrics—ferroelectrics and paraelectrics—dielectric nonlinearity can be essential even in electrical field close to 10^5 V/m that is much smaller than electrical breakdown field E_{br} . It is necessary to notice also that *optical nonlinearity* may be seen in any dielectric that is exposed to powerful laser beam as light might be self-focusing in dielectric.

However, nowadays, with the advent of technologies, it is possible to obtain high-quality thin (nanoscale) dielectric layers, in which dielectric strength is higher by two to three orders of magnitude, so the voltage of electrical breakdown greatly increases, and nonlinearity even in conventional dielectrics acquires significance.

Nonlinearity is determined not only by chemical composition, but also by structural features of solid dielectrics, and, particularly, nonlinearity is noticeable in the vicinity of phase transitions. Because of huge variety of dielectric structures and their polarization mechanisms, there is no simple way to estimate nonlinearity in all of them, so the *universal method* might be applied to calculate nonlinearity in any type of dielectric. Therefore, following analyses of dielectric nonlinearity can be estimated only *in general*.

To describe functional dependence $\epsilon(E)$, the permittivity should be presented in a form of rapidly convergent series [3]:

$$\epsilon(E) = \epsilon + \epsilon_1 E + \epsilon_2 E^2 + \epsilon_3 E^3 + \epsilon_4 E^4 + \dots \quad (7.4)$$

In dielectrics with *centrosymmetric structure*, all coefficients with *odd* powers E are equal to zero so that dependence $\epsilon(E)$ is even. Taking into account rapid convergence of series (Eq. 7.4), parameter ϵ_2 can be linked with *differential nonlinearity* parameter N_ϵ :

$$\epsilon(E) = \epsilon + \epsilon_2 E^2; \quad N_\epsilon = \frac{1}{\epsilon} \frac{\partial \epsilon(E)}{\partial E}; \quad \epsilon_2 = \frac{1}{2E} \frac{\partial \epsilon(E)}{\partial E} = \frac{\epsilon N_\epsilon}{2E} \quad (7.5)$$

In case of *noncentrosymmetric* dielectrics, both even and odd degrees in Eq. (7.4) should be considered. However, in view of rapid convergence of series (Eq. 7.4), it is enough to consider only coefficient ε_1 with which it is possible to express the differential nonlinearity parameter:

$$\varepsilon(E) = \varepsilon + \varepsilon_1 E; \quad N_e = \frac{1}{\varepsilon} \frac{\partial \varepsilon(E)}{\partial E}; \quad \varepsilon_1 = \partial \varepsilon(E) / \partial E = \varepsilon N_e. \quad (7.6)$$

Permittivity dependence on electrical field reflects the *microscopic processes* of polarization, when dielectric polarizability depends on the intercrystalline properties (internal Lorentz field F that is associated with field E).

Therefore the nonlinearity of electrical polarization dependence on electrical field always takes place; in very strong electrical field, even vacuum is nonlinear. In the vast majority of dielectrics nonlinearity can be observed only in the electric fields, comparable (or greater) to the field of electrical breakdown. The reason for this is that internal energy of electromagnetic interactions in atoms, molecules, and crystals is much higher than the energy of possible *nondestructive* impact on dielectric. However, there are exceptions [8]:

1. Because of low inertia of electrons, ordinary breakdown at optical frequencies does not occur, because there is no enough time for electronic avalanche (reason for breakdown) increase. Therefore, in strong electrical field of giant laser pulse (when intensity of this field becomes close to internal fields of atoms), the nonlinearity of optical polarization is sufficient for observations and applications, that is a physical basis of the *nonlinear optics*.
2. Near the phase transitions in dielectrics (e.g., ferroelectric transition) the forces of interaction between nearby atoms are almost compensated; therefore external influences, including applied electric field, result in significant response: permittivity is large ($\varepsilon \sim 10^4$), while the nonlinearity N_e is proportional to ε^3 .
3. In connection with the achievements of nanotechnology, it is possible to obtain high-quality thin layers of dielectrics, in which field of electrical breakdown increases by thousands of times. This implies a possibility to obtain (and to use) significant dielectric nonlinearity in case of usual mechanisms of polarization.

Dielectric anisotropy. Unlike most dielectrics that are simple isotropic insulating materials and have identical properties in any direction, the *noncentrosymmetric* dielectrics are usually characterized by the *anisotropy* of their electrical, thermal, and mechanical characteristics. In such “active” (or “smart”) dielectrics, their electrical and electromechanical parameters are quite different in various directions in the crystals or textures.

In this context, it is important to note that electrical field \mathbf{E} , polarization \mathbf{P} , and electrical induction \mathbf{D} are the *vector quantities*. In conventional isotropic dielectrics, vectors \mathbf{D} , \mathbf{E} , and \mathbf{P} are collinear; Fig. 7.1E shows these vectors in an electrical capacitor containing isotropic dielectric.

However, in the anisotropic dielectric vectors \mathbf{D} , \mathbf{E} , and \mathbf{P} differ in their orientation (at that, vector relation $\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}$ always remains true). Furthermore, it is considered to accept that vector \mathbf{P} is directed from negative charge to positive.

If dielectric is *isotropic*, this vector sum corresponds to sum of vector modules, as directions of all three vectors coincide. In such isotropic dielectric, electrical field is directed perpendicular to electrodes.

In the *anisotropic* dielectric-induced polarization is directed in accordance with internal elastic bonds of the constrained electrical charges so that vector \mathbf{P} is not parallel to vector \mathbf{E} . As a result, the direction of total electrical induction vector \mathbf{D} also has a direction different from vector \mathbf{E} (Fig. 7.1F), so that permittivity that characterizes relation between D and E ($D = \epsilon\epsilon_0 E$) is a parameter that differs in various directions.

Tensor of permittivity. In contrast to scalar parameters (such as material density, temperature, or specific heat), vectors are defined by *three* parameters—by their projections on coordinate axes. Accordingly, polar vectors \mathbf{D} , \mathbf{E} , and \mathbf{P} can be represented without vector notation, but with the subscripts m , $n = 1, 2, 3$. The transformation from one vector to another can be described as

$$D_m = \epsilon_0 \epsilon_{mn} E_n, \quad P_m = \epsilon_0 \chi_{mn} E_n. \quad (7.7)$$

It can be seen that in anisotropic media tensor-type components ϵ_{mn} and χ_{mn} are shown with *two indices*. One of these comes from the “impact” vector (in this case it is E_n), while another from the “response” vector (e.g., P_m) whose direction may not correspond to the direction of impact. Thus, to describe electrical, mechanical, and electromechanical properties of anisotropic crystals *tensors* should be used, which are physical or mathematical quantities that are transformed by *different laws* as compared to scalar quantities.

As is well known, vector algebra differs from ordinary algebra by introducing the concepts of *gradient* (e.g., $\mathbf{E} = \text{grad}\phi$, where ϕ is electrical potential), *divergence* (one of Maxwell’s equations is $\text{div}\mathbf{E} = \rho$), and *rotor* (other Maxwell equation is $\text{rot}\mathbf{E} = -\partial\mathbf{B}/\partial t$). In this case, all vectors are *first-rank tensors*, and they are recorded with *single* subscript: E_n, D_m, B_j , etc. Scalar values, according to this classification, are *zero-rank tensors* that are written with no lower indices and converted by ordinary laws of algebra, for example, $\delta Q = C_V \delta T$, that is, change in heat δQ equals to a product of heat capacity on temperature change δT .

Permittivity ϵ_{mn} and susceptibility χ_{mn} , as well as the conductivity σ_{mn} , the permeability μ_{mn} and some other parameters of anisotropic crystals and textures are *second-rank tensors*. Double letters in indices denote summation: for example, $D_m = \epsilon_0 \epsilon_{mn} E_n$ means $m, n = 1, 2, 3$. Eq. (7.7) is shorthand for three equations. In this case: $D_m = \epsilon_0 \sum_{n=1}^3 \epsilon_{mn} E_n$, in more detail:

$$\begin{aligned} D_1 &= \epsilon_0 (\epsilon_{11} E_1 + \epsilon_{12} E_2 + \epsilon_{13} E_3); \\ D_2 &= \epsilon_0 (\epsilon_{21} E_1 + \epsilon_{22} E_2 + \epsilon_{23} E_3); \\ D_3 &= \epsilon_0 (\epsilon_{31} E_1 + \epsilon_{32} E_2 + \epsilon_{33} E_3). \end{aligned}$$

Parameters ϵ_{mn} are associated with various components of vectors D_m and E_n : for example, ϵ_{13} is the material parameter that connects electrical displacement component D_1 induced in the anisotropic crystal by component E_3 of externally applied

electrical field. In isotropic dielectric, obviously, $\epsilon_{13} = 0$ and $\epsilon_{12} = 0$, while ϵ_{11} , ϵ_{22} , and ϵ_{33} are the only nonzero values, all of which are identical, and, therefore, can be written as scalar quantity: $\epsilon_{mn} = \epsilon$.

Components of permittivity tensor ϵ_{mn} might be presented in the form of matrix

$$\epsilon_{mn} = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{33} \end{bmatrix}. \quad (7.8)$$

In case of isotropic dielectric (glass, amber, plastic, nonpolarized ceramics, etc.), this matrix contains only components of main diagonal $\epsilon_{mn} = \epsilon$. However, ferroelectric ceramics, for example, after prolonged application of strong electrical field become polarized in one direction: domains are oriented by external field so that the preferred direction exists—the polar axis z . For such a case of polarized ferroelectric that is turned into a texture with highlighting axis z (usually $z = 3$). Along this axis, permittivity is ϵ_{33} while perpendicular to it is $\epsilon_{11} = \epsilon_{22}$. Corresponding matrices have the form:

$$\epsilon_{mn} \text{ isotropic diel} = \begin{bmatrix} \epsilon & 0 & 0 \\ 0 & \epsilon & 0 \\ 0 & 0 & \epsilon \end{bmatrix}; \quad \epsilon_{mn} \text{ polarized ceramics} = \begin{bmatrix} \epsilon_{11} & 0 & 0 \\ 0 & \epsilon_{22} & 0 \\ 0 & 0 & \epsilon_{33} \end{bmatrix}.$$

In the low-symmetry crystals (with a small number of symmetry elements) and in the case of arbitrary orientation of electrical field vector (not along principal axes of crystal), the permittivity tensor can have all six components (this is their maximum). From matrix (7.8) formally follows that the number of ϵ_{mn} components must be nine, but ϵ_{mn} is the *symmetric* tensor (as well as conductivity tensor σ_{mn}); this means $\epsilon_{mn} = \epsilon_{nm}$ (as well as $\sigma_{mn} = \sigma_{nm}$). The matrix (7.8) is the *symmetric* matrix as to the main diagonal. Such matrices can always be transformed to diagonal form, when all components of tensor ϵ_{mn} , except for the diagonal components, equal zero.

It is known from analytical geometry that the main (diagonal) components of matrix can be represented by the axes of *ellipsoid*. In this case, it is the *ellipsoid of permittivity*:

$$\frac{x^2}{\epsilon_{11}} + \frac{y^2}{\epsilon_{22}} + \frac{z^2}{\epsilon_{33}} = 1.$$

In isotropic dielectric $\epsilon_{11} = \epsilon_{22} = \epsilon_{33}$ and dielectric ellipsoid becomes a *sphere*. In tetragonal, hexagonal, and trigonal crystals, as well as in polarized ferroelectric (texture) components $\epsilon_{11} = \epsilon_{22}$, but they differ from the ϵ_{33} so that dielectric ellipsoid becomes the *ellipsoid of rotation*. For this type of dielectric, there are two main values of permittivity, ϵ_{33} and ϵ_{11} , which can be experimentally determined as the permittivity along the crystal main axis and perpendicular to it. Fig. 7.2 shows geometric representation of different *material tensors*; dielectric ellipsoid is shown in Fig. 7.2C [4].

As an example, hexagonal piezoelectric crystal quartz has $\epsilon_{11} = 4.5$ and $\epsilon_{33} = 4.7$, tetragonal crystal barium titanate (ferroelectric) has permittivity $\epsilon_{11} = 4000$ and $\epsilon_{33} = 200$ (at 300K), while potassium dihydrogen phosphate (KDP) at 300K has $\epsilon_{11} = 42$ and $\epsilon_{33} = 27$.

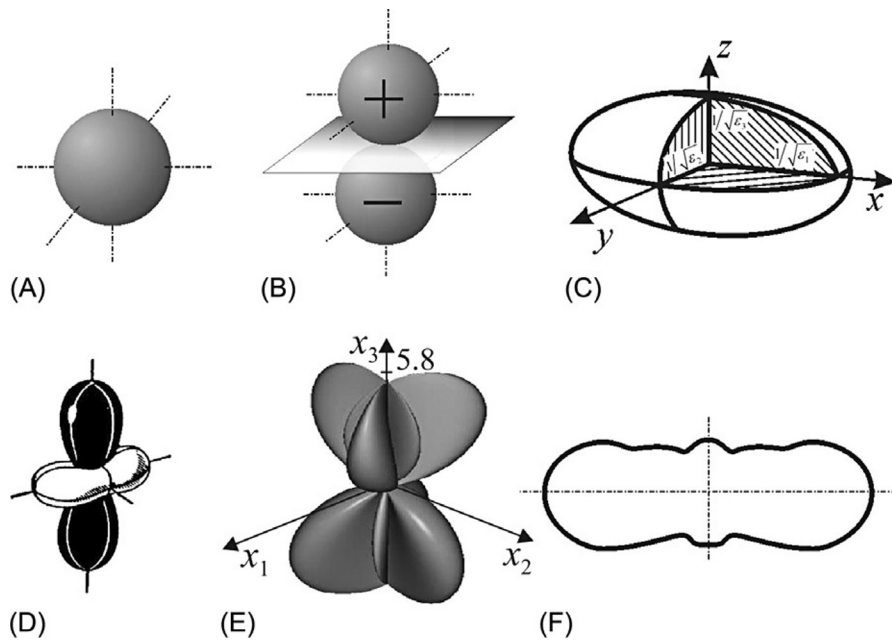


FIG. 7.2

Images of material tensors of various ranks: (A) zero-rank tensor (scalar); (B) first-rank tensor (pyroelectric coefficient); (C) second-rank tensor (dielectric constant); (D) second-rank tensor (coefficient of thermal expansion), (E) third-rank tensor (piezoelectric modulus); (F) projection of fourth-rank tensor (elastic compliance).

Crystals of low symmetry (triclinic, monoclinic, and rhombic) are characterized by dielectric ellipsoid of *general form*, and have three main values of ϵ_{mn} . For example, ferroelectric Rochelle salt is characterized by $\epsilon_{11}=2000$, $\epsilon_{22}=7.5$, and $\epsilon_{33}=7.5$. At that, the permittivity of cubic crystals (highest symmetry) looks like scalar, such is the crystal NaCl with $\epsilon = 5.6$.

For description of induced polarization (arising in dielectrics in external electric field), tensor ϵ_{mn} is used in such cases, if exposure (intensity) parameter is the electrical field E_n , while induction D_m is the response (extensive parameter). In the opposite case, a tensor of *inverse permittivity* can be introduced: $\beta_{mn} = \epsilon_{mn}^{-1}$. Dimensionless physical quantity β_{mn} is the *dielectric impenetrability*. This parameter is used, for instance, in the equations of piezoelectric effect, which will be discussed in [Chapter 9](#), because in some cases it is not enough to consider only the tensor ϵ_{mn} .

In case of piezoelectric, the permittivity is strongly dependent on *mechanical conditions*, in which crystal or texture is studied or used. When electrical field can freely deform piezoelectric, its *seeming* dielectric constant will be higher than when mechanical strains are restricted. This important effect is because piezoelectric is an electromechanical transducer; hence, electrical energy that is applied to

piezoelectric is responsible not only for electrical displacement of electrons, ions, or dipoles, but also for *macroscopic* elastic deformation of piezoelectric as a whole.

In this connection, it is necessary to distinguish the ϵ^X (permittivity of free crystal, in which there is no mechanical stress, $X=0$) and the ϵ^x (permittivity of clamped crystal, in which piezoelectric strain is absent, $x=0$), so only microscopic polarization mechanisms are possible. For the piezoelectric, the inequality $\epsilon^X > \epsilon^x$ is always true, because the ϵ^X contains additional mechanism of reversible energy storage—the elastic displacement of crystal as a whole.

Piezoelectric contribution to permittivity may be quite different. For example, accurate measurements of quartz show $\epsilon_{11}^X=4.52$ and $\epsilon_{11}^x=4.43$, $\epsilon_{33}^X=4.70$ and $\epsilon_{33}^x=4.64$; therefore in this crystal piezoelectric contribution is small: $\epsilon_{11}^X - \epsilon_{11}^x = 0.09$ and $\epsilon_{33}^X - \epsilon_{33}^x = 0.06$. However, in barium titanate crystal at 300 K components of permittivity are $\epsilon_{11}^X=4000$ and $\epsilon_{11}^x=2000$, $\epsilon_{33}^X=200$ and $\epsilon_{33}^x=90$; hence piezoelectric contribution to permittivity in this crystal is essential: $\epsilon_{11}^X - \epsilon_{11}^x = 2000$ while $\epsilon_{33}^X - \epsilon_{33}^x = 110$. Therefore the electromechanical contribution to barium titanate permittivity shows large anisotropy, and it equals approximately half of the total value of dielectric constant.

7.3 DIFFERENT MECHANISMS OF POLARIZATION

The microscopic conception of polarization mechanisms can be reduced to a few relatively *simple models* of electrical moment appearance in the electrical field. As the quantum-mechanical calculations of atomic electron shell interaction with nuclei are difficult even for some simple molecules, then, to describe electrical polarization in solids, consisting of a set of atoms, ions, or molecules, it would suffice to consider the simplest models of electrical moment arising based on *classical concepts*.

When electrical field is applied to a solid dielectric, the closely connected charges of structural units are displaced relatively to each other, whereby dielectric becomes polarized. External electrical field induces in a dielectric the *elementary electrical moments* $p=qx$, where q is a charge of constrained units and x is their mutual displacement.

Field-induced electrical moment may get contribution from:

- *electrons* displaced from their equilibrium positions in atoms;
- *ions* deviating from equilibrium state in crystal lattice;
- *dipoles* (polar molecules) changing their orientation in electrical field;
- *macrodipoles* (electrically charged radicals or complexes in the inhomogeneous structures).

Electrons, ions, and dipoles (including macrodipoles) can acquire their *electrical moment* (i.e., polarized state) through various *mechanisms*.

If particles are tightly connected in a structure, the external electrical field (or other impact) leads only to *very small* (compared with atomic dimensions) deviations from the nonpolarized equilibrium state. However, as in the process of polarization

all particles of dielectric are involved, even these small displacements of charges cause significant integral effect—the polarization. Such a mechanism is called the *quasielastic polarization*.

The main mechanisms of quasielastic polarization are shown schematically in Fig. 7.3A and B [4]. In the nonpolarized state ($E=0$) electron shells of atoms are located symmetric with respect to nuclei (Fig. 7.3A, left-hand fragment) so that the *effective center* of negative charge of electron shell coincides with positively charged nucleus. Accordingly, the elemental dipole moment is zero ($p=0$), because it is determined by a product $qx=p$ but relative displacement of charges q^+ and q^- is absent: $x=0$.

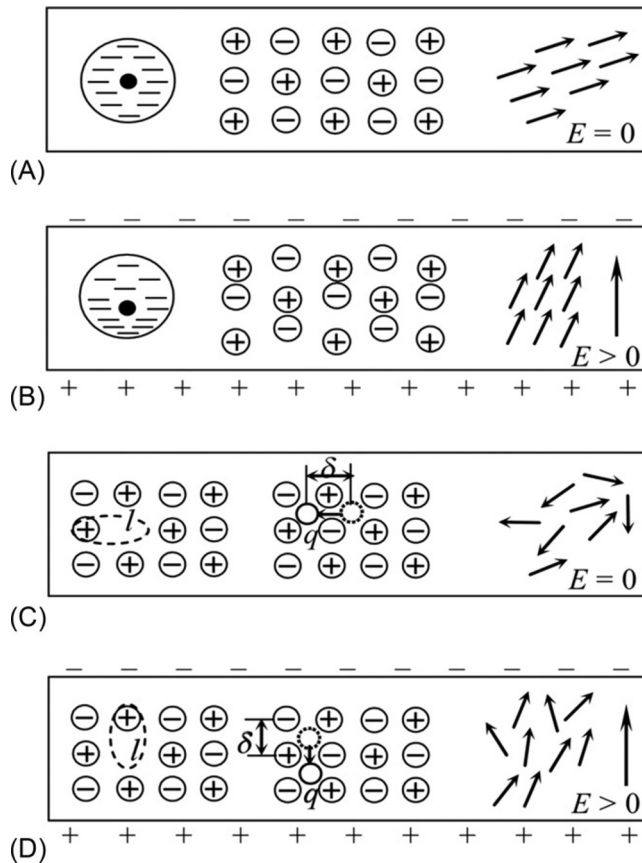


FIG. 7.3

Microscopic mechanisms of polarization: (A, B) three mechanisms of quasielastic polarization: fragments of dielectric in the absence of electric field E and when it is applied; (C, D) three mechanisms of thermally activated polarization: fragments of dielectric in the absence of electric field E and when it is applied.

If electrical field is applied (Fig. 7.3B), it influences each atom, molecule, or ion, and their electron shells distort and displace with respect to nuclei, whereby the center of negative charge shifts relatively to the positively charged nucleus so that elementary polar moment appears: $p = qx > 0$. This is the mechanism of *electronic quasielastic* polarization.

In the ionic crystal, in the absence of externally applied electric field (see Fig. 7.3A, central fragment), crystal lattice sites are balanced by cations and anions. This system of charges is electrically neutral, and it does not show any electrical moment (polarization). However, in the applied external electric field (Fig. 7.3B), cations and anions displace, forming the polarized lattice of $q^+ - q^-$ with elementary electrical moment $p = qx > 0$. By this way, the *ionic quasielastic* polarization arises, which, in the ionic crystals, has great importance.

The energy characteristics of quasielastic polarization process are shown in Fig. 7.4A. Elastic energy of bound particles (ions in crystal, electrons in atom, dipole

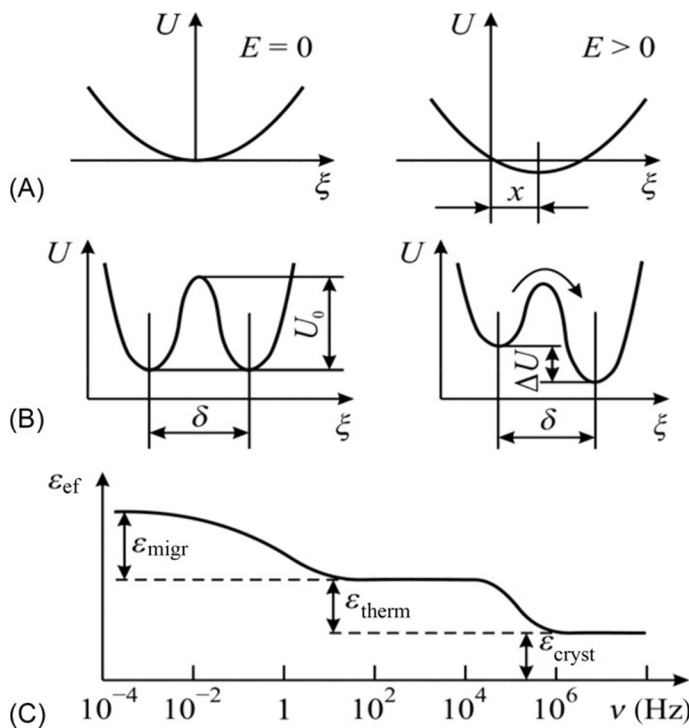


FIG. 7.4

Explanation of different polarization mechanisms: (A) energy that characterized quasielastic polarization mechanism; (B) energy diagram of thermally activated (relaxation) polarization mechanism; (C) frequency dependences of dielectric contributions from migratory, thermally activated, and fundamental (lattice) crystal polarization.

in lattice) can be given by equation $U = \frac{1}{2} cx^2$, where c is quasielastic coupling coefficient and x is elastic displacement from equilibrium position. In the absence of external influences, $x = 0$ and the charged particle is localized in the bottom of a parabolic potential well. Under the influence of electric field particle acquires extra energy $U = \frac{1}{2} cx^2 - qxE$. This energy is added to elastic energy; therefore the minimum of energy shifts to a new position $x > 0$ (Fig. 7.4A), in which particles already have elementary electrical momentum $p = qx$ and contribute to polarization. Switching off the electrical field leads to *rapid* establishment of former equilibrium position, when $x = 0$ and elastic polarization disappears.

On the right-hand fragments of Fig. 7.3A and B, the third quasielastic mechanism of polarization is shown, namely, the elastic rotation of *constrained dipoles*. It is possible only when the intrinsic polarity exists in dielectric (in the absence of external electrical field), which is observed in the noncentrosymmetric dielectrics. Ordering of dipoles in such polar lattice is conditioned by their *internal interaction* (usually it is due to mixed ionic-covalent polar bond); therefore dipoles are oriented spontaneously. External electrical field alters orientation of *all dipoles*, and thereby electrical field changes local electrical moment of a structure, that is, all dipoles obtain new orientation, induced by the field (Fig. 7.3B, right-hand fragments). This is the mechanism of *dipole elastic* polarization.

By switching off the externally applied field, all three previously mentioned mechanisms of polarization disappear *very rapidly*: dielectric returns to its equilibrium (nonpolarized) state. At that, electrons occupy their electrically symmetric position relatively to nuclei due to Coulomb forces of attraction; cations and anions will return to their stable (equilibrium) position in the crystal lattice sites by forces of electron shell repulsion. Coordinated dipoles return to their initial orientation (in which energy of oriented dipoles in given crystal is minimal). However, sometimes—in the ferroelectrics—there are some large areas, called as *domains*, which can remain in the new rotated state.

Besides the quasielastic (deformation type) polarization, the electrons, ions, and dipoles (or macrodipoles) may also participate in the mechanisms of *thermally induced* and *migratory* polarization.

In case of weak bonding in a structure, some electrons, ions, and dipoles can be greatly influenced by thermal (chaotic) motion of particles, and their motion may have influence on polarization (Fig. 7.3C and D). These figures show that such particles (usually they are impurities) are isolated in the local (nanoscale) areas; hence, they are not main structural elements of dielectric. However, these electrically charged impurities are loosely bound in the crystal lattice and they can change their position, remaining yet in the vicinity of structural defects in dielectric.

Being localized in their nanoareas these particles under the influence of thermal movement make *thermally activated jumps*, moving over a distance in the order of atomic dimensions. At that, their abrupt displacement exceeds 10^4 – 10^6 times the value of small quasielastic displacements of main structural units of a crystal (which produce quasielastic polarization). For this reason, thermally activated jumps of weakly bounded charged particles can have a significant effect on the permittivity.

The *electronic thermally activated polarization* is conditioned by weakly bounded electrons, for example, electrons that compensate charged structural defects. These defects, for example, might be anion vacancies (lack of negative ions), as shown in Fig. 7.3C, left-hand fragment. Charge compensation occurs, because crystal lattice always should be *electrically neutral*: that is, the number of negative charges in a lattice should be equal to the number of positive charges.

The electron whose charge compensates the charge of missing anion is localized near *surrounding cations*; at that, the orbit of this electron becomes large and deformed: it is stretched in the direction of anion vacancy to compensate the absent charge (Fig. 7.3C, left hand). This leads to local electrical moment formation: $p_0 = ql_0$, where l_0 is close to lattice constant (~ 3 nm). The magnitude of this moment is not determined by external electric field; this “permanent” dipole moment p_0 is thousands of times higher than possible electrical moment induced by external field during elastic polarization ($p = qx$, where x is induced elastic displacement, has the order of 10^{-5} nm).

Weakly bonded electron (localized near the anion vacancy) even at the absence of external field, from time to time, under *thermal chaotic motion* jumps from one of neighboring cation to another, overcoming energy barrier U_0 (Fig. 7.4B). At that, the direction of inserted dipole moment p_0 should be changed. Despite the fact that the quantity of these defective places in real dielectric is not large (10^{14} – 10^{20} cm $^{-3}$, as compared with concentration of basic structural units of crystal $\sim 10^{23}$ cm $^{-3}$), macroscopic polarization in crystal does not arise because all “permanent” dipoles are oriented randomly at any time.

Externally applied electrical field reduces the potential barrier (Fig. 7.4B) that results in the excessive orientation of “electron–cation” dipoles in accordance with applied field direction (Fig. 7.3D, left-hand fragment). Such is, in general terms, the mechanism of *electronic thermally activated polarization*. This polarization is called as “thermal” because jumps of electron between surrounding cations are conditioned by thermal energy of crystal. Electrical field that has relatively low impact energy $\Delta U < k_B T$ (Fig. 7.4B) leads only to a certain *redistribution* in the local electrical moments p_0 orientations.

The *ionic thermally activated polarization* mechanism (Fig. 7.3C and D, the middle fragments) is largely similar to the electronic thermal polarization mechanism. It is assumed that in crystal lattice some impurity (embedded) ions are present that usually have smaller ionic radius (e.g., positive ions). This model is close to common ionic crystals, doped by very small lithium ions (one of experimental example). It is assumed that impurity cations are located in the interstices of a structure, and their charge compensation is carried out due to increased charge of one of neighboring anions. In the vicinity of such anion, the impurity ion makes thermally induced jumps.

These jumps are hampered by potential barrier (Fig. 7.4B), because in order to change its localization the impurity ions need to overcome repulsive forces of electron shells of neighboring ions. Dipole moment p_0 is created between jumping impurity ion and fixed charge-compensating anion (that has a larger radius).

When the ion of impurity makes hopping in the vicinity of its localization, it changes the direction of electrical moment with neighboring ions. In the absence of external field, a set of such polar defects are reoriented chaotically that usually does not result in integral polar moment—polarization. (However, there are some uncommon examples, when the *interaction* between such polar defects leads to spontaneous polarization—artificial ferroelectricity.)

In the externally applied electrical field E (Fig. 7.4D), along ion hopping, a direction of asymmetry appears, and thereby macroscopic polarization arises (in this case, *thermally activated ionic polarization*). After electrical field switching off, due to the disorienting effect of thermal chaotic motion, electrically induced thermal polarization gradually disappears.

Energy barrier U_0 that should be overcome by impurity ion (Fig. 7.4B) is much greater than energy of thermal motion of particles in dielectric: $U_0 \gg k_B T$. However, the probability of thermal hopping of ions (as well as likelihood of thermal reorientation of dipoles) increases with increasing temperature. These thermally activated jumps occur on average distance δ that is defined by crystal structure, but *does not depend* on external field E (as opposed to the elastic polarization, when the size of elastic displacement is determined by a field: $x \sim E$). In case of thermally activated polarization, the external field only *changes the probability* of particle hopping over barrier. One of the potential wells, as compared to the other, becomes deeper on $\Delta U \ll U_0$, while another becomes shallow. Their difference depends on applied field: $\Delta U = q\delta E$, that is, the contribution of added electrical energy on a distance of thermal hopping.

The *dipoles thermally activated polarization* in crystals and textures can be approximately characterized by a model, shown in the right-hand part of Fig. 7.3C and D. In the absence of external field, the permanent dipoles already exist, but they are distributed randomly. Externally applied electrical field results in preferential orientation of dipoles; as a result, the electrical moment appears. In practice, realization of thermal dipole polarization in dielectrics is limited by a certain number of stable orientations of permanent dipoles (in accordance with symmetry of crystal or polar texture). In the absence of external field, these dipoles are oriented uniformly in all permitted directions, but after electrical field switching, the likelihood of orientation of dipoles in the favorable direction increases.

It is obvious that all thermal polarization mechanisms are *much slower* in comparison with elastic polarization. In case of elastic polarization, externally polarized system of elastically bounded charges after electric field removal returns to its equilibrium (nonpolarized) state very fast (at time 10^{-12} – 10^{-16} s). On the contrary, in case of thermally activated polarization, thermally stimulated electrodiffusion takes place by the jumps of “semifree” electrons (or ions) through potential barriers. It is obvious that such a process is relatively slow, and it needs a time of about 10^{-2} – 10^{-9} s. At that, the time of thermal relaxation is strongly dependent on temperature that characterizes the intensity of thermal movement.

The *migratory polarization* that is peculiar for certain active (smart) dielectrics (polarized textures) is the slowest polarization mechanism (Fig. 7.4C). When it

occurs, some of free charges can move at rather great (almost macroscopic) distance. In this case, the *accumulation* of electrical charges takes place at the boundaries of structural irregularities (between-crystallite layers, interstices, large-scale defects) that determines the *space-charged polarization*. This mechanism significantly increases *low-frequency capacitance* of electrical capacitor, containing heterogeneous dielectric. The migratory (or space-charged) polarization cannot be attributed to the microscopic mechanisms of polarization and, therefore, their “permittivity” depends on dimension of studied capacitor; in case of migratory polarization, the permittivity is called as the “effective,” ϵ_{ef} .

In piezoelectric and pyroelectric textures (ceramics), the large dipole groups are involved in migratory polarization (by orientations of different sizes domains), which also should be described by the ϵ_{ef} .

In frequency dependence of permittivity, the contribution of migratory polarization ($\epsilon_{\text{ef}} = \epsilon_{\text{migr}}$), as well as contribution of thermally activated (relaxation) polarization ($\epsilon = \epsilon_{\text{therm}}$) are shown in Fig. 7.4C. In case of high concentration of charged structural defects, these dielectric contributions may be much larger than the permittivity of pure (ideal) crystal ϵ_{cryst} . However, at higher frequencies (starting from the acoustic frequency range), the space charge has no time to be accumulated and dissipated; as a result, the migratory polarization is late, that is, the ϵ_{migr} shows frequency dispersion. In the frequency range of dielectric constant dispersion, the maximum of dielectric losses ($\tan\delta$) is also observed.

Microscopic structural defects resulting in the thermally activated polarization ϵ_{therm} demonstrate dielectric contribution at frequencies below 10^5 – 10^9 Hz, depending on temperature and type of structural microscopic defects. Dispersion (frequency dependence) of ϵ_{therm} is also accompanied by the maximum of dielectric loss.

Different inertia of various mechanisms of electrically induced polarization allows the experimental selection of their dielectric contributions, when properties of dielectrics are studied in a wide range of frequencies. This method of *dielectric spectroscopy* is suitable not only for the detection of main contributions to ϵ_{migr} (space-charge polarization or domain reorientation in textures and composites), but also for studying crystals with perfect structure, where impurities, defects, and space charges do not affect the ϵ -value. In this case, the ϵ -variance is determined only by the “fundamental” mechanisms of polarization (Fig. 7.5).

Investigation of dielectrics in a wide frequency range (10^{-3} – 10^{16} Hz) makes it possible to separate different mechanisms of polarization and to find correspondent “dielectric contributions” $\Delta\epsilon$ (Fig. 7.5). Quasielastic displacement of electrons (electron shell deformation in atoms, ions, or molecules) is the most high-frequency polarization mechanism, which has enough time to respond even at visible optical frequencies (about 10^{15} Hz). For this reason, electron shell displacement determines *optical contribution* to permittivity ϵ_{opt} (Fig. 7.5A). Only at ultraviolet frequencies, optical contribution to permittivity disappears with several resonances.

The ionic quasielastic polarization (relative shift of cationic and anionic sublattices in crystals) entirely responds to the microwaves, but this polarization shows resonant ϵ -dispersion in the *far-IR* frequency range (Fig. 7.5B). The ionic

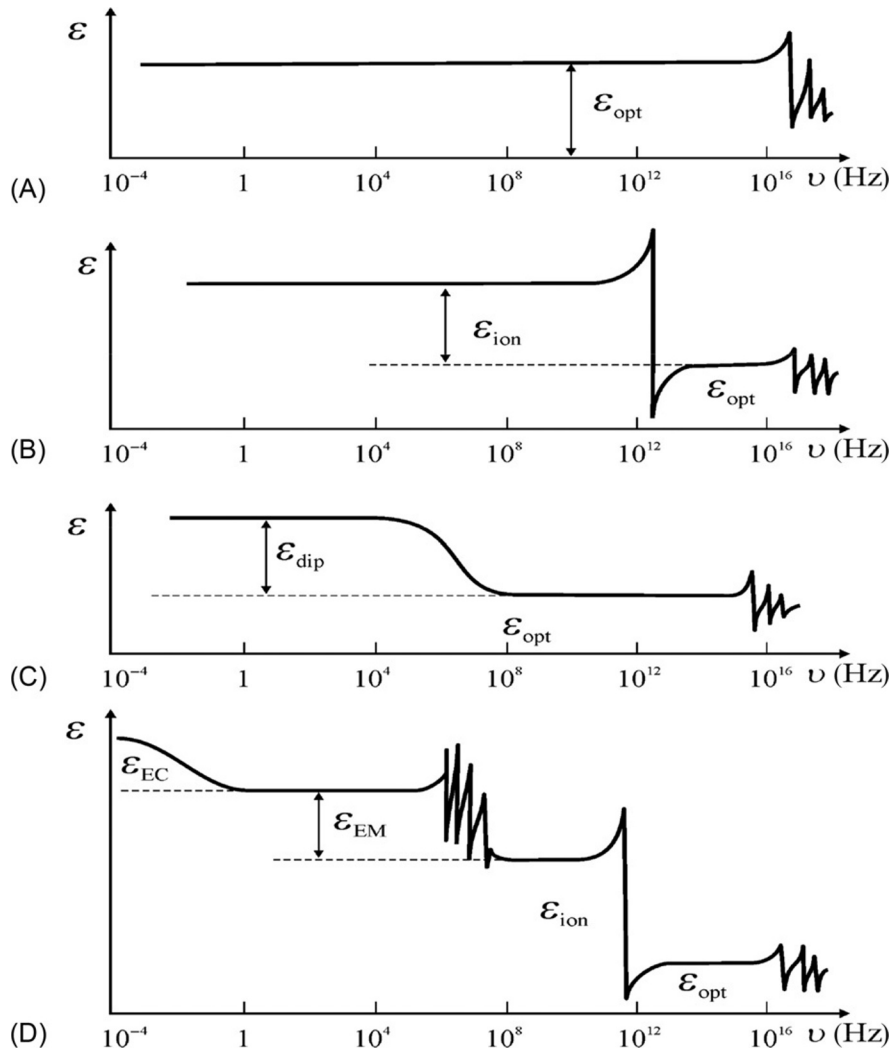


FIG. 7.5

Dielectric spectra of different dielectrics and contributions to permittivity from different polarization mechanisms: (A) optical polarization; (B) ionic polarization; (C) dipole orientation thermal polarization; (D) electromechanical and electrocaloric dielectric contributions in dielectric permittivity; ϵ_{opt} —electron displacement; ϵ_{ion} —ion displacement; ϵ_{dip} —thermally activated orientation of dipoles; ϵ_{EM} —electromechanical polarization; ϵ_{EC} —electrocaloric polarization.

polarization determines the dielectric contribution ϵ_{ion} that may be measured at microwave frequencies.

The thermally activated orientation of dipoles is a specific polarization mechanism for *polar dielectrics*. As against electronic and ionic polarization (that need

no thermal activation), the dipoles, as a rule, can be oriented by the electrical field with the support of chaotic thermal movement in a dielectric. For this reason, dipole orientation is a relatively low-frequency polarization mechanism with the contribution ϵ_{dip} that leads to *relaxation dispersion*, usually below frequency of 10^8 Hz (Fig. 7.5C). In principle, only these three mechanisms of polarization entirely correspond to the *microscopic Lorentz model*.

Most of comparatively low-frequency and radiofrequency measurements of permittivity are provided by the *capacitance* measurements of electrical capacitor prepared from the studied material. The capacitance is calculated from the value of *reactive* electrical current that occurs due to “displacement” of electrical charges under alternating voltage.

However, in the accepted definition of ϵ (from capacitance measurements), some uncertainties may occur:

1. Under electrical field influence on the *noncentrosymmetric* dielectric (piezoelectric), a part of “reactive” (returned) energy can be stored in the *mechanical form* (in the elastic deformation of piezoactive dielectric). It should be noted that methods of ϵ -determining by the capacitance measuring cannot separate the reactive current given by piezomechanical deformation from the reactive current conditioned by microscopic mechanisms of polarization. That is why a concept of the *electromechanical dielectric contribution* ϵ_{EM} should be introduced, as shown in Fig. 7.5D.
2. In the *polar dielectrics* (pyroelectrics or ferroelectrics), the energy can be partially reserved in the form of heat (electrocaloric effect). In case of low-frequency measurements of pyroelectric capacitors, it might be impossible to distinguish the contribution from the *pyroaccumulated* energy from reactive energy stored by other mechanisms of polarization, and this is the *electrocaloric dielectric contribution* ϵ_{EC} , shown in Fig. 7.5D.

It is obvious that both of these effects, being measured through a value of reactive current, are not consistent with the classic definition of *Lorentz dielectric permittivity* as a parameter for physically infinitesimal volume. Dielectric contributions ϵ_{EM} and ϵ_{EC} , obtained at the same frequency but using tested samples of different sizes, might be quite different. The point is that the manifestation of piezoelectric or pyroelectric effects is dependent on the shape and size of the test sample and from sample environment: these effects are significantly determined by mechanical and thermal conditions, at which the capacitor of piezoelectric or pyroelectric material is studied [5].

3. Many materials that are widely used in electronics consist of ordered or disordered *mixtures* of different dielectrics, as well as dielectrics with dispersed semiconductors or conductors. Measured electrical capacitance of the samples made of such mixture might be many times larger; this, again, is attributed to increase in capacitance due to “seeming permittivity.” In some cases, this is additional electrical polarization (Maxwell-Wagner mechanism in mixtures) that is caused by space-charge accumulation at the boundaries of components of a mixture or space charge in the near-electrode regions.

For describing dielectric properties of inhomogeneous materials, the concept of “effective” dielectric constant is commonly used. The value of $\epsilon_{\text{migr}} = \epsilon_{\text{ef}}$ in case of *macroscopic* Maxwell-Wagner polarization does not qualify for the definition of ϵ given as a *microscopic* Lorentz parameter. The ϵ_{migr} usually is characterized by relaxation type of dispersion, located in the frequency interval of 10^{-3} – 10^3 Hz (in Fig. 7.5, the frequency dependence of ϵ_{migr} is not shown).

4. In the *nonlinear* dielectrics (ferroelectrics and paraelectrics), the direct proportionality of reactive current to applied voltage might be violated. For example, at spontaneous polarization switching in ferroelectrics, the value of apparent permittivity varies many times reaching values of $\sim 10^5$ (depending on instantaneous values of the alternating electrical field). In this case, it is also possible to use the term “effective” (non-Lorentz type) dielectric permittivity.

All these cases are shown in Fig. 7.6, where various dielectric contributions from different polarization mechanisms are divided into two classes. Dielectric permittivity conditioned by “classical” mechanisms of polarization that satisfies the definition of Lorentz is called the *actual* (true) permittivity, in order to distinguish it from the conventional term of “effective” permittivity. In case of capacitance description in an electrical capacitor that contains heterogeneous mixtures, only effective parameter ϵ_{ef} should be used. Similarly, nonlinear polarization of ferroelectric material that is characterized by the average permittivity might be also described by effective parameter ϵ_{ef} .

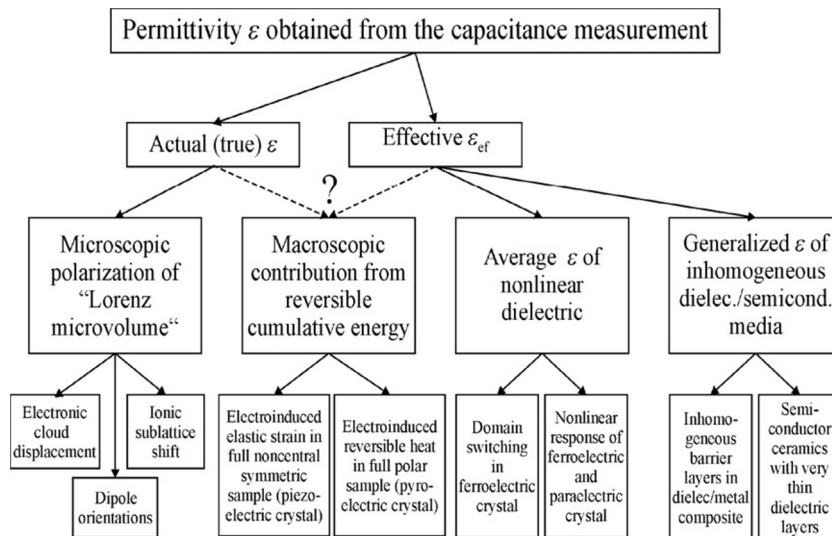


FIG. 7.6

Possible classification of dielectric contributions from various polarization mechanisms.

As it follows from dielectric spectra shown in Fig. 7.5 and from classifications given in Fig. 7.6, in the high-frequency (microwave) dielectrics only two polarization mechanisms are actual: electron shells and ionic sublattices electrically induced displacements [3]. Other polarization mechanisms are too slow to have any influence on the ϵ_{mic} (but they can add to microwave dielectric losses).

In the *homeopolar (atomic) crystals* with the structure of diamond (C, Ge, Si), as well as in centrosymmetric molecular crystals (with has no structural dipoles), permittivity is defined only by the *electronic quasielastic* polarization. This mechanism of polarization is practically noninertial, and its contribution $\epsilon_{\text{el}} = \epsilon_{\text{opt}}$ can be measured at optical frequencies; therefore, in the entire frequency range used in electronics, permittivity of such crystals is not frequency dependent.

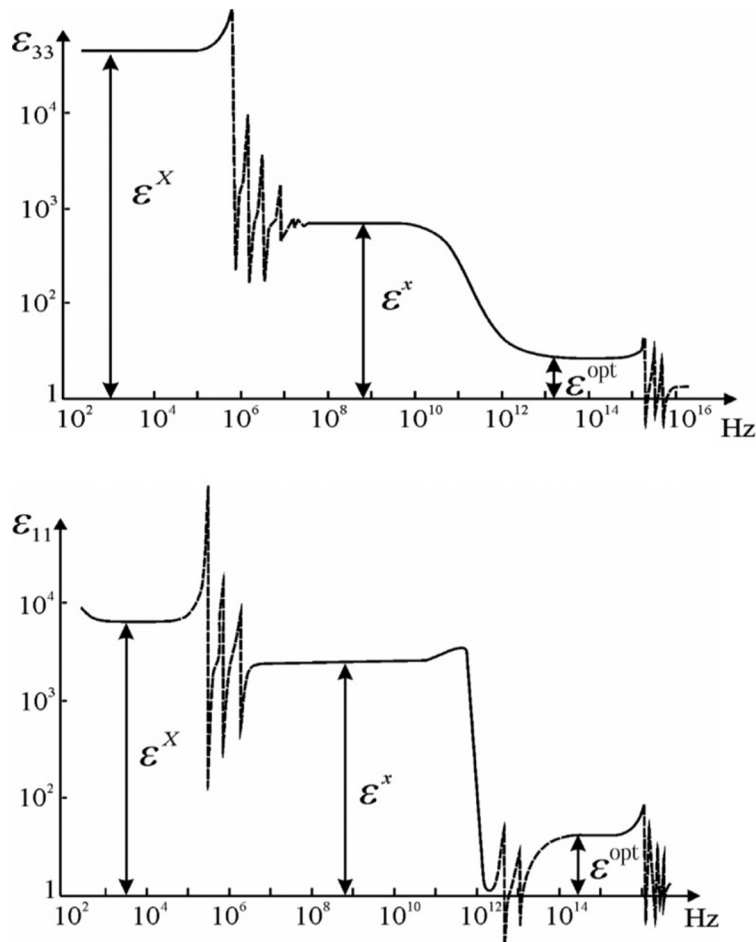
Molecular crystals have dipole-like structures, but without electron shell polarization; the low-frequency polarization is added as a contribution from the thermally activated polarization of *dipoles* (Fig. 7.5B). In this case, the dipoles are not considered as impurities, but they belong to basic structure (e.g., in the ice crystal). During frequency dispersion of permittivity, the contribution of this polarization ϵ_{dip} gradually decreases. This character of the dispersion is called *relaxation*.

In *ionic crystals* the ionic quasielastic polarization is added to always existing electronic (optical) polarization. This mechanism of polarization is also a high-frequency one, and the ϵ -variance is observed only in the IR frequency range ($\sim 10^{13}$ Hz, Fig. 7.5C). For this reason, the contribution of ionic polarization is indicated as $\epsilon_{\text{ion}} = \epsilon_{\text{IR}}$. Dielectric dispersion in this case, as opposed to relaxation dispersion, shows ϵ_{ion} resonance when measurement frequency approaches to natural frequency of ionic lattice vibrations; parameter $\epsilon(\nu)$ first increases and then reaches maximum, after it drops sharply, sometimes to the negative values. The nature *resonant dispersion* will be explained later in Section 7.4.

In Fig. 7.5 the *theoretical* frequency dependences of $\epsilon(\nu)$ are depicted, but Fig. 7.7 illustrates *experimental* frequency characteristics of two well-studied crystals—barium titanate (BaTiO_3) in the ferroelectric phase and potassium dihydrogen phosphate (KDP) near the ferroelectric Curie point [5]. In single crystals of barium titanate at temperature 300 K electromechanical contribution ϵ_{EM} is approximately equal to contributions from other mechanisms: at low frequencies $\epsilon^X \sim 4000$ but after piezoresonances the lattice contribution is $\epsilon^X = 2000$. In KDP crystals at temperature 125 K piezoelectric contribution ϵ^X almost is 100 times greater than contribution of all other mechanisms (dipole and optical).

7.4 OPTICAL AND FAR-INFRARED POLARIZATIONS

Electronic quasielastic polarization is a *general* mechanism of electrical polarization. In the external electrical field, electron shells of atoms (molecules or ions) shift relatively to positively charged nuclei (cores). As the mass of nucleus exceeds 10^4 – 10^5 times the mass of electron, this polarization actually is determined by the shift of electrons. At that, the main contribution to the induced electrical moment

**FIG. 7.7**

Examples of dielectric spectroscopy method application for analysis of various dielectric contributions: in single-domain crystal of barium titanate the piezoelectric resonant dispersion occurs at a frequency of about 1 MHz, and lattice dispersion takes place in IR region (about 10^{12} Hz); potassium dihydrogen phosphate also has piezodispersion, and dipole relaxation in microwaves as well near in 10^{11} Hz.

is given by weakly bounded electrons of the *outer shells* of atoms or ions, especially by the valence electrons. They are displaced in the electrical field to a much greater extent than more strongly bounded core electrons of atom or ion.

In case of quasielastic electronic polarization, the most important are the following two features. First, this is the *universal* mechanism of polarization, as the deformation of electron shells of atoms or ions in the electrical field occurs in all matters. Second, this is the *least inertial* polarization mechanism, as the mass of

electron is much smaller than the mass of other particles (atoms, ions, or molecules), involved in the processes of polarization. Very fast reaction allows select electronic polarization experimentally, using its contribution to permittivity at optical frequencies: $\epsilon_e = \epsilon_{\text{opt}}$. Therefore this contribution can be determined from optical refractive index: using Maxwell equations $n = (\epsilon\mu)^{1/2}$; however, at optical frequencies $\mu = 1$, so $\epsilon_{\text{opt}} = n^2$.

Response time of electronic elastic polarization is $\tau = 10^{-16} - 10^{-17}$ s. The visible optical frequency range is close to value of 10^{15} Hz; therefore the delay of electronic polarization (that determines frequency dispersion of ϵ_{opt}) should occur at higher frequencies than the visible optical range.

To simplify calculations, only elemental polarization mechanism will be considered. For this reason, electronic shell *polarizability* α_e is exemplified for hydrogen atom with the simplest Bohr's model (Fig. 7.8B). Under the action of local (microscopic) electrical field F (descending later from externally applied macroscopic field E), the electron shell of atom is displaced, so the geometric center of negative charge moves on a distance x from center of positive charge (Fig. 7.8B). Correspondingly, the electrical moment $p = qx$ is induced that is proportional to the applied local field, $p = \alpha_e F = qx$, where α_e is electronic polarizability which must be calculated.

From the model under calculation, it follows that $\alpha_e = 4\pi\epsilon_0 r_e^3$, where r_e is Bohr electronic radius and factor $4\pi\epsilon_0$ is required to record polarizability in SI; so the dimension of polarizability is $[\alpha] = \text{F}\cdot\text{m}^3$.

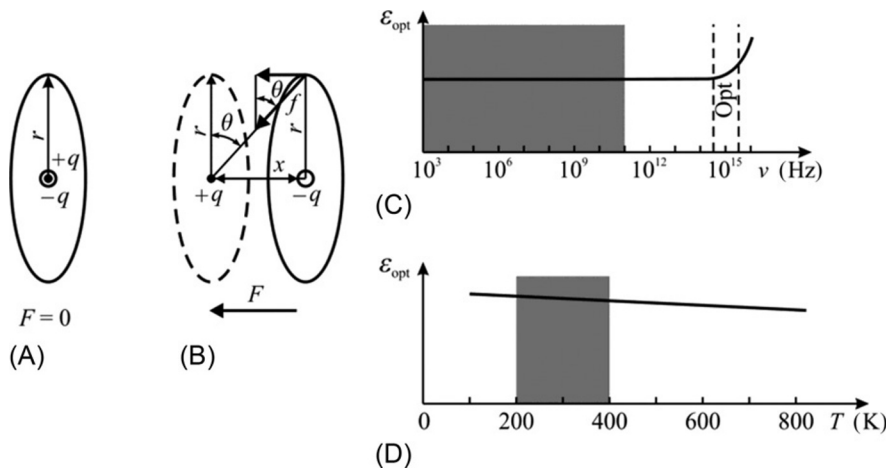


FIG. 7.8

Simplified model of electronic (optical) polarization: (A) model consisting of one electron and nucleus; (B) orbital deformation in local field F ; (C) frequency dependence of electronic permittivity; (D) temperature dependence of electronic permittivity (*shaded areas* show the ranges of technical interest in electronics).

It is known that the Bohr model is significant approximation: in fact, electronic cloud is not localized on a sphere with radius r_e , but constitutes the diffuse cloud around proton. Exact quantum-mechanical calculation yields the value of polarizability $\alpha_e = 18\pi\epsilon_0 r_e^3$ that is 4.5 times more than the calculation, obtained from a simple model. This discrepancy can be explained by the fact that the quantum-mechanical model represents electronic cloud as rather “extended.” However, in the CGSM system $\alpha'_e = r_e^3$ [5].

To evaluate *nonlinearity* of electronic polarizability in strong electrical fields, using α'_e from more simple Gauss system, its dependence on effective electrical field F can be represented as series

$$\alpha'_e(F) = \alpha_0 + \alpha_1 F + \alpha_2 F^2 + \alpha_3 F^3 + \dots,$$

that coefficients can be found from expressions $qF = cx$ and $cx = q^2 x(r^2 + x^2)^{-3/2}$, but under the condition of smallness x as compared to electronic orbit radius. As a result, the coefficients in this equation have following values:

$$\alpha_0 = r^3, \quad \alpha_1 = 0, \quad \alpha_2 = \frac{3r^7}{2q^2}, \quad \alpha_3 = 0, \quad \alpha_4 = \frac{1r^9}{8q^4}, \quad \alpha_5 = 0.$$

For estimation, it is sufficient to find the first nonzero coefficient (after main value α_0), as the nonlinearity is noticeable only in very strong fields. All coefficients at odd degrees of F are equal to zero (this result is expected in the symmetric model); therefore the nearest nonzero coefficient is positive: $\alpha_2 > 0$. It signifies that in strong electrical fields polarizability $\alpha_e(F)$ *increases*.

The nonlinearity occurs because the force that returns a system to the nonpolarized state (this force is electron attraction to nucleus) becomes reduced according to law $(r^2 + x^2)^{-1}$, while the dependence of acting electrical field on distance is linear. The orbit of electron stretches, and, therefore, polarizability α_e increases with electrical field growth. In this case, the change of electrical field direction, due to symmetry of selected model (Fig. 7.8), has no influence on polarizability. The increase in polarizability when electrical field increases leads to an increase of $\epsilon_{\text{opt}} = \epsilon_e$ and, hence, refractive index n increases in a strong electrical field even at light frequencies. This lens effect is seen experimentally and used in nonlinear optics for high-power laser radiation focusing on the dielectric.

In case of the *negative ions* (that accept electrons into their outer shell), the polarizability should be larger than in the positive ions (that give away their valence electrons). Furthermore, the polarizability of positively charged ions is less than in neutral atoms with similar electron shells. Thus, among the series O^{2-} , F^- , Ne , Na^+ , Mg^{2+} , Al^{3+} , and Si^{4+} , the value of α_e decreases. In all of these ions, the outer electron shell is similar to the shell of the inert gas neon and has structure s^2p^6 . However, the radius of the electron shell decreases systematically as nuclear charge is incremented by one, resulting in an increase of attraction of electrons to their nucleus. Therefore α_e in aforementioned series differs by more than an order of magnitude: from value of $7.4 \cdot 10^{-30} \text{ m}^3$ for ion O^{2-} down to value of $0.16 \cdot 10^{-30} \text{ m}^3$ for ion Si^{4+} .

Permittivity is determined not only by polarizability but also by the concentration of ions or atoms in unit volume of dielectric. Therefore the objective assessment of electrical moment per unit volume of dielectric is not parameter α_e , but the ratio α_e/r^3 . This ratio equals “1” only in the simplest model of hydrogen atom; however, in the majority of atoms and ions, this ratio differs from one. For some ions, such as Li^+ , Na^+ , Mg^{2+} , and Al^{3+} , the ratio $\alpha_e/r^3 < 1$, and for the majority of ions this ratio is very close to one, but for ions O^{2-} , Ti^{4+} , Pb^{2+} , and Ce^{4+} the value of $\alpha_e/r^3 > 1$.

The ability to control the value of refraction coefficient $n = \epsilon_e^{1/2}$ by introducing ions into the dielectric with higher α_e/r^3 is used in the integrated circuit technology. For example, to produce planar light guides, with the aim of refractive index increasing, in surface layer of glass or crystal the ionic diffusion of Ti^{4+} is carried out.

Dielectrics, in which electronic polarization dominates (solid dielectrics with covalent and molecular bonding), are characterized by small dependence of dielectric constant on temperature (Fig. 7.8D), and by extremely low value of dielectric loss, even at high frequencies, because $\epsilon_{\text{el}} = \epsilon_{\text{opt}}$ does not depend on frequency up to the optical range. For this reason, these dielectrics also have very small absorption not only in the optical range but also in all range. As shown in Fig. 7.8C, dielectric constant, conditioned by electronic (optical) polarization, does not depend on frequency up to optical frequencies, where ϵ_{opt} gradually starts to increase (in the ultraviolet, electronic polarization shows a resonant dispersion). In most cases, electronic contribution to permittivity somewhat decreases with temperature due to thermal expansion of a crystal (Fig. 7.3D). For this reason, the temperature coefficient of permittivity is very small and negative ($TC\epsilon < 10^{-5} \text{ K}^{-1}$, i.e., less than 10 ppm/K).

The *semiconductors of diamond structure*, such as high-resistive silicon, very often are used as microwave dielectrics at millimeter waves. Because microwave-integrated electronic circuits need dielectric substrates (where electromagnetic waves can propagate), dielectric properties of semiconductors should be considered. The nature of permittivity in the *atomic* crystals of semiconductors (Ge, Si, and C) is purely electronic. Hence the frequency change of ϵ in these crystals is possible only at optical frequencies (above 10^{15} Hz). As a result, microwave electronic *polarization* does not contribute to microwave losses (defined only by the *conductivity*). Therefore microwave absorption is seen only in at beginning of the microwave range. Owing to absence of inertia of electronic polarization, the permittivity of semiconductors does not change with frequency (Fig. 7.9).

To avoid the influence of conductivity, temperature measurements of *dielectric parameters* in semiconductors are possible only at frequencies above 300 GHz, and these experiments show an increase to some extent in ϵ with temperature.

According to energy band theory, when energy gap reduces, the permittivity should be greater. As temperature rises, the bandgap of semiconductors decreases, and correspondingly, with rising temperature ϵ increases (Fig. 7.10). However, in gallium arsenide some influence on dielectric properties can also be provided by ionic (far IR) polarization.

Ionic (far IR) polarization. Ions in molecules or in crystal lattice are electrically charged particles; therefore, like electrons, ions are shifted by external electrical field

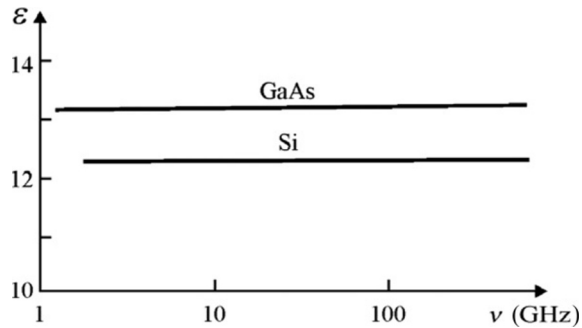


FIG. 7.9

High-frequency dependence of dielectric constant of silicon in comparison with gallium arsenide.

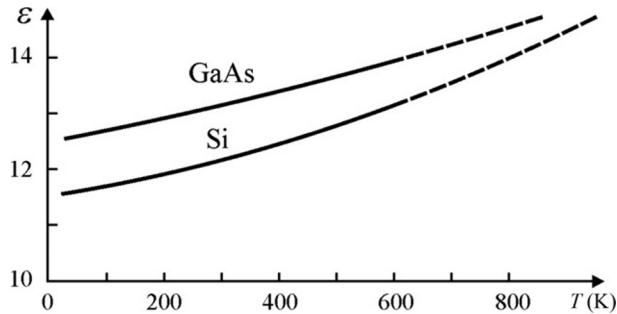


FIG. 7.10

Temperature dependence of dielectric constant of silicon in comparison with gallium arsenide on millimeter waves.

from their equilibrium position, which gives rise to induced electrical moment. When ions shift from their equilibrium position, the quasielastic returning force occurs, which, after switching off the external field quickly, returns the system of ions in the undisturbed (nonpolarized) state.

Ionic quasielastic polarization has the following features. First of all, this type of polarization is *not universal* for all dielectrics (such as electronic polarization), and it is a characteristic found in only those dielectrics in which the ionic character of bonds is expressed in the molecules or in crystal lattice. Typical representatives of dielectrics, in which ionic polarization plays a decisive role, are alkali-halide crystals (AHCs; such as NaCl). The quasielastic ionic polarization is also present in the $A^{III}B^V$ -type semiconductors (e.g., GaAs) as well as in $A^{II}B^{VI}$ -type crystals (e.g., CdS). However, in the “pure covalent” semiconductors (such as Si), ionic polarization is absent (where the mechanism of quasielastic displacement of electron shells of atoms dominates). In many *active dielectrics*—piezoelectrics, pyroelectrics, and ferroelectrics—ionic polarization is the main mechanism of electrical response to the applied field.

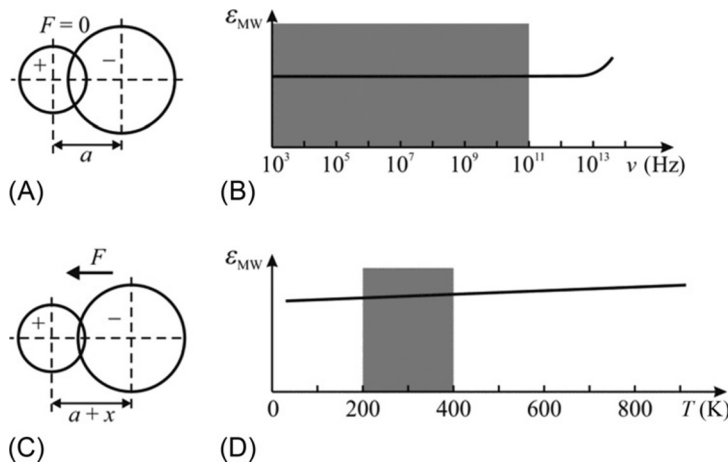


FIG. 7.11

Ionic polarization: (A) a simple model consisting of one positive and one negative ion; (B) deformation x in local field F ; (C) frequency dependence of ionic permittivity; (D) ionic permittivity temperature dependence (*shaded areas* show range of technical interest).

Another important feature of ionic polarization is the much greater response time than settling time of electronic polarization. This is conditioned by the large mass of ions as compared with that of electrons. However, the reaction time of ionic polarization ($\sim 10^{13}$ s) is still much smaller than characteristic times of thermal polarization ($\sim 10^{-6}$ s) and, even more, relaxation time of space-charge polarization ($\sim 10^{-1}$ s). As mentioned previously, using this it is possible to experimentally find the contribution of ionic polarization to permittivity while investigating permittivity frequency dependence $\epsilon(\nu)$ (Fig. 7.5B and Fig. 7.11C).

Dielectric dispersion of ionic polarization occurs in the IR *frequency range*. Therefore, even in the microwave range (MW, 10^9 – 10^{11} Hz), complete ionic polarization can occur, while thermally induced and space-charge polarizations do not occur even in a much lower frequency range.

To calculate the *polarizability* of ionic quasielastic polarization α_i , a simple model, shown in Fig. 7.11A, might be used [4]. The model has two ions, and these ions may represent two sublattices—cationic and anionic—inserted one inside the other, thus forming an ionic crystal.

This model takes into account the Coulomb attraction of ions, as well as the repulsive force, arising from partial interpenetration of electron shells. It is assumed that charges $\pm q$ are concentrated in the centers of ions; therefore, $r = r_1 + r_2$ is the distance between centers of ions. It is obvious from Coulomb law that energy of mutual attraction of ions decreases proportionally to distance between them: $q^2/4\pi\epsilon_0 r$. The repulsive energy of electron shells fast increases only in case of strong convergence of ions that is described by approximation with power function d/r^n , where parameter $n \sim 8$ – 12 depends on properties of particular pair of ions in crystal lattice. Coefficient

d can be determined from other parameters of this same model: $d = \frac{q^2 a^{n-1}}{4\pi n \epsilon_0}$. Thus interaction energy of ions is determined by the following expression:

$$U(r) = \frac{q^2 a^{n-1}}{4\pi n r^n \epsilon_0} - \frac{q^2}{4\pi r \epsilon_0}$$

The final result for ionic polarizability calculation is:

$$\alpha_i = \frac{4\pi a^3 \epsilon_0}{n-1} \approx \frac{4\pi (r_+ + r_-)^3 \epsilon_0}{n-1}$$

where r_+ and r_- are ionic radii while parameter $n \sim 10$ describes electron shell repulsion (factor $4\pi\epsilon_0$ is required to record polarizability in SI). As the overlap of electron shells is small, the distance between centers of ions is almost equal to the sum of radii of two ions. Thus, from this model, it follows that polarizability α_i is close to the polarizability of electronic elastic polarization ($\alpha_e = 4\pi\epsilon_0 r_e^3$, where r_e is the radius of electron orbital of ion).

As shown in Fig. 7.11C, permittivity, conditioned by ionic polarization, does not depend on frequency up to the far-IR range, where $\epsilon_i = \epsilon_{ir}$ gradually starts to increase (in far-IR range, ionic polarization shows the resonant dispersion). In most cases, contribution to permittivity from ionic (lattice) somewhat increases with temperature due to thermal expansion of crystal (Fig. 7.11D). The point is that at higher temperatures the distance between ions increases, so their reciprocal elastic shift in electrical field (elementary polarization) becomes larger. Due to this, the thermal coefficient of permittivity in ionic crystals is positive $TC\epsilon < +10^{-4} \text{ K}^{-1}$ ($\sim 100 \text{ ppm/K}$).

The *nonlinearity of ionic polarization* can be estimated using the same method, as in the case of electronic polarization. In the nonlinear case, it cannot be considered that relief of ionic potential is described by simple parabolic function and restoring force is proportional to strain. For the first two coefficients of series that describes nonlinear properties of ionic polarization mechanism, it can be assumed

$$\alpha_i(F) = \alpha_0 + \alpha_1 F + \alpha_2 F^2 + \dots$$

where

$$\alpha_0 = \frac{a^3}{n-1}; \quad \alpha_1 = \frac{a^5 (n+4)}{q(n-1)^2} \dots$$

While comparing this result with the nonlinear electronic polarization, it should be noted that in this model already *odd* coefficients (α_1) is nonzero; therefore, there is no necessity to take into account the next even coefficient (α_2 and more). This also means that potential well, which characterizes this model, is asymmetric [8].

Next, the question is how ionic elastic polarization becomes apparent in a variety of dielectrics. The important experimental evidence of ionic polarization presence in dielectrics and semiconductors is the far-IR *dispersion of permittivity*. It becomes apparent that ϵ_{ir} first increases (as far as measurements are close to the frequency of dispersion) and then decreases sharply (Fig. 7.5A). In the IR region strong

absorption is also observed that is very large in the “purely ionic” crystals and noticeable in ionic semiconductors with mixed ionic-covalent bonding.

In the lattice spectrum of $A^{III}B^V$ semiconductors, optically active modes in IR range are expected. In these compounds, the degree of ionicity and IR absorption is much higher than that in semiconductors of diamond structure. However, in crystals of $A^{III}B^V$ type the contribution of ionic polarization to permittivity is much smaller than that in AHCs. In semiconductors of this class, the degree of ionicity is characterized by dielectric contribution $\Delta\epsilon_{ion}$; the greater this contribution, the wider the bandgap. For example, in the InSb crystal bandgap is 0.17 eV and $\Delta\epsilon_{ion} = 1.4$, while in the InP bandgap is 1.3 eV and $\Delta\epsilon_{ion} = 3.7$. The value of $\Delta\epsilon_{ion}$ is determined according to microwave and optical measurements: $\Delta\epsilon_{ion} = \epsilon_{microwave} - \epsilon_{opt}$.

These data about IR polarization of semiconductors once again suggest that in these crystals the *mixed bonding* between atoms always has a place. Strictly speaking, all these types of bonding (molecular, covalent, ionic, hydrogen, metallic) are the idealized models of real phenomena, occurring in the material. Moreover, so-called ionic polarization, in fact, is mainly due to the shift of outer electron shells of ions. To obtain high dielectric constant, it is important to use dielectrics with more mobile and easily deformable outer electron shells.

7.5 THERMALLY ACTIVATED POLARIZATIONS

The electrons, ions, and dipoles, in addition to quasielastic polarization, may be involved in the quite different polarization mechanisms, namely, the thermally activated (relaxation) polarization. Thermal movement of particles in a dielectric may have strong influence on the polarization processes, when dipoles, ions, or electrons are *weakly bounded* in a structure of dielectric. Remaining localized in the surrounding nanovolumes, these particles, under the influence of thermal motion, can make thermally activated hopping, moving on a distance of the order of atomic dimensions.

Dipole thermal polarization mechanism was first proposed by P. Debye (1912) to explain high dielectric constant of water and other polar liquid dielectrics. At normal conditions and relatively low frequencies, dielectric constant of water equals $\epsilon = 80$, while at optical frequencies $\epsilon_{opt} = n^2 = 1.77$. Such a difference in water ϵ -value at low and very high frequencies is explained by the delay in polar molecule orientation in a fast-changing electrical field.

When external electrical field is absent ($E=0$), the dipoles are oriented randomly, and total electrical moment per unit volume is zero. If electrical field is applied ($E > 0$), then, in the process of thermal chaotic motion, *a part* of dipoles becomes oriented along the field with the result that a new equilibrium state arises—polarized. At that, the thermodynamic equilibrium is settled: due to thermal motions (vibrations, rotations) of dipoles that acquire orientation, favorable for the field. However, thermal fluctuations prevent full and stable orientation of all dipoles in the electrical field so that, on average, only the small part of dipoles becomes

oriented. The higher the electrical field strength (F), the greater the quantity of oriented dipoles per unit volume, and the bigger is thermally activated dipole polarization. Middle electrical moment, calculated per one molecule, is proportional to acting local electrical field F (if this field is not too large): $\langle p \rangle = \alpha_{dt} F$, where α_{dt} is correspondent polarizability [9].

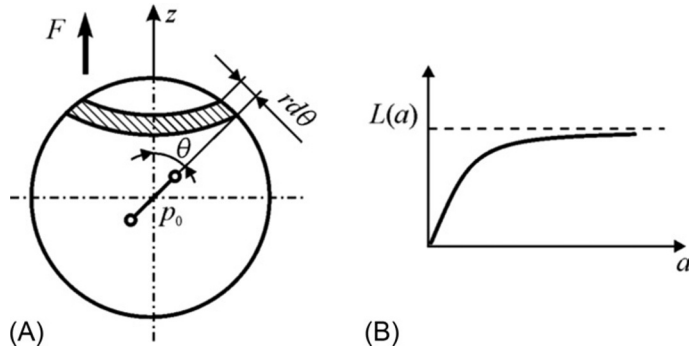
In polar *gases*, rotation of dipoles occurs freely. In polar *liquids*, the interaction of dipoles with surrounding molecules prevents free reorientation process: it manifests as a “friction” (or viscosity). In the *polar crystals*, the possibility of dipole reorientation is quite limited: usually, dipoles have only a certain number of possible stable orientations, separated by potential barriers (as shown in Fig. 7.4B). In this case, when an electrical field is absent, all dipoles are oriented uniformly in all permissible directions, but when an electrical field is applied, there is an *increase in the probability* of dipole orientation in the favorable direction. Also, it is not correct to assume dipole reorientation in a form of “mechanical rotation” of dumbbell-like polar molecules—in fact, this orientation (especially in solid dielectrics) means, for example, the *redistribution in electronic density* in molecules, or *proton hopping* between potential minima along hydrogen bonds, and so on.

The relaxation time of thermally activated polarization depends exponentially on temperature, decreasing very rapidly during dielectric heating. Under normal conditions (300 K), in dielectrics that exhibit thermally activated dipole polarization, relaxation time may be of 10^{-5} – 10^{-7} s. This means that frequency dispersion of thermal polarization takes place in the range of radio frequencies. Therefore, in dielectrics, in which this polarization mechanism becomes apparent, in radiofrequencies dielectric losses increase.

With the purpose of calculation of dipole polarization polarizability α_{dt} , it should be noted that previously (Section 7.4) elastic polarizabilities α_e and α_i calculation were considered in *elementary models* of polarization (one atom, one pair of ions, or one dipole); next, obtained results were generalized as sum of polarizabilities. However, in case of *thermally activated mechanisms* of polarization, it is necessary to consider *statistical models*, because only a few dipoles (or electrons, or ions) actually change their orientation (or they are redistributed in a volume of solid dielectric).

Thermally activated dipole polarizability α_{dt} primarily depends on relative concentration of dipoles that *dynamically* change their position (similarly, electronic thermal polarizability α_{et} depends on a fraction of actually oriented weakly bounded electrons, or ionic α_{it} depends on a certain proportion of weakly bounded hopping ions). In the case of thermal polarization, in order to calculate polarizability, which refers to a *single dipole*, it is necessary to find full dipole moment and divide it by the amount of dipoles.

In the Debye model, the reorientation of *statistical ensemble* of dipoles is considered, in which probability of given dipole orientation depends on the electrical field strength; by this way the *average value* of oriented dipoles is determined. As a studied model the spherical volume of dielectric is examined, which contains N dipoles (Fig. 7.12A). The *sphere* is chosen to simplify calculations, but this option does not limit the generality of calculation result. Dipoles have permanent electrical moment


FIG. 7.12

Thermally activated polarization of dipoles (Debye mechanism): (A) dipole moment calculation, (B) Langevin graph.

p_0 , and they are reoriented *independently* under the influence of random thermal motion. Acting on dipoles by local electrical field F changes this random orientation of dipoles into *partially oriented* state (in accordance with applied field), thus leading to induced polarization: $P = N \langle p \rangle = N \alpha_{di} F$, where $\langle p \rangle$ is average induced moment that appears in field F .

To find the $\langle p \rangle$, one needs to calculate the definite integral in spherical volume (shown in Fig. 7.12A): $\langle p \rangle = \int dp / \int dN$. Here dN is number of dipoles, oriented at the angle θ to arbitrary direction (denoted as axis z), that is, number of dipoles, that, being symbolically placed in the center of a sphere are oriented to “ring” between angle θ and angle $\theta + d\theta$. Electrical moment produced by these dipoles is designated as dp .

$$\langle p \rangle = \frac{\int dp}{\int dN} = \frac{\int_0^\pi C \cos \theta \sin \theta d\theta}{\int_0^\pi C \sin \theta d\theta} = 0$$

These integrals cover the entire volume of a sphere, that is, the angle θ varies from 0 to π . As it might be expected, in the absence of external electrical field the moment $\langle p \rangle = 0$; therefore any polarized state does not occur ($P = 0$), because dipoles are oriented randomly.

Next it should be assumed that an external electrical field is applied ($F > 0$), being directed along the axis z . Potential energy of a dipole that has own moment p_0 and is in local field F equals $U = -p_0 F = -p_0 F \cos \theta$. According to law of Boltzmann distribution, the probability of dipole orientation “in the ring” (angle between θ and $(\theta + d\theta)$) is defined as:

$$\exp\left(-\frac{U}{kT}\right) = \exp\frac{p_0 F}{kT} \cos \theta.$$

The number of dipoles, oriented in angle θ , in the electrical field changes: $dN = C \exp [(p_0 F / kT) \cos \theta] \sin \theta d\theta$, while electrical moment, produced by these dipoles, is $dp = C \exp [(p_0 F / kT) \cos \theta] p_0 \cos \theta dN$. Next, it is necessary to calculate the ratio $\langle p \rangle / p_0$, using the obtained dp and dN expressions, introducing notations $\frac{p_0 F}{kT} = a$ and $\cos \theta = x$.

$$\frac{\langle p \rangle}{p_0} = \frac{\int_{-1}^{+1} e^{ax} x dx}{\int_{-1}^{+1} e^{ax} dx} = \frac{e^a + e^{-a}}{e^a - e^{-a}} - \frac{1}{a},$$

or:

$$\frac{\langle p \rangle}{p_0} = \operatorname{ctha} - \frac{1}{a} = L(a)$$

The resulting expression is known as the *Langevin function* $L(a)$ that was first introduced in theory of paramagnetic susceptibility; graph of Langevin function is shown in Fig. 7.12B. In these calculations, no assumptions were made about the field value. Therefore the resulting solution is common—both in weak and in strong electrical fields, that is, similar for linear and nonlinear cases. Without assuming that effective field F is small, the expression obtained looks rather complicated. However, Langevin function can be expanded in a series in parameter a :

$$L(a) = \frac{a}{3} - \frac{a^3}{45} + \dots$$

In case of relatively small values of F , that is, at the condition $p_0 F \ll k_B T$ ($a \ll 1$), it is possible to use only the first term of the expansion.

Thus dipole polarizability, caused by thermal motion and electrical field, is given by the formula

$$\alpha_{dt} = \frac{p_0^2}{3kT}$$

This expression, which was first obtained by Debye, plays an important role in the theory of dielectrics. Unlike previous results (Section 7.4), the polarizability of thermally activated polarization explicitly *includes temperature*, wherein polarizability decreases with increasing temperature, because chaotic thermal movement prevents dipoles to get a fixed orientation in the external field.

The *nonlinearity* of thermally induced dipole polarization can be obtained on the basis of expansion:

$$\alpha_{dt}(F) = \alpha_0 + \alpha_1 F + \alpha_2 F^2 + \alpha_3 F^3 + \dots$$

$$\alpha_0 = \frac{p_0^2}{3kT}; \quad \alpha_1 = 0; \quad \alpha_2 = -\frac{p_0^4}{45k^3 T^3}; \quad \alpha_3 = 0.$$

As shown in these relations, the polarizability and, consequently, permittivity in dielectrics with thermally activated polarization in strong electrical fields *decrease*, because $\alpha_2 < 0$. In most dielectrics that are characterized by thermal polarization mechanisms, this nonlinear effect is experimentally noticeable, but it appears at the field strength that is close to electrical breakdown. This means that the inequality $p_0 F \ll k_B T$ is satisfactory only for a strong electrical field ($\sim 10^9$ V/m).

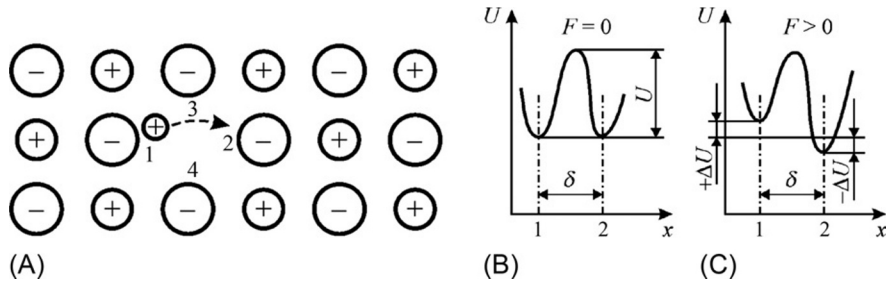
Significant nonlinearity in the acceptable fields (10^4 – 10^5 V/m) is a peculiar property of *paraelectrics* at temperatures close to phase transition of order-disorder type. However, in such paraelectrics, the mechanism of polarization is only to some extent reminiscent of the discussed thermally activated dipole polarization, as in the Debye model the interaction between dipoles is not taken into account. At the same time, in the partially ordered dipole-type paraelectrics, this interaction plays a decisive role [8].

Thermally activated ionic polarization is caused by thermal hopping of loosely coupled ions (usually, impurity ions) in the *local areas* of crystal lattice. Therefore such polarization is the characteristic of solid dielectrics with a large concentration of defects in their structure. Such are, for example, glasses, ceramics, and glass ceramics. In fact, these dielectrics have a high concentration of structural defects: glasses are generally characterized by short-range ordering in ion arrangement; ceramics have disordered boundaries between crystallites, while in pyrocerams the regular ordering in ions arrangement is broken.

However, thermally activated ionic polarization can be observed in the single crystals as well—in the vicinity of structural defects. The ions, located in the interstices, as well as the ionic vacancies (voids in a regular structure) may locally change their place under the influence of chaotic thermal fluctuations. During these movements that are limited by structural defects (e.g., dislocations), the ions overcome potential barriers and stay for a while in their new positions, representing electrical dipoles. However, in the absence of an external electrical field, this locally limited movement of charged particles is disordered, random, and usually cannot lead to a macroscopic polarization.

The external electrical field changes the distribution of ions near lattice defect sites with the result that electrically induced polarization arises. Response time of this kind of polarization (relaxation time τ) depends on temperature, peculiarities of crystal structure, and type of defects (or impurities). Typically, at ordinary temperature (~ 300 K) relaxation time equals $\tau = 10^{-4}$ – 10^{-8} s. Delay of polarization always leads to electrical energy absorption increase. Therefore thermally activated polarization may be the cause of dielectric losses at the radiofrequencies for commonly used dielectrics such as ceramics, glasses, and glass ceramics [6].

To calculate thermally activated ionic polarizability α_{ir} (with a purpose of relaxation polarization mechanism analysis), it is also necessary to use a statistical model. Thermally activated local hopping is possible only for weakly bounded ions that are localized in the vicinity of structural defects. Suppose that n_0 is concentration of such ions per unit volume: it equals 10^{18} – 10^{20} cm^{-3} that is much less than the total concentration n of ions in dielectrics (about 10^{23} cm^{-3}).


FIG. 7.13

Calculation of ionic thermally induced polarizability: (A) small positive ion of impurity is localized nearby one (1) of four possible negative ions, and it has an ability to jump in position (2); (B) in the absence of external field positions 1 and 2 are equally likely; (C) external electrical field stimulates jumps $1 \rightarrow 2$.

However, at every moment, only a part of these n_0 ions is involved in the mechanism of thermal polarization. At thermal chaotic motion loosely ions overcome some average potential barrier U , separating two (or more) possible locations of these ions. Obviously, the temporal localization of hopping ions can be maintained only in case of low temperatures: $U \gg k_B T$.

Along any particular direction in a dielectric, for example, along the x -axis (Fig. 7.13), on average, only $n_0/3$ of weakly bounded ions moves. The middle distance, separating the probable position of these ions localization, has the order of lattice constant ($\delta \approx 10^{-6}$ cm). Traditionally, parameter δ is called as the length of “free path.”

In Fig. 7.13, the equiprobable positions of ions in the potential minima are denoted as 1 and 2. The probability for moveable ion to acquire required energy is $\exp(-U/k_B T)$ that should be greater or equal to the height of barrier U (probability of ion hopping between the equilibrium position 1 and 2 at temperature T). In this case, the ion overcomes potential barrier and jumps, for example, from position 1 to position 2 (or vice versa). If the frequency of ion thermal vibrations in a lattice (Debye frequency) is ν Hz, from position 1 to position 2 (and meet) per one second next number of ions that jumps over is

$$n_{12} = n_{21} = \frac{n_0}{6} \nu \exp\left(-\frac{U}{kT}\right)$$

It is obvious that in case of such counterdiffusion electrical polarization cannot appear. If a dielectric is exposed to an electrical field F directed along the selected axis x , the probability of weakly bounded ion transition from position 1 to position 2 increases (Fig. 7.13B), while the probability of colliding transitions decreases. The fact is the height of potential barrier that ions must overcome in the first position is reduced by value ΔU , while in the second position it is increased by the same ΔU (which is the energy obtained by ion when it moves in an electrical field F on a distance $\delta/2$, i.e., $\Delta U = qF\delta/2$).

Thus electrical field disturbs uniform symmetry in the distribution of defect ions. At some time after the electrical field is applied, it turns out that $n_2 > n_1$:

$$n_1 = n_0/6 - \Delta n, \text{ and } n_2 = n_0/6 + \Delta n.$$

Such *electrostimulated local diffusion* of ions leads to polarization. As Δn represents excess concentration of ions, turned to 2 position, it is clear that electrically induced moment per unit volume of dielectric (polarization) is $P = \Delta n q \delta$.

In this model of thermally activated polarization, from total number of weakly bounded ions (n_0) only some of them (Δn) are actually jumped over the potential barrier. To determine a part of any *single* impurity ion in thermally induced polarization $P = \Delta n q \delta$, it is necessary to calculate the average elementary polarizability α_{iT} when $P = n_0 \alpha_{iT} F$. Thus the polarizability of thermal ionic polarization is

$$\alpha_{iT} = \frac{\Delta n q \delta}{n_0 F}$$

Through further calculations, it is possible to find the value Δn that depends on electrical field intensity. Moreover one can also find the time required for thermal ionic polarization settling; therefore while attempting to solve this problem, it is necessary to take into account dependence Δn on time. It is obvious that $\frac{d(\Delta n)}{dt} = -\frac{dn_1}{dt}$. As a result, polarizability of ionic thermal polarization is

$$\alpha_{iT} = \frac{q^2 \delta^2}{12kT} \left(1 - e^{-\frac{t}{\tau}}\right)$$

where relaxation time $\tau = (1/2\nu_D) \exp(U/k_B T)$. It can be seen that τ is exponentially dependent on temperature (increases very rapidly with decreasing temperature). If the electrical field is applied for quite a long time ($t \rightarrow \infty$), the thermal ionic polarization can be established as

$$\alpha_{iT} = \frac{q^2 \delta^2}{12kT}.$$

As can be seen, polarizability of ionic thermal polarization depends on temperature (decreasing with increasing temperature) because the intensity of thermal vibrations prevents the ordering of impurity ions in traps.

Nonlinearity of ionic thermal polarization. In the sufficiently strong electrical fields, the nonlinear properties of any polarization mechanism should be observed. Ionic thermally activated polarization, in this sense, is no exception. The nonlinearity should arise when a strong electrical field causes the ions to flip over through the potential barrier (in weak fields, jumps of ions are carried out by the fluctuations of thermal vibrations, and electrical field changes only their probability).

Nonlinearity might be calculated in the traditional way (as in case of elastic electronic polarization). When polarization has time to setting ($t \rightarrow \infty$),

$$\alpha_{iT}(F) = \frac{q\delta}{6F} \text{th} \frac{\Delta U}{kT}$$

passing to series expansion, $\alpha_{iT}(F) = \alpha_0 + \alpha_1 F + \alpha_2 F^2 + \alpha_3 F^3 + \dots$, the hyperbolic tangent series $\Delta U/kT$ is: $th \frac{\Delta U}{kT} = \frac{\Delta U}{kT} - \frac{1}{3} \left(\frac{\Delta U}{kT} \right)^3 + \dots$. Thus, for evaluation of thermal ionic polarization nonlinearity, we get:

$$\alpha_0 = \frac{(q\delta)^2}{12kT}; \quad \alpha_1 = 0; \quad \alpha_2 = -\frac{(q\delta)^4}{144(kT)^2}; \quad \alpha_3 = 0.$$

This means that in a strong electrical field, the so-called *saturation* of polarization is observed, and thus, dielectric constant decreases.

Relaxation time is also dependent on the electrical field:

$$\tau(F) = \tau_0 ch^{-1} \frac{\Delta U}{kT} = \tau_0 \left[1 - \frac{1}{8} \frac{q^2 \delta^2 F^2}{k^2 T^2} \right]$$

As it follows from this formula, in a strong electrical field (when forced barrier transfer by impurity ions takes place), thermally induced (hopping) polarization becomes quicker. Thus relaxation time of polarization as well as polarizability decreases in a strong field [8].

With rare exceptions, nonlinearity of thermal ionic polarization is observed only in very strong fields comparable with electrical field breakdown. However, given this consideration, nonlinearity of ionic thermal polarization mechanism has some theoretical interest. If crystal has increased permittivity, it is possible to obtain certain correlations in impurity ion orientations and, finally, to create artificial ferroelectric (an example is solid solution $(K_{1-x}Li_xTaO_3)$, where $x = 0.1-0.2$ is Li concentration).

Thermally activated electronic polarization is possible only in solid dielectrics. Suppose that in the neighborhood of certain type of structural defects in dielectric there are *weakly bounded electrons* (or electronic holes) that might be localized in two or more equivalent positions, separated by the potential barriers. Usually these electrons are captured by the crystal defects or impurity ions. These centers represent the places of irregularity in the electric charge distribution in a crystal lattice.

Captured by vacancies, electrons (holes) can lead to thermally induced polarization only when the ground state of electrons is degenerated; therefore the combination of correspondent wave functions can create dipole moments. In the absence of an electrical field but under the influence of thermal fluctuations, electron or hole successively passes from one place of location to another. Obviously, the *chaotic* movement of charges does not lead to polarization, if the external electrical field is absent. Field application stimulates unipolarity of electronic transitions and leads to being induced by a field electrical moment, that is, creates polarization.

Relaxation time of thermally activated electronic polarization is relatively long: $10^{-2}-10^{-6}$ s. This polarization is essential in many technically important dielectrics, such as rutile (TiO_2), perovskite ($CaTiO_3$), and similar complex oxides of titanium, zirconium, niobium, tantalum, lead, cerium, and bismuth. In these substances, especially in their polycrystalline state, a high concentration of defects in the crystal

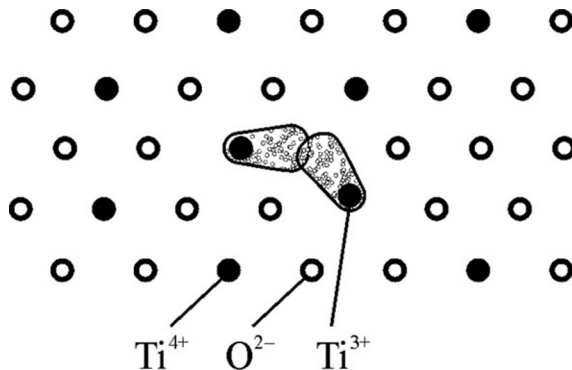


FIG. 7.14

Electronic thermal polarization in rutile: planar lattice model of titanium dioxide with oxygen anion vacancy.

structure is observed. During ceramic synthesis from the mixture of oxides (or while crystal growth), a very high temperature is used, and, therefore, the appearance of *oxygen defects*—anionic vacancies—is very likely. Electrical compensation of these defects occurs by the valence lowering of cations, located near anion vacancy. Thus appropriate conditions of electronic exchange between the neighboring cations appear that lead to polarization.

Fig. 7.14 shows one of possible cases of electronic thermal polarization in the rutile. In the selected section of crystal TiO_2 , there are three titanium ions as the neighbors to anion vacancy (in 3D crystal, there are five of such neighbors). Due to charge compensation, the oxygen vacancy is surrounded by two adjacent *trivalent* titanium ions, with each of them containing one “weakly bounded” electron in the outer shell. This model assumes that two electrons can overcome potential barriers and jump (redistribute) between five titanium ions, adjacent to oxygen vacancy.

If the electrical field is not applied, these transitions of electrons occur under the influence of chaotic thermal lattice vibrations; therefore these polar centers are oriented randomly that does not lead to total polarization. After electrical field application, the electrons around anionic vacancies are distributed asymmetrically that lead to induced polarization. In this manner, it is possible to describe electronic thermal polarization of rutile.

It should be noted that the concentration of defects in a crystal cannot be large; nevertheless, the contribution of such polarization mechanism to permittivity can be large enough, due to very high polarizability of greatly enhanced electronic orbits located near defects. Wave function of these “semifree” electrons is “smeared” in the large-enough area in the vicinity of vacancy. To estimate the reason for such high polarizability, it should be recalled that in case of electronic *elastic* polarization the value of polarizability $\alpha_e \sim r_e^3$, where r_e is electronic orbit radius.

The model, presented in Fig. 7.14, in some respects is equivalent to the mechanism of *ionic* thermal polarization. It can be argued that results of the statistical

model, obtained for ionic polarization, can be applied to the mechanism of electronic thermal (relaxation) polarization. Then, the polarizability of electronic thermal polarization is

$$\alpha_{et} = \frac{e^2 \delta^2}{12kT}$$

where δ is distance between equilibrium states of electron and e is charge of electron. In ceramic rutile (TiO_2) at ambient temperature, permittivity at frequency 10GHz equals $\epsilon_{\text{mic}} \approx 100$: this value is conditioned by fast electronic and ionic elastic polarizations [2]. However, at low frequencies (about 10kHz), depending on the concentration of anionic defects, due to electronic thermal relaxation permittivity may rise up to $\epsilon \approx 2000$.

A more convenient model for theoretical study as well as comparison with experiment would be ionic crystals with a simple structure: AHCs or A^{II}B^{VI} semiconductors, such as ZnS. The appearance of electronic relaxation polarization is associated with the excitation of color centers (F -centers). They are structural defects that arise as a result of electron localization near vacant anion sites in simple cubic crystal lattice. However, in the alkali-halide crystals with normal (low) concentration of F -centers, it is difficult to obtain reliable experimental data about contribution of electronic thermal polarization due to smallness of effect. It should be noted that color centers, such as F -centers, occur also in other crystals, such as quartz (model in Fig. 7.14 resembles F -center). Color centers are usually activated by hard radiation. It has been experimentally shown that in irradiated crystals of “smoky” quartz, relaxation polarization is observed due to defects in the vicinity of captured electrons or holes. In crystals of zinc sulfide or cadmium sulfide, thermally induced electronic polarization occurs after crystal *photoactivation* by ultraviolet radiation. At that, observed dielectric spectra are specific for the relaxation mechanism of polarization with temperature maximum of dielectric losses. In these experiments, it can be found that relaxation time depends on temperature according to relation $\tau = \tau_0 \exp(U/k_B T)$. However, thermally activated electronic polarization in ZnS and CdS crystals with impurities can be seen only after (or during) illumination of these crystals. Therefore the observed effect of dielectric constant increase is the *photodielectric effect*.

Thus thermally activated electronic polarization is related to quite a wide range of processes occurring in solid dielectrics: the photodielectric effect in the crystals of wideband luminescent semiconductor, the dielectric relaxation in ionic crystals due to color centers presence, the dielectric relaxation of electrons trapped in donor centers of oxide semiconductors, and, finally, the significant increase in low-frequency permittivity in polycrystalline materials of perovskite type and rutile.

Migratory polarization is always supported by thermal movement of charged particles. Except microscopic mechanisms of polarization, in the inhomogeneous dielectrics, electrical moment per unit volume might be originated by the *macroscopic* mechanisms of charge accumulation in areas of inhomogeneities. In some cases, this polarization plays a significant role, with respect to need for specific changes in dielectric properties and for electrical reliability control.

The migratory (space-charge) polarization is manifested only at low frequencies and causes considerable dissipation of electrical power (losses). The reason for this polarization might be the layers of different conductivity, or the presence of metallic or semiconductor particles in technical dielectrics, and so on. Migratory polarization mechanism can arise, for example, when the inhomogeneous material is placed in the electrical field: free electrons and ions of conductive and semiconductor inclusions move within each inclusion that becomes polarized region. In the same way, at the border and near-electrode layers, slowly moving ions are accumulated that also has the effect of migratory polarization. In case of migratory polarization, the movement of weakly bounded electrons or ions occurs at distances, which are much greater than crystal lattice constant.

As migratory polarization is a *macroscopic* mechanism, correspondent modeling and analysis should be carried out using equivalent electrical circuits. The simplest equivalent circuit of space-charge polarization in inhomogeneous dielectric is shown in Fig. 7.15.

Space-charge polarization is possible only at low frequencies. Therefore relatively rapid processes of thermal and elastic polarization are reflected in the equivalent circuit by capacitor $C(\infty)$, referring to sufficiently high frequency ($\omega \rightarrow \infty$), when slow polarization processes cannot appear. In addition, resistor R takes into account electrical conductivity in dielectric, while migratory polarization itself is represented in Fig. 7.15A by serial chain $r - C_a$.

To get parameters of dielectric in case of migratory polarization, it is necessary to go from equivalent circuit parameters $C(\infty)$, C_a , r , and R to ε and $\tan\delta$, which allow describing the contribution of migratory polarization into *effective* permittivity, as well as describing ε and $\tan\delta$ frequency dependence and temperature dependence.

If electrical conductivity can be neglected ($R \rightarrow \infty$), the analysis of equivalent circuit of migratory polarization essentially simplifies. The transition process for circuit $r - C_a$ (when voltage U_a on capacitor C_a decreases with time if chain is closed) is described by equation

$$rC_a \frac{dU_a}{dt} + U_a = 0,$$

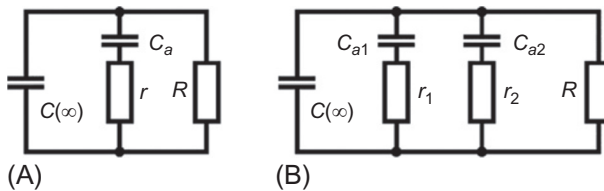


FIG. 7.15

Equivalent circuits of inhomogeneous dielectric space-charge (migratory) polarization in the case of one (A) and two (B) relaxation times.

whose solution is $U_a = U_0 \exp(-t/\tau_a)$ where $\tau_a = rC_a$. Complex conductivity (admittance) for simplified ($R \rightarrow \infty$) scheme shown in Fig. 7.15A equals

$$Y^* = i\omega \left[C(\infty) + \frac{C_a}{1 + i\omega\tau_a} \right]$$

Since the discussed model has both active and reactive conductivity, total conductivity is a complex value. Sharing its real and imaginary parts, it is possible to have

$$Y^* = \omega^2 \tau_a \frac{C_a}{1 + \omega^2 \tau_a^2} + i\omega \left[C(\infty) + \frac{C_a}{1 + \omega^2 \tau_a^2} \right] \quad (7.9)$$

Eq. (7.9) should be compared with known expression for capacitor conductivity. After this comparison, next formulas for migration polarization effective parameters can be obtained:

$$\epsilon' = \frac{C(\infty)}{C_0} + \frac{C_a}{C_0} \frac{1}{1 + \omega^2 \tau_a^2}, \quad \epsilon'' = \frac{C(\infty)}{C_0} \frac{\omega\tau_a}{1 + \omega^2 \tau_a^2} \quad (7.10)$$

Thus it is possible to designate low-frequency dielectric constant as $\epsilon(0)$ (before frequency dispersion, at $\omega \rightarrow 0$) and high-frequency dielectric constant $\epsilon(\infty)$ (after frequency dispersion):

$$\epsilon(0) = \frac{C(\infty) + C_a}{C_0}, \quad \epsilon(\infty) = \frac{C(\infty)}{C_0}$$

Dielectric contribution of migratory polarization equals $\epsilon(0) - \epsilon(\infty)$. Further consideration should be given for general scheme (Fig. 7.15A), without neglecting conductivity. Taking into account $rC_a = \tau_a$, it can be found

$$Y^* = \frac{1}{R} + i\omega \left[C(\infty) + \frac{C_a}{1 + \omega\tau_a} \right].$$

After separation of real and imaginary parts of this expression and using simple transformation, we get

$$\begin{aligned} \epsilon' &= \frac{C(\infty)}{C_0} + \frac{C_a}{C_0} \frac{1}{1 + \omega^2 \tau_a^2}, \\ \epsilon'' &= \frac{C(\infty)}{C_0} \frac{\omega\tau_a}{1 + \omega^2 \tau_a^2} + \frac{\sigma}{\epsilon_0 \omega}, \end{aligned} \quad (7.11)$$

where σ is conductivity taken into account in equivalent circuit by parameter R . Thus the model of an inhomogeneous dielectric, shown in Fig. 7.15, can be considered as sufficiently universal. When compared with the experiment, it may be that the relaxation spectrum of dispersion is more blurred than it follows from Eqs. (7.10) and (7.11). This circumstance may be corrected by inclusion in equivalent circuit several units $r_i C_{ai}$, describing space charge accumulation of in inhomogeneous dielectric with several relaxation times (Fig. 7.15B). Thanks to this method, the blurred dependence of $\epsilon(\omega)$ over a wide frequency range can be explained.

7.6 CLAUSIUS-MOSOTTI-LORENTZ EQUATION

A major problem in physics of dielectrics is macroscopic parameter calculation, namely, the permittivity ε derived from the foregoing expressions for various mechanism polarizabilities ($\alpha_e, \alpha_i, \alpha_d, \alpha_{et}, \alpha_{it}, \alpha_{dt}$) that are expressed through the molecular constants of dielectric. The ratio obtained in Section 7.1, in principle, allows calculating dielectric constant when total dielectric polarization P is known:

$$\varepsilon = 1 + \frac{P}{\varepsilon_0 E} \quad (7.12)$$

Obviously, if the dielectric shows several (k) *noninteracting* polarization mechanisms, electrical moment per unit volume of dielectric can be found from the following expression:

$$P = \left(\sum n_k \alpha_k \right) F, \quad (7.13)$$

where F is acting on particles electrical field, α_k is polarizability of k th mechanism, and n_k is particle concentration actual for k th polarization mechanism. From these models of elementary mechanisms of elastic and thermal polarization, $k = 1, 2, \dots, 6$.

Local (acting) electrical field in dielectric. It would seem that the expressions (7.12) and (7.13) can completely solve the problem of permittivity calculated through known molecular constants of dielectric—by polarizabilities. However, in order to use this expression, it is necessary to find the relationship between average macroscopic field E and acting on particles (local) field F .

Only in case of gases where molecules are spaced from each other at large distance, it can be assumed that $\mathbf{F} \approx \mathbf{E}$, and local field is $\mathbf{E} = (\mathbf{D} - \mathbf{P})/\varepsilon_0$. In the liquid and solid dielectrics, the local (acting) field is significantly different from the field E ; generally

$$\mathbf{F} \approx \mathbf{E} + \mathbf{\Gamma}, \quad (7.14)$$

where $\mathbf{\Gamma}$ is the resulting field, taking into account the impact on all other polarized particles on a given particle in a dielectric [3].

The general problem of field \mathbf{F} determination, excluding specific structural features of dielectric, is very challenging. G. Lorentz proposed a very important simplified solution to this problem, when local field \mathbf{E}_{Lor} can be found for dielectrics that have no polar molecules in their structure. At this condition, it appears that

$$\mathbf{F} = \mathbf{E}_{\text{Lor}} = \mathbf{E} + \mathbf{E}_1, \quad \left(E_{\text{Lor}} = \frac{\varepsilon + 2}{3} E > E \right). \quad (7.15)$$

Field $E_1 = P/3\varepsilon_0$ is conditioned by the action of all remote polarized particles on a given particle. Lorentz has proved that in *nonpolar* dielectrics, the influence of closely spaced particles cancels each other and can be ignored.

Using Eq. (7.15), that is good approximation for nonpolar dielectrics, it is possible to obtain a common relation between permittivity and polarizability. This ratio is named after Clausius-Mosotti, who first established the total between ε and α_k for

optical properties of dielectrics (i.e., only for special case of quasielastic electronic polarization). Later this relation was generalized by Lorentz for other mechanisms of polarization. However, in case of *polar* (liquid or solid) dielectrics, the Clausius-Mosotti-Lorentz formula cannot be applied (this is the “polarization catastrophe”).

As Lorentz approximation appears to be a very successful solution of the problem, relation (7.15) is possible to generalize for polar dielectrics as well. In the case for weak electrical fields (i.e., for linear dielectrics), the following expression can be proposed:

$$F = E + E_1 + E_2 \quad (7.16)$$

The field E_1 in this equation is the Lorentz correction, while the field E_2 takes into account the impact on polarized particles, located close to the concerned one (as already noted, for nonpolar dielectrics it can be considered $E_2 = 0$). The fields in expression (7.16) are shown in Fig. 7.16, and, respectively:

$$E = (D - P)/\epsilon_0; \quad E_1 = P/3\epsilon_0; \quad E + E_1 = E(\epsilon + 2)/3.$$

The field E_2 , in a general case, is projection of the internal field B on the direction of external field E (Fig. 7.16B); it takes into account the effect of polarized molecules adjacent to the considered particle. Local field calculations for polar dielectrics were conducted by many researchers. Solutions, obtained in some cases, are in good agreement with experimental results for individual polar dielectrics, but are not universal.

The most reliable solutions of this problem were obtained by L. Onsager, J. Kirkwood, and G. Frohlich [9]. Calculation of local fields in low-symmetry and dipole-structured dielectrics is considered one of the most difficult problems in the theory of dielectrics. It should be noted that quite a simple and, at the same time, important example of Lorentz theory application to the properties of dielectrics is the *ionic crystal* description of polarization. Developed by M. Born, the dynamic model allows not only to correctly explain far-IR polarization and electromagnetic absorption of ionic crystals, but also to establish a number of important relationships that

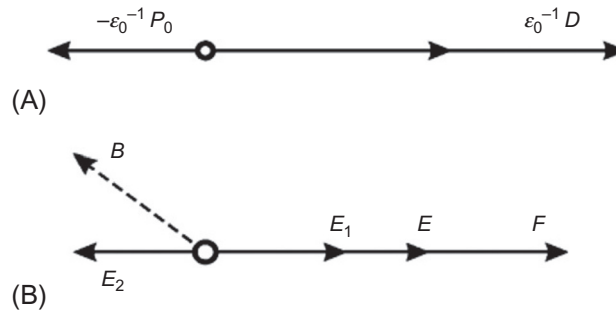


FIG. 7.16

Electrical fields in the dielectric: (A) average macroscopic field $E = D/\epsilon_0 - P/\epsilon_0$, (B) local field acting on each particle, $F = E + E_1 + E_2$.

include the ratio of Liddeyn-Sachse-Teller who have obtained the equation that connects dielectric constant, optical refractive index, and frequency of transverse and longitudinal optical phonons (see Chapter 4). Born's dynamic theory explains also specific properties of ionic crystals with very high dielectric constant.

Dielectric permittivity of gas. In gases the only elastic electronic polarization α_e and thermally activated dipole polarization α_{dt} should be considered. As interaction of molecules in gases may be ignored, it is natural to assume that the local field (acting on the molecule) differs only a little from the average macroscopic field ($F \approx E$), and static permittivity is not much greater than one. Using these relations and taking into account the foregoing assumption, the following formula can be derived for permittivity of gas:

$$\epsilon = 1 + \frac{\sum n_k \alpha_k}{\epsilon_0} = 1 + \frac{n_e \alpha_e + n_d \alpha_{dt}}{\epsilon_0}, \quad \alpha_e = 4\pi \epsilon_0 r^3, \quad \alpha_{dt} = p_0^2 / 3kT.$$

In the *nonpolar gases*, permittivity is determined by optical refractive index only: $\epsilon_{el} = \epsilon_{opt} = n^2 = 1 + n_e \alpha_e / \epsilon_0$, $\alpha_e = 4\pi \epsilon_0 r^3$. Assuming $\alpha_e \approx r_e^3 \approx 10^{-30} \text{ m}^3$, and taking into account that at normal conditions concentration of molecules in gas is $n_e = 7.7 \cdot 10^{25} \text{ m}^{-3}$, it is possible to estimate permittivity: $\epsilon_{el} = 1.0004$. Experiments show that this value is equal to 1.00055 for oxygen, 1.00027 for hydrogen, 1.00058 for nitrogen, but quite low for helium: $\epsilon_{el \text{ He}} = 1.00007$.

This is due to helium's nuclear (not molecular) structure. Therefore calculated ϵ_{el} in nonpolar gases is completely consistent with results of optical measurements of refractive index.

In the *polar gases* containing dipole molecules, such as steam H_2O , vapors of HCl , CO , NH_3 , and others, it is possible to divide dielectric contributions of electronic and dipole polarization by comparing optical refractive index n with the value of gas permittivity ϵ_{lf} measured at low frequency:

$$(\epsilon_{lf} - n^2) \epsilon_0 = \frac{N_d p_0^2}{3kT},$$

where p_0 is dipole moment and N_d is polar molecule concentration. Correspondent measurement technique is used for experimental determination of *dipole moments* of different molecules. Data on the value of dipole moment of complex organic molecules may be used for both deciphering of molecular structure and construction of correct models of these molecules.

Thus the theory of dielectric polarization of gases seems to be quite perfect. Calculation of permittivity by known molecular parameters has no fundamental difficulties, and the data obtained are in good agreement with measurements.

Lorentz model for local field calculation. A simplified method for local field calculation in homogeneous nonpolar dielectrics will be discussed later. For each polarized particle (ion, atom, molecule) surrounding them, dielectric is considered as a *continuous medium* that is characterized by certain macroscopic parameters. In fact, each particle is surrounded by adjacent particles and is influenced by the microscopic field of their neighbors. In gases, due to large distances between

molecules, the influence of these microscopic molecular fields can be neglected. However, in the condensed phases (liquid and solid dielectrics), as well as in highly compressed gases, such assumption is unacceptable.

For microscopic field calculation, Lorentz introduced a conception of *physically infinitesimal volume* that gathers around this molecule and forms a sphere of radius r . The assumption about spherical form is not critical and selected solely for simplification of calculations. The radius of Lorentz sphere is such that it is possible to take into account the influence of those particles that are located within sphere. Thus it is assumed that exclusively the influence of molecules, located in outside area, should be taken into account as polarized continuous medium. Accordingly, in the calculation, a sum of fields is used, $F = E + E_1 + E_2$, where E is average macroscopic electrical field, field E_1 describes the influence of distant molecules (located outside Lorentz sphere), and field E_2 characterizes the microscopic field of closest to the given molecule environment.

The macroscopic field E is generated by electric charges that are located *outside* the sphere and by polarization P and can be determined from Eq. (7.12): $E = \frac{P}{\epsilon_0(\epsilon-1)}$. To calculate Lorentz correction E_1 (field of polarized sphere), it should be assumed that all molecules inside Lorentz sphere are removed. Then the problem is limited to the calculation of electrical field, created *inside* the polarized dielectric sphere. The presence of associated electric charge on the surface of a sphere should be assumed—this representation is equivalent to polarization of empty sphere in dielectric (Fig. 7.17).

Electrical charge located on elementary surface dS is denoted by dq . As the integration relief, the elementary surface is selected: a ring on sphere, located at angle θ to the external electric field direction. Elementary charge dq creates the field in center of sphere: $dE_1 = \frac{dq \cos \theta}{4\pi\epsilon_0 r^2}$. The value of elementary charge, located on considered ring, is proportional to charge density and to ring surface: $dq = \rho_s dS$. The density of electrical charge depends not only on dielectric polarization value, but also on the angle of the elementary area (shown in Fig. 7.17B) that forms with macroscopic polarization direction, that is, $\rho_s = P \cos \theta$.

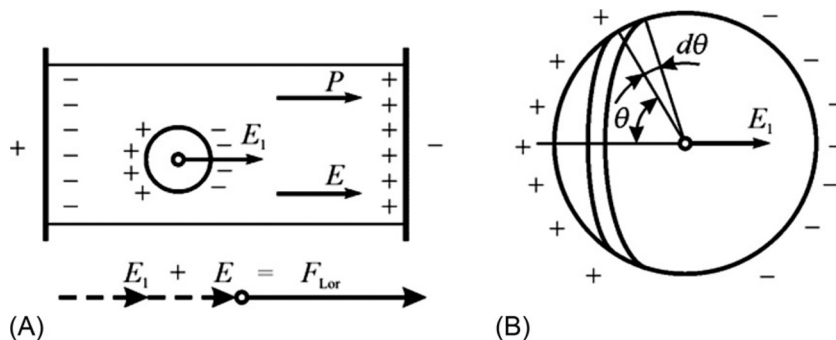


FIG. 7.17

Approximations to calculate Lorentz local field: (A) allocation of Lorentz sphere in dielectric; (B) calculation of polarized sphere field.

Surface of ring is determined by the parameters chosen in this task: $dS = 2\pi r^2 \sin\theta \cdot d\theta$. This expression should be substituted into formula for dE_1 , and after integration over entire surface of a sphere, the following can be obtained:

$$E_1 = \frac{P}{2\epsilon_0} \int_0^\pi \cos^2\theta \sin\theta d\theta = \frac{P}{3\epsilon_0}$$

Thus polarized dielectric medium that is located beyond Lorentz sphere creates in the center of sphere an electrical field $E_1 = P/3\epsilon_0$. In Lorentz approximation, the local field is determined by a sum $F = E + E_1$:

$$F = F_{\text{Lor}} = \frac{P}{\epsilon_0(\epsilon - 1)} + \frac{P}{3\epsilon_0} = \frac{\epsilon + 2}{\epsilon - 13\epsilon_0} \frac{P}{3} = \frac{\epsilon + 2}{3} E \quad (7.17)$$

It follows that Lorentz field *exceeds* the average macroscopic field in $(\epsilon + 2)/3$ times. Obviously, dense polarized media in liquid or solid dielectrics *increases* local field acting on a particle. However, in gases $\epsilon = 1$ and $F_{\text{Lor}} = E$, so the Lorentz model confirms previous calculations made for gases.

Now it is necessary to evaluate E_2 that is a field of polarized particles located *inside* a spherical cavity. In case of *polar molecules* (so-called hard, or *permanent* dipoles), the molecule concerned is affected by strong internal electrical fields induced by these dipoles, and these influences depend on a random thermal movement in the polar dielectric. In such cases, it is impossible to ignore field E_2 . However, in the nonpolar dielectrics these local fields are created only by the *induced* in external electrical field dipoles (“soft” dipoles), and in most cases the influence of their fields is *totally compensated*.

Lorentz approximation just meant a case of total compensation of local field, caused by the particles inside a sphere. This assumption, as already noted, is true for most of dielectrics. The point is that for each of polarized particles, located inside a cavity, it is always possible to find a particle whose action compensates the action of first particle. Such compensation is possible in case of disordered arrangement of atoms (or molecules) in the dielectric, that is, in case of nonpolar liquids. Note that during chaotic thermal molecular motions, some violations (fluctuations) are possible, but *on average* yet it is possible to assume field $E_2 = 0$.

A similar result can be obtained for *solid isotropic dielectrics*, for example, for many nonpolar solid-amorphous dielectrics in which $E_2 = 0$ can also be considered. However, in crystals compensation of inside-sphere local field is possible only for *highly symmetric* (simple) structures, and this is confirmed by calculations.

Lorentz approximation provides a relatively simple expression for permittivity calculation in the nonpolar and highly symmetric dielectrics, using known molecular parameters. From general expression (7.2), which binds polarization and dielectric constant, formula (7.17), in which effective field is expressed through a polarization, makes possible to obtain the Clausius-Mosotti-Lorentz equation:

$$\frac{\epsilon + 2}{\epsilon - 1} = \frac{\sum n_k \alpha_k}{3\epsilon_0} \quad (7.18)$$

The equation that is used to calculate dielectric constant of gases (when value of ϵ differs a little from unity) can be obtained from Eq. (7.18) as a special case, if $(\epsilon+2)=3$. With some approximation, Eq. (7.18) describes dielectric properties of nonpolar and weakly polar liquids and solid dielectrics with mainly electronic polarization: $\frac{\epsilon+2}{\epsilon-1} = \frac{n_e \alpha_e}{3\epsilon_0}$. Such dielectrics have ϵ value of $\sim 2-7$, which is close to the squared optical refractive index; they are nonpolar solid dielectrics (in which $\epsilon \approx n^2$), including paraffin, polystyrene, Teflon, rubber, and so on. Dependence of dielectric constant on temperature for such dielectrics is negligible; nonlinearity of these dielectrics is practically invisible.

Temperature parameter $TC\epsilon = \epsilon^{-1} d\epsilon/dT$, important for many applications, can be determined with the assumption that polarizability α_e is independent of temperature. By differentiating both sides of coerced equation with respect to temperature, and after some transformations, the following can be obtained:

$$TC\epsilon = -\frac{(\epsilon-1)(\epsilon+2)\beta_V}{3\epsilon} \quad (7.19)$$

where β_V is volumetric coefficient of thermal expansion. Indeed, in these dielectrics, temperature dependence of ϵ is due almost entirely to the thermal expansion. Resulting ratio indicates that parameter $TC\epsilon$ is negative for the nonpolar liquid and solid dielectrics, as in all of these substances $\beta_V > 0$ [6].

Thermally activated ionic and electronic polarizations, unlike the Debye dipole orientation mechanism, are possible mainly in solid dielectrics having imperfect structure. As a rule, structural defects are located sufficiently far from each other; therefore at high temperatures their interaction might be neglected.

Thus the Clausius-Mosotti-Lorentz equation can be considered as the basic relation that allows permittivity calculation from known molecular constants for most dielectrics.

7.7 DYNAMICS OF ELECTRICAL POLARIZATION

Electrons, ions, and dipoles create induced electrical moment in external electrical field (polarized status) through various mechanisms. If these particles are constrained in a relatively rigid but *elastic* structure, the influence of electrical field results in *very rapid* shifts of charged particles from their equilibrium state, and this is the quasielastic (*deformation*) polarization. On the contrary, when polarization establishes in time with a participation of thermal movement of particles (electrons, ions, or dipoles), the settling of polarization is a relatively *slow process* that is described by other dynamics (*relaxation*). Therefore, to analyze contributions of different mechanisms to a total electrical polarization, one has to distinguish between “fast” and relatively “slow” polarization processes, as well as consider electrical conductivity.

Suppose that dielectric at a certain time t_0 is exposed to electrical field E_0 , which then remains unchanged. When the field is turned on, the electrical current j arises in

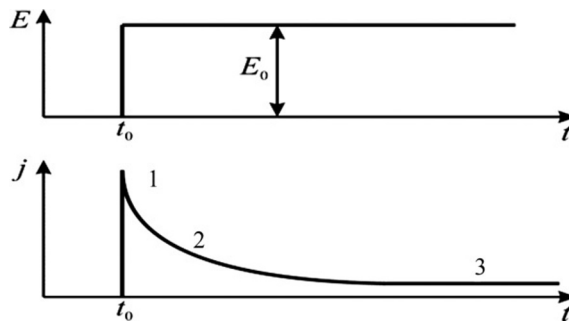


FIG. 7.18

Current density on time dependence after electric field switching on.

dielectric: its change with a time is shown in Fig. 7.18. In general, the $j(t)$ dependence can be divided into three characteristic regions. At the time of applying voltage E , the sharp and quick jump of electric current is observed due to establishment of “fast” types of polarization as well as the “geometrical capacity” charging. This peak of current takes a very short time and corresponds to a region 1 on the curve $j(t)$. Then, in the circuit containing the dielectric, characterized by the thermally induced polarization, a *gradual decrease* in current with a time occurs (region 2), which indicates the settling of “slow”-type polarization.

At that, the “fast” polarization processes correspond to different types of elastic (deformation) polarizations, while the “slow” decrease of polarization corresponds to the mechanisms of thermal (relaxation) polarization.

After some time the current, flowing through a dielectric, is reduced to a constant value: the *saturation current*. Corresponding plot 3 in Fig. 7.18 represents electrical conductivity that is usually very small in dielectrics, but it always has a finite value. Dependence of $j(t)$ that describes the decrease in current density after momentary application of a *direct* voltage can be used to calculate frequency dependence of permittivity and dielectric losses in the *alternating* voltage.

Polarization that is caused by particle thermal movement establishes relatively slowly. Its relaxation time τ depends on temperature, and at normal conditions (at 300K) equals $\tau = 10^{-3} - 10^{-9}$ s. Remember that dielectrics are used in electronics in the frequency range of 50– 10^{11} Hz, the frequencies of thermally activated molecular processes. Therefore, in many applications of dielectrics, any changes in the permittivity and losses (caused mainly by thermally activated electrical polarization) are quite undesirable. The space-charge polarization (i.e., an even more slow mechanism, $\tau = 10^3 - 10^{-3}$ s) also leads to the increase of losses on both infralow and low-frequency intervals.

On the contrary, elastic (deformation-type) mechanisms of polarization in the previously mentioned frequency range are set in time almost instantaneously; therefore these polarization mechanisms have no influence on losses in a large frequency range of 50– 10^{11} Hz.

Dynamic properties of relaxation polarization. Heat-induced jumps and reorientation of charged structural units in crystals, ceramics, or polymers (relaxation polarization) are caused by various polar groups—weakly bounded electrons (α_{et}) or ions (α_{it}) and by thermally stimulated reorientation of dipoles (α_{dt}). This relatively “slow” polarization process corresponds to the region 2 in Fig. 7.18. For these mechanisms, dielectric dispersion has a *relaxation character*: with increasing frequency, a gradual decrease in $\varepsilon(\omega)$ is observed that can be described by the Debye relaxation model, discussed later.

Further, the region 2 of general dependence $j(t)$ is considered, in which connection conductivity (region 3) is neglected, while “fast” polarization processes 1 in Fig. 7.18 are added as the contribution of high-frequency permittivity: $\varepsilon(\infty)$. If direct electrical field E_0 is applied to dielectric at time t_0 , the change of polarization over a time can be described by the expression

$$P(t) = n_0 \alpha_t E_0 \left[1 - e^{-\frac{t}{\tau}} \right],$$

where α_t is polarizability of thermally activated polarizations, while n_0 is concentration of particles, involved in the given polarization process. Since in capacitor model polarization P is *surface charge density* ρ_p , electric current is the change of this charge in time:

$$j_2 = \frac{d\rho_p}{dt} = \frac{dP}{dt} = \frac{n_0 \alpha_t}{\tau} e^{-\frac{t}{\tau}} E_0 \quad (7.20)$$

Index 2 in formula (7.20) indicates that dependence of $j(t)$ is considered only in the region 2 of Fig. 7.18, while other processes are not examined. From expression (7.18) the change in current density over time can be obtained as a result of “slow” processes of polarization: $j(t) = \frac{n_0 \alpha_t}{\tau} e^{-\frac{t}{\tau}} E$, where n_0 is concentration of particles involved in this polarization (dipoles, ions, electrons) and α_t is their polarizability. Taking into account rapid processes of polarization that dielectric contribution is referred as $\varepsilon(\infty)$, it is possible to obtain a general formula, known in literature as *Debye dispersion formula*:

$$\varepsilon^*(\omega) = \varepsilon' - i\varepsilon'' = \varepsilon(\infty) + \frac{\varepsilon(0) - \varepsilon(\infty)}{1 + i\omega\tau} \quad (7.21)$$

where $\varepsilon(0)$ is “quasistatic” dielectric constant (when $\omega \rightarrow 0$), while the difference $[\varepsilon(0) - \varepsilon(\infty)]$ represents contribution to permittivity of relaxation polarization.

Eq. (7.21) can be justified as follows. Suppose that static electrical field E_0 is turned on for a long time (enough to full settling of permanent polarization P_0). If the field is turned off at time $t = t_0$, the polarization, caused by thermal activation, gradually decreases (Fig. 7.19A). It is believed that the *rate* of polarization lowering after electrical field removal (i.e., the derivative dP/dt) is proportional to magnitude of $P(t)$. This assumption is based on a well-known principle of thermodynamics: the rate of system approach to the equilibrium state is proportional to the deviation from the equilibrium state. Thus the change $P(t)$ can be described by the following equation:

$$\frac{dP}{dt} = -\frac{1}{\tau} P,$$

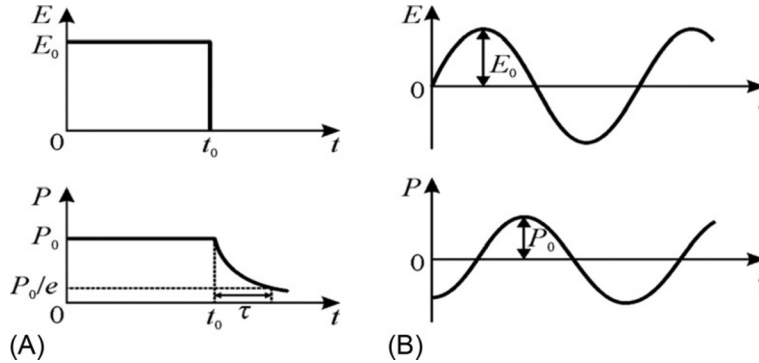


FIG. 7.19

Changing the thermal polarization with time: (A) when electric field is turned off, (B) in case of periodic field.

where $1/\tau$ is coefficient that depends on properties of dielectric and on temperature. The solution of this differential equation is $P = P_0 \exp(-t/\tau)$. From this result the meaning of τ could be found: it is a time, during which polarization is reduced by a factor of e in comparison with initial value P_0 . Parameter τ is the *relaxation time*, which describes a rate of polarization decrease after electrical field turning off (or, alternatively, τ defines the speed of $P(t)$ setting after electrical field switching on).

If a dielectric is exposed to *alternating* electrical field, $E(t) = E_0 \exp(i\omega t)$, Fig. 7.19B, the change of polarization over time can be described by the inhomogeneous differential equation of the first order:

$$dP/dt + (1/\tau)P = gE_0 e^{i\omega t}$$

where $P = (\sum n_i \alpha_i)E$ is thermally activated polarization (of electrons, ions, or dipoles), α_i is polarizability, and n_i is concentrations. Parameter $g = n_i \alpha_i / \tau$, by its dimension, characterizes the conduction, arising due to the *reactive current*.

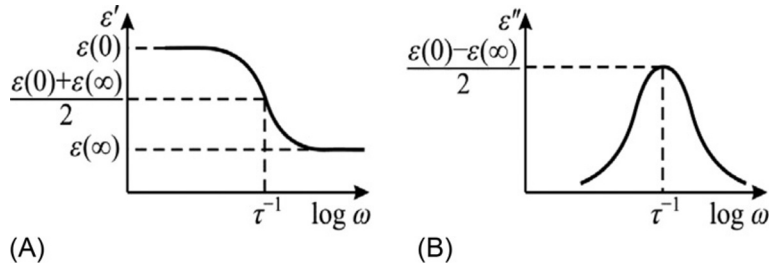
Stationary solution of this equation with assumption $t \rightarrow \infty$ and taking into account that $\epsilon = 1 + P/\epsilon_0 E$ makes it possible to obtain the formula for frequency dependence of permittivity:

$$\epsilon^*(\omega) = 1 + \frac{gt}{\epsilon_0(1 + i\omega\tau)}.$$

It is necessary to add contribution from fast mechanisms of elastic polarization that is designated $\epsilon(\infty)$. As a result, it is possible to obtain the Debye equation (7.19) that describes frequency dispersion of dielectric contribution from thermally activated (relaxation) polarization.

As it might be expected, dielectric permittivity is a *complex value*. By separating real and imaginary parts of dielectric constant in Eq. (7.21), it is possible to obtain:

$$\epsilon^*(\omega) = \epsilon(\infty) + \frac{\epsilon(0) - \epsilon(\infty)}{1 + \omega^2\tau^2}; \quad (7.22)$$


FIG. 7.20

Dispersion of permittivity for thermal (relaxation) mechanisms of polarization: (A) real part of dielectric permittivity; (B) imaginary part of dielectric permittivity (loss factor).

$$\varepsilon''(\omega) = \frac{[\varepsilon(0) - \varepsilon(\infty)]\omega\tau}{1 + \omega^2\tau^2}. \quad (7.23)$$

Eqs. (7.22) and (7.23) describe frequency dependence of real and imaginary parts of permittivity. At low frequencies $\varepsilon' = \varepsilon(0)$, while at high frequencies $\varepsilon' = \varepsilon(\infty)$, in which connection at frequency $\omega = 1/\tau$ the dielectric contribution of thermal polarization decreases exactly twice: $[\varepsilon(0) - \varepsilon(\infty)]/2$.

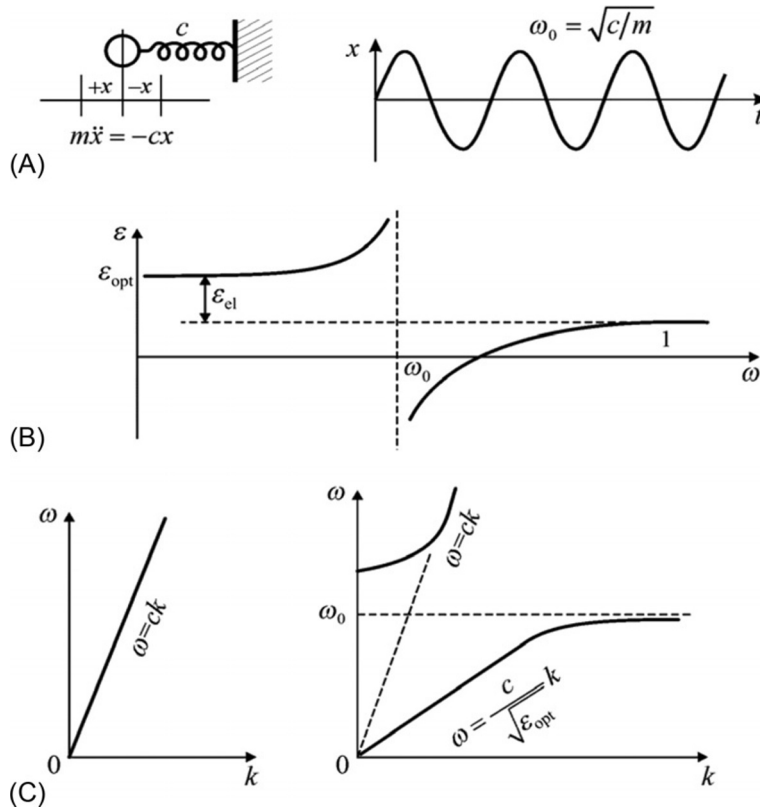
From formula (7.23) for $\varepsilon''(\omega)$ it follows that dielectric loss factor ε'' reduces at lower frequencies (when thermal polarization has enough time to settling) and at higher frequencies (when polarization mechanism is completely retarded). At this, the $\varepsilon''(\omega)$ has a maximum at frequency $\omega = 1/\tau$, which corresponds to a frequency when the dielectric contribution reduces in two steps with frequency increase (Fig. 7.20B).

Dynamic properties of optical polarization, as well as far-IR polarization, should be described using the model of harmonic oscillator (Fig. 7.21A). In this model, the particle of mass m and charge q is elastically connected with a “fixed base,” because the *outer* electronic shell shifts relatively easily as to the fixed core of atom (the core is nucleus with *deep* electronic shells).

In case of forced displacement of particle from its equilibrium position on distance $+x$ or $-x$ with overcoming elastic “spring,” a restoring force arises, which is proportional to displacement x and is directed oppositely: $f = -cx$. Parameter c is coefficient of elasticity of a bonding; in this case, it is the elastic connection of outer electrons with the core of an atom. The displacement force f balances the force of inertia $m(d^2x/dt^2)$:

$$m \frac{d^2x}{dt^2} = -cx \quad (7.24)$$

The energy of corresponding oscillator is $U = \int cxdx = cx^2/2$. This expression describes parabolic potential well (see Fig. 7.4A). The solution of Eq. (7.24) is vibrations of oscillator: $x = x_0 \cos \omega_0 t$ (or $x = x_0 \sin \omega_0 t$); the general solution should be presented in the form $x = x_0 e^{i\omega_0 t}$, where x_0 is amplitude and ω_0 is intrinsic frequency of oscillator.


FIG. 7.21

Oscillator model and electromagnetic wave dispersion: (A) oscillator; (B) ϵ -resonance in oscillator dispersion system; (C) dispersion n of electromagnetic waves in vacuum and in dielectric.

Corresponding permittivity can be found by analyzing oscillations arising at influence of periodical electrical field, by solving the equation

$$m \frac{d^2 x}{dt^2} + cx = q_0 F e^{i\omega t} \quad (7.25)$$

where right-hand side of equation is electrical force, acting on charges. At that, field F is different from applied field E : $F = E + P/(3\epsilon_0)$, where P is polarization (Lorentz approximation).

Without considering transient processes, particular solution of Eq. (7.25) for forced oscillations (of N oscillators per unit volume) can be found, using $P(t) = P_0 e^{i\omega t}$:

$$\frac{d^2 x}{dt^2} + \frac{cx}{m} = \frac{q}{m} \left(E_0 + \frac{P}{3\epsilon_0} \right) e^{i\omega t},$$

where $P = Nqx$;

$$\frac{d^2P}{dt^2} + \left(\frac{c}{m} - \frac{Nq^2}{3\epsilon_0 m} \right) P = \frac{Nq^2}{m} E_0 e^{i\omega t}, \quad P(t) = \frac{Nq^2}{m} \frac{E_0 e^{i\omega t}}{\omega_0^2 - \omega^2},$$

where $\omega_0^2 = \frac{c}{m} - \frac{Nq^2}{3\epsilon_0 m}$. Thus it is possible to describe frequency dependence of permittivity that characterizes elastic polarization at resonant dispersion:

$$\epsilon(\omega) = 1 + \frac{\Delta\epsilon}{1 - \left(\frac{\omega}{\omega_0} \right)^2}; \quad \Delta\epsilon = \frac{Nq^2}{\epsilon_0 m \omega_0^2} \quad (7.26)$$

At low frequencies the *dielectric contribution* of oscillators is $\Delta\epsilon = \epsilon_{\text{opt}}$ (Fig. 7.21B) that gradually increases as frequency approaches to the resonant value ω_0 . Above resonant frequency, the dielectric contribution decreases stepwise and becomes negative; next it again increases, crosses zero value, and at frequency $\omega \gg \omega_0$ reaches one.

Experimental dependences of $\epsilon(\omega)$ in the range of resonant dispersion are more smooth, than theoretical curve (shown in Fig. 7.21B) and calculated according to formula (7.26), where *attenuation* of oscillator is not taken into account (but it always observed in experiment).

If oscillator (7.24) describes elastic *electronic* polarization, then dielectric constant is $\epsilon_{\text{opt}} = 1 + \epsilon_{\text{el}}$. In accordance with magnitude ϵ_{opt} , light velocity inside crystal decreases: $v_{\text{cryst}} = c/\sqrt{\epsilon_{\text{opt}}}$ (where c is light velocity in vacuum).

Dependence of photon frequency on wave vector is compared in vacuum and in crystal (Fig. 7.21C). In vacuum, there is no dispersion of a light: $\omega = ck$. In dielectric at frequencies $\omega < \omega_0$ relationship $\omega = ck/\sqrt{\epsilon_{\text{opt}}}$ is true, while in the vicinity of $\omega \sim \omega_0$ dielectric dispersion occurs: phase velocity of light first decreases, because ϵ_{opt} rises with frequency, but then (in ultraviolet part of spectrum) the optical (electronic) polarization is delayed. At higher frequencies (at x-rays and gamma rays), electromagnetic waves that propagate in a crystal have velocity of light in vacuum, as any polarization mechanism at such high frequencies is absent.

Dynamic model of far-IR (ionic lattice) polarization was discussed previously in connection with phonons in solids (Section 4.3). Ionic crystals are a large and important class of dielectrics. In addition to quasielastic displacement of electron shells, the main polarization mechanism in these crystals is electrically induced *quasielastic displacements of ions*—charged particles bound in a crystal lattice.

Assuming that the oscillator model, described by expression (7.24), characterizes elastic ionic polarization (that has a much lower intrinsic frequency as compared with electronic polarization), in Eq. (7.26) it is necessary to replace $\Delta\epsilon = \epsilon_{\text{ir}}$, as ionic polarization undergoes dispersion that occurs in the *far-IR* frequency range. In addition, the $\epsilon(\omega)$ also contains fast contribution of more high-frequency electronic polarization ϵ_{opt} (optical contribution):

$$\epsilon(\omega) = \epsilon_{\text{opt}} + \frac{Nq^2 / (\epsilon_0 m \omega_{\text{TO}}^2)}{1 - (\omega / \omega_{\text{TO}})^2} \quad (7.27)$$

Here N is volumetric concentration of ions, q is ion charge, and m is reduced mass. The frequency ω_{TO} represents one of intrinsic frequencies of crystal lattice: transversal optical lattice vibration frequency.

The ionic polarization mechanism to a great extent determines ionic crystal permittivity. Below resonant frequency ($\omega < \omega_{\text{TO}}$) permittivity is determined by two contributions: $\epsilon(0) = \epsilon_{\text{opt}} + \epsilon_{\text{ir}}$. Above frequency ω_{TO} ionic polarization is delayed, and permittivity dispersion can be seen in the frequency range of 10^{12} – 10^{14} Hz. Therefore, at higher frequencies, only optical (electronic) polarization is possible.

The frequency ω_{TO} has a decisive influence on ionic crystal permittivity (Fig. 7.22). As an example, some crystals were chosen in such a way that the difference in the ϵ_{ir} constitutes about one order of magnitude. It is obvious that the less the frequency ω_{TO} the larger the dielectric constant.

Quasielastic lattice polarization in ionic crystals leads to the resonant dependence of permittivity on frequency that can be described using a dynamic model of crystal lattice. In this model, it is supposed that potential hole of each ion is well described by parabolic potential; hence ion vibrations are characterized by the harmonic oscillator model.

The simple model of “one-dimensional” crystal was considered previously in Section 4.3. In contrast to previously discussed oscillator with “fixed bearing” (Fig. 7.22A), in this model (Fig. 4.5A and Fig. 7.7) the displacement of each atom will affect the displacement of neighboring atoms such that elastic displacement will spread through the entire one-dimensional chain as elastic wave. Polarization, as usual, can be expressed in terms of induced dipole density N and dipole moment $p = qx$, so $P = Nqx$. As a result, oscillator equation takes the form

$$m \frac{d^2x}{dt^2} + \left(c - \frac{Nq^2}{3\epsilon_0} \right) x = 0$$

where intrinsic frequency of oscillator corresponds to frequency of *transverse* optical phonons

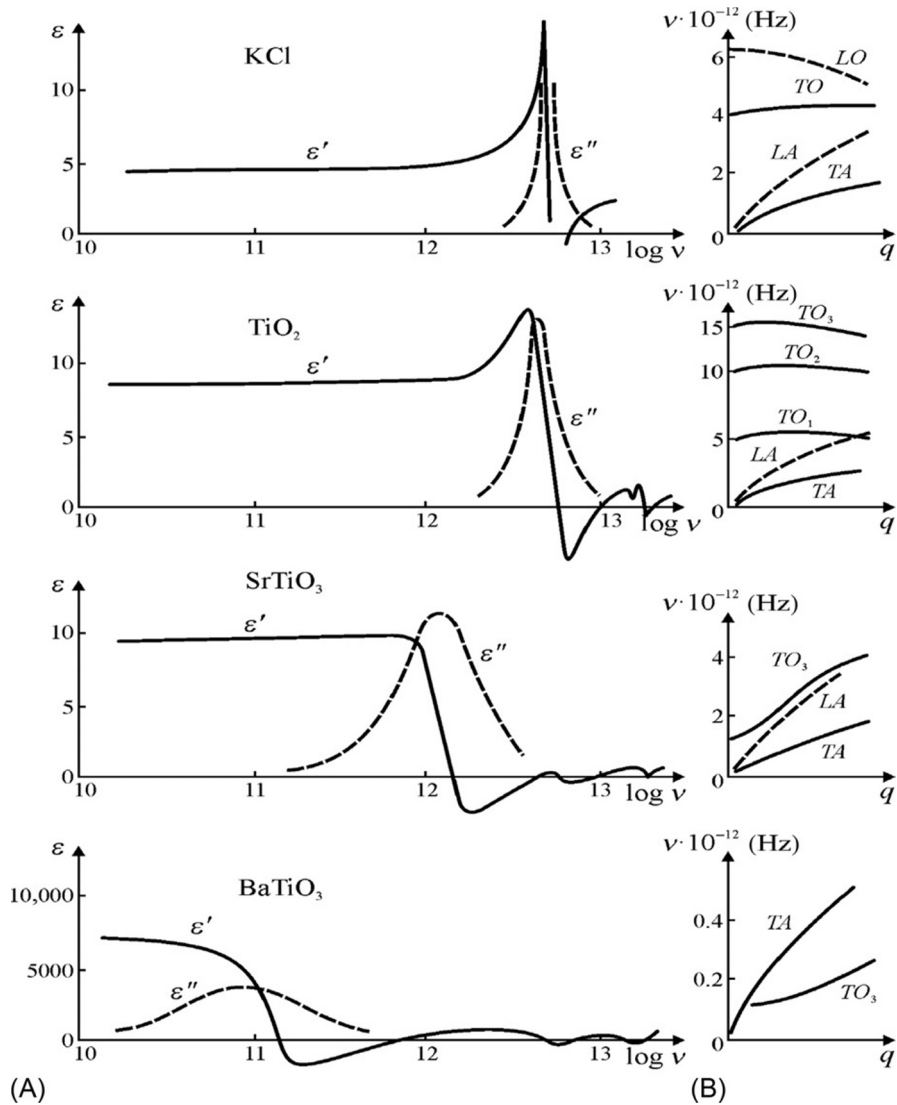
$$\omega_{\text{TO}}^2 = \frac{1}{m} \left(c - \frac{Nq^2}{3\epsilon_0} \right) \quad (7.28)$$

Thus, Lorentz field F reduces elastic constraint and, accordingly, decreases oscillator frequency $\omega_0 = (c/m)^{1/2}$ to the frequency of transverse optical phonons, that is, promotes “softening” of vibrations (reducing frequency ω_0 of oscillator: $\omega_{\text{TO}} < \omega_0$).

In case of *longitudinal* oscillations, local Lorentz field becomes significantly different, as electrical field E is directly opposite to polarization P : $\epsilon_0 E = -P$. The intrinsic frequency of such an oscillator that corresponds to the longitudinal vibrations is

$$\omega_{\text{LO}}^2 = \frac{1}{m} \left(c + \frac{2Nq^2}{3\epsilon_0} \right) \quad (7.29)$$

Therefore the frequency of an oscillator that characterizes longitudinal optical vibrations in the polarizable medium is higher than the frequency of an isolated oscillator


FIG. 7.22

Dielectric dispersion of infrared contribution ϵ_{ir} (A) and phonon spectrum (B) for some ionic crystals.

($\omega_0 = (c/m)^{1/2}$). Results, obtained in Eqs. (7.28) and (7.29), explain the locations of optical phonon branches: *LO* lies over *TO*; it is obvious that ω_{LO} in the ionic crystal is larger than ω_{TO} (see Fig. 3.5D).

The permittivity of ionic crystals depends on a difference in frequencies of longitudinal and transverse optical vibrations ω_{LO} and ω_{TO} in the *center* of Brillouin

zone. The equation, which describes IR polarization of ionic crystals, maintains the frequency of transverse optical phonons in the long-wave limit:

$$\epsilon(\omega) = \epsilon(\infty) + \frac{\epsilon(0) - \epsilon(\infty)}{1 - \left(\frac{\omega}{\omega_{\text{TO}}}\right)^2}, \quad \text{where } \epsilon(0) - \epsilon(\infty) = \frac{Nq^2}{\epsilon_0 m \omega_{\text{TO}}^2} = \frac{Nq^2}{c - \frac{Nq^2}{3\epsilon_0}} \quad (7.30)$$

This equation implies that the stronger the ionic polarization, the higher the ionic charge q and the lesser the elastic coupling coefficient c . Expression (7.30) allows quantitative calculation of IR contribution to permittivity. Indeed, ion concentration N can be found according to density of crystal, m is reduced mass of vibrating ions, q is ionic charge, and ω_{TO} is frequency of “residual” rays that can be determined experimentally by multiple reflections of IR waves from the surface of the studied crystal. Coefficient c that describes elastic coupling of ions can be calculated from the macroscopic elastic properties.

In most ionic crystals, permittivity ϵ_{mic} , measured in microwaves, is not large: $\epsilon_{\text{mic}} = \epsilon_{\text{opt}} + \epsilon_{\text{ir}} = 6\text{--}12$. Among two ion crystals (e.g., AHCs), there are no crystals with high permittivity that might be promising materials for applications in microelectronics. A comparison of electronic and ionic contributions to ϵ_{mic} for some oxides is shown in Fig. 7.23.

The more the density of crystal, the more the refractive index $n^2 = \epsilon_{\text{opt}}$ and the higher the polarization of ionic lattice, which is characterized by the difference in microwave and optical permittivity: $\epsilon_{\text{mic}} - \epsilon_{\text{opt}}$ [10]. Eq. (7.28) allows another way to write dispersion equation describing ϵ frequency dependence in the far-IR region. By eliminating parameter ϵ_{ir} , it is possible to obtain

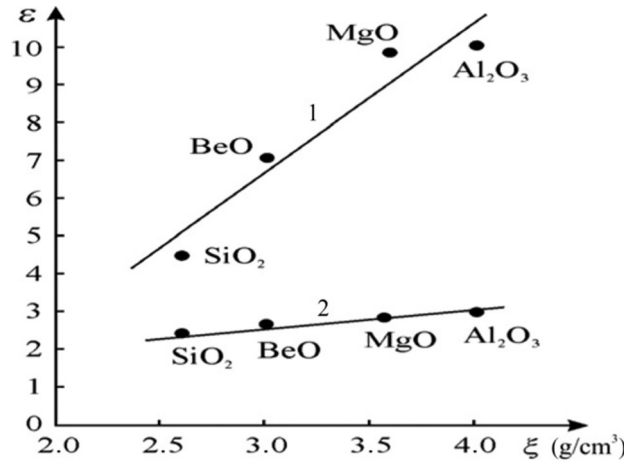


FIG. 7.23

Correlation between density ξ (g/cm³) and microwave ϵ_{mic} (1) as well as optical ϵ_{opt} (2) permittivity for some important electronic oxides.

$$\varepsilon(\omega) = \varepsilon_{\text{opt}} \frac{(\omega_{\text{LO}}^2 - \omega^2)}{(\omega_{\text{TO}}^2 - \omega^2)} \quad (7.31)$$

The obtained equation describes frequency dependence of permittivity $\varepsilon(\omega)$ that increases within all spectrum, except for the area of lattice resonance (Fig. 7.22).

As it follows from previous considerations, resonant frequency equals transverse optical frequency ω_{TO} , while permittivity at longitudinal frequency equals zero: $\varepsilon(\omega_{\text{LO}}) = 0$. By analyzing Eq. (7.30) with this assumption, we can obtain

$$0 = \varepsilon(\infty) + \frac{\varepsilon(0) - \varepsilon(\infty)}{1 - \left(\frac{\omega_{\text{LO}}}{\omega_{\text{TO}}}\right)^2},$$

and, after simple transformations, it follows

$$\frac{\varepsilon(0)}{\varepsilon(\infty)} = \frac{\omega_{\text{LO}}^2}{\omega_{\text{TO}}^2} \quad (7.32)$$

This is a very important relationship (that is referred to as Lidein-Sachs-Teller ratio) that describes the connection between total dielectric constant $\varepsilon(0) = \varepsilon_{\text{ir}} + \varepsilon_{\text{opt}}$ and frequencies of longitudinal and transverse modes. As always, $\varepsilon(0) > \varepsilon(\infty)$, the frequency ratio is $\omega_{\text{TO}} < \omega_{\text{LO}}$. However, in the covalent crystals of diamond type in the center of Brillouin zone $\omega_{\text{TO}} \approx \omega_{\text{LO}}$, so IR (ionic) contribution to permittivity is absent (such crystals are diamond, silicon, and germanium).

Electromagnetic waves that have frequency $\omega < \omega_{\text{TO}}$ extend in ionic crystal with velocity in $(\varepsilon_{\text{ir}} + \varepsilon_{\text{opt}})^{1/2}$ times less than in vacuum, where electromagnetic waves have light velocity: $c = k/\omega$. Therefore, in ionic crystal, the velocity of electromagnetic waves is reduced. At that, in the frequency interval between ω_{TO} and ω_{LO} ionic crystal-insulator *reflects electromagnetic waves* like metal; hence in this frequency range ionic crystal is not transparent (opaque). Its transparency is recovered at frequencies $\omega > \omega_{\text{LO}}$, but velocity of electromagnetic waves in ionic crystal is still less than light velocity in $(\varepsilon_{\text{opt}})^{1/2}$ times because it is slowed down by optical dielectric constant.

Obtained formulas (7.28) and (7.29) also allow to clarify the nature of permittivity temperature dependence in the ionic crystals. Because of thermal expansion, crystal decreases its density. On the magnitude of electronic part of permittivity $\varepsilon(\infty) = \varepsilon_{\text{opt}}$ thermal expansion is affected by ε_{opt} reduction with temperature increase (because of which $TC\varepsilon_{\text{opt}}$ is negative). In contrast, essential *increase* in the $\varepsilon_{\text{ir}} = \varepsilon(0) - \varepsilon(\infty)$ should be expected when temperature rises. Indeed, due to thermal expansion of crystal, the distance between ions increases and, therefore, their interaction weakens. Consequently, elastic coupling of ions (c) decreases. As c is located in the denominator of formula (7.28), if it decreases, dielectric contribution ε_{ir} increases. In most ionic crystals, contribution ε_{ir} exceeds ε_{opt} ; therefore in micro-waves $\varepsilon_{\text{mic}} = \varepsilon_{\text{ir}} + \varepsilon_{\text{opt}}$ increases with temperature, so $TC\varepsilon_{\text{mic}} > 0$.

Different temperature dependence of permittivity. The contributions to permittivity of various polarization mechanisms are compared in Fig. 7.24.

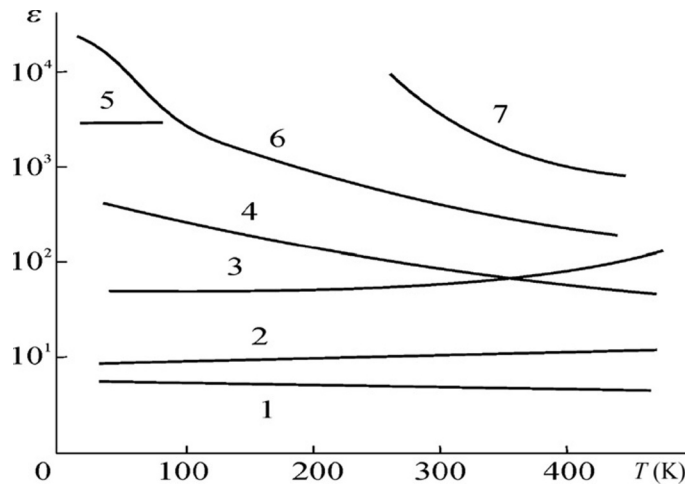


FIG. 7.24

Typical temperature dependence of permittivity for different dielectrics 1—optical (electronic) polarization; 2—optical and infrared (ionic) polarization; 3—single-domain ferroelectric; 4—“hard” type paraelectrics (CaTiO_3); 5—crystal TTF-TCNQ; 6—“soft” type paraelectric (SrTiO_3); 7—typical ferroelectric in paraelectric phase: (Ba , Sr) TiO_7 .

Optical (electronic) polarization, with less inertia, can provide extremely low microwave loss and high thermal stability: thermal coefficient of permittivity $TC\epsilon_{\text{opt}} < -10 \text{ ppm/K}$ (Fig. 7.24, curve 1).

The far-IR (ionic) polarization mechanism is possible only in crystals with ionic bonding and it is seen not only in alkali-halides but also in large classes of metal-oxides and chalcogenides. Semiconductors of the $A^{\text{III}}B^{\text{V}}$ and $A^{\text{II}}B^{\text{VI}}$ classes are also partially ionic crystals. In addition to electronic polarization, far-IR polarization produces dielectric contribution ϵ_{ir} with low microwave loss and, sometimes, it shows acceptable thermal stability of ϵ_{mic} (curve 2 in Fig. 7.24). In this case, the contribution of optical polarization ϵ_{opt} (that is described by small but negative temperature coefficient $TC\epsilon_{\text{opt}}$) partially compensates positive thermal coefficient $TC\epsilon_{\text{ir}}$.

In most of ionic crystals, permittivity ($\epsilon_{\text{mic}} = 6\text{--}12$) is not sufficient for many microwave electronic applications (resonators, filters, etc.). Indeed, the majority of alkali-halides have $\epsilon < 10$ that slightly increases with rise in temperature (Fig. 7.24, curve 2) (however, among alkali-halides, there are exceptions, namely, thallium-halides, in which $\epsilon_{\text{mic}} \approx 30$ and *decreases* with increasing temperature) [11]. Different values of microwave permittivity may be observed in oxides and chalcogenides: in them $\epsilon_{\text{mic}} = \epsilon_{\text{opt}} + \epsilon_{\text{ir}} = 5\text{--}15$. Note, that in some oxides, such as Al_2O_3 (where $\epsilon_{\text{mic}} \sim 12$), thermal stability is good ($TC\epsilon \sim 10^{-5} \text{ K}^{-1}$). Due to low losses and high thermal conductivity polycrystalline Al_2O_3 (polycor) and single crystal Al_2O_3 (sapphire) are widely used in microwave technology in spite of not very high ϵ_{mic} .

Other ionic compositions. Displacement-type paraelectrics, as well as antiferroelectrics and ferroelectrics, are strictly the materials with a very high and no-inertia

ionic-type polarization. As regards ferroelectrics, the only single-domain crystals with very high Curie temperature are capable of providing at microwaves $\epsilon_{\text{mic}} = 30\text{--}100$ (as in LiNbO_3 or LiTaO_3). Single-domain crystals of “hard” displacement-type ferroelectrics (ferroelectrics with very high Curie temperature) do not show large microwave losses ($\tan\delta < 10^{-3}$) and have positive $TC\epsilon$ (Fig. 7.24, curve 3). On the contrary, “soft” displacement-type ferroelectric crystals, such as BaTiO_3 , have increased microwave losses and they are not thermally stable. As regards displacement-type antiferroelectrics (like NaNbO_3 and PbZrO_3), they also have high permittivity and relatively low dielectric losses at microwaves, because they show absence of ϵ_{mic} dispersion. Moreover, they have rather large and positive $TC\epsilon$: in principle, these antiferroelectrics might be used for thermal stabilization of microwave paraelectrics, in which the $TC\epsilon$ is also large, but negative (curves 4, 6, and 7 in Fig. 7.10).

As shown in Fig. 7.24, curve 5, there is a theoretical possibility to obtain thermal stable dielectrics with high ϵ_{mic} but at very low temperatures. These materials are based on the fast-response electronic-phonon mechanism of polarization in dielectrics with unstable electron spectrum (quasi-1D crystals of TTF-NCNQ type). Insulating phase with the value $\epsilon_{\text{mic}} \sim 2000$ is formed in these crystals below the conductor-insulator phase transition temperature. Due to the electronic nature of this polarization ϵ -dispersion in TTF-NCNQ-type crystals is absent up to a frequency of 300 GHz.

Nevertheless, most *crystals* are too expensive for use in mass technical applications in microwave technology. In practice, only *ceramics* are widespread in an overwhelming majority of microwave devices.

The semiconductors of $A^{\text{III}}B^{\text{V}}$ and $A^{\text{II}}B^{\text{VI}}$ (as well as some other types) are partially ionic crystals. For this reason, except electronic polarization ϵ_{opt} , they have the far-IR (ionic) contribution ϵ_{ir} to microwave permittivity. This polarization mechanism can be described by Lorentz equation and it takes place in the far-IR frequency range. Thus total permittivity ϵ_{mic} of semiconductors is practically independent in all radiofrequency range. As an example, in Fig. 7.10 microwave permittivity of GaAs is shown: it is frequency-independent up to 1000 GHz.

Similar studies are conducted with gallium phosphide ($\epsilon = 7.5$), indium antimonide ($\epsilon = 17.5$), zinc selenide ($\epsilon = 7.1$), zinc sulfide ($\epsilon = 7.3$), and crystalline selenium ($\epsilon = 10.4$). Measurements were carried out at a frequency of about 10 GHz at temperature of 300 K [11]. Permittivity of all these semiconductors does not vary with frequency as in silicon (Fig. 7.10A). Microwave losses in the wide-gap semiconductors at room temperature are very small. However, at increased temperature their microwave losses become considerable.

However, permittivity of semiconductors at microwaves shows perceptible change with temperature that is possible to observe at millimeter waves: in accordance with the theory of semiconductors, permittivity increases with increasing temperature: the greater the ϵ value the smaller the bandgap. As temperature rises, the bandgap of semiconductors decreases, so their permittivity *increases* with temperature rise (Fig. 7.10B).

Table 7.1 Thin Film Dielectric Constant for Gate Layers of Field-Effect Transistors

Simple Oxides	ϵ	Complex Oxides	ϵ
Al_2O_3	9–11.5	$\text{HfO}_2\text{-SiO}_2$	10–13
BaO	31–37	$\text{La}_2\text{O}_3\text{-SiO}_2$	16–20
CeO_2	18–26	$\text{Y}_2\text{O}_3\text{-SiO}_2$	10–11
HfO ₂	20–22	$\text{ZrO}_2\text{-SiO}_2$	10–13
La_2O_3	25–30	$\text{HfO}_2\text{-ZrO}_2$	20–25
Ta_2O_5	25–45	$\text{HfO}_2\text{-Al}_2\text{O}_3$	14–17
TiO_2	80–95	LaAlO_3	~25
Y_2O_3	11–14	SrZrO_3	~25
ZrO_2	22–25	SrTiO_3	$\epsilon \leq 250$
		$(\text{Ba,Sr})\text{TiO}_3$	$\epsilon \geq 400$

Integrated dielectric films. Further progress in the miniaturization of microelectronic devices depends on the magnitude of the dielectric-gate permittivity in high-frequency transistors, as well as the dielectrics with increased ϵ could be used as storage capacitors for dynamic computer memory (DRAM) devices. Currently, these devices already work in the microwave region. Prospective materials are listed in Table 7.1.

Replacement of thin films made of silicon dioxide ($\epsilon \sim 3.7$) or silicon nitride ($\epsilon \sim 6$) by dielectrics with permittivity $\epsilon = 20\text{--}400$ reduces the size of microelectronic devices many times. In these films, only materials with electronic (optical) and ionic (far IR) polarization can be used: both these polarization mechanisms determine microwave dielectric constant ϵ_{mic} .

7.8 DIELECTRIC LOSSES AND DIELECTRIC SPECTROSCOPY

Dielectric losses, describing electrical energy transformation into heat, are important electrophysical parameters of a dielectric. The magnitude of losses and their dependence on frequency and on temperature are conditioned by the mechanisms of polarization. Dielectrics are usually studied and estimated in the alternating electrical fields. Therefore it is very important to study absorption of electromagnetic energy at various frequencies for different mechanisms of polarization and conductivity. Dynamic properties of quasielastic and thermally activated polarization are highly different (Fig. 7.25); hence the mechanisms of dielectric losses are quite various.

In case of quasielastic polarization, the variance of $\epsilon^*(\omega)$ has a resonant character with maximum and minimum in the $\epsilon'(\omega)$ dependence. In case of thermal polarization, $\epsilon^*(\omega)$ dependence shows the relaxation nature and characterized by a gradual decrease of $\epsilon'(\omega)$ with frequency. In both cases, in the frequency range of dielectric dispersion the maximum of loss factor $\epsilon''(\omega)$ is obviously observed [4].

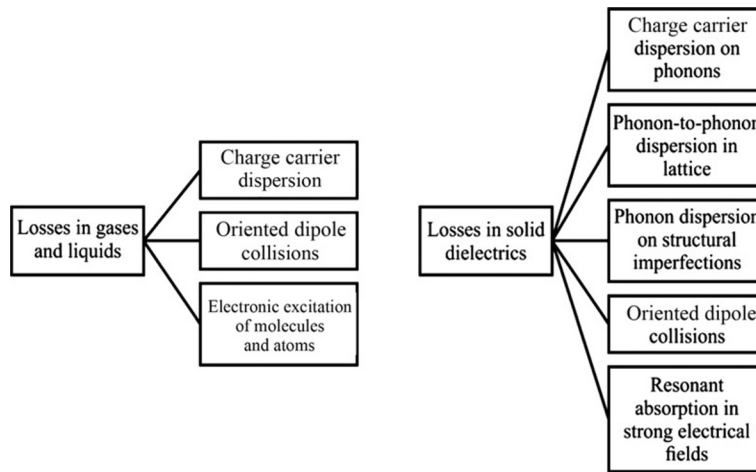


FIG. 7.25

Classification of dielectric loss mechanisms.

The equations that describe dielectric loss frequency dependence were presented in the previous discussions; they are summarized in Fig. 7.26. Dielectric relaxation at thermal and migration polarization that is described by the Debye equations (7.21) are repeated in this table wherein formulas for all basic mechanisms of losses are listed. Analysis of these formulas enables to describe the frequency dependence of parameters ϵ' , ϵ'' , and $\tan\delta$.

The simplest mechanism of losses is charge carrier scattering in a matter; in other words, this is electrical conductivity. Such losses, in one way or another, can be seen in all dielectrics—in gases, liquids, and crystals. Dispersion of charge carriers is due to their *collisions* with atoms and molecules (in disordered media), as well as to their *scattering* by lattice vibrations and structural defects (in crystals). Conductivity is the most important mechanism of electrical energy conversion into heat in the semiconductors.

A specific mechanism of energy absorption is losses, originated by polarization. In an alternating electrical field, the polarization of dielectric is always accompanied by electrical energy dissipation (any nonstationary process in real material is always more or less thermodynamically irreversible). At some frequencies, however, polarization losses can be very small, but they obviously are present.

Dielectric losses to a large extent depend on the presence of various impurities in a dielectric. In solid dielectrics, depending on the concentration of impurities or structural defects, the value of dielectric losses may vary by tens and hundreds of times, while, conditioned by defects, change of permittivity might be rather small. For this reason, dielectric losses are most *sensitive indicators* of structural defects in dielectric. Therefore the study of dielectric losses and their dependence on structural defects as well as on various factors (temperature, voltage, frequency) has considerable interest for physics and application of dielectrics.

Absorption Mechanism	Noninertia Conductivity	Thermally Activated Polarization (Relaxation Model)	Quasielastic Polarization (Oscillator Model)
Parameter:			
$\epsilon'(\omega)$	$\epsilon(\infty) = \epsilon$	$\epsilon(\infty) + \frac{\epsilon(0) - \epsilon(\infty)}{1 + \omega^2 \tau^2}$	$\frac{\epsilon(\infty) + [\epsilon(0) - \epsilon(\infty)] \left[1 - \left(\frac{\omega}{\omega_0} \right)^2 \right]}{\left[1 - \left(\frac{\omega}{\omega_0} \right)^2 \right]^2 + \Gamma^2 \left(\frac{\omega}{\omega_0} \right)^2}$
$\epsilon''(\omega)$	$\sigma/(\epsilon_0 \omega)$	$\frac{[\epsilon(0) - \epsilon(\infty)] \omega \tau}{1 + \omega^2 \tau^2}$	$\frac{[\epsilon(0) - \epsilon(\infty)] \Gamma \frac{\omega}{\omega_0}}{\left[1 - \left(\frac{\omega}{\omega_0} \right)^2 \right]^2 + \Gamma^2 \left(\frac{\omega}{\omega_0} \right)^2}$
$\tan \delta(\omega)$	$\sigma/(\epsilon_0 \epsilon \omega)$	$\frac{[\epsilon(0) - \epsilon(\infty)] \omega \tau}{\epsilon(0) + \epsilon(\infty) \omega^2 \tau^2}$	$\frac{\Gamma \frac{\omega}{\omega_0} \frac{\epsilon(0) - \epsilon(\infty)}{\epsilon(0)}}{\text{(at } \omega \ll \omega_0)}$
$p(\omega)$	$E^2 \epsilon_0 \epsilon \omega \tan \delta = \sigma E^2$	$\epsilon_0 \epsilon'' \omega E^2$	$\epsilon_0 \epsilon'' \omega E^2$
$\epsilon''_{\max}(\omega)$	—	$\frac{\epsilon(0) - \epsilon(\infty)}{2}$ (at $\omega = 1/\tau$)	$\frac{\epsilon(0) - \epsilon(\infty)}{\Gamma(1 - \Gamma^2/6)}$ (at $\Gamma < 1$)

FIG. 7.26

Basic formulas describing the various mechanisms of losses [4].

Mechanisms of dielectric losses, arising in an alternating electrical field, can be specified only by the study of dynamic properties of electrical response. It is necessary also to consider kinetic properties of molecules and atoms in a dielectric.

Dielectric loss tangent. In electrical engineering and electronics, the dielectric losses are characterized by the *loss tangent*: $\tan \delta$. However, the most general physical characteristic of losses, particularly for description of frequency dependences, is the complex dielectric constant:

$$\epsilon^*(\omega) = \epsilon'(\omega) - j\epsilon''(\omega), \quad \tan \delta = \epsilon''/\epsilon'$$

where $\epsilon' = \epsilon$ and ϵ'' is loss factor. In electrical engineering, to determine loss of electricity, parameter $\cos \varphi$ is commonly used, where φ is the angle between vectors of electrical field and current (Fig. 7.27).

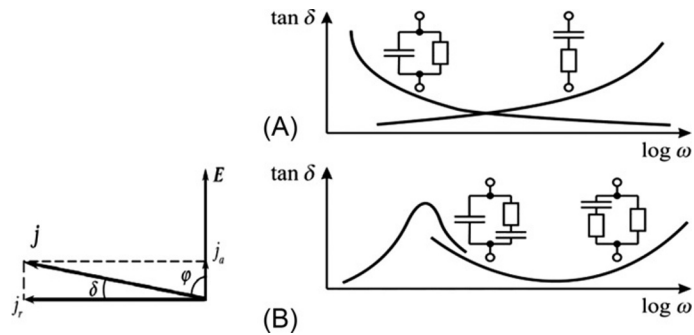


FIG. 7.27

Determination of loss tangent and its frequency dependence for different equivalent circuits that describe dielectric losses: (A) simplest representation; (B) more complicated cases.

However, to describe losses in dielectrics, this characteristic is inconvenient, as angle φ usually is a very little different from $\pi/2$. Thus, it is more comfortable to use the angle δ that supplements angle φ to $\pi/2$. Parameter $\tan\delta$ is numerically equal to ratio of active (conduction) current j_a to reactive (bias) current j_r . Parameter $\tan\delta$ is the *macroscopic characteristic* of dielectric as well as permittivity. Loss tangent dependence on temperature, frequency, electrical field, and other influences are similarly important characteristics of dielectrics as appropriate dependence of permittivity. It should be noted that $\tan\delta$ introduction as a characteristic of dielectric has physical meaning only in case of *alternating sinusoidal* electrical field.

In electronic circuits (where dielectric is very often used as electrical capacitor) to describe dielectric losses of *real* capacitor is very convenient by representing it as a combination of *ideal* capacitor and *ideal* resistor that simulates dielectric losses. Several equivalent circuits for dielectric loss description are shown in Fig. 7.27A and B.

A *parallel* connection circuit, when real dielectric is replaced by equivalent components (left-hand side in Fig. 7.27A), characterizes properties of such dielectric, in which $\tan\delta$ *decreases* with increasing frequency. Generally, this case refers to the losses, conditioned by conductivity.

On the contrary, when capacitor and resistor are *in series* connection, frequency dependence of $\tan\delta$ corresponds to polarization losses (right-hand side of Fig. 7.27A). All dependences of $\tan\delta(\omega)$ are shown in semilogarithmic scale that is usually used to describe frequency characteristics.

Thus, one or another equivalent circuit, necessary to describe real dielectric with losses, might be selected from frequency characteristics. Sometimes, it needs to use more complicated circuits that are shown with correspondent frequency dependences of $\tan\delta$ in Fig. 7.27B. In these cases, by combining various connections of capacitors and resistors, the almost complete matching of equivalent circuits with real characteristic of dielectric can be obtained. Fig. 7.27B shows equivalent circuits, corresponding to both a frequency maximum and a frequency minimum of $\tan\delta$.

Dielectric losses conditioned by conductivity can be described by a rather simple mechanism. While directed movement of charge carriers (their drift or diffusion in the external field), the carriers at their free path get energy from the electrical field. Acquired energy is spent in “collisions” that are, actually, the interactions of charge carriers with molecules and atoms, which are in a state of thermal motion. Returning acquired energy during collision, charge carriers increase the intensity of chaotic motion of particles of matter; therefore temperature of dielectric increases. For this reason, the electrical conductivity increases dielectric loss factor ϵ'' , parameter $\tan\delta$, and specific power p (energy dissipation per unit volume).

In this case, all these parameters depend only on the density of *active current*, flowing through the dielectric. Corresponding formulas are shown in Fig. 7.26. From them it follows that conductivity determines the magnitudes of ϵ'' and $\tan\delta$ mainly at low frequencies: both these parameters decrease with frequency as $1/\omega$. However, the specific power p in this case does not depend on frequency, because it is equal to the product of frequency-independent conductivity and squared electrical field ($p = \sigma E^2$). Thus, reducing ϵ'' and $\tan\delta$ with increasing frequency does not mean specific power p reduction in the dielectric with increasing frequency, as parameter p does not depend on frequency.

Frequency characteristics of considered parameters are shown in Fig. 7.28A. When no absorption mechanisms exist other than electrical conductivity, the permittivity is determined only by fast polarization processes: $\epsilon(\omega) = \epsilon(\infty)$, and it is independent of frequency. Temperature dependence of loss parameters, when the predominant mechanism is electrical conductivity, is shown in Fig. 7.28B. All of them, except $\epsilon(\omega) = \epsilon(\infty)$, show exponential increase with temperature, because, by this law, conductivity varies with temperature. It can also be seen that electrical conductivity contributes significantly to $\tan\delta$ and ϵ'' at *high temperatures* and at *low frequencies*. At very low temperatures and at very high frequencies, contribution of conductivity to dielectric losses usually is so small that it can be neglected.

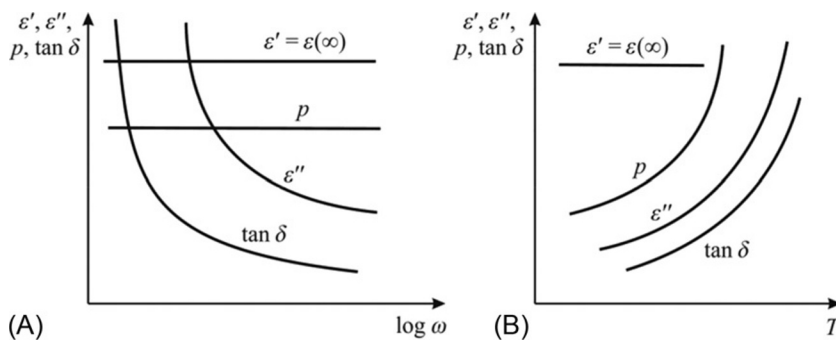


FIG. 7.28

Frequency (A) and temperature (B) dependence of basic parameters of dielectric, in which conductivity losses dominate.

In present-day electronics, dielectrics and semiconductors are widely used in the *microwave technique*. Usually their permittivity ($\epsilon = \epsilon'$) is independent of frequency. In the losses of semiconductors, the conductivity dominates; at that, conductivity is independent of frequency, including terahertz range due to high mobility of electrons. The loss coefficient $\epsilon'' = \sigma/(\epsilon_0\omega)$ and $\tan\delta = \sigma/(\epsilon_0\epsilon'\omega)$ decrease with frequency (Fig. 7.28A) because active component of current j_a is frequency independent, while reactive component j_r linearly increases with frequency and $\tan\delta = j_a/j_r$.

If in the specified frequency range, the dielectric has no relaxation or resonant ϵ -dispersions, parameter $\epsilon'(\omega)$ remains *constant*, while $\epsilon''(\omega)$ depends on conductivity σ and *decreases* with increasing frequency:

$$\epsilon^*(\omega) = \epsilon' - i \frac{\sigma}{\omega\epsilon_0},$$

This equation is not one of dispersion, inasmuch as describes only the $\epsilon''(\omega)$ dependence that is not concerned with $\epsilon' = \text{const}$. When σ is independent of frequency, conductivity *does not contribute to the real part of permittivity*.

Dielectric losses conditioned by thermally activated polarization. Relaxation polarization is caused by the local electrodiffusion process, at which weakly bounded charges are accumulated in definite localized states (or dipoles are directionally oriented). Being supported by thermal movement, this type of polarization is settled relatively slowly. Relaxation time of this polarization varies with temperature but lies in the range of 10^{-3} – 10^{-9} s. Thus, the distinguishing frequency of molecular relaxation processes in such dielectrics may be located in such a frequency range, where dielectrics are used in electrical engineering and electronics (50 Hz–100 GHz).

Dielectric relaxation at thermal and migration polarization is described by the Debye equations that are repeated in Fig. 7.26, where formulas of all basic mechanisms of losses are listed, including parameters that characterize relaxation losses. These formulas allow describing the frequency and temperature dependence of ϵ' , ϵ'' , p , and $\tan\delta$. At low frequencies ($\omega \rightarrow 0$) permittivity is denoted as $\epsilon' = \epsilon(0)$, while at high frequencies ($\omega \rightarrow \infty$) permittivity is $\epsilon' = \epsilon(\infty)$. While frequency is reduced, at frequency $\omega = 1/\tau$, dielectric contribution decreases exactly twofold (Fig. 7.29A). If conductivity is absent, loss factor vanishes ($\epsilon'' \rightarrow 0$), at very low frequencies (when $\omega \rightarrow 0$) and at very high frequencies (when $\omega \rightarrow \infty$). It is easy to show that $\epsilon''(\omega)$ has a maximum at the frequency of $\omega = 1/\tau$ when dielectric contribution to the permittivity is halved.

Frequency dependence of loss tangent also is characterized by maximum:

$$\tan\delta_{\max} = \frac{\epsilon(0) - \epsilon(\infty)}{2\sqrt{\epsilon(0) \cdot \epsilon(\infty)}}, \quad \omega_{\tan\delta_{\max}} = \frac{1}{\tau} \sqrt{\frac{\epsilon(0)}{\epsilon(\infty)}}$$

It is obvious that this maximum is observed at higher frequency than maximum of the loss factor $\epsilon''(\omega)$.

From loss power density p (formula shown in Fig. 7.26) follows that at low frequencies (when relaxation polarization is totally settled in time) specific loss power is very small (Fig. 7.29B). In the center of dispersion (when $\omega = 1/\tau$), density of loss

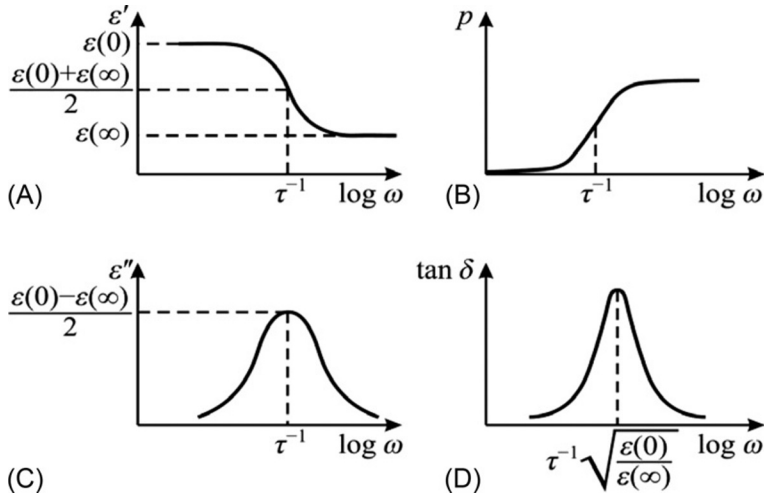


FIG. 7.29

Frequency dependence of permittivity (A), absorbed energy (B), loss coefficient (C), and loss tangent (D) for dielectrics, in which thermal polarization mechanism dominates.

power is $p = 0.5gE^2$, where parameter $g = \alpha_T \tau \epsilon_0$ is the *reactive conductivity*. At high frequencies, specific loss power p reaches its saturation gE^2 and next becomes independent of frequency (Fig. 7.29B).

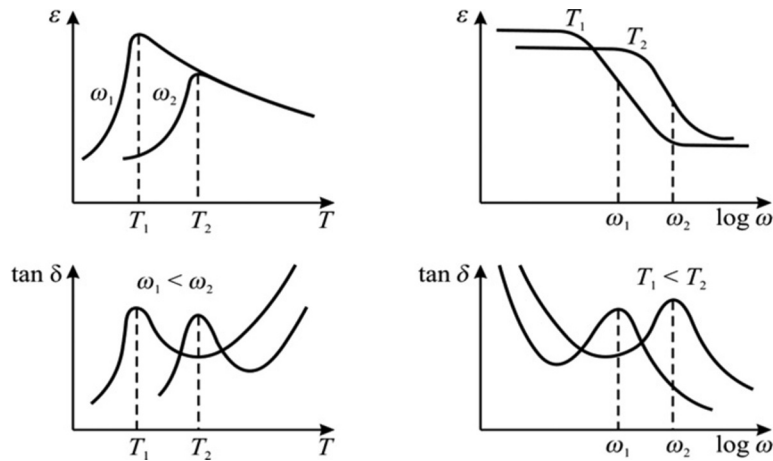
Thus, although relaxation polarization is delayed and no longer gives any dielectric contribution, specific power of losses in case of relaxation process is maximal. Therefore, for example, while high-frequency dielectrics are developing, the impurities and structural defects (that give low-frequency relaxation) are highly undesirable, because they do not appreciably affect ϵ -magnitude, but significantly increase losses.

Temperature dependences of both ϵ and $\tan \delta$ at relaxation polarization are characterized by highs. Temperature dependences of $\epsilon'(T)$ and $\tan \delta(T)$ are conditioned by temperature change of relaxation time $\tau = (2\nu_D)^{-1} \exp(U/k_B T)$. Here U is potential barrier (that charged particles need to be overcome at their thermal jumps), ν_D is Debye frequency (particles oscillation in crystal lattice), and k_B is Boltzmann constant.

Contribution to permittivity from thermally activated polarization is dependent on temperature

$$\epsilon(0) - \epsilon(\infty) = \frac{K}{k_B T}, \quad (7.33)$$

where K is Curie constant (in analogue with paramagnetic). Temperature-frequency dependence of main parameters of dielectrics in which not only thermal polarization but also electrical conductivity is clearly shown (Fig. 7.30).


FIG. 7.30

Permittivity and loss dependence on temperature and frequency in dielectrics with conductivity and thermally activated polarization.

Temperature change in permittivity in case of this type of polarization is characterized by the asymmetric maximum. Loss temperature dependence shows also maximum that is almost symmetric. Conductivity has no influence on the $\epsilon'(T)$ characteristics; however, losses conditioned by conductivity are superimposed on relaxation loss maximum. Loss tangent that is conditioned by conductivity increases with temperature but decreases with frequency.

Experimental dependences of $\tan\delta(T)$ and $\epsilon(T)$ are the basic data for determining the height of potential barrier U . For U calculation, it is sufficient to experimentally determine two temperatures T_1 and T_2 that correspond to two values of frequency ω_1 and ω_2 where peaks of losses are observed:

$$\frac{U}{e^{kT_1}} = \frac{\omega_2}{\omega_1} \frac{U}{e^{kT_2}}$$

from this it follows that

$$U = \frac{kT_1 T_2}{T_2 - T_1} \ln \frac{\omega_2}{\omega_1} \quad (7.34)$$

Theory of relaxation polarization is confirmed by many experiments.

Dielectric losses, conditioned by quasielastic polarization. Quasielastic displacement of electron shells in atoms, ionic sublattices in crystal, or rigidly connected dipoles creates a polarization. At that, restoring force is proportional to the displacement of particles from their equilibrium position. Particles that deviate from their equilibrium position can oscillate around the new equilibrium state. Therefore dynamic properties of quasielastic polarization can be described by the equation of *harmonic oscillator*, in which dielectric losses are taken into account by introducing

damping coefficient Γ . Dielectric dispersion in case of quasielastic polarization is characterized by Lorentz formula

$$\epsilon^*(\omega) = \epsilon' - i\epsilon'' = \epsilon(\infty) + \frac{\epsilon(0) - \epsilon(\infty)}{1 - (\omega/\omega_0)^2 + i\Gamma(\omega/\omega_0)}, \quad (7.35)$$

where $\omega_0 = \sqrt{c/m}$ is intrinsic frequency of oscillator, c is elastic constant, and m is oscillating mass; $\epsilon_{\text{osc}} = \epsilon(0) - \epsilon(\infty) = nq^2/\epsilon_0c$ is dielectric contribution of n oscillators; $\Gamma = \lambda/\omega_0$ is relative attenuation (λ is constant of “friction” that occurs from scattering mechanisms, listed in Fig. 7.26). The calculation formulas that permit dielectric absorption description by oscillator are listed in Fig. 7.26. Fig. 7.31 shows typical frequency dependences of ϵ' and ϵ'' for a damped oscillator.

Loss factor $\epsilon''(\omega)$ always is positive, while value of $\epsilon'(\omega)$ might be both positive and negative. At low frequencies (when $\omega \ll \omega_0$) both ϵ' and ϵ'' increase with rising frequency, while in vicinity of $\omega \approx \omega_0$ both of them have maximums. At further frequency increase, the dependences of ϵ' and ϵ'' are quite different. The real part of permittivity $\epsilon'(\omega)$ drops sharply, and at frequency $\omega = \omega_2$ reaches its minimal value (Fig. 7.31A). Then $\epsilon'(\omega)$ again increases with frequency and at sufficiently high frequencies ($\omega \rightarrow \infty$) reaches saturation: $\epsilon'(\omega) \rightarrow \epsilon(\infty)$. For quasielastic ionic

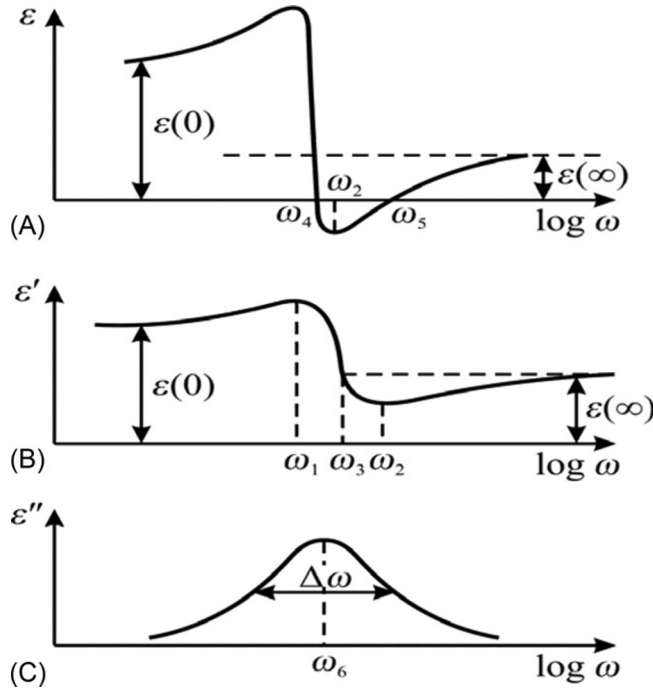


FIG. 7.31

Frequency dependence of permittivity (A, B) and loss factor (C) for resonant polarization at different parameters of equivalent oscillator.

polarization of crystals value $\varepsilon(\infty) = \varepsilon_{\text{opt}}$, that is, this is optical (electronic) polarization contribution.

Frequencies that define positions of $\varepsilon'(\omega)$ maximum and minimum can be obtained as $\omega_{1,2} = \omega_0 \sqrt{1 \pm \Gamma}$. When damping is small ($\Gamma \ll 1$), approximation $\omega_{1,2} = \omega_0(1 \pm \Gamma/2)$ is fair. Maximum and minimum values of permittivity are observed at frequencies ω_1 and ω_2 , and equal

$$\varepsilon_{\text{max}} = \varepsilon(\infty) + \frac{\varepsilon(0) - \varepsilon(\infty)}{(2 - \Gamma) \cdot \Gamma}, \quad \varepsilon_{\text{min}} = \varepsilon(\infty) - \frac{\varepsilon(0) - \varepsilon(\infty)}{(2 + \Gamma) \cdot \Gamma},$$

respectively.

Dielectric contribution from quasielastic polarization $[\varepsilon(0) - \varepsilon(\infty)]$ vanishes, when frequency is equal to ω_0 (intrinsic frequency of oscillator). In Fig. 7.31, this frequency is indicated as $\omega_3 = \omega_0$. It may be that in a certain frequency range $\omega_4 < \omega < \omega_5$ the value of $\varepsilon'(\omega)$ becomes negative; this corresponds to high dielectric strength of oscillator and its low damping. Frequency dependence of losses (ε'') in the range of dispersion is characterized by maximum at frequency $\omega = \omega_6$. If damping is small, in the vicinity of resonant frequency it is possible to assign $\omega_6 \approx \omega_0 \approx \omega_3$.

Experimental spectral studies usually give the loss factor frequency dependence: $\varepsilon''(\omega)$. The half-width of absorption line is defined by difference in frequencies at level of $\varepsilon''/2$. At small damping factor Γ , the maximum of loss is $\varepsilon_{\text{max}} \approx [\varepsilon(0) - \varepsilon(\infty)]/\Gamma$, while spectral half-width is determined as $\Delta\omega/\omega = \Gamma$. The frequency ω_6 (at which this maximum occurs) and half-width of curve $\varepsilon''(\omega)$ enable to determine main parameters of oscillator model. However, all these relations are valid only at $\Gamma \ll 1$.

In case of resonant polarization, a loss tangent is not a convenient feature, as it changes its sign in accordance with $\varepsilon'(\omega)$ (Fig. 7.31A) at the points of zeros of this function $\tan\delta$ becomes infinite. Thus dielectric absorption at resonant dispersion usually is described by loss factor $\varepsilon''(\omega)$. However, when dielectric losses are studied far away from resonance dispersion (when $\omega \ll \omega_0$), parameter $\tan\delta$ still may be convenient to describe losses.

Dielectric spectroscopy. Permittivity dispersion means its dependence on electrical field frequency: $\varepsilon = \varepsilon(\omega)$. This term is borrowed from optics, where frequency dependence of refractive index $n = n(\omega)$ is called as a dispersion. In solid-state physics, the term “dispersion” is used for quasiparticle energy ($W = \hbar\omega$) dependence on quasimomentum ($p = \hbar k$). At that, in optical and IR spectra the dependence of $W(p)$ is reduced to the dependence of frequency on wave vector: $\omega(k)$ that expresses the dependence of wave phase velocity on its frequency; finally, this corresponds to the $\varepsilon(\omega, k)$.

An important property of dielectric dispersion should be considered as the fulfillment of Kramers-Kronig relations, linking frequency dependence of real and imaginary parts of complex permittivity $[\varepsilon^*(\omega) = \varepsilon'(\omega) - i\varepsilon''(\omega)]$:

$$\begin{aligned} \varepsilon'(\omega) - \varepsilon_\infty &= \frac{2}{\pi} \int_0^\infty \frac{\varepsilon''(\Omega)}{\omega^2 - \Omega^2} \Omega d\Omega; \\ \varepsilon''(\omega) - \varepsilon_\infty &= \frac{2}{\pi} \int_0^\infty \frac{(\varepsilon'(\Omega) - \varepsilon_\infty)\omega}{\Omega^2 - \omega^2} \omega d\Omega. \end{aligned} \tag{7.36}$$

These relationships allow calculating frequency dependence of absorption $\varepsilon''(\Omega)$ from known frequency dependence of dielectric constant $\varepsilon'(\omega)$. In contrast, using frequency dependence of $\varepsilon'(\Omega)$ by analytics (or by numerical methods with computer), it is possible to determine frequency dependence of loss coefficient $\varepsilon''(\Omega)$.

The Kramers-Kronig relations are universal to describe the phenomenon of ε -dispersion and are used not only to control experimental results, but also to obtain important forecasts as regards dielectric losses. For example, using Eq. (7.36) it is possible to calculate static permittivity

$$\varepsilon(0) = \varepsilon'(0) = \varepsilon_\infty + \frac{2}{\pi} \int \frac{\varepsilon''(\Omega)}{\omega^2 - \Omega^2} \Omega d\Omega$$

Other complex parameters such as optical refractive index or complex coefficient of electromagnetic waves propagation can be also expressed in terms of complex value $\varepsilon^*(\omega)$ [4].

In low-symmetry crystals along main crystallographic directions, one can find quite different dependences for components of tensor $\varepsilon_{kl}^*(\omega)$ that form the *dielectric spectrum*. Dielectric spectra investigation is one of the important methods of studying the physical properties of dielectrics. Frequency dependence of $\varepsilon_{kl}^*(\omega)$ enables to make qualitative judgments about the physical nature of dielectric polarization and losses in a particular matter. In addition, it is possible to obtain quantitative data on characteristic frequencies (Ω_k) and dielectric contributions $\Delta\varepsilon_{kl} = [\varepsilon(0) - \varepsilon(\infty)]_{kl}$ of correspondent mechanisms of polarization.

When dielectric spectra are studied in a temperature range, that is, experimental characteristics of $\varepsilon^*(\omega, T)$ are obtained, it enables to get temperature dependences of specific frequencies (relaxation times) and other parameters of various polarization mechanisms. In some cases, there is considerable interest on the electrical field influence on properties of dielectrics, so rather complicated dependence $\varepsilon^*(\omega, T, E)$ might be established.

For a detailed study of the dielectric spectrum, it is necessary to perform dielectric measurements over a very wide frequency range. This range includes not only low-frequency interval (10^{-3} – 10^8 Hz), but also the microwave ($3 \cdot 10^8$ – 10^{11} Hz), the submillimeter (10^{11} – 10^{12} Hz), and the far-IR (10^{12} – 10^{14} Hz) frequency ranges. As permittivity is the second-rank tensor, frequency dependence of permittivity's components can be very complex, especially for ferroelectrics and other polar crystals of order-disorder type, with the reason being large anisotropy of permittivity: main components of tensor ε_1 , ε_2 , and ε_3 might be quite different, and they vary by hundreds of times. In addition, the contributions to permittivity tensor can be produced by many different mechanisms of polarization.

As an example, the *Rochelle salt dielectric spectrum* is shown in Fig. 7.32. In this type of ferroelectrics, there are two main frequency intervals of permittivity dispersion: domain wall relaxation in polar phase and dipole relaxations in all phases. Rochelle salt is characterized by two Curie points, where at low frequencies ε_1 has temperature highs of about 5000 (-18°C and $+24^\circ\text{C}$), but these highs at micro-waves decrease only to $\varepsilon_{1\text{max}} \approx 200$.

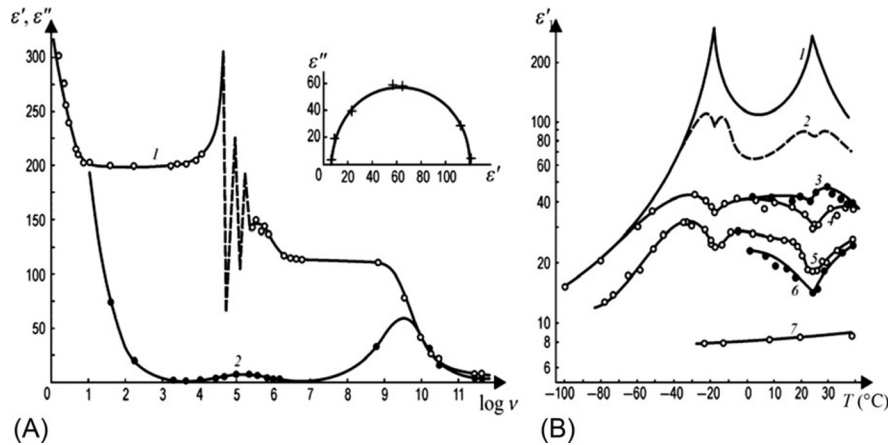


FIG. 7.32

Rochelle salt microwave study: (A) ϵ'_1 and ϵ''_1 frequency dependence at 18°C; (B) ϵ'_1 temperature dependence at frequencies (in GHz): 1—0.8; 2—5.1; 3—7.4; 4—10.2; 5—20.5; 6—27; and 7—250.

Sharp maximums of $\epsilon'_1(\nu)$ in Fig. 7.32A in the frequency interval of 10^4 – 10^5 Hz mean piezoelectric resonances, which are accompanied by a fluent ϵ'_1 -decrease near frequency of 10^6 Hz. This decrease is explained by the domain relaxation process that follows electromechanical resonances. At microwaves, Rochelle salt shows ϵ'_1 -dispersion that is accompanied by ϵ''_1 broad maximum; this dispersion characterizes structural dipole relaxation that can be described by the Debye equation. In two other main crystallographic directions (for ϵ_2 and ϵ_3), any ferroelectricity in Rochelle salt is absent and no temperature anomalies for ϵ_2 and ϵ_3 exist [5].

Dielectric dispersion is characterized by various parameters that are determined from experimental data. *Dispersion frequency* is called a frequency at which maximum of $\epsilon''(\omega)$ is observed. The *width of spectrum* is determined by the difference between frequencies, at which this maximum is reduced by a half. As *depth of dispersion* is a relative contribution from given polarization mechanism to the $\epsilon(0)$ -value, that is, this is $[\epsilon(0) - \epsilon(\infty)]/\epsilon(0)$ where $\epsilon(\infty)$ is high-frequency dielectric constant.

Therefore the main purpose of dielectric spectroscopy is experimental detection of different mechanisms of polarization and separation of their deposits to the permittivity.

7.9 ELECTROCONDUCTIVITY IN DIELECTRICS

Charge transfer in dielectrics occurs mainly in the external electrical field (current caused by temperature gradient or due to difference in free charge concentration in dielectrics usually is very small). Whereas polarization is the *total shift* of all charges

under electrical field influence, only a very *small portion* of charges (free carriers) can contribute to conductivity. Furthermore, during polarization, charge exchange between dielectric and metallic electrodes does not occur, whereas in case of the “pass-through” conductivity such an exchange is required. Therefore dielectric conductivity is highly dependent on the material of electrodes, whereas charge separation in dielectric (polarization) is practically independent of properties of electrodes (if electromotive force, EMF, is absent).

Nevertheless, a clear distinction between electrical conduction and polarization is possible to determine only in a static electrical field. In alternating electrical field, the difference between these processes becomes conditional.

Basic relations, describing charge transfer, are discussed in [Section 7.7](#). In short, current density j , that is, electrical charge, flowing per unit time through unit area that is perpendicular to field direction is equal to: $j = nqv$, where v is the velocity of orderly movement of charge carriers due to action of electric field (*drift velocity*). This velocity, typically, is much less than the velocity of chaotic charge carrier motion in material. According to Ohm’s law, failing small values of E , current density is proportional to field: $j = \sigma E = E/\rho$, where σ (its dimension is [S/m]) is specific electrical conductivity, while ρ [Ohm·m] is specific electrical resistivity. Conductivity (or resistivity) determines current density in direct electrical field; both these parameters are the quantitative characteristics of charge transport in a matter.

Electrical field acts on free charge carriers by the force qE ; therefore they would have to move with acceleration, increasing its kinetic energy. However, the path and the time of charge carrier free movement in a material are limited by charge carrier “collisions” with atoms (molecules or ions), that is, by the interaction of charge carriers with particles of matter. In crystalline dielectrics, these collisions are the interaction with phonons (lattice vibrations) or with charged impurities, as well as with other electrically active structural defects. Therefore, accelerating movement of charge carriers is interrupted, and acquired in electrical field energy is dissipated. Average time of electron free acceleration and collision is the *relaxation time* (τ), as during this time electron returns to the state of thermodynamic equilibrium with matter. The equation of motion of charge carrier with mass m under action of electrical force qE is as follows: $\frac{dv}{dt} = \frac{qE}{m}$, where dv/dt is acceleration. By separating variables and next integrating from $0 \leq t \leq \tau$, drift velocity is $v = \frac{q\tau}{m}E = uE$, where $u = q\tau/m$ is the proportionality coefficient between charge carrier velocity and electrical field, called as *mobility* of charge carriers. Mobility $u = q\tau/m$ is proportional to charge carrier relaxation time, and inversely proportional to its mass; it characterizes directed by charge transfer in the electrical field. Thus, not acceleration (as in Newton’s second law for a free body in a space), but the *drift velocity* of charge carriers is proportional to force acting on them in the condition of disordered (thermal) motion of charge carriers in a material.

Conductivity σ and mobility u of charge carriers are constrained by a simple relation: $\sigma = nqu$. With temperature change and depending on electrical field strength, the mobility changes dozens of times. Only in metals free charge carrier concentration is independent of temperature, but in dielectrics (and in semiconductors)

parameter n increases exponentially with temperature increase and it is highly dependent on impurities. Parameter σ (or $\rho = \sigma^{-1}$) determines the degree of scattering (losses) of electrical power in a matter. According to the differential form of Joule law, density of exudated heat p [W/m³] is electrical energy that is converted into a heat per unit time and per unit volume of substance: $p = E^2/\rho = \sigma E^2$. From this equation, it is possible to move on to formulas for conductance G [S], resistance R [Ohm], and dissipated power P [W] in the body of *any size and shape*. If the sample consists of a homogeneous isotropic material and applied voltage is U , then

$$G = \sigma \Xi = \Xi/\rho; \quad R = \rho/\Xi = (\sigma \Xi)^{-1}; \quad (7.37)$$

where Ξ is the geometrical parameter of a sample (dimension Ξ [m] is the reduced length). In the event of a sample that has cross section S and length l (e.g., for plane dielectric capacitor), the geometrical parameter is $\Xi = S/l$. In case of a hollow cylinder having outer diameter D , inner diameter d , and axial length l , geometrical parameter is determined as $\Xi = 2\pi l/\ln(D/d)$: such a structure has a cylindrical capacitor or section of coaxial cable. In the event of hollow spherical dielectric capacitor with inner radius r_1 and outer radius r_2 , the geometrical parameter is $\Xi = 4\pi r_1 r_2/(r_2 - r_1)$: piezoelectric ultrasonic emitter and receiver often have this shape.

Conductivity of various solid substances at normal conditions (at 300 K) covers 25 orders of magnitude: from $\sigma \sim 10^8$ S/m for good metallic conductors (copper, silver, aluminum) down to $\sigma \sim 10^{-17}$ S/m for best insulators (polymers).

Charge carrier classification. The various types of charged particles with different mechanisms of their generation (excitation) can be distinguished. Besides, there are different mechanisms of charge movement; so electrical current in dielectrics might be a complicated physical phenomenon.

The classification of dielectric conduction starts from the nature of charge carriers; in this case, there are several possible contributions to conductivity that are listed in Fig. 7.33.

In the event of *electronic* conduction, electrical current may consist of negatively charged electrons and positively charged electronic vacancies—holes. In dielectrics and in some semiconductors, *polaron*-type charge carriers are also possible, when electrons or holes are more or less bounded in crystal lattice, and, therefore, they have low mobility. The *ionic* charge transfer is typical for dielectrics. This mechanism of conduction is defined by the flow of positively charged cations or negatively charged anions, as well as by charged ionic vacancies of opposite polarity. In direct electrical field, the ionic conduction represents not only the charge transfer but also the matter transfer—*electrolysis*. Liquid dielectrics, except electronic and ionic conductivity, can also have the “molionic” conductivity, at which charge carriers constitute charged atomic groups or charged molecules—relatively large particles (*electrophoresis*). In case of the positive charge, this electromigration results in the *cataphoresis*, while in case of their negative charge leads to the *anophoresis*.

Previously, because of predominance of ionic conduction, and to emphasize this fact, in early literature dielectrics were even called as *electrolytes*. It should be noted that in solid electrolytes conductivity changes in time, especially in the direct

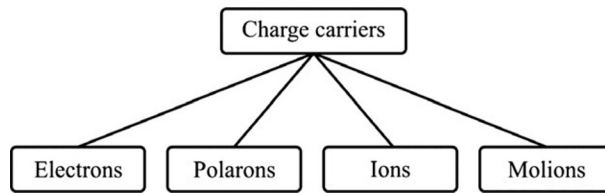


FIG. 7.33

Classification of charge carriers in dielectrics.

voltage, because of “exhaustion” of free charge carriers: the quantity of free ions in dielectrics is limited, so in case of DC these charge carriers migrate gradually to electrodes and accumulate there. In this regard, the ionic conduction is one of the causes of electrical insulator *aging* (dielectric properties change with time under the influence of electrical field). However, in the alternating electrical field ions have no time to be accumulated at electrodes; therefore in alternating field ionic conduction of dielectrics resembles a time-constant (stationary) process.

Electronic conductivity, as confirmed by recent studies, also plays a very important role in electrical charge transfer in dielectrics. Electronic charge transfer (in contrast to ionic and molionic mechanisms) is a *stationary process* not only in alternating but also in direct electric fields: between dielectric and electrodes, there is the interchange by charge carriers of same physical nature—electrons.

Two important properties of dielectrics—polarization and very low conduction—are largely *interdependent*. Electrons or holes, appearing in dielectrics as a result of various activation processes, may become less active, because they polarize by their field the nanoscale size surrounding area of dielectric, so charge carriers are forced to move altogether with these polarized nanoregions (polarons). Consequently, even a small amount of free electrons that occur in the dielectric due to thermal activation and impurities may not cause any appreciable charge transfer just because of *local polarization* around charge carrier that reduces its mobility in the electrical field.

In turn, low concentration of charge carriers and their low mobility are responsible for a *long-time existence of electrostatic field* in the dielectrics. In conductors, the electrical field is screened by free charge carriers (in metals, e.g., screening radius is approximately equal to interatomic distance).

Thus, electrical polarization contributes to emergence and existence in dielectrics a relatively stable state with *extremely low electronic conduction*. However, this stability may be broken in a dielectric by heating using high-intensity irradiation, particularly, using coherent optical (laser) radiation. Then charge carriers are generated to a very high concentration and they shield electrical field; therefore the dielectric is converted into a conductive medium.

Stability of nonconducting state of dielectrics may be also compromised by a *strong electrical field* that accelerates the unfettered electrons (or holes) up to the energy at which they can no longer be “captured” by polarization of surroundings

and acquire a “slow-moving” state. These *fast electrons* cause percussive ionization in dielectric, due to which the increase in number of free electrons is seen that ultimately gives rise to electrical breakdown, and insulator transfers into a conductor.

In some special cases, the stability of nonconducting state of dielectrics may be broken even in a weak electrical field and without their strong heating or irradiation. The reason for this is the spontaneous change in structure, associated with mutual interaction of particles, and, hence, the change in the symmetry of a crystal. In such peculiar cases, even a small change in external conditions (pressure, temperature, magnetic field, or electrical field) can lead to a spasmodic increase in conductivity (10^3 – 10^9 times), that is, the insulator turns into conductive state.

Such interruption in conductivity might be the result of phase transition, wherein because of changes in the external conditions (field, temperature, pressure) some electrons become free from their polarized surroundings and, in turn, shield the electrical field. For example, such phase transitions are observed in the oxides of transition *3d*-metals, as well as in low-dimension systems and in “superionic” conductors. Unlike the irreversible effect of electrical breakdown, these phase transitions from the insulating to the conducting state are *reversible*, because the dielectric does not undergo destruction, as it occurs in case of electrical breakdown.

A sharp increase in conductivity at strong electrical fields is also observed during *injection processes* in connection with space charge limited current (SCLC). As in the case of phase transitions of “insulator-metal” type, the unstable conductivity appears on account of charge injection that is the reversible process (unlike electrical breakdown).

Charge carrier generation. Electrical conductivity of dielectrics and semiconductors is always conditioned by *activation processes*; at that, charge carriers arise because different mechanisms cause their appearance (Fig. 7.34).

Most universal and, therefore, very important process is *thermal activation*—permanent mechanism of charge carrier appearance in dielectrics and semiconductors. At normal and elevated temperatures, thermal generation of charge carriers in dielectrics gives a major contribution to their conductivity; at that, not only thermal generation of electrons and holes are important, but also *generation of mobile ions* activated by thermal chaotic movement.

In addition to the process of thermal generation of mobile electrons and ions, there exists and constantly ongoing process of *recombination*, at which electron

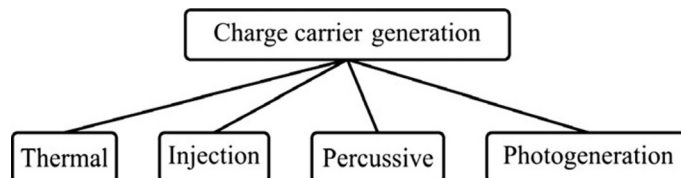


FIG. 7.34

Mechanisms of charge generation in dielectrics.

and hole (or ion and ionic vacancy) combine in the neutral structure when two types of carriers “compensate each other.” In the dielectrics (as in semiconductors) between heat generation and recombination of charge carriers, the *dynamic equilibrium* is established, which depends on the energy levels in dielectric and on temperature. Charge carriers, produced by thermal activation, are the *equilibrium* charge carriers. It is important to note that only those charge carriers can be balanced, which are formed by thermal excitation and relaxation, while other mechanisms of charge carrier generation, listed in Fig. 7.34, result in only nonequilibrium charge carriers.

The *injection* of electrons or holes in a dielectric (or semiconductor) occurs from metal electrodes; those charge carriers that are inculcated into the crystal by this way are always the *nonequilibrium* ones (injection will be discussed further in Section 7.4). The *photogeneration* of charges in a dielectric is caused by various types of irradiation; therefore this mechanism of charge carrier generation is also nonequilibrium. Finally, the *ionization by collisions* occurs in strong electrical fields, and it also leads to generation of nonequilibrium charge carriers: thanks to this ionization the concentration of charge carriers increases up to an electronic avalanche that may cause electrical breakdown (appropriate mechanisms will be discussed further in Section 7.10).

The mechanisms of charges transport are another important aspect of electrical conductivity in dielectrics (Fig. 7.35). Transfer mechanism is called as a *drift*, if upon chaotic (thermal) motion of charge carriers their directed movement (drift) in electrical field is imposed. In this movement, charge carrier spends most of its time, whereas it spends much less time in collision, capture, and scattering by other particles. In not very large electrical fields, drift velocity of charged particles is much less than their velocity in chaotic movement.

Dielectrics is such an important mechanism of charge transfer that it is considered as the *hopping* mechanism; it is peculiar for both ionic and polaron types of conductivity. According to this mechanism, charge carrier spends most of its time in the *localized state* and only a very short time does it expend for movement: it is a jump into nearby center of localization in crystal lattice. It should be noted that jump itself is a *change of quantum state* and it occurs almost instantaneously, but the number of jumps per unit time (their frequency) is quite limited.

Both in semiconductors and dielectrics, the *diffusive mechanism* of charge transfer might be also important, when in different places of a sample the concentration of

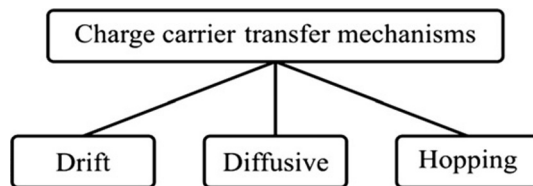


FIG. 7.35

Mechanisms of charge transfer in dielectrics.

charged carriers is different, and, due to disordered (random) motion of particles, the concentration of charge carriers in dielectric (or semiconductor) aligns. Charge carriers are gradually moving from the place of their higher concentration to the region of their lower concentration. Diffusion current can be observed in the absence of external electrical field.

In the processes of electromigration in dielectrics, the *contact phenomena* on the boundary of “dielectric-metal electrode” play an important role.

In case of ionic (usually, cationic) unchanged current at stationary conduction can be achieved only when the anode is made of the same metal that the ions transferring electrical charge in dielectric. This phenomenon can be applied, above all, in those technical devices, in which the “superionic” electrical conduction is used. For example, in $\text{Rb}_4\text{Ag}_5\text{Cl}_9$ crystal charge transport is ensured by the Ag^+ ions, and, respectively, in such a case the anode should be made of silver.

Such metal-insulator contact that provides a *free exchange* of charge carriers is called *neutral*. Otherwise, by applying constant voltage, the charge carriers are gradually depleted, and in near-electrode areas the depletion layer arises with high electrical resistance; therefore ionic current through dielectric decreases with time. Consequently, the distribution of electrical voltage inside dielectric becomes heterogeneous. Such a process is called “forming”: owing to charge carrier depletion, the strength of the electrical field in a dielectric near its contact with metal increases.

Electronic conduction greatly facilitates the exchange of charge carriers between dielectric and metal electrodes. However, in this case, current density depends on the electronic structure of contacting pair. Contact metal-to-dielectric is *neutral* (ohmic), if electronic work function ϕ of metal-to-dielectric is equal to electron work function ϕ of metal-to-vacuum: $\phi = \phi$. If $\phi > \phi$ the contact is *blocking* (locking), and if $\phi < \phi$ the contact is defined as *injection* [7].

Comparative energy diagrams for metal-vacuum (*M-V*) and metal-insulator (*M-I*) are shown in Fig. 7.36. Relative to dielectrics, two options are shown: the injection

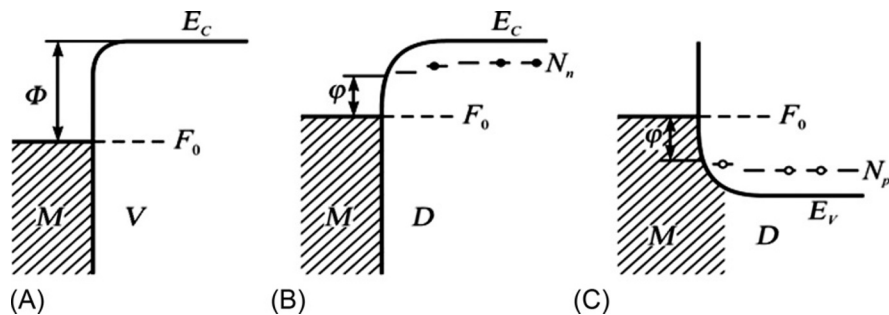


FIG. 7.36

Energy diagrams of metal-vacuum (A) and two cases of injection metal-insulator transition (B, C): F_0 , Fermi level; ϕ , electronic work function to vacuum; ϕ , reduced band bending barrier of double layer; N_n , level of small electron traps (deep traps not shown), N_p , level of shallow traps for holes.

contact that facilitates electron transition to conduction band of dielectric (bottom level of this zone is denoted as E_c), and the injection contact that promotes transition of holes in the valence band of dielectric (boundary of this zone is designated as E_v).

Ionic conductivity in dielectrics. Historically, the first dielectrics were defined as “electrolytes”—the substances with predominance of ionic transport of electrical charge (A. Ioffe). Therefore, already in early studies, a lot of attention was paid to ionic conduction in these materials, which has been studied by various methods in a wide temperature range as well as under high electrical voltage [6].

Most ions in crystal (their concentration is $n \sim 10^{22} \text{ cm}^{-3}$) are located at their lattice sites; this position is rather stable and cannot be disturbed when applied to a crystal electrical field that causes only a slight shift of ionic sublattices (that constitutes quasielastic ionic polarization). However, any crystal, almost inevitably, contains a certain concentration ($n_0 \ll n$) of impurities or defects that are relatively loosely bound in a crystal lattice. They can be located in the interstices (Frenkel defects) or represent the charged vacancies (Schottky defects). These ions are weakly constrained and cause electrical conductivity.

The movement of ionic charge carriers in direct electrical field transfers not only electrical charge, but also a *part of material*. Unlike electronic conductivity (when electrons enter into dielectric from cathode, and finally leave dielectric in anode), ionic conductivity is accompanied by the electromigration of transfer agent (mass transfer). For this reason, the value of ionic current must be time-dependent, as the concentration of available charge carriers in dielectric decreases gradually—ions are accumulated near the electrodes. Negative charge carriers—anions—are deposited and discharged at anode, while positive charge carriers—cations—are deposited at cathode and next discharged. By measuring the amount of transferred substances, it is possible to establish what type of ions is involved in electrical conductivity in different dielectrics.

The experiment (proposed by Tubandt) allows obtaining direct evidence as to the presence of ionic conduction in a substance (as well as Hall experiment indicates the presence of electronic or hole current). In this experiment, in its initial implementation, a sample of dielectric is previously sawn into two parts that are polished and pressed tightly to each other. Before being included in the circuit (that is used to determine the amount of passed charge), the weight of each part of the tested sample is measured with high accuracy. Then, for a long time (generally, at elevated temperature, when conductivity is increased) through a twofold dielectric sample the electrical current passes, while the amount of electricity is recorded. Weighing after this experiment shows how the mass of the two parts of the sample is changed. In case of cation conduction, the mass of the near-cathode part of the sample increases with reducing the mass of the near-anode part; in the event when anion conduction takes place, the results are quite the contrary. Nowadays, for Tubandt experiment realization, the radioactive isotope ions (tracers) are used whose number before and after electrical field trial is recorded by the Geiger counter.

Thus, ionic current in dielectrics is due to moving weakly constrained charged particles that can be regarded as impurities. Assume that volume concentration of moveable particles equals n_0 . In order to move in a dielectric, ions must overcome

potential barriers, that is, to overcome the forces, which bind them to neighboring particles. The probability of energy barrier U being overcome by the ionic impurities that are in random thermal motion can be determined by classic statistics: hopping probability is proportional to $\exp(U/k_B T)$, where k_B is Boltzmann constant.

The ionic conductivity calculation is similar to the ionic polarizability estimation as thermally induced polarization (discussed in Section 7.5, Fig. 7.13). In fact, both mechanisms (ionic thermal polarization and ionic electro-conductivity) are due to the diffusion of ions that is supported by electrical field. However, the magnitude of potential barrier that ion must overcome for conductivity is almost an order of magnitude higher than in case of thermally activation energy of ionic polarization.

The average number per unit volume of loosely bound charged particles that overcome per one second the potential barrier U at their jumps is $n(T) = \frac{n_0}{6} \nu \exp(-\frac{U}{kT})$, where $n_0/6$ is the number of ions moving in the positive direction of selected axis, ν is frequency of thermal vibrations of ions in lattice (Debye frequency), while parameter $\exp(-U/k_B T)$ shows probability of potential barrier U overcoming at temperature T . As thermal hopping of impurity ions at the absence of external electric field is chaotic, electrical current in case of these random hopping does not occur.

In the event that electrical field is applied to dielectric, the probability of potential barrier overcoming changes on value of $\exp(\Delta U/k_B T)$, where $\Delta U = \frac{1}{2} q \delta E$; q is ionic charge, δ is jump length (otherwise it is length of “free path”), and E is electrical field (Fig. 7.37). If external field is absent (Fig. 7.37B), loosely coupled ions randomly overcome potential barriers and jump, for example, from position 1 to position 2 (or vice versa). This thermally activated jumping (per second) makes the following number of ions:

$$n_{12} = n_{21} = \frac{n_0}{6} \nu \exp\left(-\frac{U}{kT}\right)$$

$$n_{12} = n_{21} = \frac{n_0}{6} \nu \exp\left(-\frac{U}{kT}\right).$$

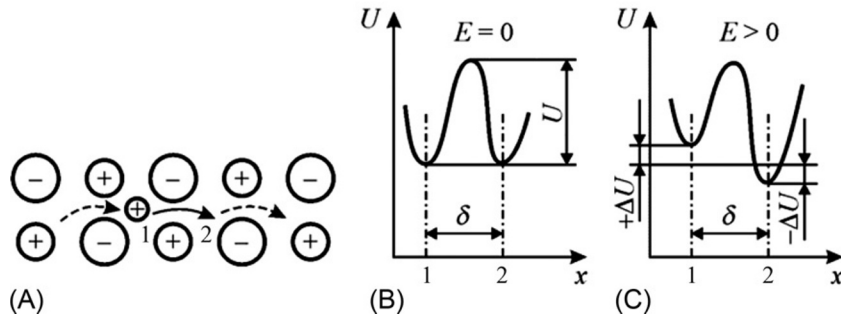


FIG. 7.37

Ionic conduction: (A) small cation jumps over interstices; (B) potential barrier in the absence of electrical field; (C) change in barrier.

Since $n_{12} = n_{21}$, in case of such counterdiffusion any electrical current does not occur. However, if one applies to dielectric the electrical field E that is directed along the selected axis x , the probability of transition of weakly bounded ions from position 1 to position 2 increases (Fig. 7.37C), while the probability of their coming from the opposite direction decreases. First, the height of potential barrier, when ions jump $1 \rightarrow 2$, reduces on ΔU , and, second, for transitions in $2 \rightarrow 1$ direction barrier it increases on ΔU . The value $2\Delta U$ represents the work, done by electrical field in the path δ of ion transference.

After electrical field application, a certain amount Δn of weakly bounded ions (from their total number n_0) can overcome the potential barrier, and thus they become involved in electrical conductivity. Average velocity of directional movement of ions is $v = \frac{\Delta n}{n_0} \delta$. This formula naturally corresponds to the dimension of velocity: [m/s], as the dimension of Δn is [$\text{m}^{-3} \text{s}^{-1}$], the dimension of n_0 is [m^{-3}], and the dimension of δ is [m]. As conductivity equals $\sigma = n_0 q v = \Delta n q \delta$, where q is charge and δ corresponds to lattice constant, calculation of conductivity is reduced only to Δn value determination. This calculation is provided in the assumption that changes of potential barrier in electrical field are much less than average thermal energy of crystal, that is, $\Delta U \ll k_B T$ (this inequality is always valid in comparatively weak electrical field). Thus calculation of conductivity is reduced to Δn value find, dependent on temperature and on electrical field strength:

$$\Delta n = \frac{n_0 q \delta v E}{6kT} e^{-\frac{U}{kT}}$$

This relationship allows to obtain the formula for specific ionic bulk conductivity:

$$\sigma = \frac{n_0 q^2 \delta^2 v}{6kT} e^{-\frac{U}{kT}} \quad (7.38)$$

which characterizes its temperature dependence.

Similar to temperature dependence of electronic conductivity, expression (7.38) represents the activation process that is characterized by energy U . In the logarithmic scale temperature dependence of conductivity corresponds to $\log \sigma = \log A - B/T$, where $A = (n_0 q^2 \delta^2 v) / 6k_B T$ and $B = U / k_B T$.

In the simplest case, when conductivity is characterized by only one type of charge carriers, this dependence is the inclined straight line (Fig. 7.38A). From its tilt, the potential barrier that weakly bound ions overcome can be determined. Usually experiments for ionic crystals give the value of activation barrier of $U = 1\text{--}3 \text{ eV}$.

It should be noted that similar temperature dependence of conductivity $\log \sigma(1/T)$ is also characteristic of electronic conductivity—as in dielectrics, so in semiconductors. In semiconductors, a sharp turn (like in Fig. 7.38B) means that at low temperatures conductivity has impurity character while at high temperatures intrinsic conductivity prevails. In the ionic dielectrics, a sharp turn in $\log \sigma(1/T)$ dependence has other interpretations. For example, at low temperatures the ionic conductivity may prevail, while at high temperatures the conductivity might have an electronic nature. Another possible case: at low temperature, the anionic conduction is observed, while at high temperature, the cationic conduction prevails [6].

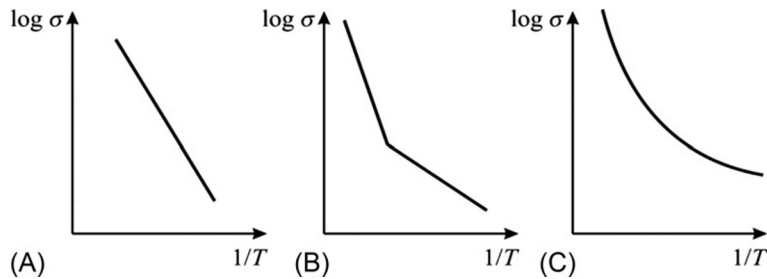


FIG. 7.38

Temperature dependence of conductivity in dielectrics: (A) one type of charge carriers, (B) two types of carriers that differ in activation energy; (C) case of continuous distribution of charge carrier activation energy.

Eq. (7.38) is derived from the assumption that only one kind of particles is involved in electrical conductivity. In general, by combining high- and low-temperature parts of σ temperature dependence, conductivity can be expressed by the following formula:

$$\sigma = A_1 e^{-\frac{U_1}{kT}} + A_2 e^{-\frac{U_2}{kT}}.$$

This not only makes it possible to explain the sharp turn in $\log\sigma(1/T)$ dependence, but it also allows to determine main microscopic parameters of ionic conductivity on the basis of experimental data. It is possible not only to find activation energies of U_1 and U_2 from the slope of line segments in Fig. 7.38B, but also to find appropriate concentration of charge carriers by the extrapolation of line segments on ordinate axis. One reason for the sharp turn in characteristic $\log\sigma(1/T)$ in glasses and in ionic crystals may be that at high temperatures electronic conductivity can be added to ionic conductivity, that is, in dielectrics the *mixed* conduction is quite possible.

However, in some dielectrics temperature dependence of conductivity may be significantly different from that described above by ordinary laws (Fig. 7.38A and B). Namely, the curved $\log\sigma(1/T)$ dependence is seen, from which it is impossible to clearly distinguish a transition from one type of conductivity to another. This case is the attribute of *weakly ordered structure* of dielectric that is usually customary for amorphous, glassy, and polycrystalline structures. The cause of the curvature of this dependence is the violation of long-range ordering in atomic arrangement. In this case, activation barriers for different mechanisms of conduction are not clearly delineated, but they are distributed in a certain energy range. For this reason, in glasses and glass ceramics, the boundary between low- and high-temperature conductivity is difficult to define.

The surface conductivity in dielectrics, typically, has *ionic character*. In solid dielectrics, due to inevitable surface moisture, oxidation, and contamination, surface conductivity becomes very significant. For its quantification, dielectric is characterized by a definite value of *surface resistivity* ρ_s . It should be noted that in the conductive materials surface current is negligibly small as compared to volumetric current, and, therefore, surface resistance of these materials is not considered.

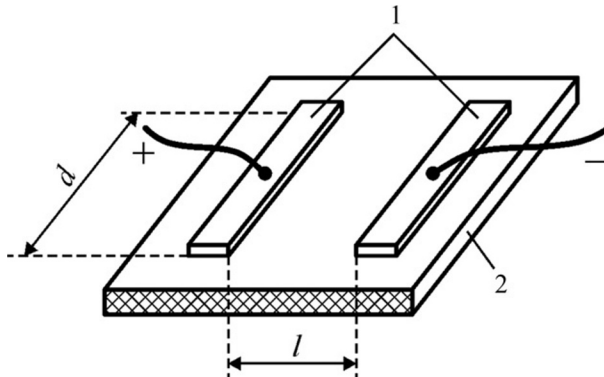


FIG. 7.39

Electrode arrangement for measuring surface resistance of material: 1—electrodes, 2—dielectric.

The surface resistivity ρ_S numerically is equal to resistance of a square locating on surface of material, while surface current (Fig. 7.39) flows through two opposite sides of this square:

$$\rho_S = R_S \cdot d/l, \quad (7.39)$$

where R_S is surface resistance of material (2) between electrodes (1) coated on surface, located at a distance l and having a width d .

The R_S is defined as Ohms/square, or [Ohm/]. It should be noted that same unit of surface resistance is a very common characteristic of conducting and semiconducting films deposited on dielectric substrate (such films are used, in particular, in microelectronics). However, determination of R_S in *bulk semiconductor* does not make sense, since in this case it is practically impossible to separate surface leakage currents from volume current. As can be seen from formula (7.39), for ρ_S determination the size of “studied square” does not matter—it can be square centimeter, square millimeter, etc. It should be noted that for bulk sample determination of specific resistance ρ the aforementioned “law of similarity” has no effect.

The surface resistivity ρ_S can be considered as a distinctive parameter of *dielectrics*, although ρ_S is dependent on temperature, humidity, and applied voltage. Water is characterized by increased conductivity, while any polar and highly porous dielectric might be highly moistened. It is sufficient to have even a thin layer of moisture on the surface of dielectric to cause significant conductivity, which is mainly determined by the thickness of this layer. At that, surface conductivity is caused not only by the presence of moisture, but also by pollutions and various defects on the dielectric surface.

The value of ρ_S in dielectrics is usually associated with the magnitude of the *contact angle* (angle of wetting) and with the *hardness* of the dielectric. The smaller the contact angle and the higher the hardness of humid dielectric, the lower the ρ_S . Depending on the contact, angle solid materials are subdivided into *hydrophobic*

and *hydrophilic*. Hydrophobic dielectrics are related to the *nonpolar* dielectrics: their clean surface is not wetted, and, therefore, when such dielectrics are placed in the moist environment, their surface electrical conductivity practically is not changed. Hydrophobic dielectrics are related to the water-soluble and polar dielectrics are related: these are mostly ionic dielectrics with wetted surface. Intermediate type of dielectrics conditionally includes the *weakly polar* dielectrics.

In case hydrophilic dielectric is placed in a wet medium, its surface conductivity significantly increases. In addition to polarity, some other contaminations may adhere onto surface that leads to additional increase in surface conductivity. It should be noted that the value of ρ_S in humidified insulation can be significantly increased with increasing temperature and after drying. To increase ρ_S in dielectrics there are different technological methods: washing in distilled water or in solvents (depending on the type of dielectric), heating to sufficiently high temperature; coating the surface by water-resistant paints or glazes, placing dielectric products in protective housings and casings, etc.

For example, surface electrical resistance of glass is determined by the resistance of its surface to swelling, especially if glass contains aqueous alkaline solution with significant conductivity. Therefore surface resistance of glass is a function of its chemical composition (on content of alkali metal-oxides) and depends on ambient atmosphere humidity. In case of relative humidity increases, surface resistance of glass rapidly decreases, while temperature increase reduces the amount of moisture in the surface layer and increases surface resistance. Typically, surface resistivity of glass is 10^{10} – 10^{12} Ohm/□.

Electronic conductivity in dielectrics. In most cases, electrical current occurs under the influence of electrical field (as well as field-induced polarization). Nevertheless, in some crystals directional movement of electrons can also be caused by temperature gradient and by gradient of concentration.

Conductivity given by directional movement of electrons in external electrical field can be found in all classes of dielectrics (gases, liquids, crystals). However, in relatively low electric fields contribution of electrons to overall conductivity is usually small. The point is that electrons that appear in a dielectric for a variety of reasons polarize the nearest environment, and, therefore, moving of electrons in the dielectric is delayed by originated local polarization. In gases and liquids, the incipient electrons usually “stick” to neutral molecules and form the charged complexes that move in the electrical field as charged particles. In solid dielectrics, some electrons are trapped by the defects of structure and cannot move, or turn into low-mobility polaron state.

However, in strong electrical fields accelerated by the field high-energy electrons do not have time to be trapped by molecules or by crystal lattice, so the contribution of electronic current to the total value of conductivity becomes predominant. Moreover, a strong electrical field promotes generation of new electrons owing to the phenomenon of ionization by collisions.

In crystals having energy gap in the spectrum of electronic states (i.e., in crystalline semiconductors and dielectrics), electron-hole type of conductivity is described

by band theory and by theory of electronic transport phenomena. In the conduction band, most dielectric crystals have definite levels of energy, and in the impulse space they formed so-called valleys. Moreover, the constant-energy surfaces even in cubic crystals are anisotropic, and, therefore, effective mass of electrons (or holes) is tensor parameter. It also determines *anisotropy* of effective mass m_{ef} , relaxation time τ , and charge carrier mobility u . However, in cubic crystals these energy valleys in impulse space are arranged symmetrically, and, therefore, effective mass can be averaged over all the valleys, so not only the m_{ef} but also parameters τ and σ may be represented by isotropic values.

The energy spectrum of electronic states is well studied only in covalent crystals of semiconductors (see Chapter 8). For majority of ionic crystal-dielectrics owing to a very large energy gap and strong electron–phonon interaction only very small electronic conductivity is seen (detailed structure of electronic spectra in dielectrics is poorly investigated). Therefore, further arguments about the nature of electronic conductivity in dielectrics are very approximate and based on the analogy with electrical conductivity of semiconductors.

Mobility of electrons in dielectrics is smaller by hundreds and thousands times than in semiconductors. For instance, in semiconductors of InSb and InAs mobility lies within the limits of $2 \cdot 10^4$ – $8 \cdot 10^4 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ while in dielectrics AgI mobility of electrons is $\sim 3 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ while in NaCl it is only $\sim 1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. Therefore electronic mobility in dielectrics is very small (1 – $10 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$), whereas in semiconductors it is 1000 times greater.

Mobility of charge carriers is related to their effective mass: $u = q\tau/m_{\text{ef}}$. Actually, electrons and holes in crystals are the quantum excited states, characterized by negative ($-e$) and positive ($+e$) charges, respectively. It should be noted that the mass of electron or hole in a crystal can be significantly different from the mass of electron in the vacuum (m_e); moreover, in crystal charge carriers mobility depends on the direction of electron or hole motion, so the mass acquires anisotropic (tensor) value. Therefore, to describe the mechanism of electronic conductivity in dielectrics and semiconductors, the concept of effective mass is used. Low mobility of charge carriers in dielectrics indicates that the effective mass of electrons and holes in dielectrics is very often abnormally high: tens and hundreds of times more than m_{ef} in metals and semiconductors.

The concentration of charged carriers is characterized by a rather strong temperature dependence. In case of high enough temperature, Fermi-Dirac distribution of equilibrium charge carriers turns into classic Boltzmann distribution. It can be shown that the concentration of charge carriers increases with temperature exponentially: $n \sim \exp(-W/k_B T)$, where W is width of forbidden band (energy gap). From this dependence it follows that temperature dependence of conductivity may be approximated by the formula $\sigma = A \exp(-B/T)$. Therefore, temperature dependence of intrinsic electronic conductivity (that is noticeable in dielectrics at elevated temperatures) in scale of $\ln\sigma(1/T)$ is characterized by an inclined line: $\ln\sigma = \ln A - B/T$ (see Fig. 7.38A).

Experimentally, in most cases, a fracture of this line is observed in the crystals (see Fig. 7.38B), and this is explained by the fact that at lower temperatures

conductivity is conditioned by *impurities*, whereas in case of high temperatures the *intrinsic* conductivity is prevailing. However, in many dielectrics temperature dependence of conductivity is significantly different from that for simple laws shown in Fig. 7.38A and B (this peculiarity will be discussed later).

The temperature dependence of charge carrier *drift mobility* indicates its difference from Hall mobility. The features of light absorption and temperature dependence of thermoelectromotive force also show different nature of electronic conductivity in most dielectrics as compared with semiconductors (such as silicon). This is difficult to explain from the standpoint of band theory. The reason for these difficulties is that electronic conductivity of solid dielectrics can be caused by the movement of polarons.

Polaron (hopping) electrical conductivity. As already mentioned, incipient electrons in dielectrics might be partially constrained. This phenomenon is especially characteristic of ionic crystals, because in this case Coulomb interaction between electrons and ions in crystal lattice is sufficiently larger than that in atomic crystals. The point is that in ionic crystal the lattice near incipient electron (or hole) becomes distorted.

In Section 7.5, the idea of polarons was introduced as constrained charged particles forming nanoregions in the ionic crystal lattice (lattice distortion) under the influence of electrical field of electron. In other words, polaron is the excited state of a lattice around electron or hole. The meaning of this term is that electron (or hole), by its electrical field, polarizes the lattice of dielectric (part of ions are shifted slightly), so electron becomes localized in the area of distortion.

Such a self-trapping electron usually occurs in a small volume (covering several unit cells), but persists for a relatively long time (with respect to atomic scale of movement). Polaron moves in crystal lattice using energy of thermal fluctuations: it makes quick “jumps” to the neighboring lattice site. The time of these “jumps” is much less in comparison with the time of electron self-trapping. In this case, together with electron (or hole) the excited region moves in a crystal that results in higher effective mass of polaron-type charge carrier.

The size of polaron is determined by the dimension of distorted area. If the area of excitation r_{pol} is far superior to lattice parameter ($r_{\text{pol}} \gg a$), it is the *large-radius polaron*. In this case, the mobility of charge carriers varies insignificantly. Another thing is if there is the *small-radius polaron* with $r_{\text{pol}} \approx a$. This is a case of strong electron-phonon interaction and is determined by ionic polarizability of crystal lattice quantified by parameter $(1/\epsilon_{\text{opt}} - 1/\epsilon_{\text{ion}})^2$. The existence of polaron state does not meet the assumption of “adiabatic approximation” (i.e., independence of electron movement on atoms or ions) that, as is known, forms the basis of electronic band model of solids.

In case of weak electron-phonon interaction, the large-radius polarons have relatively little effect on electrons (holes) mobility in a lattice. The point is that deformations of crystal lattice in this case are minor, although the size of excited area is much greater than lattice constant. Therefore the conditions for electron and hole movement for large-radius polarons differ only a little from the movement of free

charge carriers. As in the case of “zone” electron, the mobility of large-radius polarons decreases with increasing temperature.

In contrast, for the small-radius polarons, the size of lattice distortion is small, but its deformation is significant. The polarization covers only from one to three cells, but localization time is comparatively large ($\sim 10^{-9}$ s). The excited state follows beyond movement that significantly increases polaron effective mass (in slang, description of this phenomenon might be as “electron dressed in phonon’s coat”).

In quantum theory, the *indeterminacy principle* can be recorded for energy and time in the form $\Delta W \cdot \Delta t \geq h$. The *width* of energy band in case of “band” electron (or hole) in semiconductors is $\Delta W = 1\text{--}2\text{ eV}$. It follows that such electrons can be localized in the lattice at a time no greater than 10^{-15} s, as in case of a longer period electronic wave functions “blur” in the entire crystal. However, for small-radius polarons in ionic dielectrics the width of band is approximately 7–10 eV that is responsible for long-term self-trapping of quasiparticles.

The movement of polarons in crystal lattice is supported by thermal fluctuations (as in the case of ionic conductivity). The “jump” of self-localized electron (or hole) to the neighboring node in a lattice is determined by deeper potential well formation in this node. The set of such thermal jumps of the polarons is chaotic movement. However, in the external electrical field thermally activated jumps of polarons become directed: this is the *hopping conductivity*. Nevertheless, in contrast to high-mobility “band” electrons (that show the *maximum* in $u_e(T)$ dependence) temperature dependence of mobility of small-radius polarons is characterized by a *minimum* u_{pol} with the subsequent return of mobility due to the decay of the polarons (electrons liberation) that can be seen as maximum $u_{\text{pol}}(T)$ (Fig. 7.40).

A decrease in charge carrier mobility in dielectrics with increasing temperature in a wide temperature range (as well as in semiconductors at normal and elevated temperatures) is because of electron scattering on lattice vibrations. However, at lower temperatures thermal motion in lattice stimulated appearance of polarons: they arise due to thermally activated displacements of lattice around electron or hole. At that, as temperature rises, chaotic thermal motion of ions in crystal lattice becomes more intense. Therefore, during self-trapping polaron becomes overgrown by “phonon

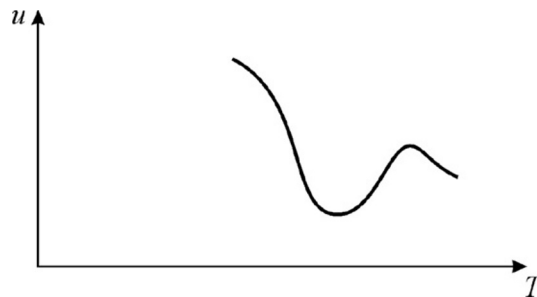


FIG. 7.40

Temperature dependence of mobility of small-radius polaron.

condensate,” that is, some closer located ions displace from their equilibrium positions. The increase in polaron effective mass decreases their mobility; for this reason, a fast decrease (down to a minimum) is observed in the dependence of $u_{\text{pol}}(T)$ (Fig. 7.40). Then, as temperature increases, stability of polarized environment around electron (or hole) begins to collapse, so mobility of electrons increases up to a local maximum in $u_{\text{pol}}(T)$ dependence.

Thus, electrons and holes in most dielectrics have low mobility and they look like being inactive. The origin of polaronic state is energetically favorable, because local polarization of crystal lattice reduces energy of charge carriers. Polarons arise on account of electron-phonon interaction, which is neglected in the band theory of semiconductors and metals. Low mobility of electrons in dielectrics results in small electronic conductivity only in weak electrical fields, when electrical field cannot change as concentration so mobility of charge carriers. However, in strong electrical fields the role of electronic conductivity increases sharply. In increased electrical field, first of all, mobility of charge carriers increases because polarons are destroyed. Second, a strong electrical field may also dramatically increase concentration of electrons in the conduction band (and holes in the valence band) owing to the *ionization* at collisions and due to *injection* of free carriers from electrodes.

Charge carrier injection and nonlinear conductivity. Ohmic conductivity is violated in case of charge carrier injection. Conductivity changes depending on the intensity of electrical field; moreover, a sudden *jump in conductivity* can be observed due to the injection of electrons or holes in a dielectric or semiconductor (also in case of phase transitions or electrical breakdown).

The process of injection occurs in relatively strong electric fields and is accompanied by a variety of nonequilibrium phenomena. The most important phenomenon is the case of SCLC. Space charge is distinguishable from usual charges: this is the excess electrical charge, distributed over a space in vacuum or in dielectric. In vacuum, space charge can be created only by electrons; in solids, the sign of space charge can be either negative or positive. Being emitted from a metal, the charge carriers form a kind of cloud in a certain volume of dielectric (in conductive medium, charges should be rapidly neutralized or screened).

Excess electrons or holes, being introduced into a crystal from electrodes by injection, can provide important information about the concentration and type of structural defects in the dielectrics or in the wide-gap semiconductors. The point is that some structural defects of crystal represent the “traps,” captivating injected charge carriers. Investigation of current-voltage characteristics of SCLC makes it possible to get information about the concentration of structural defects and energy levels of local states (capture or trapping levels) as well as about levels of charge carrier recombination.

During the *monopolar* injection, only one type of charge carriers (electrons or holes) is introduced in a crystal. In this case, injection takes place only from one of electrodes: electrons penetrate into a dielectric from cathode, or holes penetrate from anode. In case of *bipolar* injection, electrons and holes are brought into a crystal from two opposite electrodes.

The monopolar and bipolar injections of charge carriers are the nonequilibrium process when the electrical neutrality in crystal is locally violated. It should be noted that the equilibrium charge carriers (e.g., thermally activated) are generated in pairs, so electric neutrality persists. Therefore the spatial charge formation and its partial trapping are the infringement of neutrality. That is why at conditions of space-charge presence, the dependence of current on electrical field becomes nonlinear (Ohm's law is violated) [7].

The well-studied phenomenon of *electronic emission into vacuum* from heated cathode can serve as a good analogue for injection processes in crystals. Space charge in vacuum, created by emitted electrons, is located over cathode and, owing to Coulomb repulsion of electrons, *limits further emission*, which results in the nonlinearity of electrical current, flowing through vacuum diode:

$$j = aU^{3/2}d^{-2} \quad (7.40)$$

where a is constant and d is distance between cathode and anode. This expression uses the *voltage* U (not field strength E) because E has different value in the space between electrodes. For the same reason, current density is inversely proportional to the *square* (not to first degree) of distance between electrodes. Thus, the *nonlinearity of conductivity* in the condition of space charge is possible even in vacuum, at that, instead of Ohm's law, a law of "degree of the three second" takes place.

The regularities for injection of electrons (or holes) in dielectric are much more complicated than those in case of emission in vacuum. First, in crystals charge carriers *interact with lattice vibrations*, resulting in their scattering and possibility of their transfer to a slow polaron state. Second, some charge carriers may be *trapped by defects* of crystal structure that account for charge carrier localization in crystal. Finally, in contrast to vacuum diode, in the crystal a *double injection* with mutual compensation of negative and positive space charge are possible, as well as the *recombination* of electrons and holes.

In practice, the realization of capable to injection contacts looks like rather difficult task. However, the effective rate of injection can be achieved even in case of blocking contacts by using a sufficiently large electrical field. In this case, thin potential barrier "metal-insulator" allows electron tunneling. To increase the level of injection, the translucent electrode with ultraviolet illumination can be applied; hence for nonequilibrium charge carrier generation the photoelectric effect is used. As a common method to increase the intensity of injection a metallic tip can be adapted, near which the strength of the electrical field is increased. In all these cases, the density of injection current (even being limited by space charge) is thousands of times greater than the density of electrical current conditioned by equilibrium charge carriers.

Monopolar injection. In the relatively pure dielectric (with no structural defects traps), space charge depends on the Coulomb repulsion of electrons in a lattice. The *limited by SCLC* dependence on electrical voltage is given by the expression

$$j = u\epsilon U^2 d^{-3} \quad (7.41)$$

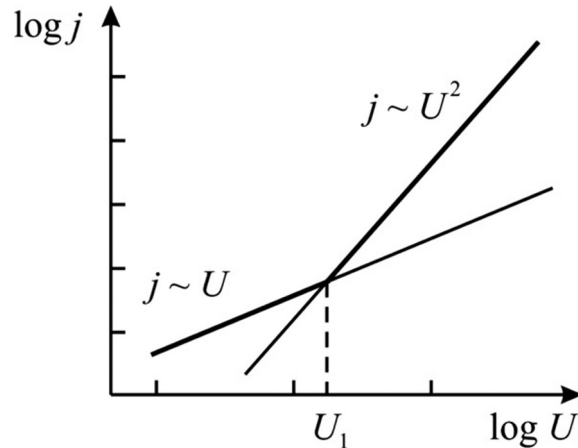


FIG. 7.41

Volt-ampere characteristic in defect-free crystals in the case of monopolar injection (here and in Fig. 7.43); each division on axes corresponds to change on the order of value.

where ϵ is permittivity, u is charge carrier mobility, and d is distance between electrodes. This is Mott-Gurney law, wherein the “non-traps” SCLC shows *quadratic* current-voltage dependence and *inverse cubic* current-thickness dependence; this law is confirmed experimentally.

Volt-ampere characteristic of a dielectric, in which there are *no traps* for electrons (or holes), is shown in Fig. 7.41. Here, except SCLC, the normal (ohmic) current conditioned by equilibrium charge carriers ($j \sim U$) is also shown. The dependence $j(U) = aU + bU^2$ consists of two easily separated sections. On the first plot (at low voltage, which is less than voltage U_1), electrical current is small and current density is proportional to voltage: this is law of Ohm. On the second plot, beginning with voltage U_1 , the dominating law is *monopolar SCLC* that is characterized by the quadratic dependence of current on voltage. It is seen that the slope of $\log j(\log U)$ characteristic becomes twice larger than that in first plot. *Strictly quadratic SCLC* low being characteristic of pure crystal can be used in analogous computing devices.

Volt-ampere characteristic of *dielectric containing defects* (traps) differs from the discussed idealized dependence. Except linear plot 1 and quadratic dependence of conductivity 2, in the event that nonequilibrium charge carriers exist, the plot 3 appears, shown in Fig. 7.42A: it is a stepwise increase in current at voltage U_2 . Then, in plot 4 the $j(U)$ dependence again becomes quadratic. Extrapolation of plot 4 to low voltages (plot 2') indicates the increase in charge carrier *mobility* above voltage U_2 .

At lower level of injection (plot 2), drift mobility of electrons (or holes) is smaller because in the vicinity of structural defects the process of continuously capturing and releasing of electrons occurs. These inhibitory traps reduce mobility of charge carriers and level of SCLC, as compared with defect-free crystal. However, above

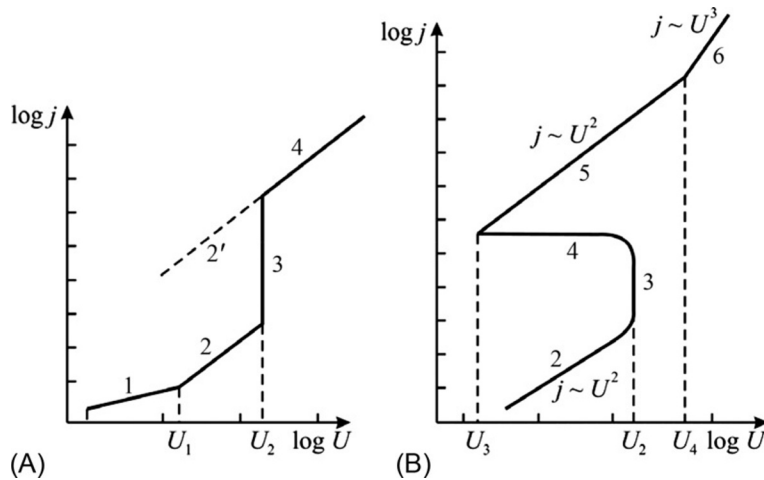


FIG. 7.42

SCLC in crystals with defects in the case of a monopolar (A) and bipolar (B) injection.

voltage U_2 all catchers-traps are filled; hence stepwise increase in current occurs at the expense of injected charge carriers that are not inhibited in traps (plot 3, Fig. 7.42). Therefore, by a value of voltage U_2 a concentration of structural defects is determined while the jump in current characterizes the depth of “sticking” levels (that are located in the band gap of semiconductor or dielectric). The closer the level N_n of “sticking” electrons to bottom of conduction band E_c , the smaller the current step in plot 3. In case of holes, the depth of “sticking” N_p is measured from the surface of valence band E_v .

The example, shown in Fig. 7.42A, corresponds to a simple case, when all defects are the same, and their level lies above the Fermi level (shallow levels), since $U_1 < U_2$. Otherwise, when $U_1 > U_2$ (the case of deep levels), the vertical jump of current density (plot 3) would start in the area of Ohm’s law (in plot 1).

This implies that in the event of monopolar injection only from the volt-ampere characteristic of SCLC it is possible to determine whether the defect traps are shallow or they are deep. If considered structure has the multiple types of defects that energy levels are located in the energy gap (forbidden zone) with different depths, the dependence $j(U)$ shows a plurality of vertical steps, each of that allows to determine both concentration and depth of corresponding traps. Finally, when the energy levels are distributed in the energy range in the energy gap, the plot 3 will not be a vertical but inclined line, while the angle of inclination enables to find distribution function of energy levels for “sticking” traps.

Data on energy characteristics of defects are important for new dielectric and semiconductor materials development, intended for use in the devices of electronic equipment. The described technique of deep-level investigation by relatively simple electrical measurements (current-voltage characteristics) gives indication

of microscopic structure of crystals. Nonlinear voltage-current characteristic of SCLC with defects can also be used in some applications, as the current can increase by thousands of times. Devices with this feature can be a basis for voltage regulators and other threshold apparatus.

The bipolar injection of charge carriers is characterized by even more complicated volt-ampere characteristics than monopolar injection. One of typical characteristics of such injection is shown in Fig. 7.42B. It is assumed that dielectric comprises only one type of shallow levels. Linear plot with Ohm's law is not shown while the plot 2 is shown with reduced trap mobility of charge carriers. The features of double injection become apparent in plots 3 and 4 of considered characteristics. With current density increases in plot 3, electrons and holes of space charge areas *interpenetrate* each other: electrons neutralize space charge of holes near anode, while holes, in turn, neutralize electronic space charge at cathode. Under these conditions, the restrictive effect of space charge is largely weakened, whereby current density continues its increase even when voltage decreases: current density in plot 4 falls from the monopolar threshold potential U_2 to a smaller value U_3 . The region of unstable current between voltages U_2 and U_3 characterizes the presence in dielectric or semiconductor *injected electron-hole plasma*.

The sharp decrease in resistance of dielectric (or semiconductor) in the plot 4 (in region of *negative resistance*) can be described as "electrical breakdown." This view looks consistent with increasing of electron tunneling near the contacts, supplying nonequilibrium charge carriers. Besides, in the vicinity of instability, the ionization by collisions may occur (caused by fast electrons) that increases charge carrier concentration. Moreover, the photoionization processes are also observed, being activated owing to intense recombination of electrons and holes. However, unlike true breakdown, when electrical current increases infinitely and crystal failure occurs, the current increase in electron-hole plasma, formed by double injection, is limited. Therefore, breakdown of the dielectric (or semiconductor), which is usual in the case of a usual breakdown, does not occur.

First, the recombination of electrons and holes prevents unlimited increase in current that is promoted by lattice defects—centers of recombination. Second, in the event of plasma, the increase in current is still limited by the space-charge effect, the influence of which is only partially neutralized by the charge carriers of opposite sign.

Thus the double injection creates in dielectric (or in wide-bandgap semiconductor) the unique case of "partial breakdown" that is not accompanied by irreversible destruction of crystal. This process can be controlled by the change of voltage or by control of electrical circuit parameters. Instability (i.e., the region of negative resistance) is typical for devices with dual injection. This phenomenon is used in various apparatus development, such as electronic equipment for switching devices, generators, etc. At recombination of nonequilibrium charge carriers the effect of *light emission* occurs, which is a basis for diode-laser operation. They use wide-band semiconductors, which are transparent in long-wavelength part of optical spectrum (near-IR wavelengths).

Returning to Fig. 7.42B, it should be noted that plot 5 of voltage-current characteristic is also characterized by quadratic SCLC (formula 7.41) that occurs in conditions of charge carriers with high mobility (analogous to plot 4 in Fig. 7.42A). However, a further increase in voltage to a value of U_4 changes the character of $j(U)$ dependence that in the plot 6 becomes cubic:

$$j = ae u_n u_p \tau U^3 d^{-5} \quad (7.42)$$

where ε is permittivity, τ is lifetime of charge carriers, a is parameter of crystal, while u_n and u_p are mobility of electrons and holes, respectively. Meaningfully, the voltage U_4 can determine the concentration of recombination levels. It is interesting to note that in the case of double injection the current dependence on distance between electrodes ($j \sim d^{-5}$) is even greater than in the case of monopolar injection. Therefore, for research and application of this phenomenon, preparing fairly thin samples or crystals using thin-film technology is necessary.

The bipolar injection is accompanied by many other options of current-voltage characteristics, which are different from those shown in Fig. 7.42B. On the peculiar form of $j(U)$ dependence, which characterizes the depth of “sticking” levels of electrons (or holes), affects mobility charge carriers, and effectiveness of their recombination. The quality and nature of injection contacts are also very important.

Thus the injection of electrons and holes from metallic electrodes into dielectric results in complicated nonlinear dependences of electrical current on electrical voltage. The study of space-charge limited current allows investigating the nature of defects in dielectric and semiconductor structures.

Frequency dependence of conductivity. Temperature dependence of conductivity, that is, the dependence $\sigma(T)$ is discussed earlier; in the case of space-charge limited current the dependence of conductivity on electrical field $\sigma(E)$ was considered. However, in many cases, for research and application, the *frequency dependence* of conductivity $\sigma(\omega)$ is also important. In accordance with charge carrier physical nature and depending on properties of dielectric, the conductivity with frequency can both increase and decrease.

The increase in $\sigma(\omega)$ is usually caused by the *delay of slow polarization* mechanisms. This effect is conditioned by the close relationship between polarization and conduction processes that can be completely separated only at direct voltage. Fig. 7.43 shows conductivity dependence in dielectrics and semiconductors in a wide frequency range (10^{-4} – 10^8 Hz) for quite different structures and chemical compositions. This community of $\sigma(\omega)$ dependence can be described by the *power law*, established by A. Ionscher [12]:

$$\sigma \sim \omega^n, \quad 0.7 < n < 1. \quad (7.43)$$

This law is peculiar to most mechanisms of charged particle local movement in the dielectrics under an alternating electrical field. The ions and polarons during their “hopping” movement between the states of self-trapping, as well as the dipoles in a process of their rotational vibrations between several equilibrium positions (separated by potential barriers) simulate conductivity while frequency increases.

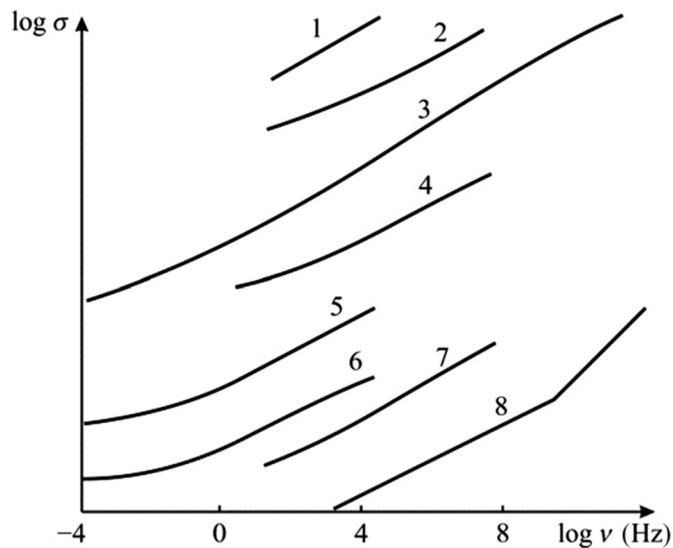


FIG. 7.43

Frequency dependence of conductivity in some dielectrics at different temperatures: 1—covalent crystal silicon (4.2 K), conductivity is caused by electron “jumping”; 2—ionic crystal Al_2O_3 (77 K); 3—molecular crystal anthracene (300 K); 4—phosphate glass $\text{P}_2\text{O}_5\text{-FeO-CaO}$ (300 K); 5—silicon monoxide (300 K); 6—thin film of stearic acid (300 K); 7—amorphous selenium (300 K); and 8—amorphous As_2S_3 , 300 K (by A. Jonscher [12]).

In the same way, other charged particles and complexes that are under electrical field move in a *confined space* and bring the contribution to frequency dependence of conductivity, described by formula (7.43). Thermally activated motion of charged particles, whose localization is determined by a set of potential minima and barriers, in external electrical field gives rise to both conduction and polarization.

At comparatively low frequencies ($\omega \rightarrow 0$), the processes of polarization predominate, as the movement of charged particles in almost constant field is limited by the potential barriers (structural defects and interfaces) that prevent total transfer of electrical charges from electrode to electrode. With frequency increase, the charged particles *do not have enough time* during a quarter of sinusoidal voltage period to reach the places of their localization but continuously follow the change in electrical field, contributing to a conductivity.

For this reason, their contribution to polarization is ceased, resulting in the dispersion (reduction) of permittivity: $\epsilon'(\omega)$ decreases. The difference in height of potential barriers and the distinction in length of charged particle free path explains the very continuous increase in conductivity $\sigma(\omega)$ in a wide frequency range (Fig. 7.43).

Therefore, at subsequent increase in frequency, that is, in more rapid changing electrical field, the *inertia* of charge carriers begins to affect, making their movement

at high frequencies impossible. Sedentary molar ions do not have sufficient time to move in the electrical field already at subsonic frequencies such that electrophoresis is studied and used mainly at direct voltage. The ionic conductivity in dielectrics is delayed already at radio frequencies; therefore this type of charge movement practically has no effectiveness at microwaves. The less inertia mechanism is electronic conductivity, but in dielectrics this mechanism, for the most part, has polaron character, so it is delayed at much lower frequencies than in the semiconductors.

Normal mechanism of electronic conductivity in semiconductors and metals (described by band theory) does not lead to σ frequency dependence over the entire frequency range that is used in electronics (up to terahertz). Nevertheless, as far as frequency increases, in the end, the movement of electrons also manifests their inertia, but $\sigma(\omega)$ -dispersion mechanism in the metals is quite another as in dielectrics (see Chapter 5, Fig. 5.2). Very small (as compared to bias current of polarization) electronic conduction of dielectrics cannot be studied in the IR and optical frequencies, but the mechanism of electronic conductivity dispersion is clearly manifested in metals.

As the permittivity of the metal is negative below the frequency of plasma resonance, the presence of free charge carriers in doped semiconductors and dielectrics *reduces* their optical refractive index on $\Delta\eta$:

$$\Delta\eta = -\frac{n_0\lambda^2 e^2}{8\pi^2 \epsilon_0 m_{\text{ef}} c^2},$$

where c is light velocity and λ is wavelength. This negative plasma contribution to permittivity is noticeable in those materials, in which effective mass of electrons is small. Such crystals are some semiconductors of A^{III}B^V type. The effect of plasmic decrease in permittivity may be used in integrated optics for planar waveguides. Thus frequency dependence of conductivity can be applied in electronics.

7.10 ELECTRICAL BREAKDOWN

Taking into account the increase in operating temperature of electronic and electrical devices, as well as widespread use in electronic thin dielectric films (that are applied in increased electrical fields), considerable attention should be paid to the mechanisms of electrical aging (degradation) and electrical breakdown in dielectrics. In a strong electrical field, owing to the increase in conductivity, the irreversible changes in electrical properties of a dielectric may occur that in solid dielectrics may be accompanied by their destruction.

The *breakdown* occurs, when the strength of electrical field reaches a certain threshold value, above which electrical durability of dielectric (that is characterized by a *small and steady electrical current*) suddenly breaks. In case of breakdown, electrical current through a dielectric increases sharply, which results in the electrical discharge (spark or arc) passing through the dielectric. If the dielectric undergoes strong electrical field of very high frequency (at microwaves), the increased

conductivity occurs in a form *local destruction*. Similar phenomena are observed in case of optical breakdown, when light-transparent dielectric can be locally damaged by the laser beam; this kind of breakdown determines maximum permissible radiation power density in the laser-based devices.

General regularities of electrical breakdown. The development in time of breakdown processes can be divided into distinguishable stages: in the first stage, dielectric loses its *electrical strength*, while in the second stage the mechanical (or thermal, or chemical) *destruction* of dielectric takes place.

The second stage in various dielectrics can occur in quite different ways depending on multitude circumstances, and it is not considered here in detail but only its general features are discussed. The second stage of electrical breakdown is developed to lesser degree, because in this case the particular physical and chemical properties of various dielectrics have significant influence on their destruction. The nature of the second stage of breakdown also depends on the properties of source of voltage: if the power of this source is large, the breakdown occurs in the form of *electrical arc*, while in the case of low power of source the breakdown looks like a *discharging spark* with substantially less destructive force.

Manifestation of the second stage of breakdown also depends on the physical state of a matter. For example, *gases* completely restore their dielectric strength within a short time after breakdown. In the *liquid dielectrics*, their electrical strength after breakdown is also almost completely restored. However, in the *solid dielectrics*, breakdown usually results in irreversible changes, even in case of low-power source of voltage: during breakdown a narrow channel (made by penetrating current) remains, and it has high conductivity after the spark goes out. In case of electrical arc, the significant destruction of solid dielectric (or organic materials) occurs.

Therefore, main attention is paid to the *first stage* of breakdown, when a balance of *small and stationary* current in dielectric (that determines reliable electrical insulation) is violated, and electrical current through the dielectric starts to grow as an avalanche. The analysis of the first stage of breakdown is based on theoretical explanation of *electrical strength* E_{br} and its comparison with empirical determination of breakdown voltage.

In the first stage of breakdown, the main physical mechanism of electrical strength loss is the *intense ionization* process arising while electrons collide with atoms, ions, or molecules. Due to this ionization, the concentration of charge carriers sharply increases that gives rise to electronic avalanches that, in turn, leads to the *electronic* breakdown. This type of breakdown is characterized by the *short duration* of preliminary processes, while dielectric strength E_{br} only depends a little on temperature, on electrical field frequency, and on the properties of environment. Electronic avalanche initiates the *streamer* (leader), which extends *very fast* via the process of *photoelectric ionization*. If the thickness of dielectric is small, electronic breakdown becomes *multiavalanches*.

Various physical and chemical mechanisms give rise to the irreversible processes of evolution in dielectrics, such as *aging* and mechanical or chemical *damage*; these processes significantly vary over a time. In case the loss of electrical strength is due

to the *rapid* electronic processes (electronic avalanches, polarons releasing, etc.), the irreversible process develops in a short time—of about 10^{-6} s. The evolution of other mechanisms of breakdown needs a much longer period.

For example, the *electrothermal breakdown* development occurs in the time interval of 10^{-2} – 10^2 s, which is much slower than in the electronic breakdown. In this mechanism, the amount of released heat (generated in dielectric under electrical field influence due to conductivity and dielectric losses) should exceed the value of heat leakage to the environment. As a result, heat balance in the dielectric might be violated, so a disturbance of *thermal steadiness* occurs owing to electrical conductivity ascending with temperature rise. In a strong electrical field, the dielectric becomes overheated, and, eventually, the breakdown takes place.

The *electrochemical processes*, which lead to *aging* of dielectric before breakdown, are developed even more slowly than heat balance violation. In solid dielectrics, different mechanisms of electrolysis are possible: metallic dendrite germination through dielectric, various electrochemical processes existent on surface and in volume of dielectric. Such phenomena are the *electrodegradation*. These processes lead to significant reduction in electrical strength and can be qualified as the *electrochemical breakdown*. The time at which electrochemical processes take place is estimated in different cases by the time interval of 10^3 – 10^8 s.

Thus evolution time of irreversible processes can be considered as one of the important parameters that can be used to distinguish between possible mechanisms of electrical breakdown. Fig. 7.44 is an example of *volt-second* characteristic. Such dependence can be obtained experimentally by electrical strength E_{br} measurement. To a tested sample, the “saw-tooth” (serrated) voltage pulses is applied being interrupted by breakdown of dielectric. To generate short-term serrated pulses, the special high-voltage impulse generators are used. As shown in Fig. 7.44, there is a great difference between the evolution time of breakdown mechanisms for electronic, electrothermal, and electrochemical types of breakdown.

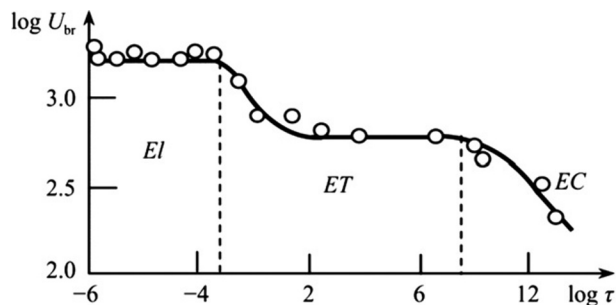


FIG. 7.44

Breakdown voltage dependence on exposure time for thin film (Al_2O_3 , 70 microns), at temperature $T = 1800\text{K}$; EI, electronic breakdown at low exposure time; ET, electrothermal breakdown, EC, electrochemical breakdown (according to S. Koykov).

Maximal voltage for breakdown is a peculiar property of the electronic breakdown; in case of electrothermal breakdown, this voltage is smaller in order of magnitude, and it is even two orders of magnitude smaller for electrochemical breakdown (compared with electrodegradation).

In dielectrics that have different states of aggregation, the mean free path of accelerated electrons is quite different, and this strongly affects the breakdown voltage. In fact, in *gases* (under normal conditions) electrical field of breakdown is less than 10^6 V/m, in *liquids* in case of electronic type of breakdown this strength reaches to the value of 10^8 V/m, while in *solids* it increases up to 10^9 V/m. In the event that electronic breakdown occurs in the *thin films* electrical strength can reach 10^{10} – 10^{11} V/m (a very important fact for microelectronic devices).

The time during which *electronic* breakdown occurs usually is very short (10^{-8} – 10^{-5} s). The large distinction in the interval of time is due to not only peculiarities of dielectrics (gases, liquids, crystals) but also differences in experimental conditions that are determined by temperature, time of exposure, sample thickness, and the rate of overvoltage. Fig. 7.45 shows the dependence of breakdown time on thickness of dielectrics in case of *electronic* breakdown. Such dependence is typical for dielectrics of quite different structure: gases, liquids, crystals. On the one hand, this similarity testifies the *identity of principal mechanism* of breakdown. On the other hand, it can be seen that mechanisms of electronic breakdown in case of diminutive distance between electrodes (including thin layers of solids) are significantly different, in which case that dielectric thickness is relatively large.

In the vicinity of certain critical length δ_{cr} the time of breakdown abruptly changes up to several orders of magnitude (Fig. 7.45). For larger and smaller values of δ , the dependence of breakdown time on the value of δ is opposite. The reason is

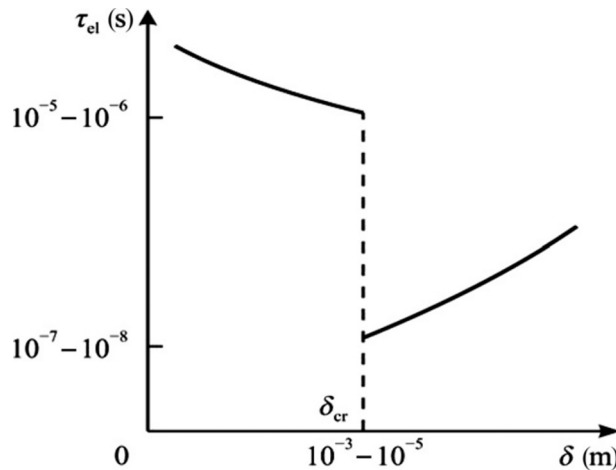


FIG. 7.45

Dependence of electronic breakdown time τ_{el} on dielectric thickness δ .

that at smaller thicknesses of dielectric the electronic breakdown has the *multiavalanche* nature, while in the event that a larger value of δ , the *one-avalanche* breakdown mechanism prevails.

Development time of electronic breakdown characterizes the acceleration of electrons in the electrical field. In solid dielectric, the ionization process by the collisions starts when the energy, which electrons gain from electrical field, becomes greater than the bandgap energy (or equal to it). A possible conception of impact (collision) ionization is as follows: accelerated by electrical field electrons release other electrons from their bonding with atoms (or ions); these secondary electrons move from the valence band to the conduction band.

Typically, in the event of ionization, instead of one “fast” electron (already accelerated by electric field), two “slow” electrons appear in the conduction band; then they should be accelerated by the electrical field and reproduce the ionization, creating already four electrons, and so on. The result is the avalanche of $2n$ electrons, where n is the number of ionization acts.

However, electrical breakdown is a very complicated physical phenomenon that depends on many properties of dielectrics—electrical, optical, thermal, mechanical, and chemical.

Electronic breakdown in crystals. Experimental studies of dielectric strength in solid dielectrics are complicated, because breakdown in solids is irreversible: in contrast to gases and liquids, in this case, for each test it is necessary to prepare a new sample, as during breakdown mechanical or thermal destruction of dielectric occurs. In addition, while studying breakdown in crystals, it is necessary to consider that the surface of solid dielectric is obviously bordered by the gas or liquid dielectric, in which electrical strength is much less than in solid dielectrics.

To prevent the flashover of samples (overlap electrical discharge), they need to have such a geometrical shape in which the path of surface discharge will be lengthened as much as possible (Fig. 7.46A). In addition, to examine only the *electronic* breakdown it is necessary to exclude the possibility of dielectric heating in a strong electrical field and to prevent the rise of electrochemical processes that lead to aging. It should be also noted that *statistical methods* are required for breakdown

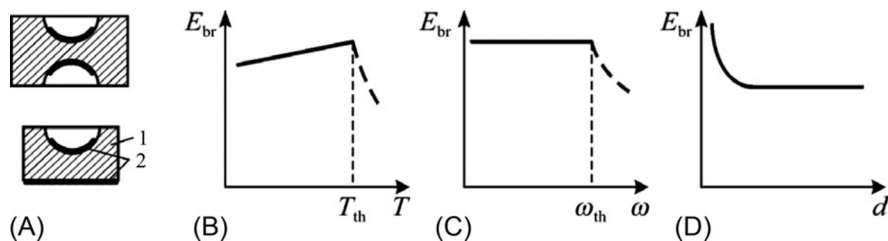


FIG. 7.46

Electronic breakdown peculiarities: (A) different forms of dielectric samples (1) and electrodes (2); strength dependences on: (B) temperature, (C) frequency, and (D) sample thickness (above T_{th} and frequency ω_{th} an electrothermal breakdown starts).

experimental data processing: therefore a large number of research samples are required. The question is that breakdown is a *random* and *local* event in its nature, and in order to make correct conclusion from the experiments it is necessary to investigate dozens of samples [13].

Thus every experimental study of solid dielectrics breakdown is a very time-consuming test. Partly for this reason, electronic breakdown has been studied mainly in single crystals in order to minimize additional complicating factors due to the nonuniform structure of samples. However, electronic mechanism is the most important mechanism of breakdown in solid dielectrics (in polycrystalline and amorphous materials as well). Of particular importance is the electronic breakdown in thin dielectric films that are widely used in microelectronics (just as inorganic dielectric film in chip elements so also in polymer films for protective coatings).

In solid dielectrics, in strong electrical fields, Ohm's law is violated and an *exponential increase* in current before breakdown is seen; however, any saturation in the $j(E)$ dependence is not observed. Electronic breakdown, except for the largest E_{br} value, is distinguishable by short time of the first stage of breakdown. In the event of one-avalanche mechanism, the processes resulting in electrical strength loss are developed during 10^{-7} – 10^{-6} s; however, in thin dielectric films it takes 10^{-6} – 10^{-4} s because the breakdown mechanism is multiavalanche.

Short time of electronic breakdown allows to experimentally separate this mechanism from other forms of breakdown in solid dielectrics. With that end in view, the volt-second characteristic with short exposition time can be used (Fig. 7.44), in which any of electrothermal or electrochemical process has no time to develop. Also note that the electronic mechanism of dielectric breakdown is peculiar to solid dielectrics at very low (cryogenic) temperatures, when the intensity of both thermal and chemical processes is greatly reduced.

In case of electronic form of dielectric breakdown, the strength E_{br} does not depend on the properties of environment, as well as on many other adverse factors. Breakdown voltage is determined by a set of electrical, mechanical, optical, and thermal properties of studied crystal, and therefore, parameter E_{br} may be considered as the most reliable (reproducible in different experiments) characteristic of dielectric properties.

This is confirmed by basic experimental data as to electronic breakdown. Fig. 7.46B–D demonstrates that the value of E_{br} is only little dependent on temperature, it is practically independent of frequency, and does not change with the variation of sample thickness (except samples of very small thickness, when the breakdown mechanism changes). Single-crystal study shows that *channels* of breakdown are *narrow*, *straight*, and *oriented* according to crystallographic axes. Therefore, breakdown strength E_{br} in crystals exhibits the *anisotropy* that is observed even in the AHCs, whose cubic structure demonstrates isotropy of all other electrical characteristics (σ , ϵ , μ) and isotropy of optical refractive index n . For example, in crystals such as NaCl, $E_{br [100]} = 1.6 \cdot 10^8$ V/m, $E_{br [110]} = 1.5 \cdot 10^8$ V/m, and $E_{br [111]} = 1.3 \cdot 10^8$ V/m [14].

It is found that the material of electrode practically has no effect on the breakdown in the AHCs. At that, before a breakdown, the dislocations in crystals appear; then (in initial stage of discharge) the channels of current flow arise, which have a thickness of about $1\ \mu\text{m}$ and increased electrical conductivity. These channels are preferably oriented along the diagonal of cubic crystal, that is, in the direction of [111] axis, and next grow through crystal with a velocity of about $10^4\ \text{m/s}$, while current density in these channels reaches $10^9\ \text{A/m}^2$. In the vicinity of these channels, an intense glow is observed that is caused by electroluminescence. Later these channels widen in their diameter up to about $10\ \mu\text{m}$, wherein the crystal melts owing to high power density.

During the study of electrical breakdown in AHCs important properties have been found: the value of E_{br} is proportional to energy of crystal lattice (Fig. 7.47A). Furthermore, with the increase in lattice parameter, the value of E_{br} decreases (Fig. 7.47B). Thus the binding force of ions (or atoms) in lattice prevents the development of electronic breakdown in solid dielectrics.

Features of electronic breakdown mechanism in ideal crystals, in which almost no defects exists (traps for electrons and holes, donor, or acceptor centers), can be explained as follows. Primarily, it should be noted how “free” electrons in such crystals can appear: first, the strong electrical field liberates electrons from their bound (polaron) state and transforms them into band-electrons state; second, owing to electron injection from metal electrode.

A possible mechanism of electronic breakdown is as follows: electrical field accelerates the released electrons, next the ionization by collisions starts, when electrons (holes) acquire a field energy that is greater (or equal) than the width of band-gap of a crystal. Speeded up by the field, electrons, owing to their interaction with lattice, free other electrons from the valence band to the conduction band. All electrons are accelerated by the electrical field, generating new electrons; therefore the avalanche of electrons with 2^n appears, where n is the number of ionizations.

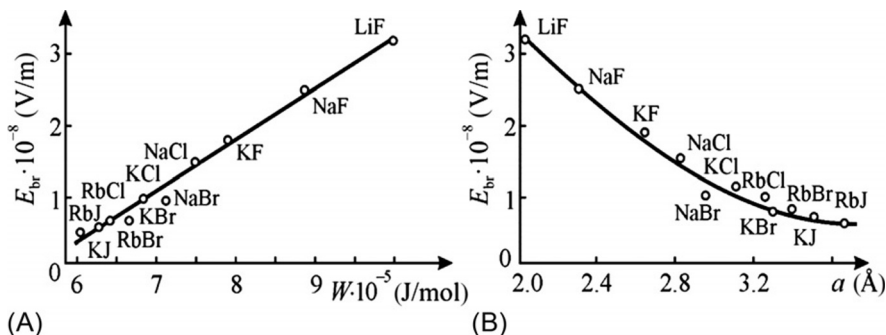


FIG. 7.47

Features of electronic breakdown in AHC: (A) proportionality between crystal lattice energy and electrical strength (by A. Vorobyov [13]), (B) connection between breakdown voltage and lattice parameter.

The value of electrical field, at which electronic avalanches are formed, is determined by the electron–phonon interaction in a given crystal.

In the defect-free crystals, acceleration of electrons is protected by crystal lattice, because during their motion electrons are scattered by lattice vibrations—phonons. It is known that probability of scattering is maximal in case of equality of impulses and energies of interacting quasiparticles. Therefore accelerating electrons actively interact just with *longitudinal* optical phonons, in which direction the impulse vector is consistent with the polarization of electronic wave (that is also longitudinal). The equality in energy is possible only when the energy of accelerating electrons reaches the value of $\hbar\omega_{LO}$, where ω_{LO} is frequency of longitudinal optical mode of lattice vibrations. Exactly at this condition, the transfer of energy from accelerating electrons to the crystal lattice is maximal, that is, there is a maximum electron energy loss when scattering by phonons takes place.

The energy of moving electron is transmitted to crystal lattice by excitation of longitudinal optical oscillations, but due to anharmonicity of vibrations this energy is distributed between other vibrational modes, turning into a heat. Generally speaking, acoustic oscillations near the boundary of Brillouin zone as well as optical vibrations also affect breakdown voltage. Destruction of crystal during electronic breakdown becomes possible due to appearance of the *elastic shock wave* that accompanies the resulting avalanche of electrons. If the electrical field becomes so large that electrons that gain energy from the field exceed a maximum of energy lost, their braking becomes small. They turn into the fast (hot) electrons, and, subsequently, are accelerating in electrical field and generate an electronic avalanche that causes the breakdown.

Thus, the higher the frequency of longitudinal optical mode, the greater the breakdown strength. Fig. 7.48 shows the dependence of breakdown voltage on the longitudinal lattice vibration frequency that is defined at the boundary of Brillouin zone. The frequency is obtained from far-IR spectra, taking into account spatial

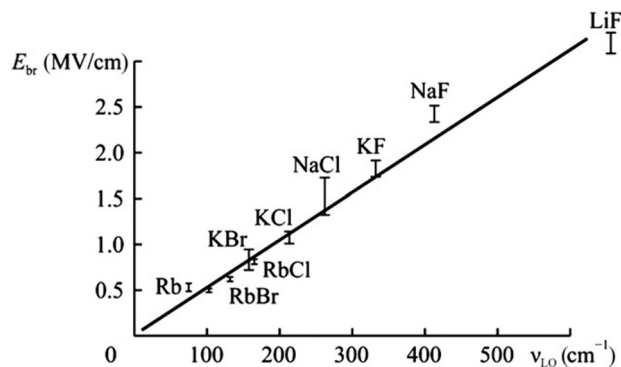


FIG. 7.48

Dielectric strength dependence in alkali-halide crystals on longitudinal optical phonon frequency, defined on the boundary of Brillouin zone ($1 \text{ cm}^{-1} = 30 \text{ GHz}$).

dispersion (i.e., depending ω_{LO} on wave number k). Breakdown strength increases in direct proportion to lattice frequency ($E_{br} \approx 0.05\nu_{LO}$, where ν_{LO} is longitudinal frequency, measured in sm^{-1}).

Thus the main criterion of electronic breakdown in crystals is obtained by comparison of energy, acquired by electron in external field and maximal energy loss of moving electron in crystal lattice.

These ideas about the nature of electronic breakdown are confirmed by experimental data. First of all, small dependence of E_{br} on temperature (Fig. 7.46B) is due to ω_{LO} small dependence on temperature. Some increase in $E_{br}(T)$ is conditioned by the increase in phonon density in LO optical branch with increasing temperature. Crystallographic orientation of breakdown channel is due to lower value of electron energy loss in [110] the direction, and, in particular, in [111] the direction, in which the channel of breakdown is preferably directed in cubic crystals. The anisotropy of E_{br} is conditioned by different values of $\omega_{LO}(k)$ in various directions of spatial dispersion in Brillouin zone. In the center of a zone where $k=0$ frequency ω_{LO} is same in all directions, but the more the value of k (closer to border of zone) the more the interaction of electrons with optical phonons. In AHCs the largest spatial dispersion (reduction $\omega_{LO}(k)$) is observed in the [111] direction; in the same direction the lowest is value of $E_{br} = E_{br[111]}$. Breakdown voltage, according to the described mechanism, should not depend on frequency of electrical field, owing to a very short time of breakdown (at least, up to frequency of 10^6 Hz that is observed in the experiment, Fig. 7.46C).

It is assumed that electronic breakdown in bulk samples is developed in accordance with *one-avalanche mechanism*, similar to streamer in gases. The avalanche creates a space charge near anode, thereby it turns into a streamer that moves from anode to cathode with a velocity of $\sim 10^4$ m/s. This explains a small time duration of breakdown (Fig. 7.45, right part of curve from $\delta = 10^{-5}$ m), but this time increases in proportion to thickness of a sample. In the case of one-avalanche breakdown, the E_{br} is practically independent of thickness (Fig. 7.46D). The theory of electronic breakdown also explains the increase in dielectric strength at very small thicknesses. This phenomenon is associated with the long time of thin film breakdown as compared with bulk samples.

There are some other theoretical assumptions as to the mechanisms of electronic breakdown in crystals. One idea is to understand how breakdown relates to polaron theory, according to which strong field releases electrons from their bound polaron state. Another conception of breakdown as manifestation of domain current instability in dielectric crystals is developed. According to this hypothesis, the formation of narrow channels of electronic breakdown may cause the *pinching* of current channels. A considerable role in the mechanism of breakdown is given also to injection processes.

In the thin films (Fig. 7.45, curve shown on the left from $\delta = 10^{-5}$ m), one avalanche cannot create near anode enough positive space charge that could initiate the expansion of a streamer. However, the needed space charge can be created by *several successive avalanches*, if they fall into the same microarea of film. Breakdown

nature becomes *multiavalanches*. Consequently, the time of development of breakdown significantly increases (by two orders of magnitude), because between avalanches that fall within the same area of anode, time delay inevitably occurs (statistical lag).

It should be noted that evolution time of multiavalanche breakdown becomes greater the smaller the film thickness. For electronic avalanches to gain enough power, thinner layer higher electrical field is required. Therefore, dielectric strength in films is greatly increased. The breakdown voltage in thin dielectric samples can exceed 10^{10} V/m. The effect of breakdown strength increasing in thin dielectric layers is used in microelectronics in structures of metal-dielectric-metal (MDM) and metal-insulator-semiconductor (MIS). Owing to electrodes, a good heat sink is guaranteed; therefore breakdown in thin film has an exclusively electronic nature.

Nevertheless, it is impossible to infinitely reduce the thickness of dielectric film while maintaining low electrical conductivity (high electrical resistance). Currently, microelectronics technology is capable of providing qualitative nitride-oxide film with thickness less than 10 nm (0.01 μ m). However, in such films electrical conductivity dramatically increases for a reason that has no relation to the breakdown: this is quantum effect of electronic *tunneling* through thin dielectric layer. Tunneling effect can be explained by the indeterminacy principle: $\Delta x \Delta p \geq \frac{\hbar}{2}$. This shows that in case of restrictions of quantum particle coordinate its impulse p becomes less certain. Impulse uncertainty can add particle energy to overcome barrier. Thus, with a certain probability, the quantum particle can penetrate through the barrier, while average energy of particle remains unchanged. Therefore, thin film of silicon oxide or silicon nitride becomes a good conductor because of their very small thickness.

It is necessary to note that effect when quantum particle penetrates through energy barrier, which frequently occurs in thin dielectric layers. In case of superconductivity, quantum tunneling through thin dielectric film corresponds to Josephson effect. In electrical engineering, the effect of tunneling of charge carriers through thin oxide film is also very important. Oxide dielectric film often covers many conductive metals (particularly, aluminum), but electrons tunneling through this film provides good conductivity in points of conductor mechanical connection (stranding wires, terminals, etc.).

However, in most cases, in electrical engineering practice the main risk is in electrochemical or electrothermal breakdown (that have much lower dielectric strength), so the study of electronic form of breakdown has mainly theoretical interest.

Optical (laser) breakdown. In a strong electrical field that arises in giant laser pulses, an optical breakdown is possible in the transparent dielectrics. The ionization of structural defects of different origin (impurities, defects caused by abrasion) contributes to laser breakdown. Near structural defects, the ionization wave or local heating is formed that destroys dielectric.

However, there might be another reason for optical breakdown: a self-focusing of light beam in the dielectric due to dielectric nonlinearity. Self-focusing of light is a phenomenon of light wave concentration in a medium, which refractive index depends on electrical field intensity. The refractive index n increases with increasing

field owing to *nonlinearity of electronic polarization* (see Section 7.4). The structure of laser pulse is such that maximal intensity of light falls on the center of a beam. Therefore, the refractive index in the middle of beam inside dielectric becomes larger than that at the edge of beam. Due to this optical inhomogeneity, the dielectric behaves as focusing lens: the thickness of laser beam is gradually reduced, and light intensity increases this effect; that is, self-focusing occurs.

In the event that self-focusing, laser beam becomes much stronger than in case of the usual method of focusing lens, the self-focusing is the concentration of electrical field. In the dielectric, a “focal point” arises, where the intensity of radiation becomes sufficient for ionization and formation of plasma: laser breakdown occurs in a limited volume of a medium where light propagates. It should be noted, however, that in case of dielectric *heating* by the laser beam an opposite phenomenon may occur—light beam defocusing (nonlinear blur of beams). In such a case, the refractive index decreases with intensity: in some dielectrics, heating and thermal expansion (decreasing medium density) reduce n .

Such is, in general terms, current understanding of electronic and optical breakdown in crystals. There are different theories detailing breakdown mechanisms; these items may indeed differ in dielectrics with different structures, as electronic and optical breakdown are rather complex physical phenomena.

Electrothermal breakdown in dielectrics. Thermal form of electrical breakdown is observed only in the *solid dielectrics* (in liquids and gases such breakdown is prevented by convection). In comparison with electronic breakdown, the electrothermal breakdown is characterized by significantly smaller (tens of times) value of E_{br} and much longer duration of preliminary processes (see Fig. 7.44). Therefore, in the event of long-acting electrical field the risk of thermal breakdown is more likely than electronic breakdown [6].

Peculiarities of electrothermal breakdown are strong dependence of breakdown strength E_{br} on sample thickness, on ambient temperature, and on frequency of electrical field (Fig. 7.49). In comparison with same characteristics of electronic

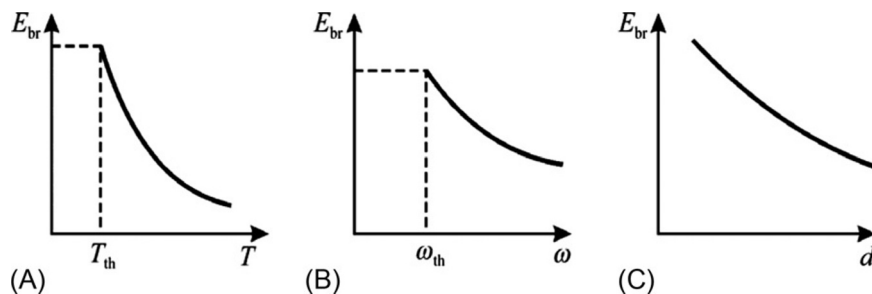


FIG. 7.49

Main characteristics of electrothermal breakdown; dependences on: (A) temperature, (B) frequency, and (C) sample thickness (below T_{th} and ω_{th} breakdown has electronic nature).

breakdown (see Fig. 7.47), electrothermal breakdown strength significantly decreases with increasing temperature. Electrical strength decreases also with frequency and with thickness of tested sample. This means that the probability of thermal breakdown occurrence in a given insulator increases with temperature and frequency.

Electrothermal breakdown occurs due to the violation of thermal equilibrium in a dielectric when it is continuously heating due to the influence of a strong electrical field. The reason is that the amount of heat, released due to dielectric conductivity (or dielectric losses), exceeds heat dissipation to the environment (thermal conductivity of dielectrics usually is small).

The breach of thermal equilibrium promotes a sharp increase in electrical conductivity and in dielectric losses with increasing temperature. Therefore, the overheating of electrical insulator increases *exponentially* with temperature rise. At the same time, heat released from dielectric into a medium is characterized by much weaker (*linear*) temperature dependence. Thus, in the dielectric placed in a strong electrical field in the areas of poor cooling the conditions for *local overheating* are created, resulting in an essential increase in local temperature.

In physics of dielectrics, only the first stage of electrical breakdown is mainly investigated, when the reason for breakdown is discovered. If the reason for *electronic* breakdown is the rise of electronic *avalanches*, in case of *electrothermal* breakdown the loss of electrical strength occurs due to *infringement of steady-state* thermal regime.

Further development of breakdown, that is, its second stage (destruction of dielectric), might occur in different ways for various structures of dielectrics. For example, it may cause a sharp increase in electronic current in thermally weakened local place or mechanical destruction of dielectric due to uneven heating that, in turn, breaks the uniformity of electrical field and causes breakdown. In dielectrics with low melting temperature during their heating, prior to any other electrofield processes occurring, a *meltdown* can occur. This phenomenon is sometimes called as thermal breakdown of *second kind*.

In the simplest case, when electrical field is direct, dielectric is homogeneous, and heat dissipation factor β is known, it is not difficult to estimate breakdown voltage U_{br} in the event of electrothermal mechanism. Thermal power Q_e that evolves in dielectric (its calorification) is determined by electrical voltage U and electrical resistance R that exponentially depends on temperature:

$$Q_B = \frac{U^2}{R} = \frac{U^2}{R_0} \exp[a(T - T_0)] \quad (7.44)$$

where T_0 is ambient temperature, R_0 is value of resistance at temperature T_0 , and a is constant factor. The value of Q_e increases exponentially with temperature (Fig. 7.50). The amount of heat that is removed from dielectric to the surroundings also increases with temperature of dielectric, but linearly:

$$Q_0 = \beta(T - T_0) \quad (7.45)$$

where β is coefficient of heat sink.

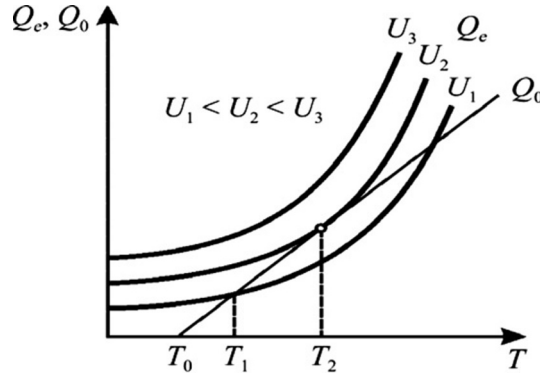


FIG. 7.50

Explanation to voltage calculation of electro-thermal breakdown.

As shown in Fig. 7.50, temperature of dielectric that is subjected to voltage U_1 can rise only to value T_1 when thermal equilibrium is settled; that is, $Q_e = Q_0$. In case of accidental deviations from this equilibrium (e.g., when dielectric would be overheated above T_1), the heat sink exceeds the heat increase, so dielectric will be cooled to equilibrium temperature T_1 .

However, in case of significant increase in voltage (e.g., to value U_3 , Fig. 7.50) the calorification in dielectric at all temperatures will exceed its heat sink. Therefore, when voltage U_3 is switched on, the overheating of dielectric obviously occurs and electrothermal breakdown inevitably comes.

Thus, in dielectric at voltage lower than U_2 the *self-stabilization* of temperature sets, while at voltage higher than U_2 the *overheating* and breakdown occur. Therefore the breakdown voltage can be calculated from the limiting conditions of thermal equilibrium: when the curve $Q_e(T)$ is in contact with the line $Q_0(T)$. This voltage is $U_2 = U_{br}$ (Fig. 7.50). From relations (7.44) and (7.45), as well as from condition of two characteristic tangency, that is, at equality of ordinates $Q_{v2} = Q_{02}$ and equality of derivatives $\left(\frac{dQ_B}{dt}\right)_{T=T_2} = \left(\frac{dq_0}{dt}\right)_{T=T_2}$, the value of *breakdown voltage* and the *maximal temperature* of sustainability can be found, at which theoretically the thermal equilibrium is still possible:

$$U_{br}^2 = \beta R/a; \quad T_{br} = T_0 + 1/a \quad (7.46)$$

It can be also shown that a time, during which process of thermal equilibrium is settled, is given by formula $\tau U^2 \approx \text{const}$, which is confirmed experimentally.

Significant dependence of breakdown voltage on the thickness of dielectric (Fig. 7.49C) is attributed to the fact that thermal conductivity of dielectrics is much lower than thermal conductivity of metallic electrodes. Therefore, by increasing test sample thickness, the overall heat transfer to environment and, thereby, the breakdown strength reduces.

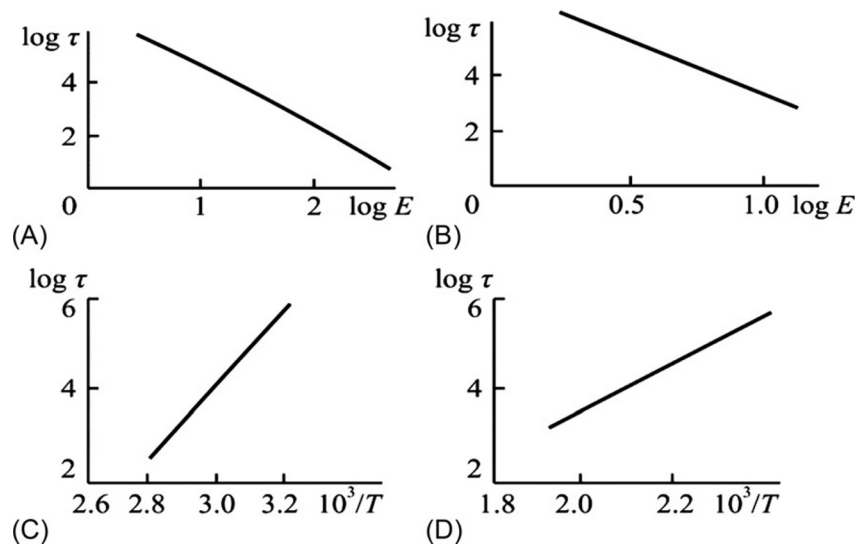


FIG. 7.51

Durability dependence on electrical field and on temperature; (A, C) polyethylene; (B, D) rutile ceramics at constant voltage (a — E in MV/cm; b — E in MV/cm).

Dependence of breakdown strength on frequency (Fig. 7.51B), that is typical for electrothermal breakdown mechanism, may be obtained from the losses of capacitor: $Q_e = U^2 C \cdot \omega \cdot \tan \delta$. From this expression it follows that in case of same calorification (which is determined also by frequency), the breakdown voltage must decrease with frequency rise as $1/\omega^{1/2}$, in full accordance with Fig. 7.49B.

It should be noted that in some experiments with *fragile dielectrics* that contain pores (gas inclusions) observed frequency dependence is much weaker: $U_{br} \sim 1/\omega^{1/3}$. Such dielectrics are, for example, ceramics. In this case, a different mechanism of breakdown becomes apparent, namely, the *electrothermomechanical* breakdown. A feature of this mechanism is that breakdown starts near overheated by *ionized gas* pores of dielectric. This overheating gives rise to uneven thermal expansion of brittle dielectric, resulting in the formation of microcracks and subsequent destruction of a dielectric.

The discussed elementary calculation (in connection with Fig. 7.50) of dielectric thermal stability in a strong electrical field is only a qualitative description of electrothermal breakdown phenomenon. At present, many theories of this mechanism of breakdown as well as the analytical methods of different dielectric structures and devices overheating are well developed in detail.

Electrodegradation (aging) of dielectrics. In strong electrical fields, conductivity of dielectrics becomes nonlinear, at that, dependence $\sigma(E)$ shows a fast increase. Here, if the electrical field does not exceed a certain threshold value, the changes in electrical properties of dielectrics remain *reversible*: while electrical voltage would be switched off, the initial properties of a dielectric are restored.

Conversely, if the electrical field exceeds this threshold, then the *irreversible* changes in properties of dielectric are observed: this is the *electrical aging* culminating in a breakdown. In this regard, it should be noted that electrophysical description of dielectric is not entirely described by the *electronic* breakdown strength E_{br} , but it is also characterized by the *electrothermal* (E_{ET}) and by the *electrochemical* (E_{EC}) breakdown strengths. However, the latest are largely dependent on many random factors, such as impurities and different external conditions; therefore, parameters E_{ET} and E_{EC} cannot serve as the fundamental parameters of a particular dielectric, as the parameter E_{br} for electronic breakdown.

While aging, the duration of electrochemical processes development, that is, the time, taken between electrical field switching on and destruction of dielectric, is called as the *dielectric durability* τ_d or, sometimes, as the “life time.” As in other types of breakdown the smaller the τ_d the higher the voltage. To electrical aging mainly organic dielectrics (polymers) are exposed, but in some cases this phenomenon is also observed in the inorganic solid dielectrics (crystals, glasses, and ceramics). Mechanisms of electrodegradation in various classes of dielectrics are quite different, but some experimental characteristics that describe aging have the *common features*.

First, the relationship between durability and electrical field for both polymers and inorganic solid dielectrics is satisfactorily described by empirical exponential formula:

$$\tau_d = AE^{-m},$$

where exponent $m = 3-4$. Fig. 7.51 shows experimental data for durability of different classes of dielectrics depending on electrical field.

Second, the reduction in durability with temperature in strong electrical fields is also described by exponential law

$$\tau_d = \tau_0 \exp(W/k_B T),$$

where τ_0 is experimental parameter and W is energy of activation.

It should be noted that same temperature dependency is usually observed for relaxation time, and this is the typical characteristic of *thermally activated processes*. Fig. 7.51C and D shows temperature dependences of durability both in the polymers and in the ceramic dielectric that are similar.

Significant difference in mechanisms of electrical aging of polymers and crystals (ceramics) primarily is due to the fact that aging of polymers occurs more rapidly in *alternating voltage*, in which connection of their durability is inversely proportional to the frequency of this field: $\tau_d \sim \omega^{-1}$. In the crystals and ceramics, in contrast, electrical aging occurs primarily at the *direct voltage*.

In inorganic dielectrics, the mechanisms of aging may be associated with charge transfer by ions and electrons. The *ionic conductivity* in direct electrical field always results in irreversible change of dielectric properties, as it is accompanied by electrotransport of matter. The flow of ionic current violates local stoichiometry of dielectric composition. In some cases, while the influence of direct current is

prolonged, even the metallic filaments germinate through the dielectric—dendrites—which ultimately may short electrodes. Obviously, these aging processes have no threshold field, and can occur at any value of voltage.

Electrical phenomena of solid dielectric aging may be associated also with the *electronic conductivity*, but in this case the irreversible changes usually occur only at sufficiently large electrical field. Such aging has been studied in detail as in ceramics and single crystals, containing titanium oxide, so in the AHCs, in which the mechanism of electronic conductivity in *strong fields* dominates over the ionic charge transfer.

The temporal change of current density can be divided into four stages (Fig. 7.52). The *first stage* that takes duration from several minutes to several hours (depending on type of dielectric and temperature) is characterized by some *decreasing* current in time. Before aging, titanium-containing ceramics, for example, demonstrates conductivity of *p*-type, but after the first stage of aging conductivity becomes of *n*-type. Therefore it is natural to assume that at the first stage of aging the influence of acceptors is compensated by donor increase.

At the *second stage*, with a rather short duration (minutes), the current increases by two or three orders of magnitude. It means that the compensation of acceptors by donors is over, but the concentration of donors continues to grow. In some cases, this effect results in such a large increase in current that the electrothermal breakdown occurs, and the stages 3 and 4 cannot be observed. In the other case, if after the second stage of aging the power would be switched off, then, after some time, the original properties of dielectric will be *restored*. This regeneration can be greatly accelerated by heating of dielectric, as well as upon application of electrical field of opposite polarity. For this reason, electrical aging of inorganic dielectrics does not occur at alternating voltage.

At the *third stage* of aging, dependence $\log j$ on $\log \tau$ for a long time (tens of hours) remains almost unchanged (Fig. 7.52). Nevertheless, the properties of dielectric in this case change *irreversibly* due to electrochemical processes. The result comes at last—the *fourth stage* that is characterized by a new increase in current and by

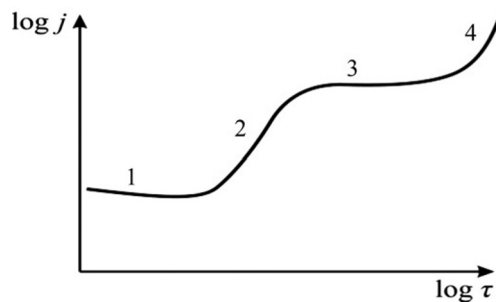


FIG. 7.52

Typical dependence of electrical current density on time while degradation in strong electric fields for inorganic dielectrics.

the breakdown. Tracing the beginning of this stage, it is possible to turn off power in a timely manner and to prevent breakdown. However, complete regeneration of dielectric properties (as it was possible after the second stage of aging) now becomes impossible.

It is assumed that at the third stage of aging the equilibrium concentration of donor is settled, depending on temperature and electrical field. The current remains constant, but some active *electrochemical processes* occur (possibly, in near-electrode regions) that prepare the injection process of holes or electrons. The fourth stage of aging always ends by a breakdown, as it results in a sharp increase in electronic current. It is assumed that this current has injection nature.

Further it has been found that in AHCs with a strong increase in electrical field electrical current increases and aging is accompanied by crystal *coloration*—appearance of so-called *F-centers*. This type of defects in crystal structure consists of anionic vacancy and electron, localized near this vacancy (*F-centers* can be created in crystal also by other methods). Conductivity of colored crystals (especially their *photoconductivity*) rises sharply, and these processes have the electronic nature.

In the titanium-containing ceramics (TiO_2 , CaTiO_3 , BaTiO_3 , SrTiO_3 et al.), as well as in single crystals of the same compounds, during electrical aging coloration is also observed. It is known that such events occur in these substances in case of a lack of oxygen. However, during the aging of solid dielectrics containing rutile the oxygen is not lost, but some other electrochemical reactions occur, resulting in the formation of color centers, involving anionic vacancies.

Thus aging of inorganic dielectrics in a strong electrical field is due, first, to the capture of electrons by the anion vacancies (reversible processes), and, second, to development of irreversible processes, preparing the injection of electrons (or holes) that lead to electrical breakdown. The prerequisite for electrical aging evolution is the presence of ionic component of conductivity. The most intensive aging occurs when ionic and electronic current components are almost equal. Although development of aging process in crystals and polycrystals can be observed as the change of current with time, physical processes of electrodegradation in crystals in many respects remain unclear.

7.11 SUMMARY

1. Most important macroscopic characteristic of dielectric is the *permittivity* ϵ . In the noncentrosymmetric crystals and textures permittivity ϵ_{ij} is a second-rank tensor; therefore due to dielectric anisotropy, permittivity can be diverse in different directions. Permittivity depends also on many conditions: in noncentrosymmetric dielectrics permittivity is different in mechanically clamped and mechanically free dielectrics; in polar dielectrics, there is difference between isothermal and adiabatic permittivity.

2. To describe polarization of dielectrics in macroscopic theory *vector of polarization* \mathbf{P} is introduced that is numerically equal to electrical moment per unit volume of dielectric and directed in accordance with the direction of electrical field \mathbf{E} and symmetry of dielectric crystal. In isotropic dielectric vectors \mathbf{P} and \mathbf{E} are collinear.
3. The *macroscopic field* in dielectric is in ϵ times less than in vacuum, being created only by the “free” charges on electrodes of a capacitor: these charges are not compensated by polarization. That is why the electrical *induction* vector \mathbf{D} is introduced in order to characterize *total* electrical charge on electrodes.
4. The vectors \mathbf{P} , \mathbf{D} , and \mathbf{E} are bound by simple relation $\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P}$ that can be obtained in the electrodynamics by averaging Maxwell’s equations for physically infinitesimal volume and time interval; the result is Maxwell equations in dielectric: $\mathbf{D} = \epsilon\epsilon_0\mathbf{E}$. For most dielectrics Maxwell equations are simplified by the assumption that relative magnetic permeability $\mu = 1$. Energy of dielectric polarization in the electrical field can be found from Maxwell’s equations as well.
5. Induced by electrical field polarized state of dielectric can be described by several mechanisms of *quasielastic*, *thermally induced*, and *space-charge* polarizations, at which the displacement of electrons, ions, or dipoles occurs in external electrical field.
6. The *quasielastic* (deformation type) polarization is almost independent of temperature, and it is the least-inertial polarization, so it determines the high-frequency and optical properties of dielectrics.
7. The *electronic quasielastic polarization* is a common mechanism of polarization, as the deformation of electron shells of atoms or ions in the electrical field occurs in all materials. In addition, it is practically no-inertial polarization mechanism, as the mass of electrons is much smaller than masses of other charged particles (ions or molecular dipoles) involved in the process of polarization. Rapid settling of electronic polarization allows selecting its dielectric contribution using optical experiments.
8. The *ionic quasielastic polarization* is peculiar to such dielectrics and semiconductors, in which ionic character in molecules or crystal lattices is expressed. Time of settling of this mechanism is significantly greater than in case of electronic polarization, as the mass of ions is larger than the mass of electrons. However, the settling time needs to establish ionic polarization ($\sim 10^{-13}$ s) is still much smaller than relaxation times of thermal and space-charge polarization of dielectrics.
9. The model of *quasielastic dipole polarization* might be applied only in case when external electrical field forcedly changes the direction of hardly oriented group of dipoles (at that, elastic force arises, tending to return previous

orientation of dipoles). For this mechanism, the existence of *intrinsic polarization* is necessary, at which all dipoles would be oriented and connected in a structure. Obviously, elastic dipole polarization can occur mostly in the polar crystals, but this polarization can also occur in the liquid crystals. Inertia of this polarization is small and comparable with inertia of ionic polarization.

10. The mechanisms of *thermally activated* (relaxation) polarization are conditioned mainly by the structural defects in dielectrics and leads to ϵ -dispersion and dielectric losses at low frequencies and radio frequencies. Thermal motion of particles in dielectric may greatly affect the polarization process, if dipoles, ions, or electrons are *weakly bounded* in dielectric structure. Remaining localized in the nanoregions, these particles under the influence of thermal motion can make thermal jumps, moving at a distance in the order of atomic dimensions and directed by external field.
11. Thermally activated *dipole polarization* (Debye's mechanism) is peculiar to polar gases, liquids, and crystals. When external electric field is absent ($E = 0$), all dipoles are oriented randomly, so electrical moment per unit volume (polarization) is zero. If $E > 0$, then thermal chaotic motion of dipoles becomes partially oriented along the field, resulting in a new state of equilibrium—polarized. In this case thermodynamic equilibrium (that is due to dipoles thermal vibrations or rotations) becomes changed and favorable for orientation in external field; however, the same thermal motions prevent full orientation of dipoles in electrical field so only some of dipoles become oriented. The higher the electric field strength, the greater is a part of oriented dipoles that causes thermal dipole polarization. Average electrical moment per molecule is proportional to the electrical field, acting on dipoles. Settling time of this polarization is relatively long ($\tau = 10^{-3}$ – 10^{-9} s), being strongly dependent on temperature.
12. The *ionic thermal polarization* is due to thermal jumps of weakly coupled ions (usually impurity ions) in local area of a lattice; thus, such polarization arises mainly in solid dielectrics that have defects in their structure. Ions located in interstices as well as ionic vacancies (voids in regular structure) may change place of their location due to the influence of thermal fluctuations. During these movements (limited by structural defects, for example, dislocations), ion overcomes potential barriers and jumps to new positions, creating electrical dipole. In the absence of external electric field locally limited movement of charged particles is disordered, random, and cannot lead to macroscopic polarization. However, in external electrical field distribution of ions in defect sites of crystal lattice changes, resulting in electrically induced thermal ionic polarization.
13. The *electronic thermal polarization* is peculiar only to solid dielectrics. Being captured by vacancies, electrons (holes) can lead to thermal polarization only in

case when ground state of electrons is degenerated and a combination of degenerate wave functions can create dipole moment. In the absence of electrical field under the influence of thermal fluctuations localized electron (or hole) moves from one to another place of its localization. However, this chaotic movement of charges does not lead to polarization, if external electrical field is absent. Switching on of this field stimulates the “unipolarity” of electronic transitions and leads to field-induced electrical moment per unit volume of dielectric, that is, increases polarization. Relaxation time of thermally activated polarization is rather long: 10^{-7} – 10^{-6} s.

14. Not all possible types of polarization are equally frequent in real dielectrics and play the same role. Moreover, it is impossible to consider various mechanisms of polarization as completely independent of each other—on the contrary, considered basic mechanisms can influence each other; so, strictly speaking, they cannot be discussed as being independent. However, in many cases this analysis is sufficiently accurate approximation.
15. Electrical field, *acting on particles* of solid dielectric (local field), is different from macroscopic field conditioned by polarization – due to polarized environment of a particle. In Lorentz approximation the important formula is obtained for dielectric constant calculating from the microscopic parameters of dielectric: $(\epsilon + 2)/(\epsilon - 1) = (3\epsilon_0)^{-1} \sum n_k \alpha_k$, where α_k is polarizability and n_k is concentration of particles that take part in polarization.
16. The *migratory polarization* is additional mechanism of polarization that may exist in solid dielectrics with inhomogeneous structure: *macroscopic* violations and impurities. This very slow polarization has no time to be set at high frequencies, but can cause significant dissipation of electrical energy (losses). One reason for this polarization appearance is the presence in dielectric some layers of different conductivity or semiconducting inclusions in technical dielectrics.
17. Electrical moment (polarized state) in dielectrics is due to different physical mechanisms. If electrons, ions, and dipoles are constrained in a structure relatively hard (but elastically), the impact of external electrical field results in a small but very fast shift of charged particles from their equilibrium state: this is the *deformation-type* (quasielastic) polarization. On the contrary, when polarization arises with participation of thermal motion of charged particles, the rise of such polarization is a much slower process, which has to be described by other dynamics. Therefore the analysis of different mechanisms of polarization should be different for rather “fast” and relatively “slow” polarization processes.
18. Polarization, which is due to thermally activated motion of charged particles, establishes relatively slowly. Its settling time depends on temperature and at normal conditions (at 300K) $\tau = 10^{-3}$ – 10^{-9} s. Dielectrics are used in technique mainly in the frequency range f 50– 10^{11} Hz; it turns out that just in this range

natural frequencies of dipole thermal motion are located. Therefore undesirable for technical application dielectric losses, as well as ε -instability are caused by *thermally activated polarization*. Migratory (space-charged) polarization is even a slower mechanism that leads to ε -instability and dielectric losses at low (sound range) and infralow frequencies. It is appropriate to note that quasielastic polarization in sound- and radiofrequency range is set almost immediately, so in the frequency range of $50\text{--}10^{11}$ Hz this very fast polarization does not change the value of ε and makes almost no effect on dielectric losses.

19. Dynamic properties of *relaxation* polarization can be described by the Debye formula that also takes into account the contribution $\varepsilon(\infty)$ of rapid processes of polarization:

$$\varepsilon^*(\omega) = \varepsilon' - i\varepsilon'' = \varepsilon(\infty) + \frac{\varepsilon(0) - \varepsilon(\infty)}{1 + i\omega\tau},$$

where τ is relaxation time, $\varepsilon(0)$ is static permittivity constant, while difference $[\varepsilon(0) - \varepsilon(\infty)]$ represents contribution to permittivity from thermally activated (relaxation) polarization.

20. Dynamic properties of ionic crystals depend on a difference between frequencies of longitudinal ω_{LO} and transverse ω_{TO} optical vibrations of a lattice; it is characterized by Lorentz dispersion equation:

$$\varepsilon(\omega) = \varepsilon(\infty) + \frac{\varepsilon(0) - \varepsilon(\infty)}{1 - \left(\frac{\omega}{\omega_{TO}}\right)^2}, \quad \varepsilon(0) - \varepsilon(\infty) = \frac{nq^2}{\varepsilon_0 m \omega_{TO}^2} = \frac{nq^2}{c - \frac{nq^2}{3\varepsilon_0}}.$$

Here n is concentration of ionic couples and m is their reduced mass. Ionic (IR) polarization appreciably affects dielectric properties of crystals: the larger is the permittivity the more q (charge of ions) and the less c (elastic coefficient of ions coupling).

21. Far-IR (ionic or phonons) polarization mechanism is possible only in the ionic crystals, for example, alkali-halide crystals. Ionic polarization is also typical for different classes of metal oxides as well as for chalcogenides. Semiconductors of the $A^{III}B^V$ and $A^{II}B^{VI}$ classes are also partially ionic crystals. Ionic polarization, as well as electronic polarization, predetermines sufficiently low dielectric losses at low and radiofrequencies (up to microwaves) and relatively thermal stability of permittivity. It is important to note that temperature dependence of electronic dielectric contribution ε_{opt} (characterized by low or negative $TC\varepsilon$) is partially compensated by ionic dielectric contribution of ε_{ir} (with small and positive $TC\varepsilon$).
22. In partially ionic semiconductors $A^{III}B^V$, typified by gallium arsenide GaAs, except electronic polarization a contribution to permittivity from ionic polarization is observed. This polarization mechanism corresponds to the far-IR frequency range (about 10^{13} Hz); therefore, below frequency of 10^{12} Hz it does

not result in ϵ -dispersion. Thus, even in the submillimeter waves range (10^{11} – 10^{12} Hz) any frequency change in permittivity in semiconductors of A^{III}B^V group (as in diamond-group crystals) is not observed. Accordingly, loss factor of high-resistive silicon and semi-insulating gallium arsenide at centimeter and millimeter waves is large enough to use these materials as dielectric waveguides or even as resonant microwave elements.

23. The *dielectric losses* are a part of electrical energy converted into heat in dielectric. In quantitative description of dielectric loss there are three main parameters that characterize losses in dielectrics: heat power density p , loss coefficient ϵ'' , and loss tangent $\tan\delta$. They are depicted as

$$p = \sigma E^2, \quad \epsilon^* = \epsilon' - i\epsilon'', \quad \tan\delta = \epsilon''/\epsilon', \quad \tan\delta = \sigma/(\epsilon_0\epsilon'\omega), \quad \sigma = \epsilon_0\omega\epsilon'',$$

where ϵ^* is complex permittivity that consists of real (ϵ') and imaginary (ϵ'') parts while σ is specific conductivity.

24. Frequency and temperature dependences of dielectric losses are determined by distinctive physical mechanisms that define electrical energy dissipation in dielectrics. The most important of these mechanisms are the *conductivity* and the *delay* of thermally activated polarization.
25. Mechanisms of dielectric losses are quite different. In case of elastic polarization dispersion of complex permittivity $\epsilon^*(\omega)$ has *resonant character*, so frequency dependence $\epsilon'(\omega)$ demonstrates not only maximum, but also minimum, while loss factor $\epsilon''(\omega)$ in the vicinity of resonant dispersion shows a rather sharp peak. In case of thermal polarization permittivity dispersion has *relaxation nature* that is characterized by gradual frequency decline in $\epsilon'(\omega)$ and by broad maximum of $\epsilon''(\omega)$.
26. The magnitude of losses as well as their dependence on frequency and temperature indicates certain features of polarization mechanism. Microscopic source of losses might be as *conductivity* so the *anharmonicity* in lattice vibrations. As a rule, the influence of conductivity at microwaves is minimal. The primary cause of anharmonicity in crystalline dielectrics is the asymmetry in electron density distribution along atomic bonds. This is conditioned by differences in the electronegativity of atoms. Sometimes, distinction of different atoms in their electronegativity might be large. Atom with higher electronegativity strongly attracts electron-pair bond, so its true charge becomes more negative. Atom with lower electronegativity acquires increased positive charge. Together these atoms create the nonsymmetric polar bond.
27. Dielectric losses at high frequencies might be the footprints of:
- conductivity;
 - slow polarization mechanism (electronic defects, ionic defects, various kinds of dipoles, etc.) described by the Debye model of relaxation;

- different kinds of structure imperfections that result in frequency-diffused dielectric spectrum depicted by the Cole-Cole model;
 - resonant mechanisms of fast polarization described by the Lorentz oscillator model;
 - in noncentrosymmetric dielectrics and semiconductors microwave loss factor governs by very high-frequency polar mechanism of losses.
28. Frequency dependence of dielectric absorption is convenient to describe by complex permittivity: $\epsilon^* = \epsilon' - i\epsilon''$. Delay of polarization leads to ϵ' decrease with frequency accompanied by ϵ'' maximum. In case of thermally induced polarization ϵ^* frequency dependence is described by Debye relaxation equation, and for quasielastic polarization is described by resonant Drude-Lorentz equation.
 29. Dispersion of dielectric constant is *correlated change* of real and imaginary parts of ϵ^* with frequency alteration. Main property of ϵ^* dispersion is a fulfillment of Kramers-Kronig relations, to which should satisfy any dispersion equation. In a wide range of frequencies and temperatures as well as in different crystallographic directions in dielectrics typically observed several dependences of $\epsilon^*(\omega, T)$ that form the dielectric spectrum.
 30. Depending on charge carrier physical nature, electrical conduction in dielectrics might have electronic, ionic, polaronic, and molar-ionic nature. Mechanisms of charge transport in electrical field can be divided into the *drift* (electrons, holes), the *hopping* (small-radius polarons, ions), and the *diffusion* mechanism (electrons, polarons, ions).
 31. The value of conductivity depends on concentration and mobility of charge carriers. High polarizability, on the one hand, reduces conductivity, as charge carriers are slowing down due to their self-trapping by shift of surrounded ions (in solids), or by carrier link with molecules and particles (in liquids). On the other hand, high polarizability and increased dielectric constant weakens Coulomb interaction of charged particles and thus increases the probability of charge carrier generation, that is, leads to rise of concentration and thus to conductivity increase.
 32. Conductivity in dielectrics depends on many factors: $\sigma(T, E, \omega)$. It increases exponentially with increasing temperature and in higher electrical field as new charge carrier generation. In nonideal dielectrics conductivity increases with frequency rise in the range of 10^{-2} – 10^8 Hz due to polarization lag. With further frequency increase (at very high frequencies), conductivity is reduced by reason of charge carrier inertia.
 33. Directional movement of *ions* transfers not only electrical charge, but also a part of material. In contrast to electronic conductivity, when electrons enter into a material from cathode and transfer charge to anode through a crystal, in case of

ionic conductivity the migration of charge is accompanied by the “mass-transfer.” For this reason, ionic current at direct voltage must be time dependent, as the concentration of charge carriers gradually decreases: ions are accumulated at near-electrode layers. Negative charge carriers—anions—are deposited and discharged at anode, while positive charge carriers—cations—are deposited and discharged at cathode. By measuring of the amount of transferred substance, it is possible to determine what kind of ions involved in the conductivity in different ionic dielectrics.

34. The *surface conductivity* of dielectrics typically has ionic character. It is due to inevitable hydration, oxidation, and soiling of surface in solid dielectric. In this case, the dielectric is characterized mainly by the surface resistivity that is numerically equal to resistance of a square, located on the surface of material when current flows between two opposite sides of this square.
35. Strong *nonlinearity* in current-voltage dependence is observed in case of *injection* of nonequilibrium charge carriers in a dielectric. In the event that injection is *monopolar* current is limited by *space charge* (SCLC). In perfect crystals, SCLC is characterized by *quadratic* volt-ampere dependence. In crystals with defects, their volt-ampere characteristic has kinks and threshold, dependent on electrical field. In case of *bipolar* injection, the effect of space-charge *neutralization* may be observed that results in *negative differential resistance* in the volt-ampere characteristics. Current limiting by SCLC depends on peculiarities of dielectric (semiconductor). Owing to SCLC analyses, many important physical parameters of material can be obtained, such as drift mobility of charged carriers and density of impurities states in the bandgap.
36. Conductivity depends on *frequency*. In accordance with the physical nature of charge carriers and depending on properties of the dielectric, conductivity can both increase and decrease with frequency rising. The increase in $\sigma(\omega)$ is usually caused by a delay of slow polarization mechanisms. This effect demonstrates the interdependence of polarization and electrical conductivity that can be completely separated by only at direct voltage.
37. In a strong electrical field at a certain threshold, the *electrical breakdown* occurs in the dielectric (or semiconductor): electrical current suddenly increases, accompanied by electrical discharge (spark or arc). The main physical mechanism of the first stage of breakdown (when dielectric loses its strength) is electronic ionization by the collisions that results in a sharp increase in concentration of charge carriers while *electronic avalanche* is formed. This type of breakdown in dielectrics (or semiconductors) is the *electronic breakdown*.
38. Main features of electrical breakdown is not only a *rapid increase* in current with voltage rising, but also the *critical character* of $j(E)$ dependence at $E = E_{br}$ (current aspires to “infinity”). Electronic breakdown is characterized by a very

short evolution process when dielectric strength is gone. At that, breakdown voltage is practically independent of temperature, frequency of electrical field, and properties of environment. Typically, electronic avalanches initiate a streamer that moves with the support of *photoionization*. In case of small thickness of dielectric (films), electronic breakdown becomes *multiavalanches*.

39. The *electronic breakdown* in crystals, except for great value of E_{br} , is distinguishable by a *short time* of the first stage of breakdown. Processes that lead to dielectric strength loss in bulk samples happen during time of 10^{-7} – 10^{-8} s (one-avalanche mechanism), or in the thin films during 10^{-5} – 10^{-6} s (when multiavalanche mechanism of breakdown occurs). The value of E_{br} only depends a little on temperature, frequency, and sample thickness (except samples that have very small thickness; in thin films mechanism of breakdown changes). Channels of breakdown in crystals are *straight* and *oriented* accordingly to crystallographic axes.
40. In *thin dielectric films* a single electronic avalanche cannot create sufficiently intense *positive space charge* near the anode, which can initiate the *streamer*. However, such space charge can be created by several successive avalanches, if they strike at same area of dielectric film. Therefore the character of breakdown becomes the *multiavalanches*. As a result, the evolution time of breakdown increases by two orders of magnitude because between the avalanches that fall into same area near anode a *time delay* is seen (statistical delay). The effect of electrical strength increase in thin layers is used in microelectronics in metal-dielectric-metal (MDM) and metal-insulator-semiconductor (MIS) structures, where, thanks to electrodes (that provide *good heat sink*), the breakdown has electronic nature. However, the infinite reduction in dielectric layer thickness with a preservation of low conductivity is impossible. The point is that when films thickness becomes less than 10 nm (0.01 μm) electrical conductivity increases sharply due to *quantum tunneling* of electrons through thin dielectric layer.
41. The *optical breakdown* in light-transparent dielectrics is conditioned by structural defects in dielectric and by *self-focusing* of laser beam. Near the defects the ionization wave is formed that results in *local heating*, which destroys dielectric. Self-focusing of light is a phenomenon of field concentration in the nonlinear medium whose refractive index depends on field intensity. Owing to nonlinear change of electronic polarization, the refractive index of dielectric increases with increasing field. The structure of laser pulse is such that maximum of light intensity is located at the center of a beam; as a result, lens effect occurs and dielectric becomes locally damaged.
42. The *electrothermal breakdown* is the loss of thermal stability in solid dielectrics owing to calorification in strong electrical field caused by electrical conductivity or dielectric losses. Overheating might occur so rapidly

that it cannot be compensated by heat-dissipation processes. In case of electrothermal mechanism of breakdown, the electrical strength decreases with temperature and with frequency rise.

43. The *electrochemical* reactions in solid dielectrics that lead to irreversible changes of their properties and significantly reduce electrical strength (as well as reliability of electrical insulation) are the *aging* (degradation). In *inorganic* dielectrics, aging takes place mostly in the *direct* electrical field and consists in the formation of structural defects that contribute to the increase in electrical current. In the *polymers* aging occurs mainly in the *alternating* voltage due to the occurrence of partial discharges in gas inclusions that result in erosion of polymer films surface, or dendrite-type appearance in the volume of polymeric insulation.

REFERENCES

- [1] L.H. Van Vlack, *Materials Science for Engineers*, Addison-Wesley Publishing Co., Reading, Massachusetts, 1975.
- [2] N.P. Bogoroditsky, Y.M. Volokobinsky, A.A. Vorobyov, B.M. Tareyev, *Theory of Dielectrics*, Energia, Moscow, 1965.
- [3] Y.M. Poplavko, *Dielectric Spectroscopy of Solids: Basic Theory and Method Application*, LAMBERT Academic Publishing, Saarbrücken, 2013.
- [4] I.S. Rez, Y.M. Poplavko, *Dielectrics: Basic Properties and Applications in Electronics*, Radio e Svyaz, Moscow, 1989.
- [5] Y.M. Poplavko, *Dielectrics at Microwaves: Thermostable Ceramics, Ferroelectrics, Films, Composites*, LAMBERT Academic Publishing, Saarbrücken, 2014.
- [6] G.I. Skanavi, *Physics of Dielectrics*, Gosfizmatizdat, Moscow, 1958.
- [7] N.F. Mott, *Metal–Dielectric Transitions*, Taylor and Francis Ltd, London, 1974.
- [8] I.S. Vortyagin, Y.M. Fomichev, Y.M. Poplavko, Features of dielectric non-linearity in paraelectrics, *Ukr. J. Phys.* 60 (4) (2015) 339–350.
- [9] G. Fröhlich, *Theory of Dielectrics*, Clarendon Press, Oxford, 1958.
- [10] R.E. Newnham, *Properties of Materials. Anisotropy, Symmetry, Structure*, Oxford University Press, Oxford, 2003.
- [11] Y.M. Poplavko, Y.I. Yakimenko, Y.V. Didenko, in: *Physical mechanisms determining microwave dielectrics properties; Part 1: Thermal stability nature, Part 2: Dielectric losses nature*, International Conference on Electronics and Nanotechnology, Kiev, ELNANO-2015, 2015.
- [12] A.K. Jonscher, *Dielectric Relaxation in Solids*, Chelsea Dielectric Press, London, 1983.
- [13] A.A. Vorobyov, G.A. Vorobyov, *Electric Breakdown and Destruction of Solid Dielectrics*, Vishaya shkola, Moscow, 1966.
- [14] A.R. Von Hippel, *Dielectrics and Waves*, John Wiley, New York, 1954.

Semiconductors

CONTENTS

8.1 Defining Characteristics and Classification of Semiconductors	409
8.2 Fundamentals of Band Theory of Semiconductors	414
8.3 Intrinsic and Extrinsic Semiconductors	427
8.4 Kinetic Processes in Semiconductors	438
8.5 Optical Phenomena in Semiconductors	452
8.6 Semiconductors in Magnetic Field	465
8.7 Nanoscale and Quantum-Dimensional Effects	485
8.8 Summary	501
References	508

Physical properties of semiconductors are studied most thoroughly in comparison with metals and dielectrics. The main stimulus for semiconductor investigation is the wide production of semiconductor devices and integrated circuits. In addition, in semiconductors, many interesting and important physical effects can be analyzed, which cannot be observed in other substances. These effects are caused by the peculiarities of electronic band structure in different semiconductors, including small energy gap (some foundations of band theory, which is used in physics of semiconductors, are reviewed in [Section 4.6](#)).

8.1 DEFINING CHARACTERISTICS AND CLASSIFICATION OF SEMICONDUCTORS

Semiconductors are large class of solids whose conductivity σ occupies the intermediate position between conductors and dielectrics: in metals $\sigma > 10^5$ S/m, while in dielectrics $\sigma < 10^{-10}$ S/m (at room temperature). A characteristic feature of semiconductors that distinguish them from metals is *increasing conductivity* with temperature rise. In contrast, metals' conductivity decreases with temperature. In semiconductors, exponential increase in $\sigma(T)$ dependence is observed in a wide temperature range:

$$\sigma = \sigma_0 \exp(-E_g/2k_B T),$$

where E_g is activation energy of electrons. As seen from this formula, near-absolute zero (when $T \approx 0$) semiconductor shows insulator property: $\sigma \approx 0$.

Such temperature dependence of conductivity in semiconductors is a result of breach of electronic bonds in crystal lattice, so that part before bounded electrons that is proportional to $\exp(-E_g/2k_B T)$ turns into free charge carriers. Electronic bonds in semiconductors can be destroyed not only due to lattice thermal motion, but also by various external influences: light illumination, stream of fast particles, strong electrical field, and so on. Semiconductors differ from conductors by strongest dependence of conductivity on impurities and various types of radiations [1].

For these reasons, semiconductors are characterized by high sensitivity of conductivity to external influences, as well as to the concentration of impurities and defects in crystals, because activation energy of electrons, localized near impurities or defects, is significantly less than activation energy of electrons in the regular crystalline lattice of semiconductor. The ability to manage conductivity in a wide range by electrical and magnetic fields, temperature change, introduction of impurities, etc., is the basis of many and varied applications of semiconductors.

Semiconductor materials can be classified on the basis of their chemical composition (inorganic or organic) and on the basis of their structure (crystalline, amorphous, and liquid semiconductors, the latter are not considered here). In most practical applications, inorganic crystalline semiconductors are used, which are divided into several basic groups [2].

Monoelement semiconductors. In Mendeleev's periodic table of elements, it is possible to find only a few *simple elements*, which are semiconductors. The most important monoelement semiconductors are found among the elements of group IV: carbon (C in graphite, diamond, and other forms), silicon (Si), germanium (Ge), gray tin (α -Sn). Whereas graphite properties are similar to conductors ($E_g < 0.1$ eV), natural diamond (C) is close to dielectric ($E_g > 5$ eV). However, artificially grown diamonds with impurities acquire the properties of semiconductors. Most important semiconductors from group IV are Ge and Si that have crystal lattice of diamond type. Together they can form continuous series of solid solutions Ge-Si that also have semiconductor properties.

Boron (B) can be included in the monoelement semiconductor of group III. Group V includes phosphorus (P), arsenic (As), and antimony Sb, while group VI is presented by sulfur (S), selenium (Se), and tellurium (Te). Iodine (J) belongs to group VII of semiconductors.

Chemical compound semiconductors. Crystalline structure of many semiconductor compounds is characterized by tetrahedral coordination of atoms as well as by diamond-type lattice. Therefore, these compounds are called as diamond-like semiconductors. Most scientific and practical interest is seen in the *binary* compounds ($A^{III}B^V$ and $A^{II}B^{VI}$) that at present are very important materials in optoelectronics. Modern semiconductor technologies use also *ternary* and even more complex chemical compounds.

Diamond-type semiconductors have similar properties and they can create isovalent solid solutions between them. In these solid solutions, by changing the composition of a compound, it is possible to control (smoothly and in very wide range) important properties of semiconductors, including band gap and charge carrier

mobility. This opens an additional opportunity for optimization of device parameters and enables better coordination of physical characteristics of various components of electronic equipment.

Binary compounds. Firstly, there are of about 30 crystalline compounds of three- and five-valence elements of $A^{III}B^V$ type; of these, the most used are GaAs, GaSb, InSb, AlAs, GaN, and others. Secondly, there are known (and widely applied in electronics) crystals of $A^{II}B^{VI}$ group: ZnS, ZnSe, CdS, CdSe, HgSe, HgTe, and others; they are also about 30.

Many others binary compounds belong to semiconductors: A^IB^{VII} (SuCl, AgBr, ...), A^IB^{VI} (Cu_2O , CuS, ...), A^IB^V (KSb, K_3Sb , ...), $A^{II}B^{IV}$ (Mg_2Sn , Ca_2Si , ...), $A^{III}B^{VI}$ (GaS, In_2Fe_3 , ...), $A^{IV}B^{VI}$ (PbS, PbSe, ...), $A^{IV}B^{IV}$ (SiC), A^VB^{VI} (Sb_2Te_3 , Bi_2S_3 , ...).

Ternary compounds may also have semiconductor properties. Most important among them are cuprites ($CuAlS_2$, $CuInS_2$, $CuSbS_2$, $CuFeSe_2$), compounds based on zinc ($ZnSiAs_2$, $ZnGeAs_2$), lead ($PbBiSe_2$), cadmium ($CdHgTe$), and others. Ternary compounds of $A^{II}B^{IV}C^V$ type usually are crystallized in a lattice of chalcopyrite; also they can form solid solutions with each other. Considering their electronic properties, compounds $A^{II}B^{IV}C^V$ resemble analogues of $A^{III}B^V$ compounds. Typical representatives are $CdSnAs_2$, $CdGeAs_2$, and $ZnSnAs_2$.

$A^{III}B^V$ compounds are most important representatives of binary semiconductors. They are the closest analogue of silicon and germanium. The $A^{III}B^V$ crystals are formed by the interaction of elements of subgroup III (boron, aluminum, gallium, and indium) with elements of subgroup V (nitrogen, phosphorus, arsenic, and antimony). It should be noted that bismuth and thallium compounds relate to the toxic materials.

Semiconductors of $A^{III}B^V$ group are classified by their metalloid element. Accordingly, nitrides, phosphides, arsenides, and antimonides are distinguished. Except nitrides, all $A^{III}B^V$ compounds crystallize in the lattice of *cubic sphalerite* (zinc blende). Note that semiconductors of sphalerite structure show *piezoelectric* properties. In case of nitrides, the wurtzite structure of hexagonal type is typical; this structure demonstrates not only piezoelectric but pyroelectric properties as well.

Regardless of lattice type, each atom of group III element is in tetragonal environment by four atoms of group V element, and vice versa. Sphalerite structure, unlike the structure of diamond, has no center of symmetry. This feature leads to the differences in physical properties of (1 1 1) and ($\bar{1}\bar{1}\bar{1}$) faces, which (theoretically) are presented by atoms III and V groups. Different treatment of mentioned faces is detected during etching, oxidation, and crystal growth.

All $A^{III}B^V$ compounds are characterized by *mixed covalent-ionic* chemical bonds (ionic component proportion is 15%–20%). Sometimes, chemical bonds in these compounds are treated as a special type of chemical bond, called as *donor-acceptor* bond. This term means that three of four covalent bonds (by which each atom is embedded in lattice) are formed with socialization of three valence electrons (from A^{III} and B^V), while the fourth connection is created by nonsegregated pair of valence electrons of atoms B^V . The formation of these types of bonding corresponds to

energetically favorable transition of electrons from atom B^V to such energy state, which is shared with donor (atom B^V) and acceptor (atom A^{III}).

Many $A^{III}B^V$ semiconductors can form continuous series of solid solutions between them—triple and even more complex ($Ga_xAl_{1-x}As$, $GaAs_xP_{1-x}$, Ga_xIn_{1-x} , $Ga_xIn_{1-x}As_yP_{1-y}$, and so on), which are very important in modern electronics.

$A^{II}B^{VI}$ compounds. Semiconductors of the $A^{II}B^{VI}$ group are characterized by a crystal structure type of sphalerite or wurtzite (rarely—by structure of NaCl type). The relationship between atoms in the lattices has a *covalent-ionic character* (proportion of ionic component is 45–60%). For $A^{II}B^{VI}$ compounds, the polymorphism and polytypicism of cubic and hexagonal modifications are peculiar.

The most important members of this group of semiconductors are crystals CdTe, CdS, ZnTe, ZnSe, ZnO, and ZnS. Many compounds of $A^{II}B^{VI}$ type can form together continuous series of solid solutions, typical representatives of which are $Cd_xHg_{1-x}Te$, $Cd_xHg_{1-x}Se$, and $CdTe_xSe_{1-x}$. Physical properties of such compounds are largely determined by the nature of intrinsic point defects that have low ionization energy and show high electrical activity.

The $A^{IV}B^{VI}$ compounds are of great practical importance. They crystallize in the NaCl-type structure or in *orthorhombic* structure, and they also are characterized by covalent and ionic types of chemical bonds. Most important representatives of this type of compounds are semiconductors PbS, PbSe, PbTe, and SnTe; among them, the most common are solid solutions $Pb_xSn_{1-x}Te$ and $Pb_xSn_{1-x}Se$. Compounds such as $A_2^{III}B_3^{VI}$ also have technical interest; many of them have a crystalline structure of *sphalerite* type, in which one-third of cationic nodes are empty; typical representatives are Ga_2Se_3 , Ga_2Te_3 , and In_2Te_3 . Among compounds of group VI with transition metals and rare-earth elements, there are many semiconductors with ionic character of bonding, characterized by *ferromagnetic* (or antiferromagnetic) properties.

The silicon carbide (SiC) is a single chemical compound, formed only from elements of group IV. Silicon carbide shows semiconducting properties in all its structural modifications, both in the β -SiC (sphalerite structure) and in the α -SiC (hexagonal structure, which has of about 15 varieties). This is one of most heat-resistant and wide-band crystals of semiconductor materials used.

Amorphous semiconductor materials. Typical representatives of amorphous semiconductor materials are glassy semiconductors—chalcogenides and oxides. Chalcogenides include alloys of Tl, P, As, Sb, Bi with S, Se, and Te. They are characterized by a wide range of values of electrical conductivity, low temperature of softening, and high resistance to acids and alkalis. Typical representatives are solid solutions As_2Se_3 - As_2Te_3 and Tl_2Se - As_2Se_3 .

Oxide glassy semiconductors, such as the compounds of V_2O_5 - P_2O_5 - RO_x type (R is metal of group I–IV), are characterized by electrical conductivity of 10^{-4} – 10^{-5} $\text{ohm}^{-1}\text{cm}^{-1}$. All glassy semiconductors have the hole-type conductivity and also exhibit *photoconductivity* and *thermoelectromotive* properties. When they are slow cooled, they usually turn into the crystalline semiconductors.

Another important class of **amorphous** semiconductor materials is solid solutions with hydrogen: hydrogenized noncrystalline semiconductors are α -Si:H,

$\alpha\text{-Si}_{1-x}\text{C}_x\text{:H}$, $\alpha\text{-Si}_{1-x}\text{Ge}_x\text{:H}$, $\alpha\text{-Si}_{1-x}\text{N}_x\text{:H}$, and $\alpha\text{-Si}_{1-x}\text{Sn}_x\text{:H}$. A number of features have been discovered and identified in the conductivity mechanism of amorphous inorganic semiconductors. Hydrogen, which is contained in these semiconductors, has high solubility and “locks” a significant number of free connections that are characteristic of amorphous semiconductors. As a result, the density of energy states in the band gap is sharply reduced; therefore, creation of p - n junction is made possible.

Organic semiconductors (Fig. 8.1) are solid substances that have electron or hole conductivity. Organic semiconductors are characterized by a system of intermolecular valence bonds. Charge carriers in organic semiconductors are formed by the excitation of electrons, delocalized within the polymer chains. Activation energy, required for the formation of charge carriers in organic semiconductors, decreases with an increase in connections of polymer molecule, and it can reach a level of middle heat energy.

Most organic materials are electrical insulators with a value of conductivity at room temperature in the range of 10^{-9} – 10^{-14} S/cm. Usually they are represented by molecular structures, which do not have a system of covalent bonds. Therefore, quantum interactions between highest molecular orbitals of neighboring molecules are weak, and the valence zone, formed by these interactions, is found as very narrow. The conductivity band that occurs during such weak interactions between orbitals is also very narrow; therefore, energy gap appears just as the same as the area of free molecules.

In order to increase conductivity and, hence, awake semiconducting properties, it is necessary to reduce the gap between orbitals. This can be done by adding some heteroatoms with a pair of valence electrons (e.g., polyacetylene, polyaniline, or polyaromatic) to organic compounds. This reduces electron transitions between conduction and valence bands and enables these substances to get properties of semiconductors.

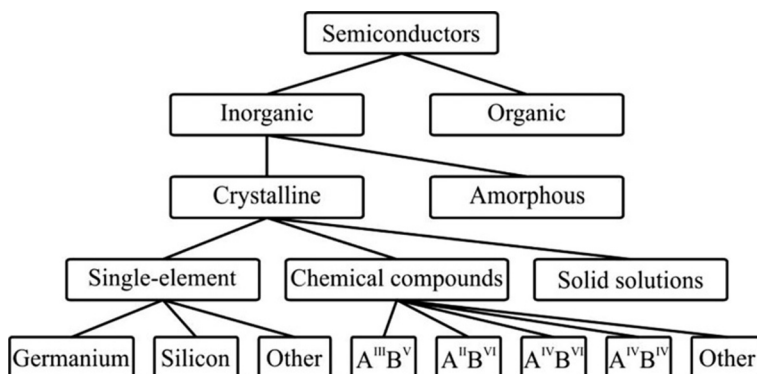


FIG. 8.1

Classification of semiconductors.

Organic semiconductors might have formed of single crystals, polycrystals, amorphous powders, and films. The values of resistivity ρ at room temperature in organic semiconductors are found from 10^{18} ohm cm (in naphthalene and anthracene) up to 10^{-2} ohm cm (in ionic-radical salts). The highest conductivity is seen in the ionic-radical salts, mostly based on anion radical tetracyanogen-hinodimerhyl. They exhibit electrical conductivity that is close to the conductivity of metals. In the organic semiconductors having low electrical conductivity, the phenomenon of *photoconductivity* is usually observed [3].

In disordered organic semiconductors, the usual band electronic transport cannot be realized, and charge carriers (polarons) are transported via jumps between localized states, scattering at every jump (hopping conduction). Jumping is facilitated by polaron interaction with phonons, making charge mobility increase with temperature, but it remains very low ($u \ll 1 \text{ cm}^2/\text{V s}$).

The boundary between hopping and band electronic transport mechanism is determined by mobility value; in case of hopping transport $u < 0.1 \text{ cm}^2/\text{V s}$, but in the case of band transport, mobility is higher: $u > 1 \text{ cm}^2/\text{V s}$. Highly ordered organic semiconductors, such as anthracene and pentacene, have charge carriers with increased mobility that is in addition temperature independence, including polycrystalline thin films of pentacene. This property favors the following argument: activated by temperature hopping the charge transport mechanism can be implemented for high-quality thin films of pentacene.

8.2 FUNDAMENTALS OF BAND THEORY OF SEMICONDUCTORS

According to quantum theory of solids (see Chapters 4 and 5), in the ground state of semiconductors, all allowed energy levels of electrons are occupied. In total, these levels form the *valence band*. In order to “liberate” electrons and cause electrical current, it is necessary to supply crystal energy from the outside (e.g., to heat a crystal). Due to this influence, electrons will be thrown from the *valence band* over the *band gap* to energy levels of the *conduction band* (zone of free electrons).

It is necessary to recall that in semiconductors between valence band and conduction band there is an intermediate energy band (band gap, or forbidden zone). The peculiarity of this area is the complete absence of permitted levels. Band gap equals to the energy that is needed to add to transfer electron from its bound state in the valence band into a free state in the conduction band; usually, the band gap value in semiconductors equals 0.5–2.5 eV.

Elementary theory of electronic energy bands in semiconductors takes into account a fact that mass of atoms (or rather atomic core) is much greater than mass of electron, and, therefore, the nature of motion of these particles is quite different [4]. Heavy atoms can even be considered as fixed; however, it is assumed that their movement can be taken into account by a certain averaged field. In this case, the energy of atoms and the energy of electrons in a crystal can be found from *independent* Schrödinger equations. It means that Hamiltonian (function, defined by

generalized coordinates and impulses) of initial Schrödinger equation for crystal breaks into the *sum of Hamiltonians* for heavy and light subsystems.

Wave function will be the product of two factors, one of which depends on the atomic coordinates, and the other only on the coordinates of electrons. This approach is called *adiabatic approximation*. Its main result is the *wave function* that describes the behavior of valence electrons in a crystal. This wave function can be found from the equation whose parameters are spatial coordinates, describing the instantaneous position of atoms [5].

Thus the adiabatic approximation is based on the fact, that $m_e \ll M$, because atoms are relatively heavy particles (approximately 10^4 times more than electron). For this reason, electron can, without inertia, monitor the instantaneous position of nuclei in the process of its movement, while influence of nuclei is reduced to accounting of the *average field*. Due to inequality of masses, it is considered to hold nuclei as being fixed in their position. This makes it possible to separate the nuclear and the electronic subsystems, and independently analyze processes taking place in these subsystems.

However, within adiabatic approximation, any information about thermal vibrations of atoms (phonons) is lost, as well as the interaction of electrons with thermal vibrations of atoms (electron-phonon interaction) is not considered. The energy spectrum of thermal vibrations of atoms (phonon spectrum) is determined by solving the semiclassical equations, as atoms are heavy particles and quantum effects in this case are expressed very poorly. Nevertheless, the information about electron-phonon interaction is essential for understanding the theory of electron scattering of phonons.

The operational Schrödinger equation for the steady state of a system of atoms and electrons: $H\Psi = E\Psi$ has as many solutions (with wave functions Ψ and energy values E) as many nuclei and electrons are in the solid. The density of solid is 10^{22} – 10^{23} atoms in cm^3 , about the same as that for electrons. Therefore the exact solution of Schrödinger equations in such conditions is impossible. The general principle of solving equations for a large number of interacting particles lies in the necessity to transform the system of Hamiltonians in such a way that it will be the simple sum of individual Hamiltonians. Then solving only one equation would suffice $H_k\Psi_k = E_{kk}\Psi_k$, that is, to have a whole range of its solutions, and get specific information about what occurs in the system of valence electrons and atoms. Electrons of inner atomic shells usually are considered together with their nuclei, as they cannot participate in the formation of valence bonds in solids [5].

Next, the multielectron problem should be reduced to *one electron*, because in the adiabatic approximation wave function of electrons in a crystal would depend on 10^{23} variables. With this in view, the *average effective field* is introduced into a consideration, which replaces the interaction of electrons. The energy of electron is a sum of kinetic energy, potential energy in lattice field (interaction with nuclei), and potential energy in average effective field, created by all other electrons.

The one-electron approximation consists in the *middle-field theory* that supposes that at some point of interatomic space r_i the i th electron is conditionally fixed. Other electrons create at this point a certain potential $U(r_i)$. A similar procedure is

performed to another point in a space, but only within the nearest neighboring atoms [5]. As crystalline structure has translational symmetry, configuration of potential field is *periodically repeated*. Thus the sequence is determined by the atomic potential field in a space, generated by all electrons except *i*th. As electrons are the *indistinguishable* microparticles, *i*th electron can be like any other. With these assumptions, it is enough to solve only *one equation* to determine the wave functions and the energy spectrum of all electrons in a crystal.

Thus the Hamiltonian of electron system is divided into the sum of Hamiltonians, which depend only on coordinates of individual electrons. At that, electronic wave function of system is the linear combination of single-electron wave functions. Schrödinger equation is represented by a set of equations for finding the wave functions of individual electrons. The middle effective field (*self-consistent field*), by which interelectron interaction is represented, demonstrates that there is no need to calculate but more important to show that by this way the multielectron task can be reduced to the *one-electron task*.

Therefore, because of adiabatic approximation and self-consistent field, introducing the task is reduced to the problem of a single electron behavior in a potential field of crystal lattice [6].

Now, let us consider the *type of wave functions* of electrons in the crystal. Each wave function describes the quantum state of the electron, and, being in a steady state, the electron has certain energy. In other words, every wave function is always the function of coordinates and it always includes the options that determine the value of energy E . Taking into account translational properties of electronic wave function, it can be shown that in case of spatial shift in the periodic structure on the vector of translation \mathbf{n} the electronic wave function is multiplied by the phase factor $e^{i\mathbf{k}\mathbf{n}}$. The eigenfunction of translational operator is

$$\Psi(\mathbf{r}) = \varphi(\mathbf{r})e^{i\mathbf{k}\mathbf{n}}$$

This plane-modulated wave is the *Bloch wave*. So, starting from a position that Schrödinger equation for electron in crystal can be reduced to a system of linear homogeneous algebraic equations, in which energy is included, it is possible to conclude that the *energy spectrum* of electrons in a crystal has *band character*, consisting of prohibited and allowed energy bands (Fig. 8.2) [1].

The existence of a set of different functions $E_1(k), E_2(k), \dots$ constitutes the foundation of band theory of solids. These bands can be closed partially or completely (as for metals, see examples in Figs. 4.26 and 4.27 in Chapter 4), but may not overlap. This means that bands can form regions of energy, which do not have any electron under any conditions (as for dielectrics, see Fig. 4.26). They are bands of forbidden energy (or band gaps). It can be shown that the greater the number of higher energy, the wider the area of permitted and prohibited zones, and the narrower the energy intervals (when $E \rightarrow \infty$ spectrum becomes continuous).

One of the important consequences of wave function translational properties is the presence of physically equivalent points in electronic states. As is known from quantum mechanics, the physical meaning of a square modulus of wave function of

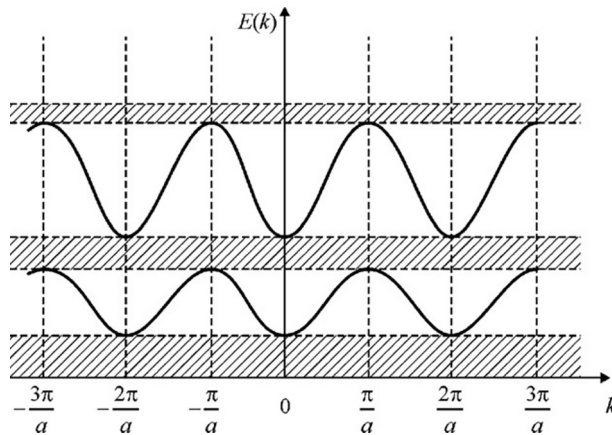


FIG. 8.2

Permitted and prohibited (shaded) bands of energy.

coordinates $|\Psi(\mathbf{r})|^2$ is the *probability* of electron being at the point with coordinates \mathbf{r} . In case of electron shift on a vector \mathbf{n} , the squared modulus $|\Psi(\mathbf{r} + \mathbf{n})|^2$ does not change. This means that in points \mathbf{r} and $\mathbf{r} + \mathbf{n}$ the probability to find electron is the same.

Translational properties of electron wave function become apparent only when wave vector \mathbf{k} has *real value*. In case of imaginary values of \mathbf{k} , the wave function of electron will characterize *localized* electron states at a point \mathbf{r} , and its square module is the probability to find electron only in this point. The consequence of translational symmetry structure with long-range order placement of atoms is the presence of *physically equivalent points* in the interatomic space. Translational properties of wave function of electron in the periodic structure are realized only in case of real values of wave vector. The solutions of stationary Schrödinger equation for electrons in a crystal form the *modulated plane waves*—the Bloch waves [4].

In Chapter 1, the relationship of direct and reciprocal lattices was considered. It should be recalled that direct lattice is a *physical space*, in which the position of each atom is determined by the direct lattice vector \mathbf{a} . The inverse lattice is the *impulse space* with a dimension of the *inverse length*; geometric configuration that defines the reciprocal lattice is vector \mathbf{b} . For direct simple cubic lattice, the reciprocal lattice is also a simple cubic lattice; for face-centered cubic lattice, the reciprocal lattice is a space-centered lattice; and for direct cubic space-centered lattice, the reciprocal lattice is a face-centered lattice.

For a diamond-type structure, the unit cell of reciprocal lattice is the cut octahedron. Simultaneous translation of wave function in the direct space on the vector \mathbf{a} , and in the reciprocal space on vector $2\pi/\mathbf{a}$, does not break the translational properties of wave function [1]:

$$\psi\left(\bar{r} + \bar{a}, \bar{k} + \frac{2\pi}{a}\right) = \psi(\bar{r})e^{i\left(\bar{k} + \frac{2\pi}{a}\right)\bar{a}} = \psi(\bar{r})e^{i\bar{k}\bar{a}}.$$

Thus wave function shows double periodicity: the period a in direct lattice and the period $2\pi/a$ in reciprocal lattice. The basic period of wave vector \mathbf{k} in the inverse lattice is: $0 \leq k \leq 2\pi/a$. However, for convenience, this period is made *symmetric*: $-\pi/a \leq k \leq +\pi/a$. All possible values of wave vectors can be found within this basic period; for this reason, it is called the *coerced zone* of wave vectors. Symbols “−” and “+” indicate that there may be two identical electronic waves, which move opposite to each other.

If one chooses in spatially infinite periodic structure the one-dimensional (1D) (linear) chain of identical atoms with length L , the distance between atoms coincides with the period of potential field. The minimal possible length of electronic wave is $2a$, while the maximal length of wave is $2L$. Thus, the length of electronic wave has such a range of values:

$$2a \leq \lambda \leq 2L.$$

The most important fact in this formula is that the length of electrical waves is the *discrete* being divisible by $2a$. Wave discontinuity means that the wave vector $k = 2\pi/\lambda$ and the impulse $p = h/\lambda$ have discrete values. The discrete spectrum shows the number of allowed values of k within a coerced zone. For a chain of atoms $L = Na$ (where N is number of atoms in a chain), the linear dimension of coerced band in the reciprocal lattice is $2\pi/a$.

For three-dimensional (3D) space $k = 2\pi m_i/a_i N_i$, where $i = 1, 2, 3$ (that implies number in coerced band), the number of atoms in a bulk is $L^3 = L_1 L_2 L_3$, where $N = N_1 N_2 N_3$ is the number of atoms in the volume L^3 . If $L^3 = 1$, then N is concentration of atoms. Thus 3D impulse space in the coerced band takes a certain volume, but unlike real physical space it has a dimension of *inverse length* and, therefore, it is called the *unit cell of inverse lattice*. Elementary cells tightly fill the entire space and form the inverse lattice. The number of wave vector within the band is defined by the number of atoms and their valence, that is, each valence state of each atom contributes one permitted value of a wave vector. The equal value of k can have only two electrons with opposite spins.

The peculiarity of quantum mechanical description of physical systems is that each steady state of system meets the complete set of physical quantities that have in this state certain values, but proper functions of corresponding operators coincide with the wave functions of a system.

The discussed items correspond to properties of electron in a crystal. Another important characteristic of electron is a value that is similar to impulse. As electron is a quasiparticle, its impulse characterizes electron interaction with its crystalline environment, so the state of electron in a crystal cannot be described by the “normal” impulse.

Quasi-impulse. In the ideal crystalline structure, without action of external fields, the full energy of electrons is independent of time. Full energy and other physical

quantities that are stored at electron movement are called *integrals of motion*. Potential field of ideal structure has translational symmetry, but it has the gradient character. In turn, gradient of potential corresponds to a certain internal periodic force, acting on the electron and changing its impulse. This means that in an ideal periodic structure the force field exists, but energy and impulse are independent of time [5]. As a consequence, electronic waves in the periodic ideal structure are stationary non-damped. The impulse p , which is associated with the periodic potential, is the *quasi-impulse*.

If crystal is subjected to the external field, the quasi-impulse of electron is changed under this field influence. The point is that the external field excites a system of electrons from their thermodynamic equilibrium by the changing of quasi-impulse vector modules and ordering their destinations. The quasi-impulse should also vary at any violations of potential field. Structural defects cause the *internal force field*. This means that in nonideal periodical structure electronic waves are scattered; this scattering often is caused by such defects as thermal vibrations of atoms (phonons) and by the neutral or charged impurity atoms.

Thus, one of the consequences of potential energy translational symmetry and translational properties of the electronic wave function is the *integral of movement*: invariable in a time quasi-impulse. Quasi-impulse invariability in time in the perfect periodic structure means that electronic waves are stationary, and scattering does not occur. Defects of structure that violate its periodicity change the quasi-impulse and lead to electronic wave scattering. As mentioned earlier, the electron scattering in lattice is the physical reason for the *limited conductivity* in metals (see Section 5).

Energy spectrum of electrons. For crystal structure formation from N atoms, all valence energy states are N -fold degenerated, so a single value of electron's energy E_a meets $\psi_{a1}, \psi_{a2}, \dots, \psi_{aN}$ wave functions. The valence electrons in a crystal can be viewed as the perturbation of degenerated system, which wholly or partly removes its degeneracy. Under these conditions, valence atomic levels split into N separate levels, $E_1 \rightarrow \psi_1, E_2 \rightarrow \psi_2, \dots, E_N \rightarrow \psi_N$, each of which has its own wave function. The set of energy levels creates the *energy band*.

Splitting is easily seen for both “ s ” and “ p ” valence states, regardless of whether they are occupied by electrons or not (Fig. 8.3). Thus the s - and the p -bands are formed, and hybridization in these bands (sp^3 -hybrid) is also possible.

Energy bands creation in a system of interacting electrons can be explained by the necessity to keep them within a solid body, because electrons are negatively charged particles, and between them the electrostatic repulsion exists. The stability of valence bond system is provided not only by the attraction of electrons to positively charged nuclei, but also by magnetic gravity of electrons (that have opposite spins), which partly compensates electrons' electrostatic repulsion. Such compensation is possible only for the *pair of electrons* with same energy, while other electron pairs must have different energy (Pauli principle) [2].

Solutions of Schrödinger equation acquire formation of Bloch waves (see Fig. 8.2) for *real values* of wave vector k , and corresponding value of electrons energy form the allowed energy bands: for bonded electrons, this is the *valence band*;

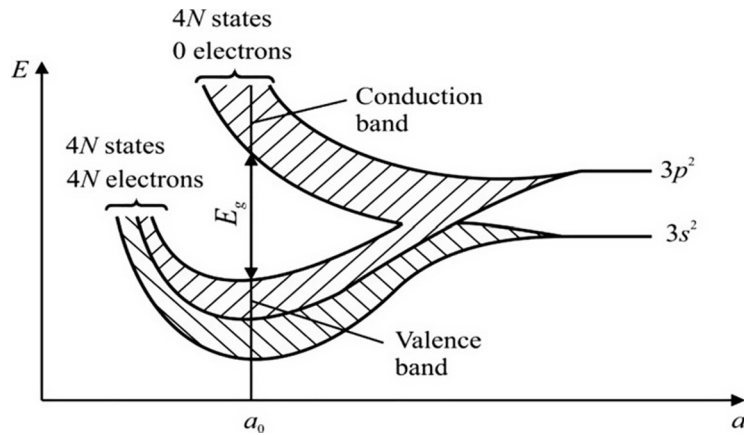


FIG. 8.3

Valence level splitting in silicon and energy band formation; compared with bands in metals: sodium, Fig. 4.23 and copper, Fig. 4.24.

for the first excited state of valence electrons this is the *conduction band*. The term “allowed energy levels” means that only for those ranges of energies the nonlocalized solutions of Schrödinger equation are possible within a crystal. It is worth recalling that in semiconductors and dielectrics between valence band and conduction band there is *energy gap* E_g (in other words, *prohibited zone* or *forbidden zone*), that is, the energy range, where wave solutions are impossible, because they correspond to the *imaginary values* of wave vector k . Bands for dielectrics, metals, and semiconductors are compared previously in Section 3, while the E_g value at temperatures 0 and 300 K for some semiconductors is shown in Table 8.1

From a physical point of view, the presence of energy band gap, located between the valence band and the conduction band (Figs. 8.2 and 8.3), means that on the boundaries of these bands the *total internal reflection* of electronic waves occurs; that is, electrons themselves cannot go beyond permitted bands. Band gap energy can be seen as a barrier between the levels, related to occupied electronic states and free electronic states, which correspond to the conduction band [2].

There are no usual electronic states in the band gap, but the *localized states* may exist in it. The wave functions of localized states should look similar to the radial functions of electron in a hydrogen atom: $\psi(\vec{r}) \sim r^2 e^{-r/\bar{r}}$. Square module of this function (probability to find electron with coordinates \vec{r}) has a maximum at this point, and

Table 8.1 Band Gap in Some Semiconductors

Crystal	InSb	Ge	Si	GaAs	PbS	PbTe
$E_g(0)$, eV	0.23	0.74	1.17	1.52	0.29	0.19
$E_g(300)$, eV	0.17	0.66	1.12	1.42	0.41	0.31

decreases with increasing distance from it. Usually, the energy of local states inside the band gap is created by the defects (impurity atoms, vacancies, dislocations, and so on).

The basic properties of electron energy (as a function of wave vector) in the permitted bands are [2]:

- unambiguity—electron cannot simultaneously be in different states;
- parity—energy should be independent of electronic wave direction and frequency;
- periodicity—all possible values of wave vector are contained in the area which looks like a unit cell of reciprocal lattice.

In the neighboring cells, wave vectors are repeated with the period $\pm\pi/a$. Exit beyond the unit cell of reciprocal lattice cannot give new energy values; that is, the adjacent cells repeat the same energy value. It should be noted also that such a feature of electron energy, as the *quasi-continuity*, wave vector in the crystal of limited size has discrete values. Accordingly, the *energy spectrum should be discrete*. The states number in band is $N_a \approx 5 \times 10^{22} \text{ cm}^{-3}$, where N_a is a typical concentration of atoms in semiconductors and metals.

The size of the energy gap in semiconductors, typically, is 0.5–2.5 eV; accordingly, the difference in energy between neighboring states is 10^{-22} – 10^{-21} eV. Such a very small energy difference of neighboring levels might be compared with an average energy of thermal motion: at a temperature of 300 K, it equals 2.6×10^{-2} eV. The number of energy states, experienced by electrons in their thermal motion, is very large: it is about 10^{20} . That is why the energy spectrum is *quasicontinuous*.

Filling of permitted bands. In metals, the *s*- and the *p*-valence bands overlap—there is no band gap. Taking into account the Pauli principle, the capacity of bands is doubled; therefore in metals the filling of permitted band is *partial*, and free electrons exist under any conditions (see Section 5).

In semiconductors and dielectrics in their ground state (at temperature $T \approx 0 \text{ K}$), there are no free electrons: the valence band is *completely filled* and the conduction band is *completely empty* (Fig. 8.3). If temperature $T > 0 \text{ K}$, then some free electrons in conduction band appear—they are born by the thermal ionization of some electrons located in the valence band [4].

However, thermal ionization processes in semiconductors can create no more than 10^{20} cm^{-3} carriers, which is in two orders of magnitude less than the capacity of allowed energy bands. This *slight filling* of bands leads to the concentration of free charge carriers exceptionally in the area of minimal energy (near the *bottom*) of the conduction band, or near the maximal energy (near the *ceiling*) of the valence band; that is, near the *extreme* values of energy. Therefore for practical application of semiconductors there is no need to examine the whole energy range of permitted bands, because most interesting are values of energy near the extremes.

Unlike the conduction band, to which free electrons are *added* due to thermal (or photo) ionization, the valence band, by contrast, *loses* electrons as a result of the

same process. In the places of incomplete valence bonds the *holes* appear; that is, nonoccupied by electron energy states. The availability of free energy states allows valence electrons to migrate *between noncompleted levels* of valence bonds. The spatial movement of electrons in the conduction band (as well as movement of holes in the valence band) is ordered by the external field that creates electrical current.

However, in the external electrical field, a peculiarity in hole behavior is observed: negatively charged electrons move *against* the direction of a field while holes move *in* the direction of a field that is like the *positively charged particles*. Just this characteristic behavior of holes that partially fill the valence band is the basis for introduction of a concept of *positively charged* quasiparticles—holes that have charge “+e.”

Effective mass. In connection with a specific filling of the bands (just *near extremes* of energy), energy dependence of wave vector $E(k)$ in the conduction band can be studied only in the vicinity of absolute minimum of energy (Fig. 8.4A). Under such conditions, energy dependence on wave vector has the form of a parabola:

$$E(k^2) \approx \frac{1}{2} \left(\frac{\partial^2 E}{\partial k_x^2} k_x^2 + \frac{\partial^2 E}{\partial k_y^2} k_y^2 + \frac{\partial^2 E}{\partial k_z^2} k_z^2 \right).$$

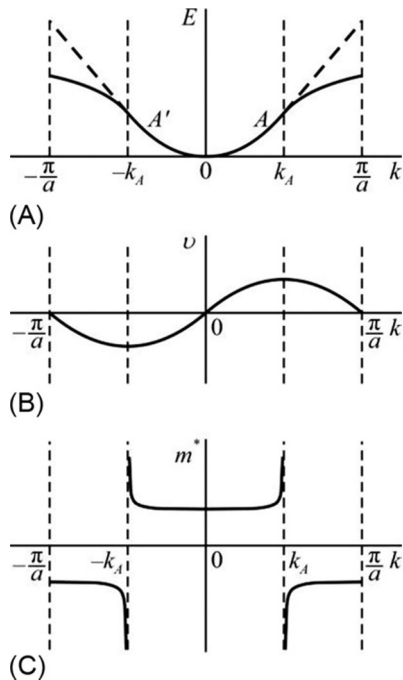


FIG. 8.4

Electron's parameter dependence on wave number; (A) energy, (B) velocity, (C) effective mass.

For free electron in a vacuum in the 1D case is $E(p) = p^2/2m_e$, while impulse is $p = \hbar k$; therefore: $E(k^2) = \hbar^2 k^2/2m_e$. Correspondingly, in the 3D space:

$$E(k^2) = \frac{1}{2} \left(\frac{\hbar^2}{m_e} k_x^2 + \frac{\hbar^2}{m_e} k_y^2 + \frac{\hbar^2}{m_e} k_z^2 \right).$$

If one compares these expressions for energy of electron (in crystal and in free space) using dimension, it is possible to get the following expression:

$$m_{ii}^* = \hbar^2 \left(\frac{\partial^2 E}{\partial k_i^2} \right)^{-1},$$

this parameter has dimension of mass and is the *effective mass*.

For free electron in a vacuum, which is defined by energy $E = \hbar^2 k^2/2m_e$, electron mass corresponds to its *rest mass*. The effective mass of charge carriers in the semiconductor is a specific parameter of charged quasiparticle: m_e^* that is different from the rest mass. It is not a mass in its usual sense, because it does not specify either gravitational or inertial properties of electron. The magnitude m_e^* may be as bigger so less than mass of free electron m_e ; moreover, the sign m_e^* can have the positive as well as the negative value.

In the first Brillouin zone in the absence of external field, the electron is situated in the bottom of conduction band. If crystal is subjected to the external field E under which electron accelerates, its kinetic energy increases, leading to electron transition to the higher energy level in the conduction band.

In case of small values of the wave number k when the curve $E(k)$ is parabola ($E(k) = \hbar^2 k^2/2m_e^*$, Fig. 8.4A), the velocity of electron $v = \hbar k/2m_e^*$ increases *linearly* with k increasing (Fig. 8.4B) while electron's effective mass $m_e^* = \hbar^2 (\partial^2 E / \partial k^2)^{-1}$ remains almost constant (Fig. 8.4C). As the wave vector increases (distance from zero k rises, Fig. 8.4A), dependence $E(k)$ already is not a parabola; correspondingly, the velocity of electron grows sharply near k_A (Fig. 8.4B). In the point A (point of inflexion), first derivative dE/dk shows maximum while second derivative $d^2 E / dk^2$ vanishes. It means that effective mass m_e^* increases (Fig. 8.4C) [2].

For wave vector values of $k > k_A$ the effective mass of electron changes its sign and *becomes negative*. The velocity of electron in case of $k > k_A$ reduced, although external force direction remains. At $k = \pi/a$ (the boundary of Brillouin zone) electron feels Bragg reflection (k jumps to the value $-\pi/a$). Further, the electron is accelerated in the opposite direction to external force, and its velocity varies from zero to the maximum, while effective mass m_e^* changes up to $-\infty$. At the point A' the sign of effective mass is changed to the positive, and electron is accelerated in the direction of external force.

In other words, in the middle of Brillouin zone ($k = \pm \pi/2a$) the sign of curvature of $E(k)$ dependence changes, and it passes through a zero, while corresponding mass value ($m_e^* \rightarrow \infty$) loses its physical meaning. Therefore, the approximation of effective mass is valid only for those charge carriers that are located *near the bottom of the conduction band* (or near the ceiling of the valence band).

In case of crystal anisotropy, the dynamic properties of electrons during their movement show anisotropy. Consequently, the effective mass is a tensor of the second rank. This anisotropy is most manifested in two directions; correspondingly, the surface of constant energy has a form of *ellipsoid of revolution*. In this case, the effective mass divides into effective mass in the longitudinal direction m_t^* and effective mass in the transverse direction m_t^* .

In the absence of anisotropy, the surface of constant energy looks like a sphere, and effective mass has scalar value. Table 8.2 represents the experimental values of electrons' effective mass in silicon and germanium.

Comment. Indication m_{dn}^* is effective mass for rotating ellipsoid axes. Physical content of m_{dn}^* is ellipsoidal isoenergetic surface, where minima of absolute energy are replaced by one equivalent energy states of sphere, whose radius is proportional m_{dn}^* . Within single energy minimum, the effective mass is anisotropic; however, because of symmetric spatial placement of energy minima, the average value of all masses of ellipsoid looks isotropic. For the *valence band* of silicon and germanium, the situation with effective mass of holes is slightly different, because in the subzones just as the *heavy* so also the *light* holes exist. In this case, the isoenergetic surface represents deformed areas, but it can be approximated by conventional fields [5].

Therefore:

- Effective mass is introduced in consideration, when there is a *parabolic* dispersion law; in the other case, mass might be dependent on energy, but in this regard the negative effective mass is possible. Effective mass is inversely proportional to the *second derivative* of expression $E(k^2)$, which describes the *curvature* of function at the points of extreme. The inverse curvature value is the radius of curvature. As a result, the effective mass is directly proportional to the radius of curvature of the $E(k^2)$.
- Curvature of function at the point of extreme might have signs: “–” at the maximum and “+” at the minimum. For holes, the minimum of energy exists in the ceiling of the valence band; therefore, the direction of energy axis in the valence band should be chosen in the opposite direction to its direction in the conduction band. Under these conditions, effective mass of holes is positive. Therefore, the hole is a quasiparticle with the charge “+e” and positive effective mass. Inasmuch as the valence band in the silicon originates as the mixture of $3s^2$ - and $3p^2$ -orbitals, the holes in silicon have different effective mass: the *light*

Table 8.2 Components of Effective Mass Tensor for Silicon and Germanium

Semiconductor	Mass of Electrons				Mass of Holes		
	M	m_t^*	m_ℓ^*	m_{dn}^*	m_{pB}^*	$m_{p\Pi}^*$	m_{dp}^*
Silicon	6	$0.19m_e$	$0.98m_e$	$1.08m_e$	$0.52m_e$	$0.16m_e$	$0.58m_e$
Germanium	4	$0.082m_e$	$1.59m_e$	$0.56m_e$	$0.34m_e$	$0.04m_e$	$0.35m_e$

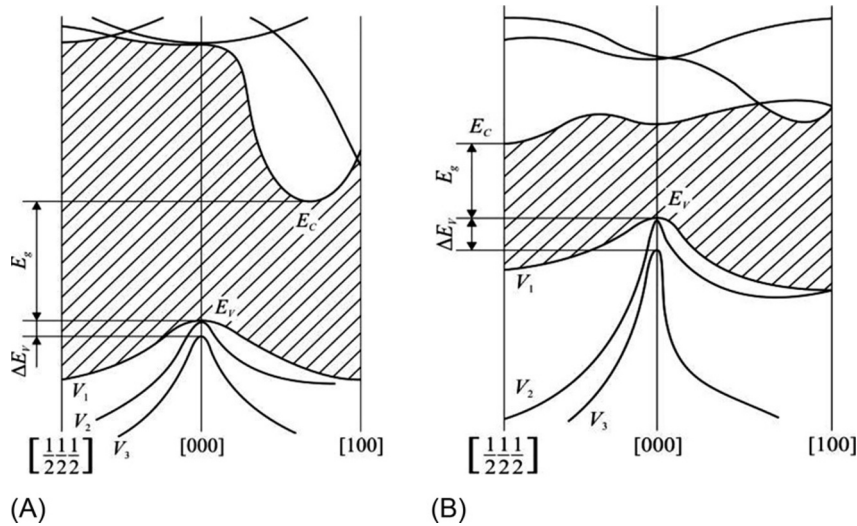


FIG. 8.5

Bandgap energy profiles: (A) silicon, (B) germanium.

holes, when the curvature of the $E(k^2)$ dependence is small, and the *heavy* holes, when this curvature is large (Fig. 8.5A).

- In order for effective mass to play the same role as the actual mass of electron, it should be not dependent on energy and on the wave vector that can be changed under the influence of external forces. The parabolic dependence satisfies this condition, as second derivative for such dependence is constant.
- Tensor nature of effective mass illustrates the fact that electron in a crystal lattice does not move as a usual particle with the rest mass but as a quasiparticle, whose mass depends on the direction of crystallographic axes of crystal.

In the approximation of effective mass, the knowledge of exact configuration of periodic potential in crystal is not necessary; it would suffice to replace the actual mass to the effective mass in the equations of motion of electrons and holes. If one fixes a certain value of energy, it will be possible to get the quadric equation that corresponds to the ellipsoid of general form:

$$\frac{k_x^2}{a_x^2} + \frac{k_y^2}{a_y^2} + \frac{k_z^2}{a_z^2} = 1,$$

where $a_i = \sqrt{\frac{2m_i^* E}{\hbar^2}}$ is the semiaxis of ellipsoid: $i = x, y, z$.

For high symmetry crystals (typical for semiconductors of a cubic structure), the absolute extreme is located in the center of conduction band; at that, the absolute extreme is only one, while isoenergetic spherical surfaces are spheres, and effective mass is scalar. However, absolute extreme can be found in the intermediate point on one of the main axes of symmetry of cube ($\langle 100 \rangle$, $\langle 111 \rangle$, $\langle 110 \rangle$), and then the

isoenergetic surfaces are ellipsoids, in which the rotation axis is the main axis of the cube. Therefore, effective mass has two components: m_t^* and m_ℓ^* (Table 8.2).

Based on these general positions, let us consider the band structure of mostly used semiconductors.

Conduction band of silicon and germanium. The energy profiles of band structure of silicon and germanium are shown in Fig. 8.5. Similar to silicon, in the germanium the minimum and maximum of energy correspond to different values of a wave vector. For this reason, these crystals are the *indirect bandgap semiconductors*.

The scanning of the bandgap energy profile is shown in Fig. 8.5 by shaded area in a reciprocal lattice space. The absolute minimum of energy in silicon is placed on the axis $\langle 100 \rangle$; the coordinates of minimums are $k_{\min} \approx 2/3(\pi/a)$, the number of extremes is six, correspondent effective masses are $m_t^* = 0.19m_e$ and $m_\ell^* = 0.98m_e$, and ellipsoids anisotropy is $m_\ell^*/m_t^* = 5.16$.

Band structure of $A^{III}B^V$ semiconductors. Most of these compounds have $k_{\min} = k_{\max}$ so they are the *direct bandgap semiconductors* (except GaP and AlP). Therefore, conduction band at $k_{\min} = 0$ is characterized by spherical isoenergetic surface.

As an example, Fig. 8.6 shows the band structure of GaAs. Scalar effective mass of electrons has wide margins: $0.01m_e < m_n^* < 2m_e$. In the conduction band, two minimums of energy can be seen: the absolute minimum (in the center of coerced band, with effective mass $m_n^* = 0.067m_e$) and the nonabsolute minimum (on coerced band

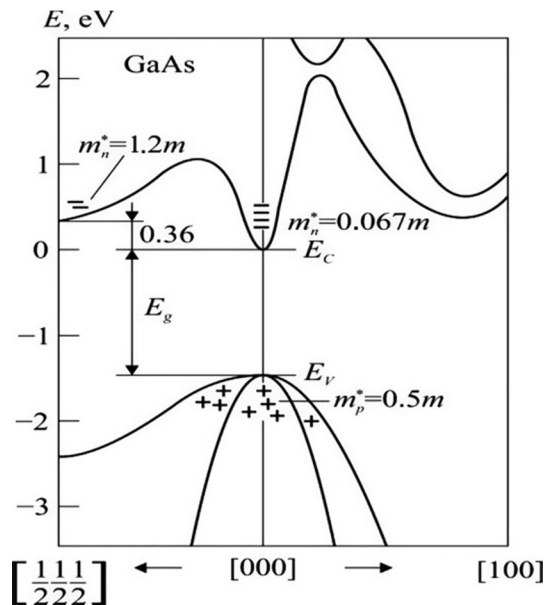


FIG. 8.6

Energy profile of gallium arsenide.

boundary in the direction $\langle 111 \rangle$; it is higher by 0.36 eV relatively to the absolute minimum. The isoenergetic surface for nonabsolute minimum is ellipsoid, the number of full ellipsoids is four, and the effective masses are $m_t^* = 0.075m_e$, $m_\ell^* = 1.9m_e$; $m_\ell^*/m_t^* = 25.3$.

The attention given to nonabsolute minimum is because in the strong electrical fields it is filled up by electrons with significantly increased effective mass. Under certain conditions, this fact can be used to generate microwave oscillations (Gunn effect).

All $A^{III}B^V$ compounds are characterized by a marked deviation of $E(k^2)$ dependence from parabolic law, which means the m_n^* dependence on electron's quasi-impulse that changes under the influence of an electrical field. The structure of valence band of $A^{III}B^V$ compounds, generally, is the same as in Si and Ge: energy maximum is located in the center of coerced band and two sub-bands can be seen—for heavy and for light holes; the isoenergetic surfaces are spherical, while the effective masses are scalar.

Therefore, the concept of effective mass greatly simplifies the description of electron behavior in a crystal. Effective mass is such a mass that must be attributed to the electron in its motion in a crystal under the influence of external forces that have the form of Newton's second law. Although effective mass has the dimension of normal mass, in general, it is characterized by the components of second-rank tensor. The concept of effective mass is very useful in solid-state physics, particularly, in the physics of semiconductors.

8.3 INTRINSIC AND EXTRINSIC SEMICONDUCTORS

The intrinsic semiconductors practically have no electrically active impurities, so the concentration of free charge carriers in them is determined only by temperature and inherent to this semiconductor energy of valence electrons, that is, by the band-gap E_g value. Intrinsic semiconductors are also known as *pure* (or *undoped*) semiconductors, that is, as perfect semiconductor crystals, which are free from defects and impurities of other elements [7].

Covalent bond is characterized by a pair of electrons belonging to two neighboring atoms. The number of covalent bonds that can create an atom is limited to the number of electrons missing in the outer shell of the atom to create the stable completed configuration of nearest inert gas. For example, monoatomic semiconductors of Group IV of the periodic table of elements (Ge and Si) create crystal lattice of diamond type with tetrahedral covalent bonds, that is, each atom highlights four electrons for covalent bonds. Fig. 8.7 shows a model of Group IV semiconductor lattice, when all bonds are filled; then any free electrons in a crystal are absent; accordingly, the conductivity of such a crystal should be zero in its ground state ($T = 0$).

Ideally, when thermal motion is absent, free electrons in the lattice cannot exist, because all valence electrons are involved in the bonds. In a real case at $T > 0$, the fluctuations of atomic thermal motion can *break covalent bonds* in some places of a

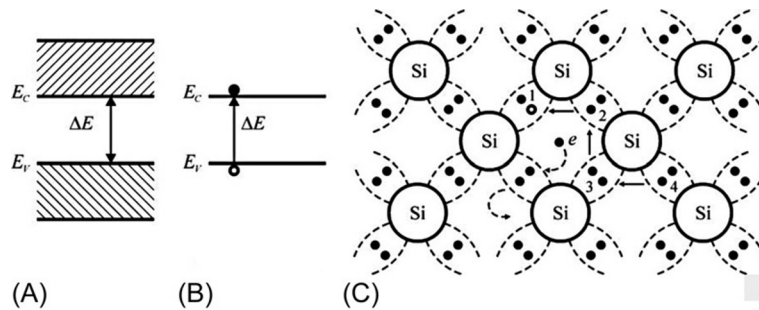


FIG. 8.7

Band diagram (A, B) and flat model (C) of intrinsic crystal lattice of semiconductor.

crystal and release the electrons, which are now eligible to participate in conductivity. Hence, in order for valence electron to become the conduction electron with charge “ $-e$ ”, it must get some *activation energy*, equal or greater than E_g (sufficient for covalent bond destruction). In turn, the nonfilled bond can be seen as the charge carrier with positive charge “ $+e$ ”, and it also can move freely through a crystal by a transition through valence electron bonds between neighboring (filled) bonds.

The energy of random thermal motion that destroys electronic bonds between atoms equals $k_B T$, where $k_B = 1.38 \times 10^{-23}$ J/K is Boltzmann constant. As stated earlier, at room temperature of $T = 300$ K the energy of thermal motion equals $k_B T \approx 0.026$ eV. Note that this value of $k_B T$ characterizes only the *average* thermal vibrational energy of the atoms [7].

Due to a random (chaotic) thermal motion in a crystal, at any given time there are some atoms, whose energy becomes much more than the average value, including those whose energy is greater than the value E_g . A number of these atoms are relatively small, but they exist, and some electronic links between atoms are ragged by *thermal ionization*. A probability of such an event in the statistical physics is defined by the exponential dependence: $\exp(-E_g/2k_B T)$, where E_g is the energy of valence bonds destruction.

It is easy to calculate the probability of the ionization at temperature of 300 K equals 3×10^{-2} for InSb crystal ($E_g = 0.18$ eV), equals 8.5×10^{-10} for Si crystal ($E_g = 1.1$ eV), and it is only 1.4×10^{-12} for GaAs crystal ($E_g = 1.4$ eV). Therefore, the increase in E_g of about eight times (while comparing indium antimonide with gallium arsenide) reduces the probability of thermal ionization of valence bonds by 2×10^{10} times. Note that a significant change in the probability of thermal ionization of valence bonds will be seen in the semiconductor when temperature changes. For example, by cooling Si to the temperature of “dry ice” (solid CO_2 that is evaporated at 216 K = -57°C) results in thermal ionization probability decrease 4000 times.

The valence bonds can be broken in semiconductor not only by thermal ionization but also due to absorption of light, if the energy of light photon is $h\nu \geq E_g$ (this

type of valence bond destruction is *photo-ionization*). The energy E_c in Fig. 8.7B is the lowest energy level of conduction band, that is, the minimal energy is correspondent to the bottom of the conduction band, while E_v is the upper energy level of the valence band (at the ceiling of the valence band); difference $E_g = E_c - E_v$ is the band-gap. The arrows on the band charts show possible electronic transitions between bands.

The valence bond violation has the following main consequences [7]:

- to deliver electrons into the conduction band, where they are able to change their energy and change spatial position under the influence of external fields; in other words, to create the opportunity for electric current;
- to release some energy states in the valence band that also gives the opportunity for holes to change their energy and spatial position, therefore, also to create electrical current.

The energy states vacating the valence band should be considered in more detail. From the covalent binding orbit *one electron is removed, but another remains*. This incomplete bond that is allowed but not occupied the electronic state is the *hole in the energy spectrum*. In the absence of external electrical field, similar to the hole, the electron also moves within a crystal randomly. This occurs because the neighboring electron of covalent bond, using energy of thermal vibrations of lattice, can fill the incomplete covalent bond; as a result, the hole from a position 1 moves to the position 2 (Fig. 8.8C). For the same reason, the hole can go further. Thus the movement of holes, actually, is carried by the electrons moving between the covalent bonds.

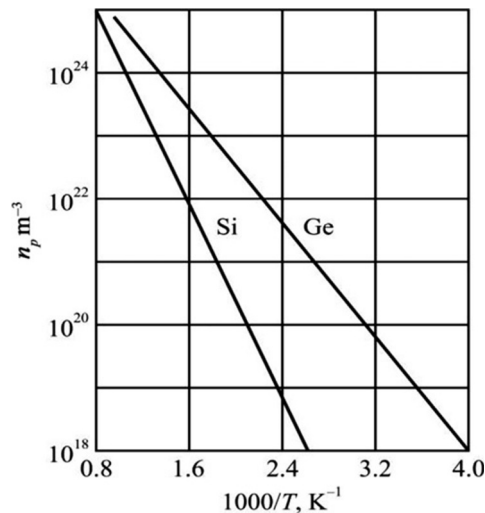


FIG. 8.8

Temperature dependence of intrinsic concentration in silicon and germanium.

In other words, the absence of negative charge of electron in the atomic orbit *disturbs the neutrality* of covalent interatomic space: it becomes *positively charged*. There is no difference, in principle, in the movement of which particle—electron (as material carrier with negative charge) or hole (partially filled valence band, which is positively charged carrier)—gives rise to conductivity.

In support of this approach, it is possible to give also the following argument: the electrical field creates a flow of electrons that are directed against the field, but the electrons leave behind them the holes that move in compliance with field, so they look like positive charges. In any case, the hole is an empty place; therefore regardless of properties attributed to hole, it cannot be the *material* positive charge carrier. However, in nature, another *material* positive charge carrier exists—a *positron*; hence, to avoid confusion, the hole is the *quasiparticle with positive charge*, which can exist only as a partially filled valence band.

Therefore eventually torn covalent bonds form a pair of free charge carriers—electron in the conduction band, and hole in the valence band, which also contributes to electrical conductivity of intrinsic semiconductor. The concentration of charge carriers is denoted as follows: $n_i = n = p$ (“i” index derived from the English “intrinsic” = own). The concentration in some semiconductors at temperature of 300 K is presented in Table 8.3 [2]. The concentration is proportional to the probability of thermal ionization of electrons from the valence bonds: $n_i \sim \exp(-E_g/2k_B T)$. The analysis of this exponential function shows that it varies monotonically with temperature; correspondingly the n_i varies, by the exponential law throughout the temperature range.

The intrinsic conductivity of semiconductors is the sum of their electronic and hole conductivity components: $\sigma_i = en_i(u_n + u_p)$, where u_n and u_p are the *mobility* of electrons and holes, respectively. Usually $u_n > u_p$, because holes are “heavier” than electrons. The resistivity $\rho_i = 1/\sigma_i$ for some semiconductors at temperature of 300 K is also given in Table 8.3. This table shows how radically E_g affects intrinsic (own) charge concentration and resistivity. As stated earlier, the movement of negatively charged electrons along empty vacancies against field direction is equivalent to the movement of positively charged unfilled bonds in the opposite direction.

Total intrinsic conductivity consists of electron and hole components. Obviously, in the pure semiconductor the concentration of electrons (n) and holes (p) is equal, as

Table 8.3 Properties of an Intrinsic Basic Semiconductor at Temperature of 300 K

Semiconductor	Ge	Si	GaAs
E_g , eV	0.66	1.12	1.42
n_i , cm^{-3}	2.4×10^{13}	1.45×10^{10}	1.8×10^6
$\rho_i = 1/\sigma_i$, Ohm cm	47	2.3×10^5	10^8

these free carriers are *created in pairs* due to covalent bond breaking. However, the mobility of electrons and holes is different, because mechanisms of their motion in the electrical field differ significantly.

Total conductivity σ in crystal is

$$\sigma = \sigma_n + \sigma_p = e_n u_n n + e_p u_p p,$$

where indices n and p refer to electronic and hole components, respectively, while $n_i = n = p$ is the concentration of electron-hole pairs.

Thus the semiconductors in which conductivity is because of a break of their own covalent bonds in a lattice are called as intrinsic, and carrier concentration n_i is the *intrinsic concentration*. This concentration increases with increasing temperature; the less the semiconductor activation energy E_g (which determines energy gap), the more the intrinsic charge carrier concentration at given temperature.

Generation and recombination. When it comes to formation of intrinsic carriers, the terms *thermogenesis* and *photogenesis* are used. In this regard, it should be noted that thermogenesis is continuous in time process, and the question is: Why charge carrier concentration is fixed at a certain temperature? The answer to this question can be obtained if one considers relating an opposite process—*recombination*. In a physical sense, recombination is a reconstruction of valence bonds due to electrons fastening in a lattice. To restore valence bonds, at least three conditions should be taken into account: spatial electron must “*meet*” the hole, it should *lose energy*, which is more than E_g , while the *spin states* of free electron and electron in the incomplete valence bonds must correspond to the Pauli principle; all these events have to occur simultaneously [2].

Just the simultaneous fulfillment of these conditions requires a lot of effort and delaying the recombination process in time. The point is that electrons of conduction band and holes of valence band move independently with the average thermal velocity $vT \approx 5 \times 10^5$ m/s (at $T = 300$ K). This velocity is important, because the “meeting” of electrons and holes should be realized at regular intervals of time. The loss of energy occurs by its transfer from charge carriers to thermal vibrations of atomic lattice; therefore at the time of meeting, atoms should take an energy $\geq E_g$, but this is not always possible. Finally, electronic spins all the time “switch” at electron collision with each other and with the oscillation of atoms; so the implementation of the third condition also requires some time.

When temperature increases, the rate of generation increases: the number of electron-hole pairs, generated in the unit volume per one second, rises, wherein more number of vacancies in the valence band appear, that is, the rate of recombination also increases: the number of electron-hole pairs, which recombine in the unit volume per one second, rises. If the temperature is fixed, a balance between the rate of generation and the rate of recombination will be established, as a result of which the concentration of free charge carriers will be independent of time. Due to the fact that rates of generation and recombination increase synchronously, valence bond violation during charge carrier thermogenesis does not cause a destruction of the semiconductor as a solid.

Concentration of electrons and holes in intrinsic semiconductor. At steady state, the dynamic equilibrium exists: the number of pairs of free “electrons-holes” released every second in the semiconductors by thermal generation is equal to the number that “die” (consolidate in a crystal) by recombination.

Therefore, a certain *temperature equilibrium* in the concentration of electrons and holes in semiconductor is established. For calculation it should be determined how many free carriers are generated in unit volume per unit time, and how many of them are neutralized by the recombination. The number of electron-hole pairs (K_1) generated every second in unit volume of semiconductor equals

$$K_1 = \alpha \exp(-\Delta E/k_B T),$$

where α is proportionality coefficient, different for various semiconductors. The number of charge carriers that recombine every moment per unit volume (K_2) is

$$K_2 = \beta n_i p_i = \beta n_i^2 = \beta p_i^2,$$

where n_i and p_i are concentrations of electrons and holes, respectively, in the unit volume of semiconductor. These concentrations are proportional to the probability of occurrence that the number of free electrons and holes is equal, because charge carriers are always generated and recombined by pairs.

Temperature dependence of carrier concentration in germanium and silicon is shown in Fig. 8.8.

The equilibrium concentration of electrons and holes in the intrinsic semiconductor is determined by the numbers K_1 and K_2 equating:

$$\alpha \exp(-\Delta E/k_B T) = \beta n_i^2 = \beta p_i^2..$$

From this equation, it is possible to find

$$n_i = p_i = (\alpha/\beta)^{1/2} \exp(-\Delta E/k_B T) = (A/B)^{1/2} \exp(-\Delta E/k_B T).$$

In all semiconductors, the values of constants A and B are well known; they are measured in cm^{-3} . At room temperature ($\sim 300\text{ K}$), these values are within approximately 10^{17} – 10^{19} cm^{-3} .

Extrinsic (doped) semiconductors. The special characteristics of semiconductors can be obtained only in high-purity crystals. In *usual engineering*, a crystal that has one foreign atom per 1000 own atoms (i.e., concentration of impurities equals 0.1%) can be considered a pure substance. From a *chemical point of view*, a substance can be considered quite clear if it has one foreign atom per 10^5 own atoms. Some additives to semiconductor can relatively easily give free electrons or holes. Therefore, from a chemical point of view, a very clean germanium (with impurity concentration 0.001%) at room temperature will contain impurities, which can create 4000 times more electrons compared with intrinsic germanium. Similarly, in silicon, 0.001% of impurity concentration increases charge carriers 10^7 times, while in the gallium phosphide it is 10^{17} times.

Thus the *purity of semiconductors should be many times higher than the usual standard with “chemically pure” material.*

It looks impossible to eliminate impurities completely; however, well-developed technology enables to obtain semiconductor crystals of *very high purity*. Nevertheless, even in the ultrapure silicon, in which impurity concentration is no more than 10^{12} cm^{-3} , the amount of charge carriers induced by impurities at room temperature has been detected to be about 100 times greater than intrinsic concentration of electrons and holes.

The *extrinsic* is a semiconductor that is doped, that is, in which the doping agent has been introduced, giving its different electrical properties from intrinsic (pure) semiconductor. One of most characteristic features of semiconductors is their very high sensitivity to small amounts of impurities. Addition of foreign substances in the quantities of one atom per million (or even per billion of own semiconductor atoms) can markedly change its properties. Doping results in such concentration of free charge carriers, which determines properties of a semiconductor: the *concentration* and *type* of charge carriers.

Intrinsic semiconductors have the *same* number of electrons and holes; therefore, there is *no dominant type of conduction*—electrons or holes. However, to create microelectronic devices (diodes, transistors, integrated circuits, etc.) only semiconductors with the *adjusted type of conductivity* should be used. For this purpose, in the intrinsic semiconductor material the additives are embedded by *replacing* basic atoms.

Using doping technology, following important problems are solved:

- receiving certain type of charge carriers—electrons or holes;
- prescribed concentration of charge carriers;
- adjustment of charge carriers relaxation time;
- guaranteeing necessary temperature range for temperature-independent charge carrier concentration.

To solve the first problem, one must have a source of electrons or holes, which do not depend on intrinsic carrier concentration. With this aim, in Group IV semiconductors such additional atoms should be introduced, whose *valence is different per one*: these are atoms of Group III or V elements. Group IV semiconductors are characterized by covalent type of bonds, so four orbits with two electrons in each can be achieved. Accordingly, into the semiconductor compounds of $A^{III}B^V$ the atoms of elements of Group II or VI should be introduced.

Acceptors. In the atoms whose valence on one is *smaller* than Si or Ge valence, one electron is missing in a full bond, so main atoms of a crystal must be replaced by elements turning to the *negatively charged ions* (Fig. 8.9B). This replacement requires some energy costs that are realized by thermal ionization (Fig. 8.9A). Thus the total number of valence electrons decreases, while in the energy spectrum of valence band the unfilled energy states appear. Therefore, the introduced holes appear whose number equals to quantity of imbedded three-valence atoms. Such impurities of replacement are called *acceptors*. Most acceptors are boron (to Si), gallium, and indium (to Ge), while for GaAs the acceptor is zinc [7].

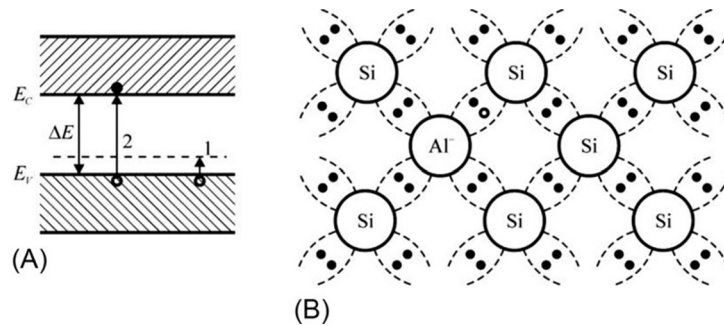


FIG. 8.9

Band diagram (A) and simplified model (B) of p -type semiconductor (silicon doped with aluminum).

In Fig. 8.9A, by the energy level 1 the state of acceptors is marked. During ionization, the atom of acceptor takes the electron from the valence band; due to this, in the valence band hole appears. Because activation energy of acceptors is much smaller than the energy gap $\Delta E = E_c - E_v$, the conductivity activated by doping appears at much lower temperatures than intrinsic conductivity (this ionization is shown by number 2 in the figure). In the doped semiconductor, if concentration of acceptors is sufficient, in a certain temperature range the intrinsic conductivity is much less than the extrinsic one ($\sigma_{\text{intrinsic}} \ll \sigma_{\text{doped}}$), because intrinsic ionization requires much higher energy than ionization of impurities.

Donors are those embedded in semiconductor impurity atoms, in which the valence is bigger per one than a valence of basic atom; therefore, the excess electrons appear, which are not involved in the formation of valence bonds. As these atoms in a structure of semiconductor are surrounded by four connecting orbits, the attraction of excess electrons to atoms is significantly weakened, which facilitates their release into the conduction band by thermal ionization (Fig. 8.10A). As a result, free

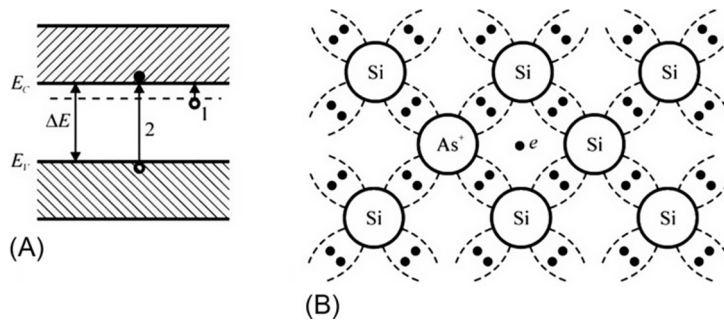


FIG. 8.10

Band diagram (A) and simplified model (B) of n -type semiconductor crystal (silicon doped with arsenic).

electrons and positively charged ions are generated (Fig. 8.10B); their quantity as many as doping atoms were imbedded. These impurities are the *donors*. Most common donors for Si are phosphorous, antimony, and arsenic; for Ge, they are antimony and arsenic; and for GaAs, tellurium.

Extrinsic conductivity is *unipolar*, that is, charge carriers of only one sign are created, unlike the bipolar intrinsic conductivity, which involves carriers of both signs. Another important difference between extrinsic conductivity and the intrinsic one is the significantly lower value of activation energy of donors and acceptors [7].

The second problem, which is solved by doping, is the receiving of desired concentration of charge carriers. At first glance, this problem seems quite simple: the quantity of free electrons (or holes) will be the same as imbedded donors (or acceptors). However, this simplicity exists only at first glance: the *lower limit* of doping depends on the intrinsic carrier concentration, while the *higher limit* of doping depends on the possibility of dissolution of impurities in a crystal.

Any semiconductor has its own concentration of electrons and holes. If the semiconductor is not doped, intrinsic concentration is minimal at given temperature: $n_{\min} = p_{\min} = n_i$. To have an electron (or the hole) type of conductivity, the concentration of impurity (that gives additional electrons or holes) should be at least 10 times greater than n_i .

As for the maximal concentration, there is also a certain limit. Impurity atoms distort the structure and the potential field of doped semiconductors at any concentration of impurities. Experience shows that semiconductor material still retains its basic properties, if no more than *one* of impurity atom from a *100* of basic atoms is replaced. Atomic density used in microelectronic semiconductors is about 5×10^{22} atoms/cm³, the maximum permissible concentration of impurities is about 10^{20} cm⁻³, and the same is maximal concentration of charge carriers: $n_{\max} \approx 10^{20}$ cm⁻³. As a result, possible concentration of charge carrier should be within $n_i < n < 10^{20}$ cm⁻³.

Usually in semiconductors, there are charge carriers of both signs. If $n = p$, then, as already noted, the semiconductor is intrinsic. Impurities make a semiconductor extrinsic (doped). If the concentration of impurity atoms-donors is N_d and acceptor concentration is $N_a \sim N_d$ (i.e., identical or similar in order of magnitude), the semiconductor is called as totally or partially *compensated*. However, in the more common cases, $n \neq p$. At that, if the concentration of one type of charge carrier is larger than the other type, they are the *majority carriers*. Charge carriers whose concentration is less are known as *minority carriers*. The majority carriers in the *n*-type semiconductors are electrons, while in the *p*-type semiconductors they are holes.

Thus conductivity of semiconductors at absolute zero temperature should be zero as all valence electrons are bound in their atomic orbits, and cannot carry electrical current. However, excitation energy of electrons in semiconductors is not very big, and, therefore, at a relatively low temperature, due to lattice thermal motion, some electronic bonds become broken. As a result, free electrons (conduction electrons) and holes appear. Concentration of conduction electrons (n) and holes (p) exponentially increases with temperature. In Fig. 8.11 the temperature in the abscissa is given

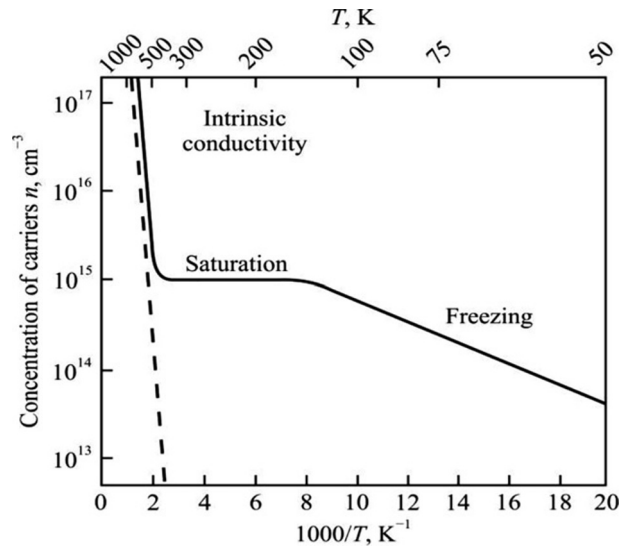


FIG. 8.11

Free charge carrier concentration temperature dependence studied in silicon doped by donors (about 10^{15} sm^{-3}).

in the *inverse coordinate* (T^{-1}), while in ordinate charge carrier concentration it is given in the logarithmic scale.

In the high-temperature range (higher than $\sim 500 \text{ K}$, top portion of the graph in Fig. 8.11), the *intrinsic* conductivity in semiconductor dominates and very rapid temperature change in concentration of charge carriers is observed. In the low-temperature range (below $\sim 200 \text{ K}$, in middle of graph in Fig. 8.11), the conductivity is caused by impurities, but it gradually reduces in the range of doped carrier “freezing.” In the temperature range commonly used in semiconductor devices, the charge carrier concentration is *almost constant*, and, therefore, semiconductor devices demonstrate a rather stable work in the saturation range (also shown in Fig. 8.11). Indeed, due to the doping, in the usual application temperature interval ($t = -60$ to $+80^\circ \text{C}$, i.e., $T \approx 200$ – 350 K , or $10^3/T \approx 5$ – 3), charge carrier concentration remains practically constant.

According to the charge carrier concentration, conductivity of semiconductors increases exponentially with temperature rise. Correspondingly, electrical resistance decreases (Fig. 8.12). It is evident, that a range of variation of resistivity of semiconductors is very large: it covers five orders of magnitude in size. Therefore, a rather simple idea can clearly explain temperature dependence of electrical conductivity of semiconductors by increasing concentration of free charge carriers, excited due to thermal ionization of atoms—as the basic atoms, so the atoms of impurities. However, by measurement of electrical conductivity of semiconductor it is impossible to determine its type, that is, to conclude what exactly—electrons or holes—take part in

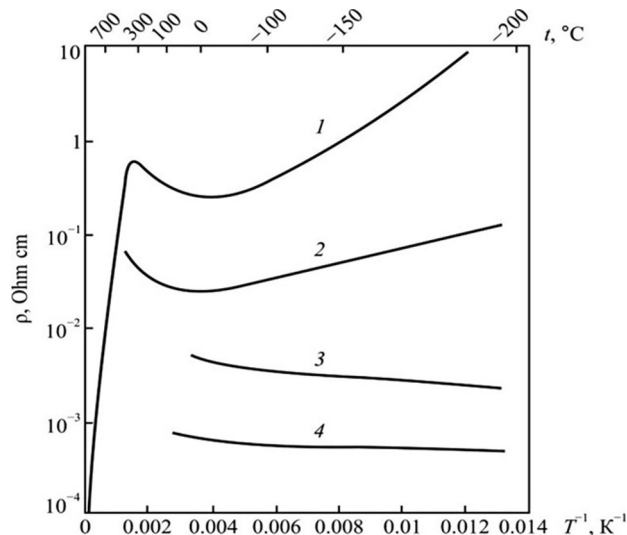


FIG. 8.12

Temperature dependence of silicon resistivity with different phosphorus content doses: 1— $4.8 \times 10^{17} \text{ cm}^{-3}$; 2— $2.8 \times 10^{18} \text{ cm}^{-3}$; 3— $4.8 \times 10^{19} \text{ cm}^{-3}$; and 4— $4.8 \times 10^{20} \text{ cm}^{-3}$.

the conductivity. Predominant conductivity mechanism (electrons or holes) can be verified experimentally by Hall's effect as well as by the sign of thermoelectromotive force (Seebeck effect, see Section 8.3).

In case of Hall's effect, in the magnetic field under the influence of Lorentz force, the electrons deviate from their natural direction of motion and move aside to one of the faces of studied sample; so a transverse electrical field occurs. This process takes place as long as Lorentz electrical field is compensated by charge carriers on the side edges. In the *n*-type semiconductors, the majority of carriers are electrons, and they create at this sample the facet, to which electrons are directed, the *negative* charge. In the *p*-type semiconductors, holes, for the same reason, create the *positive* charge. Experimentally measured *sign* and *magnitude* of potential difference can determine not only the sign of charge carriers, but also the value of their concentrations.

The sign of charge carriers is also possible to determine using the thermoelectromotive effect. This method of determining conductivity type is called the *hot probe method*. The sample is fixed on the metal plate (i.e., the "cold" contact) and then it is touched by the heated probe (i.e., the "hot" contact). In the closed electrical circuit with "cold" and "hot" contacts, the galvanometer is included. Direction of thermoelectric current (its sign) indicates what major carriers are in the sample. In the case of *n*-type sample, the negative potential is created on the "cold" contact [2].

Detailed explanation of this effect will be given in Section 8.4; in short, between "hot" and "cold" contacts of sample, the temperature gradient is created. If the majority of charge carriers are electrons, then they diffuse from the hot area of a sample

into the cold, as their thermal velocity will be higher. Therefore, the “cold” part of a sample becomes enriched by the majority carriers and the negative space charge appears in it. Accordingly, the “hot” area of the sample is depleted of electrons and has the positive charge. If the type of conductivity were to be changed, the sign of thermoelectromotive effect that can be indicated by the arrow of galvanometer will be reversed.

8.4 KINETIC PROCESSES IN SEMICONDUCTORS

Exceptionally labile electronic structure of semiconductors demonstrates a large amount of kinetic effects, many of which are used in the development of electronic devices. Some kinetic phenomena are shown in Fig. 8.13.

Electrical conductivity. With the approximation of effective mass, equations of electron movement in crystal formally coincide with Newton’s second law: acceleration is proportional to force $a_{cr} = -eE/m^*$ (the sign “-” denotes that electrons move against electrical field). This makes it possible to argue that in a crystal, under certain conditions, electrons have dynamic properties similar to electrons in vacuum, but with *another inert mass*.

Electron dynamic processes in semiconductors consist of not only in acceleration and kinetic energy, but also in loss of energy, received from the external field, which is dissipated at electron scattering. In the process of scattering, electrical energy is transmitted into lattice vibrations that are accompanied by heating (known as a Joule heat); so the temperature of semiconductor increases under electrical current flow.

When electron kinetic description is given, it is necessary to calculate drift velocity of electrons v_d , if time, during which field electrons are accelerated, is known. This time is the *free path time*, and usually it is denoted by letter τ . After this time, the electron completely loses its energy, and a new period of acceleration is started. To determine the *average drift velocity*, acceleration should be multiplied by time τ :

$$v_d = -(e\tau/m^*)E.$$

This expression shows that drift velocity is directly proportional to electrical field. The proportionality ratio, denoted by letter u , is the *mobility*:

$$v_d = -uE.$$

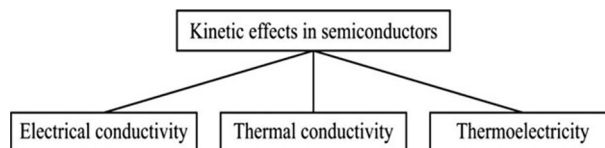


FIG. 8.13

Classification of main kinetic effects in semiconductors.

In its physical meaning, mobility is drift velocity in electrical field of 1 V/m (or 1 V/cm). For example, at temperature of 300 K mobility of electrons in the silicon is: $u_n = 1450 \text{ cm}^2/\text{V s}$.

Electrical current is directed movement of charged particles in electrical field. The *density* of electron drift flow is the product of electron density n on the velocity v_{dn} that determines the number of electrons crossing the unit area per unit time. Each electron has the negative charge “ $-e$.” Current density $j = env_{dn} = enuE$ is directly proportional to the electrical field (Ohm’s law). Proportionality factor $\sigma = enu$ is the *specific conductivity*. In its physical content, conductivity σ is current density in electrical field of 1 V/m (or 1 V/cm).

It is appropriate to remind that classification of solids is based only on their electrical conductivity and semiconductors can be determined as a matter, which occupy intermediate value of conductivity between dielectrics and metals; however, the most important fact is that conductivity of semiconductors increases with temperature rise.

Generation and recombination of charge carriers. At fixed temperature of semiconductors, a certain concentration of free charge carriers (n and p types) is formed as a result of thermal ionization of valence bonds and impurities; this concentration is the *equilibrium concentration* and it remains permanent. The independence of equilibrium concentration on time (at invariable temperature) is conditioned by the equality of rates of thermal generation and recombination.

Thermal ionization process is characterized by the speed G_T that has dimension $[\text{m}^{-3} \text{s}^{-1}]$: this is number of free charge carriers that are generated in unit volume per unit time. By the same dimension, the rate of *recombination* R_T is characterized. In a state of thermal equilibrium, the equal exchange by charge carriers takes place between the conduction band and the valence band; the same exchange occurs between bands and impurities levels. This exchange is accompanied by change of carrier energy through the absorption and birth of *thermal vibrational* energy quanta (phonons). During thermal generation, the electrons absorb phonons (total energy of phonons reduces), while during electron recombination, the birth of phonons occurs (total energy of phonons restores). A condition of thermodynamic equilibrium of interchange is $G_T = R_T$.

At temperature 300 K, in a state of thermal equilibrium, characteristic frequency of interchange conversions is about 10^{13} Hz, and the average time of a single charge carrier existence in their free state (depending on purity and quality of crystal) is $10^{-4} - 10^{-11}$ s. During this time, the electron, moving with a velocity of $\sim 2.5 \times 10^5$ m/s, “finds” the appropriate hole with correspondent spin direction (or find ionized donor or find nonionized acceptor), and localizes for some time. Following interactions in the lattice of semiconductor results in the birth of electrons (with phonons absorption), and everything starts from the beginning.

Free charge carriers in the conduction band (or holes in the valence band) can be obtained not only by thermally activated generation but also with other energy sources, such as the *light energy*, the energy of *electrical field*, the energy of *radiation*, and others. In these cases, *nonthermal* generation happens.

Therefore total concentration of electrons (or holes) consists of two parts: the thermal (equilibrium) carriers n and p and the excess (nonequilibrium) carriers Δn and Δp , conditioned by nonthermal energy sources. Excess charge carriers violate the energy balance and balance of concentration in the conduction band or in the valence band. At the time of phonon relaxation, the excess charge carriers acquire *energy equilibrium* within the band, but *equilibrium in concentration* remains disturbed.

The stationary equilibrium (independence on time of excess charge carriers concentration) can be set at condition, when the rate of nonthermal generation G and the rate of recombination of excess carriers R are identical: $G=R$. This condition, however, differs from thermodynamic equilibrium conditions: $G_T=R_T$. The fact is excess charge carriers occur not because of absorption of phonons, and, therefore, do not change total energy of phonons. The energy that they have to lose through the recombination often is beyond the phonon energy spectrum; therefore, this energy cannot be directly transformed into thermal vibrations of atoms.

Recombination is a process of restoring equilibrium concentration, broken by the *nonthermal* mechanisms of generation. Recombination comes down ultimately to restore the state of thermodynamic equilibrium of concentration. The energy of each individually taken electron, during quantum transitions, changes instantly, but the *number* of transitions per unit time (i.e., frequency) is limited, so the restoring of equilibrium concentration requires a certain amount of time.

The rate of recombination R is the number of electron-hole pairs, which recombine per unit time in unit volume. In these processes, the recombination rate is

$$R = -\frac{d\Delta n}{dt} = -\frac{\Delta n}{\tau} = -\frac{n - n_0}{\tau},$$

where Δn is concentration of nonequilibrium carriers and τ is relaxation time, during which the excess concentration of charge carriers in the process of recombination reduces by 2.7 = "e" times (Fig. 8.14).

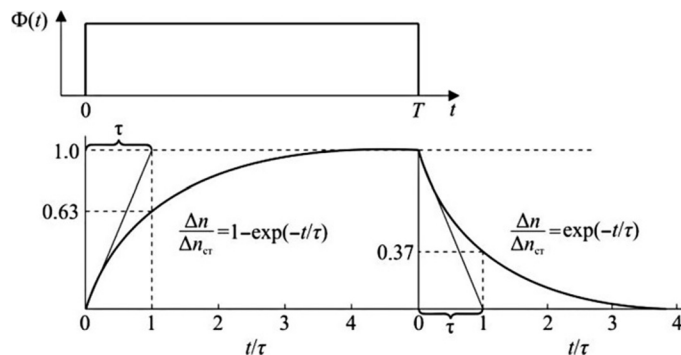


FIG. 8.14

Relaxation processes under rectangular light pulse.

Graphic representation of relaxation process in normalized coordinates is shown in Fig. 8.14 for conditions when nonequilibrium carriers are generated by rectangular light pulse $[F_{cy}](t)$ with duration T . At the time of light pulse, charge carrier concentration increases and gradually reaches saturation; after light switching off, it gradually falls to the equilibrium concentration.

Defined in this way (with exponential decrease), parameter τ is average time of recombination process called the *lifetime of excess charge carriers*.

The principle of operation of almost all electronic devices is based on the phenomenon of injection of nonequilibrium charge carriers that is realized by the impact of external forces on a crystal. For this reason, just the recombination rate determines the *operation speed* of devices. The higher the recombination rate, the higher the operation frequency of device. This speed is characterized by the *lifetime* of charge carrier τ , that is, characteristic time, after which excessive concentration of charge carriers is reduced by “ e ” times through their recombination. Thus the average time τ characterizes the existence of excessive concentration and it depends on temperature, as well as on mechanisms of *recombination* in the semiconductor.

During recombination, the energy, obtained at generation, returns to the lattice or environment. Mechanisms of recombination can be classified based the ways of energy returning, released during the act of charge carrier capturing by crystal during the recombination.

Recombination can be distinguished in two ways [8]:

- “band-to-band” recombination, at which excessive electrons are directly transferred from the conduction band to the valence band (this is characteristic of the *direct-gap* semiconductors);
- recombination through deep impurity levels, located in the bandgap of semiconductor (this is characteristic of the *indirect-band* semiconductors).

The most *likely mechanisms* of recombination in semiconductors are:

- phonon recombination (direct energy transmission to lattice vibrations);
- radiating recombination (when energy released in the form of quantum of electromagnetic radiation);
- impact recombination, when energy originally is delivered to nearby free electron (or hole), which then gives the excess energy to atomic lattice vibrations or to other charge carriers.

Comments. Recombination through the localized centers, sometimes, is called as Shockley-Read-Hall recombination (or trap-assisted recombination). In this case, the electron, during transition between bands, passes through the localized state, created in bandgap by deep-level traps. This trap coordinates electron and hole impulses. Such a process of recombination is dominant in silicon and in other *indirect bandgap* semiconductors. In the direct bandgap semiconductors, the *radiative* recombination can also occur that is accompanied by spontaneous emission of photons, whose wavelength corresponds to the released energy [5].

Charge carrier scattering. Drift mobility of charge carriers in electrical field depends on the free path length, and, thus, it is defined by the scattering processes while electrons move in a semiconductor. The scattering process lies in the distortion of charge carrier trajectory under the influence of forces acting on the electron or hole from the center of scattering.

If the center of electron scattering is the positive ion (e.g., ion of *donor*), the force of dissipation will be Coulomb potential. Similarly, the negative ion of *acceptor* strongly dissipates the holes. If center of scattering is the *neutral atom* of impurity, free electron collides with this atom and knocks another electron out of this atom. Then, the first electron remains in the atom, while the knocked electron, receiving the energy, moves by altered trajectory. As electrons are indistinguishable, the act of electron sharing can be considered as the act of electron's trajectory changing while scattering. A characteristic feature of scattering on neutral atoms is the *independence* of relaxation time on energy and temperature.

The process of electron scattering by thermal lattice vibrations is considered as collisions with phonons. As the number of phonons is determined by temperature, the charge carrier scattering depends on the temperature. Structural defects in crystal lattice also can serve as centers of electron scattering: just as dislocations so also vacancies; moreover, the electron-on-electron scattering is also possible. In the actual semiconductors, multiple scattering mechanisms can be found simultaneously, and the contributions of each of them can greatly vary with temperature and with concentration of impurities.

Electron and hole mobility. Free charge carriers always are in a state of chaotic motion that occurs with a high speed. Under normal conditions, the average speed of electron's chaotic motion is about 10^5 m/s. If the concentration of free carriers in semiconductor is distributed unevenly, a diffusion occurs: the flow of charge carriers from the area of higher concentration to the region of lower concentration.

Under electrical field influence on thermal chaotic traffic, the directed drift of charge carriers is superimposed. The average drift velocity in the weak electrical field is proportional to field strength (Ohm's law). However, in the strong electrical field, charge carrier drift velocity saturates because carrier velocity *cannot exceed* the velocity of thermal motion.

The ability to increase electron velocity by the increase in electrical field is described by the *drift mobility* of charge carriers. This mobility is very crucial for high-frequency devices as it defines the frequency limit of semiconductor applicability at very high frequencies.

Charge carrier mobility depends on temperature and on defect concentration (especially, on the *charged* impurities, such as donors and acceptors). Fig. 8.15 shows temperature dependence of mobility with the example of silicon. This temperature dependence is characterized by a maximum, which is located at relatively low temperatures. Two basic processes of charge carrier scattering affect mobility u : the scattering on ions u_{ion} (charged impurities) and the scattering on lattice vibrations—phonons u_{phon} :

$$1/u = 1/u_{ion} + 1/u_{phon}.$$

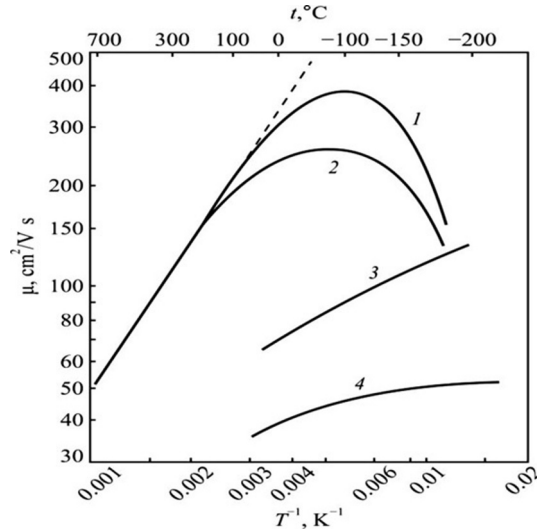


FIG. 8.15

Experimental dependence of drift mobility of electrons in silicon with different impurities, phosphorus dosed: 1— $4.7 \times 10^{17} \text{ cm}^{-3}$; 2— $2.7 \times 10^{18} \text{ cm}^{-3}$; 3— $4.7 \times 10^{19} \text{ cm}^{-3}$; 4— $4.7 \times 10^{20} \text{ cm}^{-3}$; dotted line indicates the static dependence $\mu(T) = 15 \times 10^5 T^{-3/2}$.

The *first component* influences mobility mainly at low temperatures, and its temperature dependence follows the law $\mu_{\text{ion}} \sim T^{+3/2}$. Indeed, according to experiments, in the temperature range of -200 to -50°C charge carrier mobility increases significantly (Fig. 8.15, curves 1 and 2) for crystals with low concentration of impurities (if there are too many impurities, this growth is not seen). Increasing mobility when temperatures decrease is explained by the increasing of charge carrier velocity. Indeed, in case of high velocity the time of charge carrier interaction with defects decreases, thus mobility increases.

The *second component*, namely, charge carrier scattering on phonons, conversely, rapidly decreases with increasing temperature: $\mu_{\text{phon}} \sim T^{-3/2}$. Thus, total dependence $\mu(T)$ is characterized by a maximum (Fig. 8.15).

The Fermi level. To determine the number of free charge carriers in a semiconductor, it is necessary to know the number of energy levels (states) in the conduction band actually occupied by electrons, and the number of free levels (states) in the valence band. Depending on temperature and energy, the probability of finding an electron in given energy level $w(T, E)$ is defined by the Fermi-Dirac energy distribution function

$$w(T, E) = \{1 + \exp[(E - E_F)/k_B T]\}^{-1},$$

where k_B is Boltzmann constant, T is absolute temperature, and E_F is Fermi level (see Section 4.4). From this formula, it can be seen that distribution function for the level $E = E_F$ at $T \neq 0$ equals $w = 1/2$. Thus, the Fermi level is such an energy level where the probability of its filling at given temperature (other than absolute zero) is equal to $1/2$.

In *metals* at temperature $T = 0$, the Fermi level divides conduction band in a half—on the filled part of a band, and on the empty part of a band, without any energy gap between these parts (see Section 4.4). In the *semiconductors* in the absence of external influence and $T = 0$, the valence band is completely filled, while the conduction band is free of electrons. It can be concluded that the Fermi level in the semiconductors is located *inside the energy gap* (in the forbidden zone).

Only at first glance, this conclusion contradicts the definition of Fermi level (as a level, the probability of filling at a temperature, other than zero, equals to $1/2$). In fact, Fermi-Dirac distribution function is valid only for the *allowed* energy states; therefore, this conclusion does not mean that electrons obviously have to be on the Fermi level.

Calculations show that in the *intrinsic* semiconductor the Fermi level is located almost in the *middle* of the band gap:

$$E_F = (E_c + E_v)/2 + 2/3 k_B T \ln(m_e^*/m_p^*),$$

where m_e^* and m_p^* are effective masses of electron and hole, respectively. The concept of “effective mass” makes it possible to describe the movement of “free” charge carriers in semiconductors like moving charged particles excluding periodic field crystal lattices. For electrons located near the bottom of conduction band, their acceleration in the free path is proportional to attached force.

It is worth recalling that effective mass is introduced as a proportionality coefficient between force and acceleration, similar to Newton’s second law. Mass of electron in the lattice might appear to be even less than the mass of electron in free space. During electron motion in a crystal in the absence of external field, its total energy remains constant. Through periodicity of field action in crystal lattice, only the ratio between potential and kinetic energy changes periodically. At that, the average velocity of the electron remains constant.

Under the influence of external electric field, movement of electron can be changed in such a way that much work of the external field is spent for potential energy increase; then, under the influence of the electrical field the velocity of electron might increase less if electrons have a mass equal to mass in a free space. Thus, during movement of electron (under influence of external field), the change of its kinetic energy surpasses the work of force (due to potential energy partial transition into kinetic energy), and electrons can have a velocity of particle lighter than electrons in free space.

It is already established that the Fermi level in the *intrinsic* semiconductor is allocated *near the middle* of the energy gap, and it depends only on the ratio of effective masses of electron and hole (Fig. 8.16A). In the *electronic*-type semiconductor, the Fermi level is placed *closer to the bottom* of the conduction band (Fig. 8.16B), while in the *hole*-type semiconductor it is closer to the *ceiling* of the valence band (Fig. 8.16C). However, with increasing temperature, intrinsic conductivity begins to prevail over the impurity conductivity, and the Fermi level moves to the middle of the energy gap (Fig. 8.16B and C).

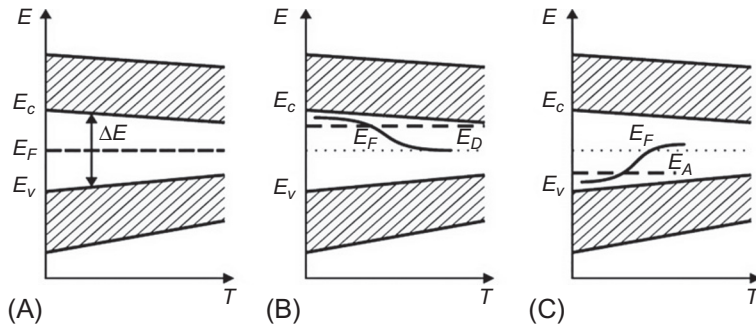


FIG. 8.16

Regulation and temperature dependence of Fermi level in a semiconductor: (A) intrinsic type, (B) electronic type, and (C) hole type.

Knowing Fermi level position (which is *electrochemical potential* for metals and semiconductors) is crucial for the development of various semiconductor devices and microelectronic circuits [1].

Semiconductors in strong electrical field. Electrical field affects both the mobility and the concentration of charge carriers. There are several explanatory mechanisms for charge carrier concentration increase in a strong electrical field: electrothermal ionization, electrofield ionization (tunnel effect), and ionization by the collisions.

Electrothermal ionization mechanism is seen mostly at low temperatures, when concentration of electrons in the conduction band is determined by the probability of donor level releasing (or flipping-over electron to acceptor levels). In case of electrical field influence on electrons located in the donor level, except Coulomb attraction to positive ion (donor), the additional force acts: $F = -qE$, which can help to break away the electron from a donor, and the electron becomes free. This increases the probability of electron transition from donor levels into the conduction band, which means an increase in charge carrier concentration and increased conductivity of a semiconductor.

At higher temperatures, when donors (or acceptors) are totally ionized, the main role in the increase in charge carrier concentration is played by the phenomena associated with ionization by collisions, as well as electrostatic (tunneling) ionization in the strong electrical fields.

Thermal conductivity (see Fig. 8.13) in semiconductors is due to the tendency of a system to take up the state closer to *thermodynamic equilibrium*, which is a result of temperature equalization. As it was shown previously in Section 3.8, if the temperature gradient exists in a crystal, the energy flow occurs, which is directed opposite to temperature gradient: $\Delta Q = -\lambda \text{ grad } T$.

Kinetic coefficient $\lambda = \Delta Q / \text{grad } T$ is the *thermal conductivity*. It is the amount of energy passing in unit time through unit cross section of sample being created by temperature gradient. In solids, thermal energy is usually transferred by the electrons

and by the elastic waves—phonons. In metals, electronic heat transfer dominates, while in dielectrics, phonon mechanism prevails.

The heat transfer in semiconductors may be implemented by many mechanisms [5]:

$$\lambda = \lambda_{\text{phon}} + \lambda_e + \lambda_{\text{bp}} + \lambda_{\text{phot}} + \lambda_{\text{exc}}$$

Designations:

λ_{phon} —phonon thermal conductivity, when heat transfer is caused by thermal vibrations of crystal lattice atoms or ions;

λ_e —electron (or hole) thermal conductivity, when heat transfer is due to free charge carriers;

λ_{bp} —bipolaron thermal conductivity, which is due to movement of electron-hole pairs;

λ_{phot} —photon thermal conductivity, which is due to heat transfer by electromagnetic radiation;

λ_{exc} —exciton thermal conductivity due to movement of excitons.

The mechanism of *phonon thermal conductivity* is universal for all solids; it was discussed in Section 3.6 in detail. Coefficient $\lambda_{\text{phon}} = (1/3)C \cdot v \cdot l = (1/3)C \cdot v^2 \cdot \tau$, where C is heat capacity, v is average velocity of photons, l is average free path of phonons, and τ is free run time. Temperature dependence of phonon thermal conductivity shows a maximum at temperature close to $0.1 \theta_D$, below this maximum the λ_{phon} fast decreases.

The *electronic thermal conductivity* mechanism dominates in metals, and it is associated with conductivity by Wiedemann-Franz law that was considered earlier in Section 3.8. The ratio $\lambda_e/\sigma = L \cdot T$ (where L is Lorentz number which is same for all metals) witnesses that thermal conductivity is directly proportional to absolute temperature T .

Bipolar, exciton, and photon thermal conductivity mechanisms are specific to semiconductors.

The *bipolar thermal conductivity* is due to *intrinsic* electrical conductivity of semiconductor crystals. The number of electrons and holes near the hot end of a semiconductor is larger than that near the cold end, and this causes the *diffusion* of electron-hole pairs from the hot end to the cold end of a sample. At that, on the hot end of a sample, energy is absorbed as it is required to form electron-hole pairs, while on the cold end of a sample, energy is released during recombination of electron-hole pairs. This energy consists not only of kinetic energy of electrons and holes, but also of energy required to move electrons from the valence band to the conduction band that is equal to the width of the band gap.

In most semiconductors $E_g \gg k_B T$; for this reason, the energy, which is transferred by electron-hole pairs, is much more than energy that is transferred by each of charge carriers in the event of impurity conductivity. Thus, in the case of intrinsic electrical conductivity, additional heat flow appears, hence the extra thermal

conductivity λ_{bp} occurs—due to bipolar diffusion. It is obvious that bipolar conductivity depends on the concentration of electron-hole pairs and on bandgap width.

Regarding bipolar thermal conductivity, the following expression is proposed, which is similar to electronic thermal conductivity:

$\lambda_{bp} = L_{bp}\sigma T$, where L_{bp} is similar to Lorentz number (see Section 2.8), but it is specific to the bipolar thermal conductivity [8].

In semiconductors of complex energy band structure, by the analogy with processes of charge carrier electro-transportation, similar contributions to heat transfer of various types of charge carriers should be considered, as well as in the case of interband scattering. For example, in the p -type semiconductors (GeTe, SnTe, PbTe, Cu_2Te), where in the valence band *two sub-bands* exist (for light holes and heavy holes, Fig. 8.17), the contribution to thermal conductivity from heavy holes increases with temperature rise. Thus, while temperature increases, *diffusion of charge carriers* from sub-bands makes an additional contribution to the bipolar thermal conductivity.

At the hot end of a semiconductor, the concentration of carriers in the sub-band of light holes is larger than that in the cold end. During carrier diffusion from sub-band of light holes into sub-band of heavy holes, the energy releases, which is equal to energy of spacing maxima of these sub-bands. Therefore, the additional thermal conductivity occurs by the diffusion of charge carriers.

Photon thermal conductivity is important in those semiconductors in which absorption coefficient in the thermal radiation range is small, and, therefore, photons have a big free path. Therefore, the contribution of heat transfer by electromagnetic radiation should be taken into account.

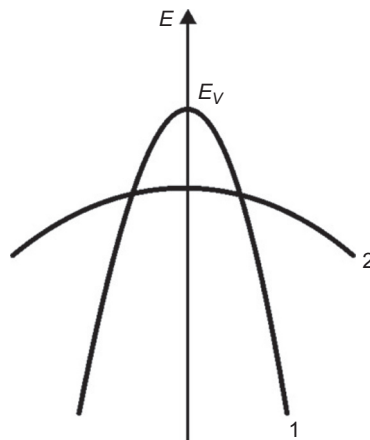


FIG. 8.17

Energy structure of a semiconductor with a complex valence band.

Thermal conductivity by excitons is caused by their diffusion in the case of temperature gradient existence. This process becomes important when semiconductor is inclined to the formation of excitons.

Thermoelectric effects (see Fig. 8.13) are a set of phenomena, by which temperature difference creates the electrical potential, or electrical potential creates the difference in temperature. In the case of heat flow, which is partially caused by the charge carrier movement, an electrical field appears. Since both average energy and concentration of charge carriers increase with temperature rise, the stream of free charge carriers arises, being directed by temperature gradient ∇T .

Thermoelectric effects are various phenomena in crystals associated with charge carrier transfer in a condition of temperature gradient. They include the Seebeck effect (thermoelectromotive force origination), the Peltier effect (electrostimulated change in temperature), and the Thomson effect (heat transfer by charges flow) (Fig. 8.18).

According to Seebeck effect, in a closed electrical circuit composed of heterogeneous conductors EMF (thermoelectricity) occurs, if contacts have different temperatures. In this case, a circuit consisting of two different conductors is the *thermocouple*. The value of thermoelectric power depends on the material of conductors and on temperatures of hot (T_1) and cold (T_2) contacts. In case of not very large temperature difference, thermoelectric power can be considered as being proportional to difference in temperature:

$$U = \alpha(T_2 - T_1),$$

where α is thermoelectric ability of contacted couples (Seebeck coefficient). The coefficient α is determined by conducting materials, and also it is dependent on temperature (in some cases, α even changes its sign with temperature).

In metals, in which speed of charge carriers is very weakly dependent on temperature (through electronic gas degeneration), thermoelectricity effect is rather small, but, nevertheless, sometimes it is used to *measure temperature*. In some semiconductors, the thermoelectromotive force reaches thousands of microvolts per degree, which makes such materials useful in thermal generators for the direct conversion of heat into electricity. Thermoelectricity is applied also to create very small and very precise temperature sensors (in particular, required in the computers).

Various mechanisms of thermoelectricity can be distinguished:

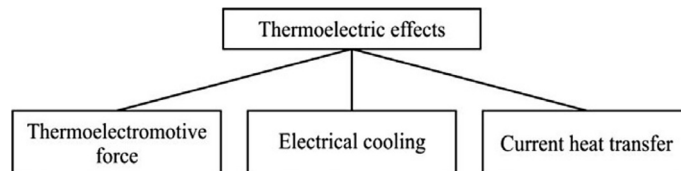


FIG. 8.18

Classification of thermoelectric effects in semiconductors.

- *Temperature dependence of electron average energy.* If there is a temperature gradient along a conductor, the electrons, located in the hot end, acquire higher power and *velocity* than the electrons located in the cold end; in addition to velocity, in semiconductors the *concentration* of conduction electrons increases with temperature. The result is a flow of electrons from the hot end to the cold end; as a result, on the cold end negative charge is accumulated, while on the hot end uncompensated positive charge remains. This process of charge accumulation is continued as long as the *potential difference* that appears does not allow electrons to flow in the opposite direction, thus establishing the balance. Thermal EMF that occurs according to this mechanism is bulk thermoelectric power (or volume EMF).
- *Temperature dependence of contact potential difference.* Contact potential difference is caused by the difference between Fermi energies of contacting conductors. When contact is created, the electrochemical potentials of electrons become similar; therefore, difference in contact potential that occurs equals to

$$U = (F_2 - F_1)/e,$$

where F is Fermi energy and “ e ” is charge of electron. As a result, on contact, being localized in thin near-contact layer, electrical potential U exists. When a closed electrical circuit from two contacting metals is created, this potential appears on both contacts. Electrical field will be directed in the same way in both contacts—from the larger F to the smaller one. This means making a roundabout way along a closed circuit, in one case, the bypass will be on the field, and another against the field. The circulation of vector E will be zero.

If temperature of one of contacts changes, then, as Fermi energy depends on temperature, potential U also changes. However, in this case the contact potential also changes; therefore electrical field will be changed in one of contacts, and vector E circulation becomes nonzero: in the closed circuit, the EMF appears, which is the *contact EMF*. If both thermocouple contacts have the same temperature, then contact EMF and bulk thermoelectric effect will disappear [8].

- *Phonon capture.* If in the solid a temperature gradient exists, the number of phonons moving from the hot end to the cold end will be higher than reverse moving. As a result of collisions with electrons, the phonons can capture electrons, so the cold end of a sample acquires negative charge (for hot end, positive charge), as long as the potential difference does not counterbalance the effect of capturing. This potential difference is the third component of thermoelectric power.

The effect of electron capturing by phonons seen in metals can be seen in semiconductors too. However, it is usually considered that in semiconductors charge carriers interact only with *long-wave phonons*. Therefore, the greater the effect of carrier capturing by phonons, the stronger the interaction of phonons with carriers and the greater relaxation time of long-wave phonons. Thus, at low temperatures, this capturing should occur stronger than that at high temperatures, as in the first case phonon distribution slowly returns to equilibrium.

It has been found experimentally that at low temperatures this additional thermoelectric power may be hundreds of times greater than conventional thermoelectric effect, caused only by the effect of temperature gradient on charge carrier movement. In case of high temperatures, the role of phonon-to-phonon scattering increases and electron capturing effect disappears.

It should also be noted that in the magnetics one additional thermoelectric component exists, when electrons are captured by the magnons.

The Peltier effect in some ways appears to be the opposite of Seebeck effect, and lies in the fact that, when current passes through a conductor (or semiconductor), some heat Q_P is released or absorbed in the contacts (depending on current direction), at that Q_P is proportional to the amount of electricity $I \cdot dt$, passed through contact:

$$Q_P = P \cdot I \cdot dt.$$

where P is kinetic coefficient of the Peltier effect.

When external electric current coincides with the direction of thermoelectric current, this contact is cooled. The analogy with Seebeck effect means that electrical current circulation through a circuit consists of two different substances; in the absence of initial temperature gradient ($\nabla T = 0$) the current cools one contact and heats another contact. It should be noted that in the case of Seebeck effect external electrical field is absent, but the temperature gradient exists ($\nabla T \neq 0$).

The Peltier effect can be explained on the basis of metal-semiconductor energy diagram. For definiteness, a contact "metal- n -type semiconductor" is considered in the condition, when work function of electrons in metal is greater than that in semiconductor: $F_m > F_s$. The cause of thermoelectric effect is that *average kinetic energy* of electrons (that are involved in electrical current creation) in metal and in semiconductor is *different*. It is important that in the metal only those electrons are involved in charge transfer whose energy is located near Fermi surface. However, in the n -type semiconductor, current can be carried by electrons, located in the conduction band. The energy of electrons in the conduction band of semiconductor is *greater* than the energy of electrons in the metal at the Fermi level on a value $E_c - F_m$.

Under the influence of the external electric field, directed in such a manner that electron transition occurs *from semiconductor to metal*, the higher-energy electrons of a semiconductor penetrate into the lower Fermi level of the metal; next, during collisions with atoms of metal, these electrons return their excess energy. The heat, released in this way, is a Peltier heat. As the electrons reach thermal equilibrium as a result of small number of collisions in the vicinity of contact, almost all Peltier heat is released just in the contact. In case of the opposite direction of external electrical current, the electrons can move from the metal to the semiconductor *only by overcoming energy barrier* $E_c - F_m$. To do this, the electrons have to *get energy from lattice*, causing the *cooling* of conductor in the contact area.

When two metals are in contact, the Peltier effect is so small that it is indistinguishable against the background of ohmic heating. Therefore, this effect has practical applications only with the contact of two semiconductors [3].

Thus, the electrical current actually transfers heat from one side to the opposite side of the *Peltier element* and creates the temperature difference. When the heated side of Peltier element is additionally cooled, for example, by a fan or heat sink, the temperature of the cold side becomes lower. In the single-stage element, depending on the type of element and current value, temperature difference can reach up to about 70 K.

The Thomson effect is the heat transfer by a current, flowing through *homogeneous* material in which *temperature gradient is created*. In the volume of a conductor, a certain quantity of heat is absorbed or released (depending on current direction), that is proportional to current strength, time, and temperature gradient:

$$Q_T = \tau_T dt / \nabla T,$$

where τ_T is the *kinetic coefficient* of the Thomson effect.

The physical cause of the Thomson effect can be explained by examining heat transfer by free charge carriers in the external electrical field. Such conditions are considered, when along the conductor, through which electrical current flows, the temperature gradient exists, and direction of this current corresponds to electrons moving from the hot end to the cold end. Electrons, moving from hot to cold areas, give excess energy to the surrounding lattice, wherein heat is released, and the conductor is heated. If the direction of current were to be reversed, the electrons will move from the cold area to the hot area, acquiring energy from the lattice. Obviously, this case corresponds to absorption of heat.

The Thomson effect in a heterogeneous semiconductor can occur in the absence of external electrical current. If the conductor were to be heated unevenly, the charge carrier concentration will be greater in the area where temperature is higher. Therefore the temperature gradient will result in carrier concentration gradient, and, hence, there will be the *diffusion current*. The separation of charges creates internal electrical field [8].

In the absence of current flow ($I = 0$), the following expression can be obtained for electrical field appearance in the case of thermoelectric effects:

$$E = \Delta F / e + \alpha \nabla T.$$

This ratio describes all three thermoelectric effects. It should be noted that in terms of mechanism of heat release or absorption, the Peltier and the Thomson effects are similar. The Thomson effect is the release of energy in a thermoelectric field: $E_\alpha = \alpha \nabla T$, while the Peltier effect is energy release in the an electrical field: $E = 1/e \Delta F$. Between kinetic coefficients of all thermoelectric effects there are relationships, which can be obtained from thermodynamics;

$$\begin{aligned} \alpha &= P/T; \\ \tau_T &= -T \cdot d\alpha/dT. \end{aligned}$$

Thus, all thermoelectric phenomena are closely related to each other, and can be described by only one parameter α determined experimentally.

8.5 OPTICAL PHENOMENA IN SEMICONDUCTORS

Optical phenomena include a large group of effects occurring in crystals during their interaction with electromagnetic radiation of optical wavelengths. It is assumed that on crystal the incident electromagnetic irradiation (light) acts with wavelength λ and intensity $I_{\text{inc}}(\lambda)$. By investigating reflected light intensity $I_{\text{ref}}(\lambda)$ and intensity of light, passing through a sample $I_{\text{pass}}(\lambda)$, it is possible to study processes that occur in a crystal under light illumination [3].

To describe optical effects, such characteristics are introduced:

The reflection coefficient $R(\lambda)$ that is given by ratio

$$R(\lambda) = I_{\text{ref}}(\lambda)/I_{\text{inc}}(\lambda),$$

which describes the part of light, reflected from crystal, to the incident light; $R(\lambda)$ is the dimensionless value that very often is expressed in percentages. Reflection coefficient is described by the relationship:

$$R(\lambda) \approx \left[(n-1)^2 + k^2 \right] / \left[(n+1)^2 + k^2 \right]^2,$$

where n and k are real and imaginary parts of the *complex refractive index*: $n^* = n - ik$. Both parameters n and k are dimensionless, and depend on dimensionless complex dielectric permittivity: $\epsilon^* = \epsilon' - i\epsilon''$. Refractive coefficient equals $n = (\epsilon\mu)^{1/2} \approx (\epsilon)^{1/2}$, because in optical range magnetic permeability $\mu \approx 1$; correspondingly, absorption index is $k = \epsilon''/2n$.

In *metals* photon energy is spent on the excitation of free electrons that are characterized by the quasicontinuous energy spectrum. Light absorption in metals occurs only on surface (within a few atomic layers), and it is accompanied by the reradiation of photons. This explains the large (over 95%) light reflectivity of metals.

Semiconductors usually have optical refractive index $n=3-4$, while reflection from semiconductors in *far-infrared* illumination is characterized by a value $R(\lambda)=25\%-40\%$. However, in the visible optical range, when energy of incident light quanta exceeds the bandgap of semiconductor, the *fundamental absorption* is observed with $k \gg n$. The point is that on the surface of the semiconductor under the influence of light a large concentration of excited free charge carriers appear; thus optical reflectance of semiconductor is $R(\lambda) \approx 90\%$; they show a “metallic luster” and are almost opaque. The main condition of fundamental light absorption in semiconductors is that photon energy $h\nu$ is sufficient for valence bond photo-ionization: $h\nu \geq E_g$, where E_g is bandgap of semiconductor.

The absorption coefficient α is determined by the index of absorption k and light wavelength λ : $\alpha = 4\pi k/\lambda$; it is characterized by light intensity absorbed in a sample having unit of thickness (Bouguer-Lambert law):

$$I_x = I_{\text{inc}}(1 - R)e^{-\alpha}, \quad \alpha = x^{-1} \ln \{ [I_{\text{inc}}(1 - R)]/I_x \}.$$

The value of $(1 - R)$ is a fraction of light that passes through illuminated surface into the sample (taking in account reflection coefficient R). When passing any layer of

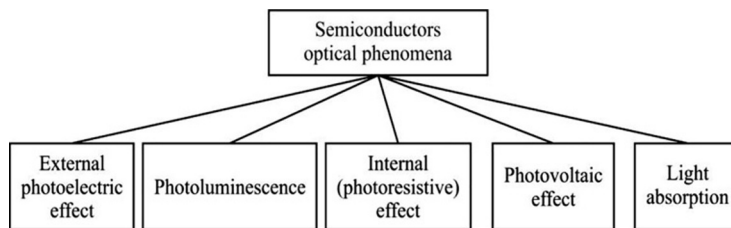


FIG. 8.19

Optical phenomena classification of semiconductors.

material, the intensity of luminous flux decreases to a certain part, dependent only on thickness. It is assumed that the loss of photonic beam in the media is independent of light intensity and on thickness of absorbing layer. Absorption coefficient α has the dimension of *inverse length*, measured in cm^{-1} .

During crystal illumination, the energy of incident photons is transmitted into the semiconductor, resulting in quite different physical or chemical processes (Fig. 8.19).

The external photoelectric effect in the optical range in semiconductors is insignificant (unlike metals). In order to demonstrate this effect, it is necessary that energy of absorbed quantum is sufficient to release electrons from the valence band and throw them outside. This process requires considerable energy; therefore, external photoelectric effect in semiconductors under light irradiation becomes noticeable only at frequencies, which are much higher than the frequency of visible light. Typically, a part of very high-frequency radiation in total incident solar radiation is relatively small; therefore, external photocurrent in conventional semiconductors is small.

Photoluminescence is secondary radiation, conditioned by quantum excitation of matter by light; it is explained by electron quantum transitions in atoms or ions from the excited state to the basic state (or to less excited state). During absorption of a light in semiconductors, a pair of free electrons and holes appear. These free charge carriers are in the excited states only for a certain period of time τ (lifetime), and then, usually, electrons and holes recombine. The energy that is released during recombination might be reradiated as photons. This is the phenomenon of photoluminescence.

Secondary radiation arises when a system is in the nonequilibrium state. Return to an equilibrium state can be different. The luminescence that occurs only during the excitement is the *fluorescence*, while another that continues for some time *after* excitation is the *phosphorescence*.

There are some specific mechanisms of photoluminescence. *Spontaneous* luminescence occurs, when, at first, nonradiating transition from the excited level to the basic energy level occurs; at that, further radiation takes place. This type of luminescence is a characteristic of the impurity states in solids. *Forced* luminescence is defined as a process that occurs after energy absorption, that is, at transition to

intermediate (metastable) level, from which further transitions take place. *Recombination* luminescence occurs as a result of particles reassociation (they were departed due to absorption of exciting energy). Recombination luminescence is an important characteristic to study defects or impurity centers in semiconductors.

Therefore, luminescence refers to *nonequilibrium* optical phenomena that are important characteristics of semiconductors and also used in manufacture of correspondent devices [8].

The internal photoelectric effect is caused when an illumination of a number of free electron-hole pairs appears near the surface of a semiconductor. This occurs when photon energy is sufficient for tearing off electrons from atoms and throwing them over from the valence band to the conduction band. This absorption of light is *intrinsic absorption*.

Study of absorption spectrum is an important experimental method of investigating solid band structure, especially for semiconductors. Light absorption shows a sharp increase when energy of photons equals or exceeds the energy gap between the valence and conduction bands (i.e., forbidden energy of semiconductor). At that, the impulse of photon h/λ (where $\lambda \sim 10^{-4}$ cm is the average optical wavelength) is very small as compared with the impulse of electron in crystal h/a ($a \sim 10^{-8}$ cm, lattice parameter); therefore, impulse of electron is practically unchanged during photon absorption.

During illumination by a light, the concentration of electrons and holes greatly increases, resulting in *electrical conductivity* of semiconductor increasing vastly. Under the influence of external factors in pure monocrystalline, semiconductor-induced conductivity exhibits an *intrinsic* character, because it is caused by the excited state of semiconductor atoms. Optical absorption of solids is usually described by the frequency dependence of absorption coefficient α .

The photoresistivity (and photoconductivity) of semiconductor is the consequence of internal photoelectric effect, and it is conditioned by light generation of charge carriers at the expense of intrinsic or impurity absorptions. Regarding semiconductor illumination and the following light absorption, one photon usually generates one electron-hole pair. As in the light beam, there are a large number of photons, thereby significantly increasing electrical conductivity during illumination. The effect of photoconductivity is used to create a wide class of electronic devices.

Total conductivity σ_{ph} under light irradiation and absorption can be presented in two parts:

$$\sigma_{\text{ph}} = \sigma_0 + \Delta\sigma_{\text{gen}} = e(n_0u_n + p_0u_p) + e\Delta n(u_n + u_p),$$

where σ_0 is *dark* conductivity (stationary), $\Delta\sigma_{\text{gen}}$ is conductivity caused by light absorption, and Δn is charge carrier density, generated by light. With regard to practical application, photoconductivity should be the dominating mechanism: $\Delta\sigma_{\text{reH}} \gg \sigma_0$. This can be achieved in the intrinsic or in the weakly doped semiconductors.

Charge carrier generation by light depends on a value of *quantum yield* η , which is a number of free charge carrier pairs that single photon creates during light absorption. Quantum yield depends on the frequency of falling light (Fig. 8.20). In germanium, the photogeneration starts with energy $E_g = 0.68$ eV (Fig. 8.20A), and in the

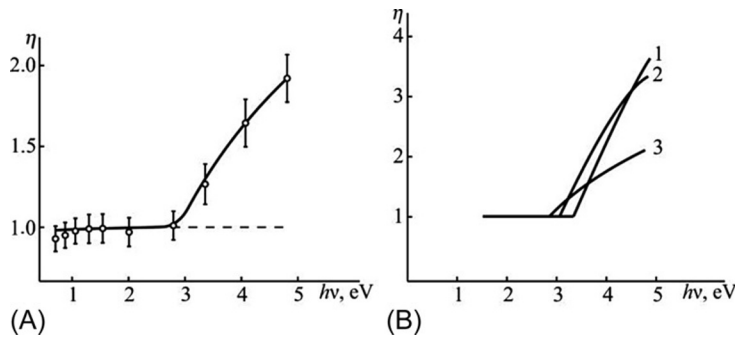


FIG. 8.20

Spectral characteristics of photoelectric effect; internal quantum yield at different temperatures: (A) in Ge at 300 K; (B) in Si: 1—at 100 K, 2—300 K; 3—at 400 K.

frequency range of 1–2.7 eV the quantum yield is constant value $\eta = 1$, that is, one photon generates one electron-hole pair. The excess energy (relatively E_g) is dissipated with phonon generation. With further increase in energy, one photon generates more than one pair of charge carriers. Excess energy is sufficient for both *interband impact ionization* of valence bonds and formation of additional electron-hole pairs; therefore, quantum yield increases: $\eta > 1$.

The effect of temperature change on quantum yield is shown, for example, for silicon in Fig. 8.20B. With increasing temperature, the limiting power of early interband ionization shifts toward the smaller values of photons energy that can be explained by the bandgap decrease with temperature [8].

Thus, in case of internal photoelectric effect of valence bonds, ionization becomes possible, creating the *secondary electron-hole pairs* that increase quantum yield. When $\eta > 1$, energy of photons corresponds to violet light waves.

The photovoltaic (Dember) effect is bipolar diffusion of charge carriers generating EMF. The nature of this effect can be explained in the following way: light is absorbed in the surface layer of the semiconductor while average thickness of this layer is equal to the track length of photons (light cannot penetrate deep into the volume of semiconductors). Thus, the surface layer of a semiconductor becomes a source of nonequilibrium electron-hole pairs. This random generation of charge carriers is irregular and decreases exponentially (Bouguer-Lambert law).

As a result, common bipolar (electron and hole) diffusion occurs directed to the depth of a semiconductor due to gradient of excessive concentration of charge carriers. As diffusion streams of electrons and holes spread in the *same direction*, these streams could compensate each other—in case of identical values of electron and hole diffusion coefficients. However, because the rate of electron diffusion is greater than the rate of hole diffusion, electroneutrality is violated: electrons have greater mobility (and higher diffusion coefficient), thus charges become spatially separated. The space charge appears, and, as a result, the electrical field is directed along the flow of separated charges.

As a consequence of such dynamic polarization between illuminated surface and the opposite (dark) surface of a semiconductor sample with thickness d , the difference of potential U_D and static electrical field E_D appears, which is called the *Dember field*:

$$E_D = U_D/d.$$

Potential difference U_D for germanium and silicon is from tens of microvolts to several millivolts.

Thus, in simplified definition, the Dember effect is violation of electrical neutrality in a semiconductor during light absorption that is explained by different mobility of electrons and holes.

Light absorption mechanisms. Light absorption is the decrease in optical radiation intensity during light passage through a matter and interacting with it, causing light energy conversion into other forms of energy. Usually, absorption coefficient does not depend on light intensity, but it is different for various wavelengths [8].

With the assumption of exponential law of light attenuation, in the depth x of a sample parameter α can be interpreted as an indicator of photon absorption in the unit of sample thickness, while value α^{-1} is the *average track length* of photons in a sample. In case of large values of *refractive index* ($n \gg 1$), practically total reflection occurs: $R \approx 1$. If absorption is very large, the reflection also will be nearly absolute. For this reason, a mirror-type luster in the metals is seen, while for most of semiconductors the dim shine is peculiar, because the absorption coefficient of material in visible spectrum is very large ($\alpha \geq 10^4 \text{ cm}^{-1}$).

The features of spectral dependencies in different parts of optical spectrum are due to different mechanisms of light wave interaction with a crystal, mainly, thanks to different mechanisms of light energy absorption. Therefore, a theory of optical phenomena in semiconductors is developed conformably to analyze light absorption spectra.

In various spectral ranges, the prevailing mechanisms of absorption should be identified. As the light absorption is associated with photon energy transformation into another form of energy in a crystal, the classification of absorption mechanisms is conditioned by energy states of semiconductor lattice (Fig. 8.21). This diagram lists only basic mechanisms of light absorption.

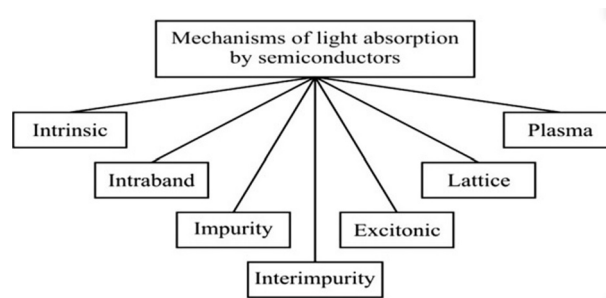


FIG. 8.21

Mechanisms of light absorption by semiconductors.

The *intrinsic* (or fundamental) absorption is conditioned by electronic transitions between allowed energy bands. Under these conditions, the interaction of photons with electrons in the valence band occurs, that is, with own electrons of atoms that make up the crystal lattice. Photons of certain energy are capable of giving their energy to electrons and detaching them from atoms, transferring electrons to the higher energy levels. In this case, photons are absorbed in a crystal.

During intrinsic absorption, the transition of electron can be *direct*, if the wave vector of electron remains the same as well as the electron and originated hole have the same quasi-impulse. The *indirect* transition occurs with the involvement of phonons, to which the excess impulse of electron is delivered. During frequency dependence of coefficient α study, the energy gap of a semiconductor can be defined by the *edge* of intrinsic absorption [2].

The absorption by the *free charge carriers* is due to electron (or hole) transitions *inside* allowed bands, as well as between the sub-bands of the approved band. Thus, this absorption occurs when photons react with free charge carriers in the permitted bands. Photon energy is spent on transition of charge carriers to the higher energy levels. Under the influence of electrical field of light, the charge carriers perform *oscillatory movements*, synchronously with the light field, and return accumulated energy during collisions with lattice sites.

The *impurity absorption* is due to electron (or hole) transitions between allowed bands and impurity levels in the forbidden band. In case of impurity absorption of light, its photons interact with impurity atoms, ionizing or exciting them. The *inter-band absorption* can be explained by the transitions of electrons (or holes) between the impurity states in the forbidden band.

The interaction of photons with impurity atoms has a resonant character. In semiconductors, the absorption of photons can be created also by the excitons (bound electron-hole pairs) that randomly move through a crystal. The *excitonic absorption* is due to exciton generation that makes a significant contribution directly near the fundamental absorption edge, as assessment of energy by excitonic state is small.

The absorption of light by the crystal lattice (*phonon absorption*) can also occur in semiconductors. It manifests itself in the far-infrared region of a spectrum. Phonon absorption is caused by the absorption of light wave energy by atom vibrations in a crystal by the birth of new phonons in a lattice. In Fig. 8.21, the *plasma absorption* is also mentioned; this is energy of light wave absorption by the electron-hole plasma, which leads to transition of plasma into higher quantum state.

Almost all mechanisms of absorption, which are caused by different electrons (or holes) transitions, are accompanied by absorption or emission of *phonons*. The need for phonon participation in the processes of light absorption is associated with the implementation of the *law of impulse conservation*. The fact that a large change in electron (or hole) impulses during some transitions usually can be caused by small impulses of photons (which are absorbed at these transitions) that require the participation of phonons, which can have quite large impulses.

Therefore many mechanisms of absorption, listed in Fig. 8.21, are the combination of different mechanisms, which involved electrons and holes, as well as

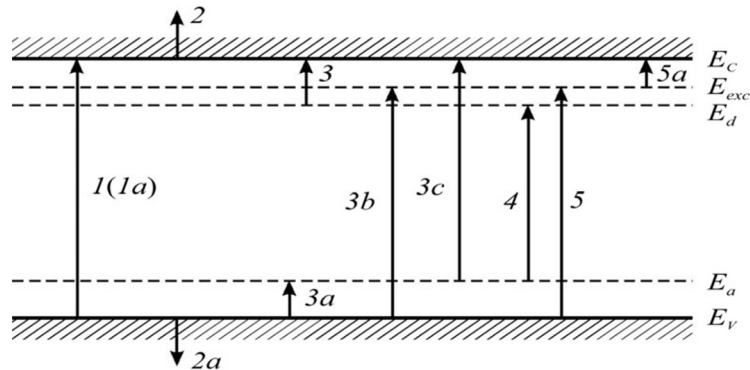


FIG. 8.22

Scheme of electronic transitions during optical absorption: 1, 1a—intrinsic absorption; 2, 2a—absorption by free charge carriers; 3, 3a—absorption by impurities of closest zone; 3b, 3c—absorption by impurities; 4—interband absorption; 5—excitonic absorption by excitation; 5a—excitonic absorption by optical decay of excitons.

phonons. A diagram of electron transitions, responsible for different absorption mechanisms (1–5), is shown in Fig. 8.22.

Next, basic mechanisms of light absorption will be discussed in more detail. To assess the role of various absorption mechanisms in different parts of spectrum, the energy of electronic transitions, occurring in semiconductors, should be compared [8].

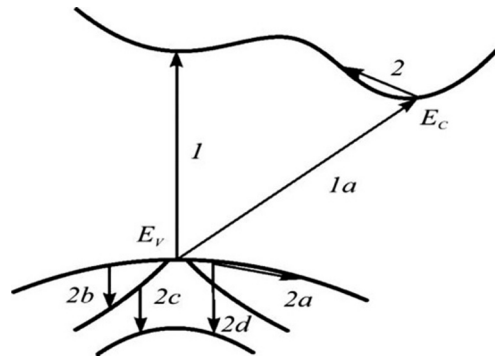
The fundamental (intrinsic) absorption is dominating absorption mechanism in semiconductors in near-infrared and in visible ranges of spectrum. The main condition of fundamental absorption is that the energy of photons should be sufficient for valence bond photoionization:

$$h\nu \geq E_g,$$

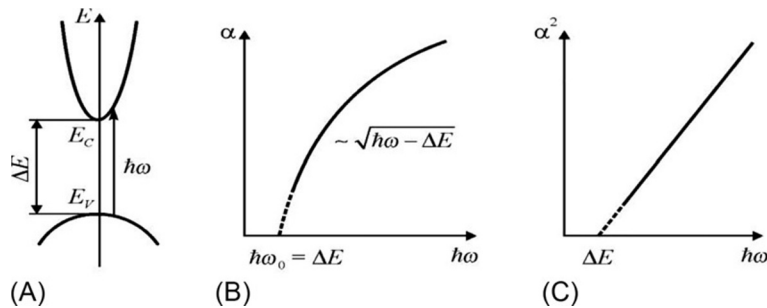
where ν is frequency of incident photons, E_g is bandgap, and h is Planck constant. Sign of equality in this expression determines the *edge* of intrinsic absorption: $\lambda_k = hc/E_g$, where c is speed of light.

Absorption in the short-wave part of spectrum is accompanied by the bipolar generation of nonequilibrium electron-hole pairs. The energy and impulse conservation laws for electronic transitions are determined by the energy band structure of a semiconductor. The electron and hole transitions in the k -space are depicted in Fig. 8.23. Intrinsic absorption of light is shown by transitions 1 and 1a. The transition 1 can be realized without significant change in wave vector of electrons, and, therefore, this absorption is the *direct intrinsic absorption*.

Transition 1a occurs with significant change in electron wave vector. This change in germanium and silicon has the order of the Brillouin zone size (i.e., $k \approx 10^8 \text{ cm}^{-1}$).


FIG. 8.23

Electronic transitions in k -space describing optical absorption: 1 —direct transition; $1a$ —indirect transition; 2 and $2a$ —intraband (selective) absorption of free electrons and holes; $2b$, $2c$, and $2d$ —selective absorption by free holes.


FIG. 8.24

Electronic transitions (A) and absorption spectra (B, C) in direct-band semiconductor.

This effect is *indirect intrinsic absorption*. As already noted, at indirect absorption phonon involvement is necessary.

From the examination of this transition, it becomes clear that absorption is not possible in case of small energy of quanta ($h\nu < \Delta E$). As a result, absorption spectrum should have a recession, called the *absorption edge* at the frequency of photons $\nu \approx \Delta E/h$.

The direct intrinsic absorption is characterized in detail in Fig. 8.24A, which shows the band structure of direct-band semiconductor in the neighborhood of $k = 0$. Fig. 8.24B demonstrates the typical character of frequency dependence of absorption; in semiconductors, it corresponds to parabola of $\frac{1}{2}$ degree.

Observed for direct transitions, parabolic dependence $\alpha(\hbar\omega)$ is a result of energy band parabolic form. Therefore, deviations in $\alpha(\hbar\omega)$ dependence from parabolic

character indicate the nonparabolic dispersion law of valence band (or conduction band). In case of $m_n^* < m_p^*$ the function $\alpha(\hbar\omega)$ gives important information about the shape of the conduction band, while in case of assumption $m_n^* \ll m_p^*$ (often implemented to narrow-band semiconductors), it is possible to assume that deviation of $\alpha(\hbar\omega)$ dependence from parabolic law clearly indicates nonparabolic form of the conduction band.

By observing absorption spectrum of the direct-band semiconductor, it is possible to determine the width of its energy gap, extrapolating curve of absorption to the region where $\alpha = 0$. To increase the accuracy of experiment, it is necessary to extrapolate the $\alpha(\hbar\omega)$ dependence in such coordinates, where it becomes straight. Obviously, the absorption spectrum needs to be built in the coordinates $\alpha^2(\hbar\omega)$. In Fig. 8.24C, this method is used for allowing direct transitions. It is shown that band-gap can be determined by linear extrapolation to zero [8].

Indirect intrinsic transitions are characterized by different forms of absorption edge (this case is observed in silicon and germanium). Corresponding valence band diagram has a form, shown in Fig. 8.25 in the impulse space. The valence band maximum E_v is located in the center of Brillouin zone, while the minimum of conduction band E_c is located on the border of Brillouin zone (or very close to border) so that change in electron quasi-impulse at transition from E_v to E_c must be very large—close to the size of a Brillouin zone.

The impulse of absorbed photon $p_{\text{phon}} = \hbar k_{\text{phon}}$ is small being conditioned by photon energy $\hbar\omega$. Assessing $\hbar\omega \approx \Delta E \approx 1$ eV, the value for k_{phon} can be obtained, $\sim 10^4 \text{ cm}^{-1}$, which is much smaller than the size of Brillouin zone. Thus the indirect transitions of electrons cannot be realized only with photon absorption, because the law of impulse conservation will not be implemented. This law is valid only in case of *phonon absorption involvement* (with impulse of large enough magnitude). Then, for indirect transitions, a law of impulse conservation (neglecting small quantity of photon impulse) takes the following form:

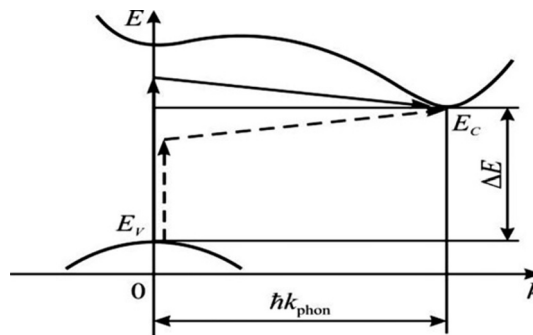


FIG. 8.25

Direct light passes through intermediate virtual state; dotted line shows transition with phonon absorption, while solid arrows show transition with phonon generation.

$$\hbar k' - \hbar k \approx \pm \hbar k_{\text{phon.}}$$

The “plus” sign corresponds to phonon absorption simultaneously with photon (at the same time), while the “minus” sign corresponds to the birth (generation) of phonon while photon absorption.

The law of photon energy conservation in case of indirect transitions takes the form:

$$E_{k'} - E_k = \hbar\omega - \hbar\omega_k,$$

where $\hbar\omega_k$ is energy of phonons participating in absorption. Two different values of energy for a particular point $E_0 = \hbar\omega_0$ can be observed at the process of absorption or emission of phonons. As the indirect transition should engage several number of particles (electron, photon, and phonon), as compared during direct transitions (electrons and photons), the probability of indirect transitions and, thus, absorption coefficient must be lower than that for direct transitions. This can be explained by the *conditional dividing* of absorption process in case of indirect transitions on two stages, as shown in Fig. 8.25.

In the first stage, the electron-absorbing photon jumps over from the valence band by the direct transfer into the *virtual state* in the conduction band (that is not related to violation of energy storage law). In the second stage, the electron moves from the virtual state of the conduction band into the final state—to the minimum E_c , releasing or absorbing a phonon. The probability of transition will be determined by multiplying density of states.

The probability of phonon absorption process should be proportional to the number of phonons, given by Bose-Einstein function, while the probability of phonon emission process should be proportional to the function that determines probability of phonon emission. Therefore, the absorption coefficient for allowed indirect transitions should be determined by a complex dependence.

To interpret the $\alpha(\hbar\omega)$ dependence in case of indirect allowed transitions, the $\alpha^{1/2}(\hbar\omega)$ dependence should be built, as shown in Fig. 8.26. The absorption spectrum

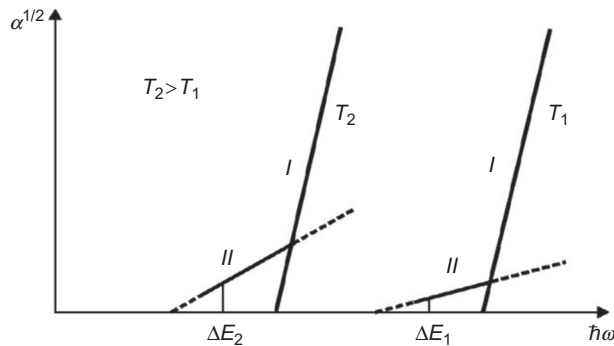


FIG. 8.26

Determination of bandgap energy and photons involved in indirect transitions.

can be represented as a sum of two linear dependencies, which by extrapolating $\alpha \Rightarrow 0$ give two energy values: $\hbar\omega_1 = \Delta E - \hbar\omega_k$ and $\hbar\omega_2 = \Delta E + \hbar\omega_k$. The point that is located in the midway between $\hbar\omega_1$ and $\hbar\omega_2$ corresponds to bandgap ΔE . Fig. 8.26 shows dependencies at two different temperatures, taking into account the bandgap temperature change. Note that contribution from transitions that occur with phonon absorption decreases with temperature decline.

If temperature is below Debye temperature, the transitions with phonon absorption practically do not occur, because phonon concentration in crystal is very small. This case is shown in Fig. 8.26 at temperature T_1 . On the contrary, with increasing temperature (T_2) phonon absorption in transitions increases.

Using absorption spectra, temperature shift of band edge can be observed for direct transitions as well (Fig. 8.27). The absorption coefficient in case of direct transitions sharply increases, because the probability of involving only two particles—electron and photon—increases.

There is another important opportunity of absorption edge shift, in addition to temperature shift. Suppose that absorption is investigated in the noticeably *doped by donors* direct-band semiconductor, whose band diagram is shown in Fig. 8.27A. Then, in case of *electron degeneration*, the Fermi level enters into the conduction band, and states, which lie below the Fermi level F , are all nearly filled. Electron transition from the valence band with photon absorption $\hbar\omega = \Delta E + (F - E_c)$ is impossible, because the relevant states in the conduction band are already occupied by electrons. Because of such absorption, its edge will be shifted toward the higher-energy photons; Fig. 8.27B shows this shift in highly doped indium antimonide.

The interband absorption of light in semiconductors by charge carriers is possible in the presence of free charge carriers. Transitions occur within the band (electrons in the conduction band, Fig. 8.27A, or holes in the valence band) with absorption of photons and phonons. Transitions 2, 2a 2b, 2c, and 2d, which have been shown schematically in Fig. 8.22, occur during absorption of light energy by free charge carriers in the

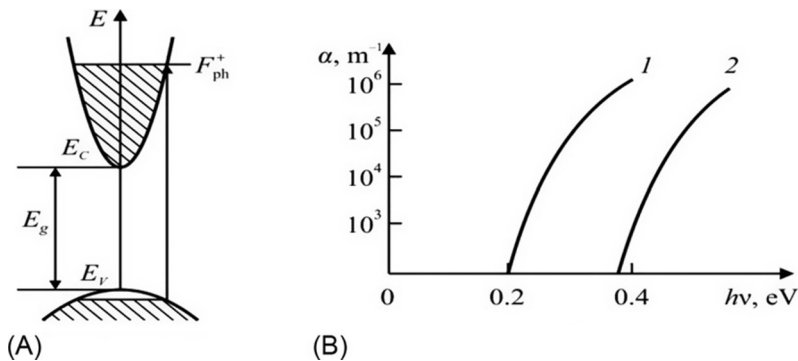


FIG. 8.27

Shift of absorption edge in InSb spectrum: (A) band diagram; (B) absorption edge of undoped (1) and strongly doped (2) samples.

permitted bands. Transitions 2 and 2a occur with change in the wave vector that necessarily requires participation of phonons. The spectral dependence of absorption for interband transitions of electrons and holes as appears to be a monotonous curve and is called *nonselective* (indiscriminate) absorption by free carriers.

If free charges at light absorption move within a permitted band and change their dispersion law (as shown for holes transitions 2b, 2c, and 2d between the sub-bands in the valence band), the absorption spectrum has structural features in the form of highs and lows alternating absorption, and it is the *selective absorption* by free carriers. The selective absorption by free charge carriers can take place without phonon participation, and it can be direct. Studies show that contribution to absorption from the free charge carriers increases strongly in the near-infrared region of spectrum.

Laws of conservation of electron energy and wave vector are executed only at simultaneous participation in the processes of absorption at phonon scattering on ionized impurities. The fact is impulse of photon in a crystal is negligibly small and does not meet the energy transmitted to electron. However, during scattering, the impulse of charge carrier varies widely.

Coefficient of absorption by free electrons (α_n) is directly proportional to their concentration n and to the wavelength, while it is inversely proportional to the average relaxation time τ and the effective mass m_n^* . Such laws have a simple physical explanation: the greater the concentration n , the greater the probability of electron-photon interactions; the greater the wavelength λ , the lower the energy of photons, and, therefore, a change in electron wave vector module, which facilitates the implementation of conservation impulse law.

The absorption by impurities is also shown in Fig. 8.22 (transitions 3, 3a, 3b, and 3c), including between-impurity absorption (transition 4), which takes place with a participation of local impurity states. Impurity atoms can be changed from neutral to the ionized state, as in transitions 3 and 3a, and, vice versa, from ionized to the neutral state, as in transitions 3b and 3c. In case of between-impurity transitions, the state of impurity levels also changes.

Spectral plots, which contribute to absorption transitions 3–3a and 3b–3c, are spaced far enough apart for small-depth impurity states. Transitions 3–3a in this case will contribute to the absorption in far-infrared part of spectrum, while transitions 3b–3c and transition 4 will contribute to the absorption near the fundamental absorption edge. If impurity-type absorption involves deep impurity states, these transitions will give a contribution to absorption beyond the edge of absorption, in the long-wavelength range of a spectrum.

The absorption by excitons can be significant in certain semiconductors. Excitonic transitions 5 and 5a in Fig. 8.22 are shown as conditional, because the description of excitons is a problem of interaction between *two particles*: electron and hole; therefore the level E_{exc} , shown in Fig. 8.22, only conditionally represents the state as “one particle” being electron-hole pair.

Intense light absorption, which is associated with formation of excitons (Fig. 8.28), can be observed at lower energies, that is, at lower frequencies of light, than interband absorption, because the binding energy of exciton is small in

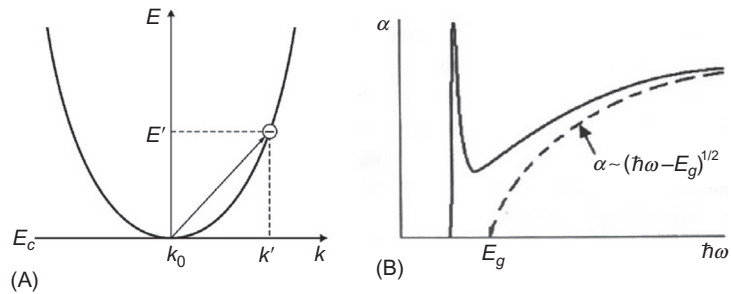


FIG. 8.28

Absorption of light: (A) by free electrons, (B) by excitons (*solid curve*—excitonic absorption; *dashed curve*—actual absorption).

comparison with E_g . Therefore the range of exciton absorption is located close to the fundamental optical absorption, but differs from lower frequency absorption. Interband absorption in the semiconductor, as shown in Fig. 8.28B, describes smooth dependence of absorption coefficient on energy of photon, while the excitonic absorption begins as a sharp sawtooth ascending energy, less than a threshold energy of interband absorption. It should also be noted that light absorption is accompanied by a sharp increase in photoconductivity, while during excitonic absorption of light the photoconductivity is absent [2].

The exciton is characterized not only by its hydrogen-like structure, but also by its wave vector, impulse, angular momentum, and so on. Consequently, the exciton is the excited state of a whole lattice. Because of lattice periodicity, this excitation can move from one atom to another. The size of excitons in the semiconductors is *much larger than lattice constant*. For example, in the germanium crystal, radius of exciton is 800 times larger than the Bohr radius of hydrogen atom. Such excitons with large radius are the *Wannier-Mott excitons*.

If the interaction between electron and hole is large enough, the radius of interacting pair can be commensurable with lattice parameter. The exciton in this case is the excited state of *individual* atom that easily moves through a crystal. Such an exciton of small radius is the *Frenkel exciton*. In semiconductors, the Wannier-Mott excitons are mainly observed, while the Frenkel excitons are peculiar to ionic and organic crystals. The excitonic states are seen in the optical spectra as comparatively narrow line near the absorption edge. Therefore, an account of electron-hole interaction can lead to a bound state (excitons) and a nonbound state. This is a reason for alteration of $\alpha(h\nu)$ dependence near the edge of intrinsic absorption (Fig. 8.28B).

Thus, in semiconductors during photon absorption the excitons (bound electron-hole pairs) can be created, which randomly move through a crystal. In the event of collisions with impurity centers, the excitons can either disintegrate creating electron and hole or recombine and transfer into atom of excited state. In the first case, the exciton needs thermal energy, while in the second case, either photon radiation occurs or energy of exciton transfers to semiconductor lattice in the form of a heat.

Phonon absorption, associated with light energy transition into the lattice vibrations, occurs in the spectral range, corresponding to the phonon energy: this is the *far-infrared region*, which only partially overlaps with the region of absorption by free carriers.

Plasma absorption occurs at sufficiently high concentrations of free charge carriers, and it has an important feature, so-called *plasma resonance*. As the absorption coefficient in the neighborhood of plasma resonance is very large, it is usually investigated not in the absorption spectrum but in the *reflection spectrum*.

At the frequency of plasma resonance, the *minimum of reflection* is observed, whose frequency position is associated with concentration and effective mass of charge carriers. The plasma minimum in the reflection spectrum of semiconductors and metals can be located near (or in the middle of) the infrared range of a spectrum.

In Fig. 8.22 it would be impossible to display as phonon, so plasma absorption as well as energy levels shown in these figure correspond to the *one-electron approximation*; therefore in diagrams, drawn in Fig. 8.22, the energy of lattice vibrations or energy of plasma (mixed electrons and holes system), in principle, is impossible to be shown.

8.6 SEMICONDUCTORS IN MAGNETIC FIELD

In this section, the influence of magnetic field on the movement of electrical charges (electrons and holes) is examined in the conductors and semiconductors if they are diamagnetic or paramagnetic substances. The effect of magnetization in this conditions is insignificant ($\mu \approx 1$), but when external magnetic field B is applied to conductors, and, especially, to semiconductors, many interesting and important electronic effects occur due to Lorentz force impact on the moving electrical charges.

Free charge carriers (electrons or holes) under the influence of constant magnetic field directed on z -axis in case of indefinitely free path of electron (with no scattering) would form its closed curve in the xy -plane perpendicular to the direction of magnetic induction B_z . Main parameters of this trajectory are the rotation frequency $\omega_c = eB_z/m^*$ with radius of orbit $r = v/\omega_c$ (v is average linear velocity) and period of rotation $T = 2\pi/\omega_c$. Thus, it is possible to assume that the effect of magnetic field is reduced to rotation of velocity vector projections in the plane xy without changing average modulus of velocity.

According to thermodynamic equilibrium state, all directions of linear velocity of charge carriers are equiprobable; therefore, magnetic moment of closed orbits is compensated. In addition, electronic spins in covalent bonds are also compensated, so the total spin moment practically equals zero. Only in case of very large value of external magnetic field, this spins compensation may be violated. This phenomenon can be observed as the paramagnetism of electronic gas in metals and semiconductors. However, even at these conditions, the change in electron energy in the external magnetic field is negligible.

This situation varies, when external magnetic field acts on the conductor (or the semiconductor) that is found in thermodynamically *nonequilibrium conditions*, when, due to many reasons (electrical field gradient, temperature gradient, lighting) a *directed* movement of electrical charges takes place. At that, mobility of electrons is usually higher than the mobility of holes.

The mechanism of charge transfer (drift, diffusion) and the velocity of electrons (holes) flow is also important. It should be noted that this velocity is not equal for *all* charge carrier flow: some of them are relatively *slow* (“more cold”) charge carriers, while others are much *faster* (“more hot”) charge carriers. In case of nondegenerated semiconductor, distribution of charge carrier on their velocity is given by the Maxwell law:

$$f(v) = \left(\frac{\beta}{\pi}\right)^{3/2} v^2 \exp(-\alpha v^2),$$

where $\beta = m^*/2k_B T$ while m^* is effective mass of charge carrier and k_B is Boltzmann constant.

The distribution function for velocity $f(v)$ is shown in Fig. 8.29. There are *three* velocities, peculiar in this case: the most probable velocity v_{\max} , the average velocity $\langle v \rangle$, and the mean square velocity $\sqrt{\langle v^2 \rangle}$. As shown in Fig. 8.29, Maxwell distribution of charge carriers is asymmetric as to most probable v_{\max} , because with increasing velocity the relative number of charge carriers decreases slower. Note that magnetic force influence is related exactly to the mean square velocity.

Galvanomagnetic effects. At the *joint action* of electrical and magnetic fields on conductors and semiconductors, a number of *kinetic phenomena* occur, such as difference in electric potentials, electrical conductivity change, and thermal conductivity change. Besides, some changes in electrical and thermal properties can be seen as in the *longitudinal* direction (along which electrical field is applied), so in the *transverse* direction. Such phenomena are usually called the *galvanomagnetic effects* (Fig. 8.30).

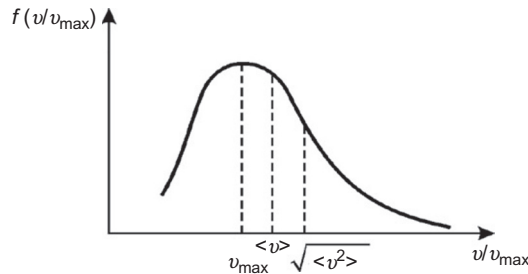


FIG. 8.29

Charge carrier Maxwell distribution on velocity in nondegenerate semiconductors.

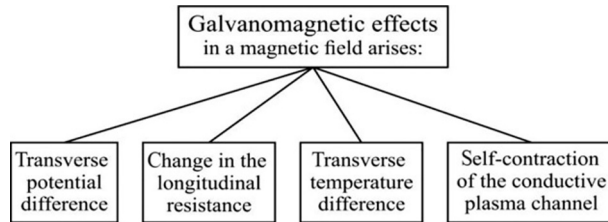


FIG. 8.30

Basic galvanomagnetic effects.

When placed in the magnetic field conductor, a transverse potential originates, known as *Hall's effect*, while the longitudinal resistance change in magnetic field is the magnetoresistive effect (this effect, sometimes, is also called *Gauss effect*). The appearance of transverse difference in temperature in magnetic field is the transversal galvanothermomagnetic *Ettingshausen effect*, while the longitudinal difference in temperature is the galvanothermomagnetic *Nernst's effect*. The self-induced compression of a channel with conducting plasma under magnetic field is the *pinch effect* [3].

Relative to magnetic field, galvanomagnetic effects are divided into even and odd. The effect is called as “odd,” if its direction changes when magnetic field switches its sign to opposite. Hall's effect and transversal galvanothermomagnetic effect belong to “odd” effects. In case switching magnetic field direction, the sign of effect does not change, then such effect is “even” that includes magnetoresistance effect and longitudinal galvanothermomagnetic effect.

All galvanomagnetic phenomena can be studied as in the *adiabatic* so in the *isothermal* conditions. The effect is adiabatic, if the studied sample cannot exchange its energy with the environment; therefore, in sample the temperature gradient (∇T) occurs. In case of isothermal conditions, on the contrary, the energy exchange takes place, and $\nabla T = 0$. Thus, Hall's effect and magnetoresistance effect are the isothermal ones as they are usually considered under the isothermal conditions.

It should be noted that galvanomagnetic effects can be described by the movement of charged particles, considering the *crossed* electrical E and magnetic B fields. As is known from electrodynamics, in such electromagnetic conditions the Lorentz force occurs, acting on electrical charge:

$$\mathbf{F}_{Lor} = e(\mathbf{E} + [\mathbf{v}_d \mathbf{B}])$$

In the crossed fields ($\mathbf{E} \perp \mathbf{B}$) charge carriers move along a *cycloid* that is a result of adding of two kinds of particle motion:

- rotation in a circle of radius $r_0 = m^*E/eB^2$ under crossed electrical and magnetic fields;
- moving in electrical field with drift velocity $\mathbf{v}_d = u\mathbf{E}$, where u is the drift mobility.

Thus the imposition of magnetic field to a conductor, in which electrical current flows, changes the trajectory of charge carriers; at that, the greater their

“twisting” in circular orbit, the stronger the magnetic field. The criterion of magnetic field value is a ratio between radius of curvature r_0 and middle free path λ of charge carrier.

If radius r_0 of electron’s trajectory curvature is much greater than mean free path λ ($r_0 \gg \lambda$), then in such fields the middle free time (relaxation time) is less than period of rotation. In this case, relaxation time is valid only on a part of electron complete rotation in magnetic field, that is, electron’s circular orbit is disconnected. Therefore electron movement is distorted only a little, and such magnetic field is called as *weak*.

Conversely, if $r_0 \ll \lambda$, then the magnetic field is *strong*, because it twists charge carriers so much that they significantly change their trajectory. Under this condition, relaxation time is greater than the rotation period, and, therefore, charge carrier has enough time to make several complete rotations in the magnetic field; hence scattering mechanism of charge carriers will be different from that in a weak magnetic field.

The concept of weak or strong field depends not only on “external” factor (magnitude of magnetic field B), but also on the mobility of charge carriers in crystal, that is, on peculiar properties of conductor (or semiconductor). It may be that the same magnetic field for one value of mobility looks as weak, but for another it is strong. For example, in the germanium at rather high temperatures (400 K), the mobility of electrons is $u_n \approx 0.3 \text{ m}^2/\text{V s}$ and the magnetic field strength of 10 kOe meets the criteria of a weak field. However, the same magnetic field for germanium at low temperature (about 10 K), when the mobility of electrons is high ($u_e \approx 100 \text{ m}^2/\text{V s}$), is considered as a strong field.

In metals under normal conditions, mobility of electrons typically is $u_e \leq 0.01 \text{ m}^2/\text{V s}$. Therefore, galvanomagnetic effects in metals as a rule correspond to the criterion of weak magnetic field.

When joint action of magnetic and electrical fields is examined, it should also take into account the *distribution* of charge carrier on their velocity and their energy. Usually, in many experiments, only *mean* velocity of electrons is taken into consideration. However, in the *magnetic field* the difference between fast (“hot”) electron behavior and slow (“cold”) electron behavior might have important meaning.

Hall’s effect was discussed in short previously in [Section 5.2](#). This effect is the appearance of Hall’s difference in potential when placing a conductor or semiconductor with a current in the crossed electrical E and magnetic B fields. In conditions of joint action of electrical and magnetic fields perpendicular to each other, the mobile charge carriers are turned by the Lorentz force in the third direction—perpendicular to the directions of both fields, see [Fig. 5.3](#) in [Section 5.2](#). In the previously considered case (electrical field E_x is directed along the sample while magnetic induction vector B is perpendicular to sample), electron bends aside by Lorentz force $\mathbf{F}_{\text{Lor}} = e[\mathbf{v}_d \mathbf{B}]$ from the initial direction of movement to one of lateral edges of sample. This corresponds to imagination about turning of trajectory of electron under magnetic field acting.

When charge carriers deviate to lateral edges of a sample, the transverse electrical field E_y occurs. If the sample in this direction is open circuited, then the redistribution

of charges results in a prohibitive force—field intensity in the direction of y -axis (axes are indicated in Fig. 5.3). This process will go on as long as electrical field E_y becomes large enough to compensate the force that causes deviation of charge carriers to lateral edges.

In equation (5.9), it needs to substitute: $E_{Lor} = 0$; from this the expression for Hall transverse electrical field follows:

$$E_y = v_x B_z = R_H j_x B_z.$$

It can be seen that Hall's field is directly proportional to current density j_x and to magnetic induction B_z . The constant of proportionality R_H is Hall's coefficient (or Hall's constant); in weak magnetic fields, it does not depend on magnetic induction value.

Hall's constant is independent of scattering mechanism: $R_H = (ne)^{-1}$. After transformation it is possible to obtain: $E_y = (1/n_e e c) j_x$. Since charge of electron "e" and light velocity c are well known, while values of j and B are measured directly, this formula allows to determine the number of *electrons per unit volume* n_e both in the semiconductor and in the conductor. For this reason, Hall's effect is widely used in the investigations of semiconductors.

In other words, in n -type semiconductor major charge carriers are electrons. Deviating from the electrical field direction by Lorentz force, they create on the lateral surface of sample (to which Lorentz force is directed) a negative space charge. In p -type semiconductors, for same reason, the holes create a positive space charge on the side surface of a sample. Experimentally measured sign and magnitude of Hall potential between sample side surfaces open a possibility to determine not only sign of majority charge carriers, but also their concentration.

In the case of *mixed* conductivity, when in charge transfer as both electrons and holes are involved, an analysis of Hall's effect is difficult. As it can be seen from the shown formula, Lorentz force has the same direction for electrons and holes. Therefore, electrons and holes deviate to one side. However, mobility and concentration of electrons and holes are different; in this case, in the created space charge electrons and holes cannot completely compensate each other. In the degenerated semiconductors, as in metals, those electrons are involved in electrical conductivity located in highest levels of energy. Therefore, in this case, it is possible to ignore electron distribution in energies.

If one compares Hall's effect in semiconductors and metals, it should be noted that concentration of electrons in metals is higher in several orders of magnitude than in semiconductors. Therefore, the Hall field in metals is much less than in semiconductors. Therefore, in the *Hall sensors* of magnetic fields, only semiconductors should be used. As Hall's effect is widely used in magnetic field measurements, it should be noted that the largest value of Hall field E_y can be obtained in *nondegenerated doped semiconductors*, provided when the concentration of majority charge carriers is, at least, in order of magnitude higher than the concentration of minority charge carriers.

Among the different effects discussed, there are also the *anomalous* Hall's effect, the *quantum* Hall's effect, and the *spin* Hall's effect.

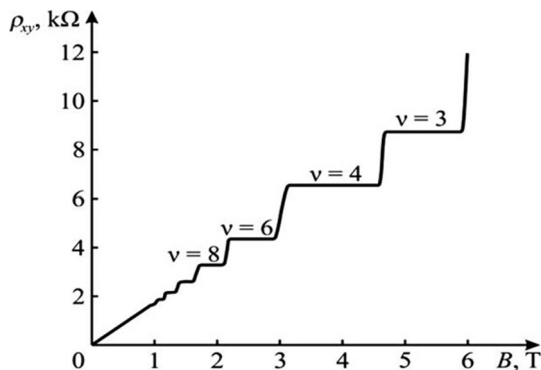


FIG. 8.31

Hall resistance dependence in strong magnetic field: quantum numbers n correspond to distance between Landau levels; see also thermomagnetic effect in Fig. 8.38.

Electrical field appearance in conductive material perpendicular to direction of current (which passes through a sample) may be observed even *in the absence* of magnetic field; this phenomenon is called the *anomalous* Hall's effect. It is quite similar to Hall's effect, but occurs without external magnetic field. Anomalous Hall's effect can be observed in conductive materials having *intrinsic magnetization*.

If the *flat* conductor is placed into the *strong* magnetic field (this is the case of two-dimensional [2D] electronic gas), the quantum effects become apparent that results in appearance of *quantum Hall's effect* with quantization of Hall resistance (Fig. 8.31). In much more strong magnetic fields, the *fractional quantum* Hall's effect manifests itself, which is associated with radical restructuring of internal structure of 2D "electronic liquid."

The magnetoresistance (or magnetoresistive Gauss effect) is the change of electrical resistance in a conductor when magnetic field is applied. In general, under magnetic field influence the effect of current change should be observed keenly. Therefore, any conductive material, to some extent, has to show magnetoresistance. However, relative change of resistance in *semiconductors* can be hundreds of times more than that in metals [7].

The quantitative characteristic of resistance change in magnetic field is the factor

$$K_{\rho m} = \frac{\rho(B) - \rho(0)}{\rho(0)},$$

where $\rho(B)$ is resistance in applied magnetic field and $\rho(0)$ is resistance in the absence of magnetic field.

Magnetoresistivity of conductive material depends on sample orientation in the magnetic field. This is because magnetic field does not change the projection of charged particle velocity *along* magnetic field direction, but through Lorentz power only twists free path of carrier in plane, *perpendicular* to magnetic field. This

explains why transverse magnetic field shows stronger impact on resistance than the longitudinal one.

Magnetoresistance is discussed earlier in the metals (see Section 5.2). Without magnetic field, the charge carrier moves between collisions along a *straight line*. In the external magnetic field (perpendicular to current), free path of electron becomes a *cycloid* with the length l . Therefore, at free time (time between two collisions) the way along electrical field will be $l_x \approx l \cos \varphi$, that is, less than l in the E_x direction. This corresponds to the *decrease in drift velocity* (i.e., to decrease in mobility), and, therefore, to resistance increase. Therefore, the relative difference between resistance at magnetic field's presence and resistance when magnetic field is absent is the magnetoresistance.

In *doped semiconductors* with one type of charge carriers, taking into account statistical distribution of velocities, theory gives the following expression for kinetic coefficient of magnetoresistance:

$$K_{\rho m} = Cu^2B^2,$$

where u is mobility, B is magnetic induction, and C is a factor that depends on scattering mechanism. Based on this formula, the effect of magnetoresistance is the *even* effect (B^2). Resistance measurement in magnetic field makes it possible to find charge carrier *mobility*, if scattering mechanism is known.

Based on the magnetoresistance effect, the *magnetic field sensors* are created. To select materials for such sensors, charge carrier mobility u is of crucial importance: as seen from this formula, magnetoresistance is proportional to u^2 . Therefore, sensitivity of magnetoresistive sensors is entirely dependent on the mobility of charge carriers. Because $u_n > u_p$ only *electronic semiconductors* have an advantage. However, parabolic dependence of magnetoresistance on induction makes calibration of magnetometers (that use Gauss effect) difficult.

At present, in the nanostructured “magnetic” semiconductors, the giant and the colossal magnetoresistance effects are found.

The transversal galvanothermomagnetic effect (Ettingshausen effect) is the occurrence of temperature gradient in the direction perpendicular as to magnetic field \mathbf{B} so to current density vector \mathbf{j} , which flows toward the applied electrical field E .

$$\nabla_{\perp} T = A_{\perp} [\mathbf{j} \times \mathbf{B}],$$

where A_{\perp} is kinetic coefficient of transversal galvanothermomagnetic effect. The sign of temperature gradient $\nabla_{\perp} T$ varies with magnetic field direction, so this effect is the *odd* one.

The physical nature of the Ettingshausen effect is close to the magnetoresistance effect: various influences of Hall field and magnetic Lorentz force on the *fast* and *slow* charge carriers. For charge carriers whose velocity is greater than average velocity, the magnetic component of Lorentz force exceeds the influence of Hall transverse electrical field E_y , so “hotter” charge carriers will deviate to one of the edges of a sample. For charge carriers whose velocity is less than average velocity,

the influence of Hall field is stronger than Lorentz force, so “colder” charge carriers deviate to opposite edge.

Obviously, faster charge carriers, accumulating near one of edges, will give their energy to the crystal lattice, and that this edge becomes heated. The opposite edge that accumulates slower charge carriers cools down, as at returning to thermodynamic equilibrium the “colder” charge carriers will pick up the energy from crystal lattice. Thus, along the direction that is transverse to both magnetic and electrical fields (along Hall’s field E_y), the transverse temperature gradient $\nabla_{\perp} T$ arises.

The longitudinal galvanothermomagnetic effect (Nernst’s effect) is also associated with diverse influence of Hall’s field and Lorentz force on the charge carriers moving with different velocities.

Twisted by the magnetic field, slower charge carriers will be stronger than faster charge carriers. Due to various deviations of the “warmer” and the “colder” charge carriers, their contribution to energy (i.e., energy transferred along electrical field) in the direction x is different for opposite edges of sample. Faster (“warmer”) charge carriers are accumulated near the sample edge along x direction in which they move and heat it. Slower (“colder”) carriers, because of their further slowing by the magnetic field, will assemble on opposite faces along the x direction, causing its cooling.

Thus, along the direction of electrical field, *longitudinal temperature gradient* occurs:

$$\nabla_{\parallel} T = A_{\parallel} j_x B,$$

where A_{\parallel} is kinetic coefficient of longitudinal galvanothermomagnetic effect. The sign of longitudinal temperature gradient is not dependent on the magnetic field direction.

The compression effect (pinch effect of current channel narrowing) is the self-compression of electrical discharge, that is, the channel of *plasmic* electrical current becomes squeezed into a “cord.” This effect can be observed in conductional environment and it is due to action of own magnetic field (induced by the same current).

During plasma compression, the charge carriers form a quasistationary narrow current channel, in which repulsive pressure of plasma is balanced by contracting magnetic pressure; then usually in this case oscillations arise with final break of a current. The Pinch effect is investigated mainly while discharge occurs in the gas, but it has some importance for plasma formation in solids, particularly, in case of highly degenerated electron-hole plasma in the semiconductors, where the pinch effect can be used to study peculiarities of charge carrier transport.

Depending on current direction in the plasma column, it is possible to distinguish the *z-pinch*, when plasma compression occurs in longitudinal current in plasma with the *azimuthal magnetic field* creating, and the *θ -pinch*, when magnetic field is created by *external current* and interacts with induced currents in plasma. Compression of plasma is observed not only in cylindrical current flows, but also in thin layer configurations of flat plasma current.

The mechanism of compression effect can be considered with the example of *z-pinch*. Power lines of magnetic field, generated by current, have the form of

concentric circles, whose plane is perpendicular to the axis of current. The resultant electrodynamic force that acts on volume unit of a medium with current is radially directed to axis of current channel and causes its compression. The squeezing effect of current can be explained as a consequence of Ampere law as to *magnetic attraction* of separated *parallel filaments* with a current, in which electrical current passes in the *same direction*. According to the value of current, the magnetic pressure on movable plasma is so great that current channel starts to reduce its section, and this is the pinch effect.

In stationary condition, there should be a balance between kinetic pressure (which seeks to expand plasma “cord”) and electrodynamic force that compresses it. The equilibrium condition, in general, is:

$$-\frac{dp}{dr} = \frac{1}{c} jH,$$

where j is current density, H is magnetic field at distance r from the axis of plasma cord, p is pressure, and c is light velocity. Relationship that links current strength with average pressure in the plasma cord is

$$I^2 = 200pR^2,$$

where I is current, p is average pressure in cross section of plasma cord, and R is radius of plasma cord [8].

Therefore, the pinch effect appears in the current channel, such as a cylinder filled by the conducting material. Electrical field is applied to the opposite ends of a cylinder and acts along its axis. Magnetic field lines have the form of concentric circles, whose plane is perpendicular to the axis of the cylinder. Magnetic power is directed to the cylinder axis and tends to compress conducting medium.

The pinch effect takes place both in the solid-state plasma (assuming equal concentration of charge carriers of *opposite signs*) and in the low-temperature plasma. In semiconductors, pinch has a *magnetothermal* character. The point is that, as a result of magnetic compression and thin plasma cord appearance (which focuses almost all current), almost all power is released in the channel of a cord. When duration of current pulse is long enough, the temperature of crystal lattice in the pinch channel increases, and equilibrium concentration of plasma also increases. In case of strong heating of a lattice, the equilibrium plasma that is formed by thermal ionization plays a significant role in the overall balance among carriers. This stage of pinch effect is called as magnetothermal. In the case of very high power, the magnetothermal pinch goes into the *thermal pinch*, which has the nature of *electrical breakdown*, accompanied by the melting of crystal lattice in location of plasma cord.

As already noted, the pinch effect can occur only in the *bipolar plasma*, when in semiconductors mobile charge carriers are present with *different sign* of charge carriers (electrons from the conduction band and holes from the valence band). In the monopolar plasma, Coulomb forces of space charge prevent even low spatial redistribution of charge carriers. Main factors that prevent strong compression of electron-hole plasma are the ambipolar diffusion, as well as the recombination of

bulk carriers. For this reason, the pinch effect can occur only in the semiconductors with *very high mobility* of charge carriers and *large time* of their recombination; with the aim of pinch effect investigation and application most commonly InSb, Ge, and BiSb are used.

Thermomagnetic effects. Magnetic field can change not only electrical conductivity. As the *thermal conductivity* is also associated with the flow of charge carriers, in a magnetic field at the presence of temperature gradient some thermomagnetic phenomena should appear due to distortion of charge carrier trajectory.

In Section 8.4, some thermoelectric effects are considered (Thomson's, Seebeck's, and Peltier's effects). Therefore, next only the specifics of thermal and electrical interactions in semiconductors will be considered. Physical basis of thermomagnetic effects is the interaction of charge carriers of conductor (or semiconductor) with magnetic field when external *electrical field is not applied* to a semiconductor.

Thermomagnetic phenomena, as shown in Fig. 8.32, are the effects of transversal electrical field appearance (*Nernst-Ettingshausen* transverse effect), the longitudinal electrical field occurrence (*Nernst-Ettingshausen* longitudinal effect), the transverse temperature gradient (*Righi-Leduc* effect), and the longitudinal temperature gradient (*Mudgee-Righi-Leduc* effect) originations.

The transverse electrical field (Nernst-Ettingshausen effect) appears in semiconductor in the direction, perpendicular to both magnetic field B and temperature gradient ∇T . This effect is proportional to temperature gradient and to magnetic field B :

$$E_{\perp} = A_{\perp t} \cdot \nabla T \cdot B,$$

where $A_{\perp t}$ is kinetic coefficient of transverse effect that depends on intrinsic properties of material. In the metals and degenerated semiconductors, this effect is very small, because the value E_{\perp} strongly depends on degeneration. For this reason, investigation of transverse electrical field, originating in magnetic field and temperature gradient in sample, is used mainly during study of nondegenerated semiconductors.

Physical meaning of transverse electrical field appearance at temperature gradient and magnetic field is as follows. Thermal velocity of those charge carriers that move from the hot end is higher than velocity of charge carriers that diffuse in the opposite direction; therefore, they have different relaxation times. As magnetic

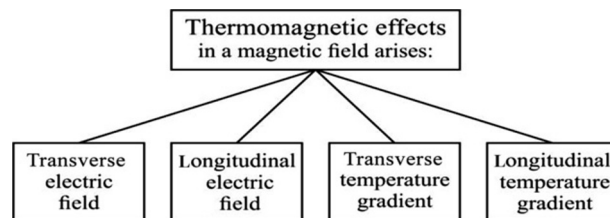


FIG. 8.32

Main thermomagnetic phenomena in semiconductors.

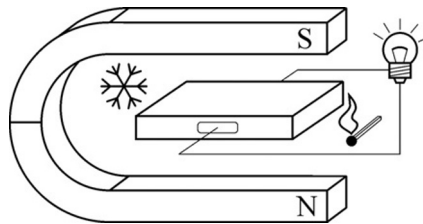


FIG. 8.33

Illustration to Nernst-Ettingshausen effect.

component of Lorentz force acts differently on fast and slow carriers, they also are bended by the magnetic field at different angles. Thus, the flow of charge carriers to the sides of a sample will not be the same, thereby transverse electrical field occurs (Fig. 8.33).

The transversal Nernst-Ettingshausen effect occurs with the same reason as Hall's effect, that is, it is a result of charged particle deviation by the Lorentz force. The difference, however, is that in case of Hall's effect the stream of particles occurs due to their drift in the *electrical* field, because of charge carrier *thermal diffusion*.

An essential difference also is the fact, that, unlike Hall's effect, the sign of E_{\perp} is *independent* of charge carrier sign. Indeed, when drifting occurs in the *electrical field*, the change in a sign of charge carriers leads to the change in the direction of a drift, which gives the change in the sign of Hall field. In case of transversal Nernst-Ettingshausen effect, the diffusion flow is always directed from the heated end of a sample to its cold end, regardless of charge carrier sign. Lorentz force direction for both positive and negative particles is mutually contradictory, but the direction of flow of electrical charge in both cases is the same.

In metals and semiconductors, Nernst's effect is caused by charge carrier *relaxation time* (determined by electron interaction with crystal lattice), as well as by charge carrier *energy* (or velocity), and, therefore, this effect is very sensitive to scattering mechanism. During the investigation of transverse effect, the nature of charge carrier mobility can be established: if mobility is known, it is possible to discover charge carrier scattering mechanism.

Relaxation time for fast electrons is greater than that for slow ones; for this reason, one edge of a sample charges negatively (i.e., transverse effect has positive sign: $E_{\perp} > 0$). This case shows that scattering of charge carriers occurs on the *acoustic phonons*. If conditions of electron movement in a sample are such that relaxation time of charge carrier *decreases* with the increase in velocity, the opposite edge of sample charges negatively, thus Nernst's effect has a negative sign ($E_{\perp} < 0$). This case is typical for the mechanism of charge carrier scattering on the *ionized impurity atoms*. It should be noted that the sign of Nernst's effect does not depend on the carrier sign, but depends on the *mechanism of their scattering*.

Nernst's effect is actively used in the investigation of Cooper pairs in *superconductors*. For example, in superconducting amorphous films $\text{Nb}_{0.15}\text{Si}_{0.85}$ through to

extremely small free path, electron contribution to Nernst's effect is negligibly small. At that, the contribution of Cooper pairs is quite significant; therefore experiment permits directly measuring their contribution. By such a way, the existence of Cooper pairs (and, correspondingly, the *local superconductivity* existence) was proved at temperatures significantly higher than the point of phase transition to superconducting state.

The longitudinal electrical field of Nernst-Ettingshausen occurs along the temperature gradient ∇T in the transverse magnetic field. However, as noted earlier, even in the absence of magnetic field ($B=0$) along longitudinal temperature gradient ∇T a thermoelectromotive field exists: $E_{B=0} = \alpha_{B=0} \nabla T$. For this reason, the additional field, occurring along longitudinal direction at magnetic field application, can be described as

$$E_{\parallel} = E_{\parallel}(B) - E_{\parallel}(0) = A_{\parallel} \cdot \alpha(0) \cdot \nabla T \cdot B^2,$$

where A_{\parallel} is kinetic coefficient of longitudinal effect, and $\alpha(0)$ is thermoelectric coefficient. In weak magnetic fields, the square dependence of E_{\parallel} on magnetic field is seen.

The physical nature of Nernst-Ettingshausen longitudinal effect is explained by the fact that magnetic field, deflecting electrons, reduces their average velocity, and, thus, reduces the energy transfer in this direction. At magnetic field absence of ($B=0$), thermoelectromotive field is determined by a difference of fast v_1 and slow v_2 electrons along temperature gradient: $v_1(0) - v_2(0)$. In the magnetic field this component changes, at that, its change is dependent on Hall's effect, and, therefore, on relaxation time τ .

Thus, in the n -type semiconductors, thermoelectric power increases, if relaxation time decreases with increasing electron energy (scattering on acoustic phonons). The peculiarity of thermoelectric power value, depending on scattering mechanism, in p -type semiconductors is same as in n -type semiconductors.

For example, if τ_2 for slow electrons is greater than for fast electrons (τ_1), then the relative velocity change

$$\frac{v_1(B)}{v_1(0)} > \frac{v_2(B)}{v_2(0)}.$$

Then $\alpha(B)$, which is determined by the difference, will be greater than $\alpha(0)$. Thermo-EMF in the magnetic field increases. If the relaxation time increases with increasing energy, $\frac{v_1(B)}{v_1(0)} < \frac{v_2(B)}{v_2(0)}$ and, consequently, $\alpha(B) < \alpha(0)$, that is, the thermo-EMF in the magnetic field decreases.

Thus, in the electronic semiconductors, thermoelectric power increases if the relaxation time decreases with increasing electron energy (scattering by acoustic phonons), and decreases if the relaxation time increases with the increase in the electron energy (scattering on the ionized atoms of the impurities). The nature of the change in magnitude of thermo-EMF in dependence on scattering mechanism for hole-type semiconductors is the same as for electron-type semiconductors.

The transversal temperature gradient $\nabla_{\perp} T$ (Righi-Leduc effect) occurs in the magnetic field in semiconductors when heat flow exists in them. The value of temperature gradient is proportional as to magnetic field so to longitudinal (“basic”) temperature gradient $\nabla_{\parallel} T$:

$$\nabla_{\perp} T = A_{\parallel t} \cdot B \cdot \nabla_{\parallel} T,$$

where $A_{\parallel t}$ is kinetic coefficient of transverse effect.

In a certain sense, this effect is the thermal analog of Hall’s effect; at that, the role of external electrical field E plays heat flow $\nabla_{\parallel} T$, directed in the same line, while instead of transversal electrical Hall field the transverse temperature gradient $\nabla_{\perp} T$ arises.

This effect, like other thermomagnetic phenomena, is due to the fact that free path of charge carriers is bent in the magnetic field by Lorentz force. In the process of their diffusion, charge carriers transfer a heat; in the absence of magnetic field, this heat flow is directed from the hot end to the cold end of a sample. When the magnetic field is switched on, the diffusion flux is bent by Lorentz force at a certain angle, and, due to this mechanism, the transverse temperature gradient occurs.

The physical nature of heat transfer is similar to the nature of electrical charge transport: it is conditioned by the fact that fast (“hotter”) charge carriers under the influence of magnetic field are deflected to one side, while the slow (“colder”) charge carriers to the opposite direction. Therefore one edge of a sample is heated, while the opposite edge of sample is cooled.

Righi-Leduc thermal transverse effect is positive in p -type semiconductors while in the n -type semiconductors it is negative.

The longitudinal temperature gradient $\nabla_{\parallel} T$ (Mudgee-Righi-Leduc effect) occurs in the magnetic field toward existing temperature gradient. This effect is the change of electronic contribution to the thermal conductivity $\Delta\xi$ due to reduction of charge carrier free path along the heat flow by twisting of their trajectories (changing velocity in the direction of temperature gradient).

When contributions into heat flow from “hotter” and “colder” charge carrier change, the temperature difference along the direction of heat flow also changes. Thus, the *additional* (to existing) temperature gradient occurs. The longitudinal effect is described as thermal conductivity ξ change due to heat transfer by electrons:

$$\Delta\xi/\xi = [\xi(0) - \xi(B)]/\xi(0).$$

Theoretical calculations show that change in thermal conductivity depends on the *square* of magnetic field and is proportional to existing temperature gradient without the magnetic field:

$$\Delta\xi = A_M \cdot B^2 \cdot \nabla T,$$

where A_M is kinetic coefficient of additional longitudinal temperature gradient. In semiconductors, this longitudinal effect is much greater than that in metals. In this way, thermal conductivity measurements in the magnetic field can separate electronic part of thermal conductivity from its phonon part.

As can be seen from these descriptions of various galvanomagnetic phenomena, these effects are very sensitive to charge carrier interaction with lattice defects; therefore, these phenomena are used to study *mechanisms of scattering* of charge carriers in semiconductors. The kinetic factors of thermomagnetic effects can be expressed through a combination of appropriate magnetoelectric coefficients, which, however, are dependent on the value of magnetic field B .

The research of thermomagnetic phenomena in semiconductors is used to determine the type of conductivity, the mobility of charge carrier, as well as to make clear mechanisms of charge carrier scattering. In the magnetic electroconductive material, at present, such research is used to clarify Neel and Curie points. All explanations of thermal conductivity change under the influence of magnetic field are based on consideration of electron flow and how this flow is deflected by the Lorentz force.

Magneto-optical effects. Usually these effects are studied and used in the *strong* magnetic fields. Effects will be considered when parallel monochromatic light falls *perpendicular to the surface* of solid; light is partially reflected and partially passes through material without absorbing. (The light that is absorbed in the semiconductor gives rise to different photovoltaic processes: internal photoelectric effect, Dember effect, and other effects discussed earlier in [Section 8.4](#).)

Magnetic field, applied to the semiconductor, gives birth to a variety optical effects ([Fig. 8.34](#)).

The photoelectromagnetic effect (Kikoin-Noskov effect) is due to the bipolar diffusion from illuminated surface of semiconductor. When an illuminating semiconductor is exposed by magnetic field directed perpendicular to the light propagation in a crystal, the *electromotive force* arises, as in case of Dember effect. Diffusive stream of excess (light-generated) electrons and holes moves deep into the semiconductor due to the gradient of concentration, and it is turned by magnetic field to opposite sides of a sample.

The photoelectromagnetic effect looks like Hall's effect, but in case of Hall's effect the magnetic field turns the flow of charge carriers caused by the external electrical field. The difference of photoelectromagnetic effect from the Hall's effect lies in the fact that electrons and holes are *single-directional diffusion fluxes*, so that magnetic field deploys them to various edges of a sample. A situation occurs when

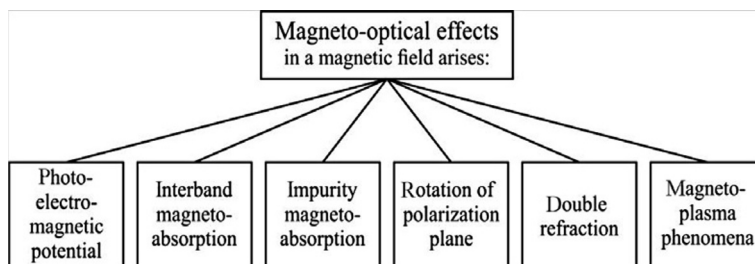


FIG. 8.34

Main magneto-optical effects in semiconductors.

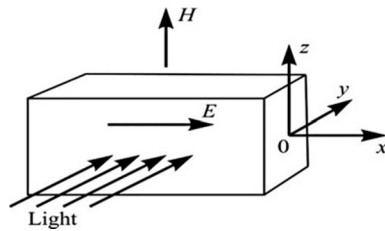


FIG. 8.35

Illustration to the photoelectromagnetic effect.

electron-hole pairs are separated not only by different diffusion velocities, but also by oppositely directed magnetic forces.

If magnetic field H is directed along the axis z (Fig. 8.35), while the light beam and the diffuse stream of charge carriers are directed along the y -axis, the magnetic field deflects electrons and holes into different directions, causing spatial charge separation. If the ends of a sample are close circuited, the current j_x appears; if these ends are open, the photo-EMF occurs.

Unlike the Dember effect, the photoelectromagnetic effect is not caused by the obligatory difference in the mobility of electrons and holes. The photoelectromagnetic effect can be observed as in case of the intrinsic absorption of light as well as in case of the impurity-type light absorption. Photomagnetic voltage is proportional to magnetic induction and to luminous flux, and inversely proportional to concentration of equilibrium carriers. This effect is observed at any value of mobility of electrons and holes, and by this property this effect differs significantly from the Dember effect. At equal conditions, the photoelectromagnetic voltage is higher in weakly doped and intrinsic semiconductors.

By investigation of photoelectromagnetic effect characteristics, it is possible to obtain information about the *band structure parameters* and the states of impurities in the semiconductors. Photoelectromagnetic effect opens the possibility to determine, first, the *lifetime* of charge carriers, and, second, the *rate of surface recombination*. Moreover, this effect can be used to study these important parameters of semiconductors even in case of short lifetimes of charge carriers.

Most of the photomagnetic effects listed in Fig. 8.34 are caused by the *quantization of energy levels* of electrons and holes in the strong magnetic field. In case of such quantization, electronic spectrum of semiconductors cannot be considered as quasicontinuous. This phenomenon is related to the *cyclotron resonance* in semiconductors. It is possible to create such conditions in a crystal that are similar to the conditions in a *cyclotron*. If one would place a crystal in the constant magnetic field and irradiate it by high-frequency electromagnetic radiation, that frequency is equal to the cyclotron frequency, and the resonance in electromagnetic radiation absorption will be observed.

From the frequency of resonant absorption, it is possible to find cyclotron frequency $\omega_c = eB/m^*$ using effective mass of charge carriers m^* . It should be noted

that mass, found from cyclotron frequency, matches the effective mass of charge carriers only for spherical surfaces of constant energy. If constant energy surface is the ellipsoid, the components of effective mass tensor in the main axes of ellipse equal to m_{xx}^* , m_{yy}^* , and m_{zz}^* . In this case, effective mass m^* , calculated from cyclotron frequency ω_c , is called the *cyclotron effective mass* that depends on the angle between magnetic field direction and the axes of constant energy of ellipsoid.

Therefore, very important information can be obtained about the form of dispersion law in the permitted bands of electron energy. The Schrödinger equation solution implies that the movement of electrons in plane, perpendicular to magnetic field, is quantified. Corresponding energy levels are *Landau levels*. The distance between Landau levels with quantum numbers n and $n + 1$ will meet the energy:

$$\Delta E = \hbar\omega_c = \hbar eB/m^*.$$

Thus electrons in the conduction band (and holes in valence band) in a strong magnetic field are not characterized by the quasicontinuous spectrum. The spectrum of permitted bands in this case is converted into *discrete* Landau levels; at that, the distance between levels is determined by magnetic induction B and by the value of effective mass m^* of charge carriers. The quantization of electronic energy in the magnetic field leads to a number of resonant and magneto-oscillation phenomena, some of which are listed in Fig. 8.34. The effects of spectrum quantization can be seen experimentally, if a peculiar condition is met:

$$\hbar\omega_c > k_B T,$$

that is, at very low temperatures and very strong magnetic fields.

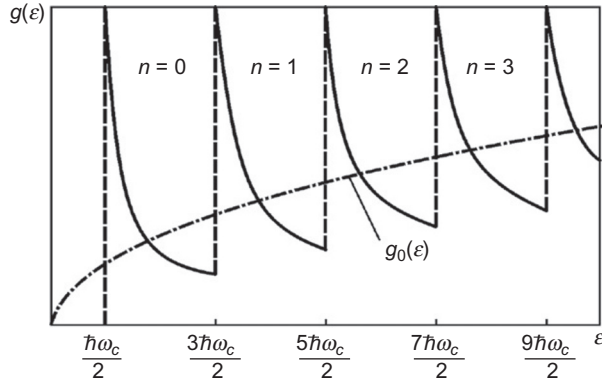
By substituting to this formula all constants given that $\omega_c = eB/m^*$, it is possible to find the magnetic induction:

$$B(\text{Gs}) = (m^*/m)10^4 T.$$

At liquid helium temperature, the effect of electronic spectrum quantization in the magnetic field can be found at $B > 40 \text{ kGs}$ if $m^* \approx m$. In case when $m^* = 0.1m$, the effect of quantization is probable in the magnetic field of $B > 4 \text{ kGs}$. It is clear that research on magneto-oscillations and resonance effects is possible only at low temperatures with the use of strong magnetic fields. Dispersion law in magnetic field is shown in Fig. 8.36.

Dispersion law looks like a number of parabolas; for the branch with quantum number $n = 0$ at energy $\hbar\omega_c/2$ sharp peak is seen above the origin $g_0(\varepsilon)$ corresponding to the bottom of the conduction band in the absence of magnetic field. The function of density of states is the sum of hyperboles, and each of them meets parabola's dispersion law. Near the size of energy that corresponds to Landau parameter $k = 0$ the density of states turns into infinity like δ -function. It can be shown that number of states in any finite energy interval is finite.

The important feature of band structure of semiconductor in the magnetic field is a shift of band extremes. The bottom of the conduction band rises on the value $\hbar\omega_c/2 = e\hbar B/(2m_{cn}^*)$ while the bottom of the valence band goes up on $\hbar\omega_c/2 = e\hbar B/(2m_{vp}^*)$.


FIG. 8.36

Function of density of states in a strong magnetic field; *bar-dashed curve* shows function of density of states in the absence of a magnetic field.

The quantization in magnetic field results in a number of *magneto-optical effects* due to the *interband impurities* and the *intra-band* optical transitions. Next the most important of magneto-optical effects are listed with their brief description.

The interband magneto-absorption is seen as interband light oscillations in the magnetic field. This effect is caused by the transitions between Landau's levels in the valence band as well as in the conduction band.

In the case when light frequency changes, the absorption coefficient at direct light transition oscillates, if the energy of photon is larger than the energy of bandgap. Maximums in the absorption spectrum correspond to the transitions between Landau levels in different areas with a selection rule $\Delta n = 0$. These oscillations were observed in germanium, indium antimonide, and other high-charge mobility semiconductors. If the oscillation peaks dependent on magnetic induction are obtained, then, by extrapolating to value $B = 0$ it is possible to find the *bandgap*. This is one of *most accurate* methods of its determination. Moreover, from the slope of straight line $\hbar\omega(B)$, the reduced *effective mass* of charge carriers can be determined:

$$(m_r^*)^{-1} = (m_n^*)^{-1} + (m_p^*)^{-1},$$

that is, if one of the masses, m_n^* or m_p^* , is known, it is possible to find another.

If the transitions are indirect in the range of interband magnetoabsorption, the number of steps is observed (that becomes eroded with increasing temperature). Indirect absorption involves phonons, so the selection rule $\Delta n = 0$ will not be mandatory.

The magnetoabsorption by impurities occurs during optical transitions of electrons and holes from the ground state to the excited state (Zeeman effect on impurity levels), as well as during the transitions between impurity states and the Landau levels in the permitted bands.

The spectrum of magnetoabsorption shows oscillations, such as during interband transitions (Fig. 8.37). From the distances between peaks in magnetoabsorption

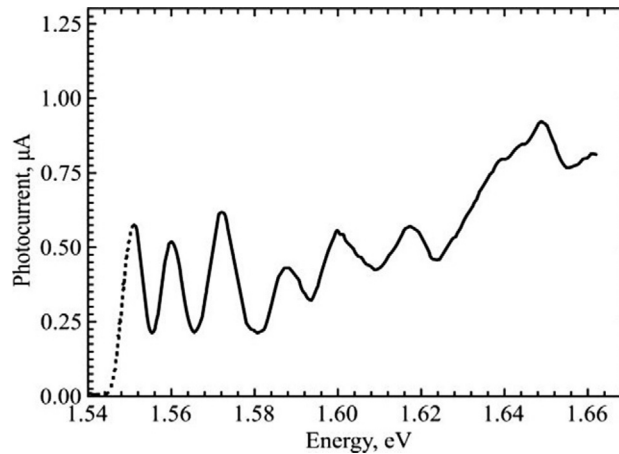


FIG. 8.37

Magnetic absorption spectrum of GaAs/AlGaAs perpendicular to electrical field (10^3 – 10^4 V/cm) at magnetic field (8T) at liquid helium temperature.

oscillation spectrum, it is possible to determine effective mass of charge carriers in the permitted bands, to which these transitions are related (at that, the interband magnetoabsorption gives only reduced effective mass).

The magneto-optical effects, which include the magnetogyration and the gyration of the plane of polarization (Faraday effect), occur on the free charge carriers (interband) and determined by the difference in the optical paths of two opposite circular polarization of electromagnetic waves [8].

The effect is observed when the plane-polarized wave passes through a crystal placed in the constant magnetic field, which is *parallel* to the direction of wave propagation. As a result, the plane of light polarization, after passing through crystal, becomes turned on the angle φ that depends on the distinction of light frequency from the cyclotron frequency ω_c and on the thickness of a crystal. Near the cyclotron frequency, the angle φ changes its sign to opposite.

When electromagnetic wave frequency is large ($\omega \gg \omega_c$), the angle φ is inversely proportional to the square of frequency and directly proportional to magnetic field: $\varphi = B/(\omega^2 m^*2)$. Experimentally obtained dependence of $\varphi(B)$ or $\varphi(\omega)$ allows to find the *effective mass* of charge carriers.

Recently, Faraday effect has been registered in multilayer graphene (Fig. 8.38). It is assumed that the rotation angle will be about 0.01 radians, but in fact it was found to be 0.1 radians (about 6°). The magnitude of angle in terms of a single layer of atoms shows that graphene is ahead of all its “opponents.”

Both Faraday effect and related magneto-optical Kerr effect are widely used in the optical communications, data storage devices, and computing systems. Discovered peculiarities of graphene make it possible to create unique devices. In practice, however, fairly large angles of rotation of polarization plane (45°) are required, the

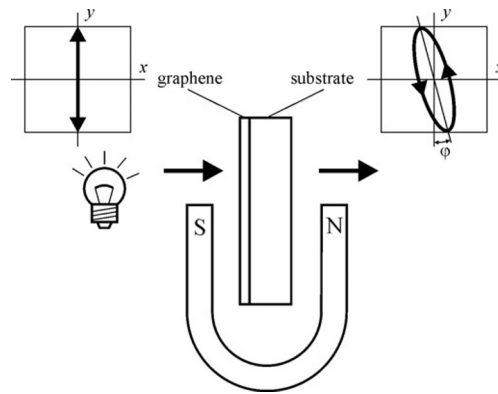


FIG. 8.38

Faraday effect in graphene.

implementation of which requires about 10 layers of graphene. However, this material absorbs infrared radiation leading to weakening in signal of devices.

The birefringence (*double refraction*) that occurs in thermomagnetic field (Voigt effect) can be also considered as the effect related to Faraday effect on free charge carriers. This effect arises at such circumstances when the magnetic field is *perpendicular* to the direction of light propagation.

After passing through the semiconductor, placed in the magnetic field, the *linearly polarized* light turns into the *elliptically polarized* light. At that, the phase shift θ appears between components of electrical vector E_{\parallel} and E_{\perp} , which can be measured. Comparing angle φ (plane of polarization rotation in Faraday effect) and angle θ (in Voigt effect), it is possible to find the value $\varphi/\theta = \omega_c/m^*$. Thus, the study of Faraday effect and Voigt effect permits direct determination of charge carrier *effective mass*.

Faraday and Voigt effects can occur not only at intraband transitions, but also during the interband transitions. The interpretation of experimental results in this case is more complicated than in case of free charge carrier intraband transition. In case of interband magneto-absorption (at interband Faraday effect), observed oscillations are dependent on photon energy at constant magnetic field. The spectrum of the magnetoabsorption for indirect transitions is stepped, while in case of Faraday effect oscillations look as a series of highs, providing advantage to accuracy of measuring.

The magnetoplasma phenomena are seen in the crystal, placed in constant magnetic field when there is an interaction of light with crystal at the frequency close to *plasma frequency*. The manifestation of this class of phenomena is different—depending on the method of experiment.

Basic methodology of experiment is *reflected light* monitoring near the plasma frequency of a crystal placed in the magnetic field, and to measure the *rotation* of light polarization plane.

While exploring plasma reflection spectra without magnetic field, it is possible to obtain the ratio of charge carrier concentration to effective mass n/m^* , if the plasma frequency ω_{pl} and dielectric permittivity are known. The error in the effective mass m^* determination in this case largely depends not on the accuracy of plasma resonance measurement, but on the accuracy of Hall measurements (to determine carrier concentration n) and on the accuracy of permittivity measurement. If plasma reflection is studied at constant magnetic field, the effective mass can be determined directly by the shift of plasma minimum without knowledge of charge carrier concentration.

The magnetoplasma resonance method can be used as a contactless rapid method for determining concentration and mobility of electrons in the *thin film samples* and in *mesostructures* deposited on high-resistive substrates. Generally, the magnetoplasma effect in semiconductors is experimentally investigated in the infrared region of spectrum when $\omega\tau \gg 1$. The frequency of plasma resonance $\omega_{pl}^2 = ne^2(m^*\epsilon_0\epsilon_r)^{-1}$ is determined using concentration of electrons n and does not depend on the size of a sample. Measured frequency dependence of magnetoplasma reflection in the magnetic field (that is directed perpendicular to plane of sample) permits determining as the effective mass m^* so the relaxation time τ of electrons.

In experiments with ultra-high frequency (UHF), the dimensional resonances (*helicon*) can be seen that arises at the circumstances when thickness of a sample is equal to integer half-waves. Measurement of wavelength at resonance is used to determine charge carrier concentration. At that, the thickness of a sample should be much greater than the depth of skin layer at zero magnetic field.

Optical phenomena in the crossed electrical and magnetic fields. At these conditions, oscillations of light absorption for *interband transitions* can be observed. In case of allowed direct transitions, the highs of optical absorption oscillations in magnetic field are shifted to lower energies, when perpendicular electrical field is applied.

When studying magnetoabsorption and oscillations in the crossed fields, it is possible to find the sum ($m_n^* + m_p^*$), while by observing the oscillations in magnetic field it is possible to find reduced effective mass m_r^* . Thus the joint study of these effects allows the *direct determination* of m_n^* and m_p^* . Note that such experiments should be provided at the conditions of not too large magnetic fields.

Different electrical properties at oscillations, on their physical nature, are adjacent to the magneto-optical phenomena in strong magnetic fields, because these effects are caused by the quantized states of charge carriers in the bands conditioned by strong magnetic field:

- oscillations of *magnetic susceptibility* (de Haas-Van Alphen effect);
- oscillations in *light transparency* in crystals (Shubnikov-de Haas effect);
- oscillations of *ultrasound absorption* in crystals placed in magnetic field;
- *thermoelectromotive power* oscillations in magnetic field.

Oscillations of magnetic susceptibility and thermoelectric power are caused by the *change of thermodynamic potential* in the magnetic field. Oscillations in light

transmission while magnetic field changing are observed in degenerated semiconductors and caused by *sharp change in the density* of states in permitted bands while the Landau quantum level passes through the Fermi level. Observations of oscillations can give information on the dynamic properties of charge carriers in the vicinity of the Fermi level, that is, to restore the shape of Fermi surface in metals and degenerated semiconductors.

8.7 NANOSCALE AND QUANTUM-DIMENSIONAL EFFECTS

The dependence of properties of a solid on its size is observed in many cases. For example, in piezoelectrics the frequency of piezoelectric resonance and magnitude of dielectric constant depend on the size of a sample. Similarly, the magnetic permeability of ferromagnetic films and the dielectric permittivity of ferroelectric films strongly depend on their thickness.

As for the *electrical conductivity* of low-dimensional nanosized materials, in addition to quantization effect of electronic energy spectrum, the wave properties of particles begin to get affected. The *coherence length* of electronic wave in a solid at normal temperature has a magnitude of several nanometers. Therefore, at distances of 1–10 nm, the wave properties of electrons begin to be detected. This is expressed by the fact that when a substance is taken in small quantities, it cannot always be clearly attributed to isolators, conductors, or semiconductors. For example, some chemical elements taken in an amount of, say, 20, 50, and 100 atoms will consistently pass the stages of isolator, semiconductor, and conductor, respectively [9].

Dimensional effect occurs if the length of a body at least in one dimension becomes comparable with some critical magnitude l_{cr} . For classical dimensional effects l_{cr} is the classical value, such as the diffusion length, the length of the free run of electrons, etc. During charge carrier movement or oscillation in the semiconductor nanostructures, the *wave nature* of electrons and the discreteness of energy spectrum vividly manifest themselves. There is a *quantum mode of transfer* of charge carriers, in contrast to the classical regime when free path of electrons is much smaller than the size of a system and electrons are regarded as classical charged particles.

This section deals with the basic fundamental physical phenomena that determine transfer of charge carriers in the *nanosized* structures. As the size of nanostructures is comparable to free path of electrons, charge carriers can freely pass through nanostructure without scattering on defects, impurities, phonons, etc. In such structures, the phase of noninteracting electron waves is maintained throughout the path and therefore effects of phase interference are observed.

Potential wells, barriers, and tunneling. To analyze the properties of quantum systems, their *energy models* should be used. Nanoparticles in their discrete spectrum can be compared with atoms. Based on known energy distribution, the method of *potential curves* allows determining dynamic parameters of moving particle [10].

According to classical mechanics, the particle can overcome the potential barrier only if its energy *exceeds* the potential energy on the top of a barrier. The quantum particle behaves completely differently: the difference from zero *probability* exists to penetrate the barrier, even if energy of quantum particle is less than the height of potential barrier. Such quantum-mechanical effect is the *isoenergetic tunneling*.

From the point of view of classical physics, in case of a tunneling effect, the law of energy conservation is violated. But for quantum physics of the tunneling effect, it is possible to overcome the “deficit” of particle energy over a *short time* $\Delta t = \hbar/(2\Delta E)$ in accordance with the uncertainty relation. Therefore, in terms of quantum physics, if during this time the particle can tunnel through the barrier, then law of conservation of energy is not violated. The point to be noted is that after the tunneling the particle *retains* its energy. The tunneling effect lies at the basis of tunnel diode operation and is widely used in other nanoelectronic devices.

Another feature of a quantum particle behavior in the potential well of nanosize and atomic size is the *discreteness of energy spectrum*. Discreteness of energy spectrum of electrons is the basis of functioning of many nanoelectronic structures. The configurations of real potential wells and barriers in the “microworld” depend on the geometric features of fields that form of these wells.

Behavior of microparticle in potential well can be determined by solving Schrödinger equation under corresponding initial conditions. This is a differential equation: for its solution it is necessary to know how potential energy U of microparticles depends on coordinates, that is, the need to set the function $U(x, y, z)$. Here and thereafter, it will be assumed that this function does not depend on time (since microparticle is in the stationary state).

To find out main features of microparticle behavior in the potential well, it is enough to consider the 1D case $U(x)$ and the potential well of *rectangular shape* (Fig. 8.39A). The width of potential well is denoted by l , and its depth is U^0 . In this

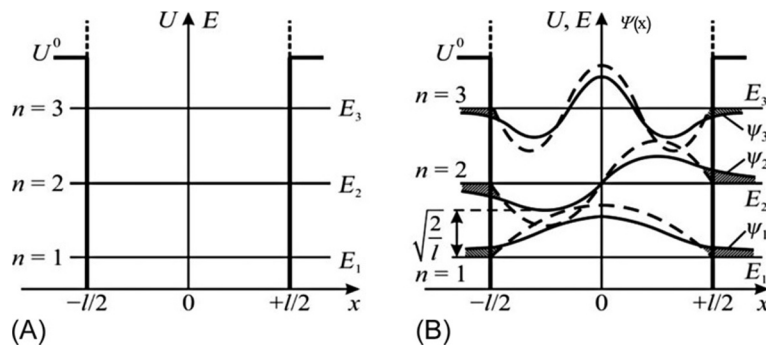


FIG. 8.39

One-dimensional potential well: (A) energy spectrum of microparticle in potential well; (B) wave functions $\Psi(x)$ for three states of microparticles ($n=1, 2, 3$) in potential well (*dotted line* shows distribution function in case of infinitely deep well).

model, the width of barriers that limits potential well to the right and to the left is infinite. The depth of well U^0 can vary from a finite value to infinity (if the well would be considered as indefinitely deep).

The solution of Schrödinger equation (result of which is graphically illustrated in Fig. 8.39B) enables determining the *energy spectrum* of microparticle, that is, a complete set of its energy values E and the wave function $\Psi(x)$, whose module square $|\Psi(x)|^2$ is the probability density of finding microparticle at point x .

It can also be shown that on the width of well l it is possible to put the integer number of de Broglie half-waves: $l \approx n(\lambda/2)$. For the infinitely deep well, the wave function is expressed in the terms of trigonometric functions: through cosines in case of odd n ($n = 1, 3, 5$) and sinus in case of even n (these solutions in Fig. 8.39B are shown by the dashed curves). It can be seen from the figure that in this case the amplitude of de Broglie wave at points $x = \pm l/2$ vanishes. Therefore when $U^0 \rightarrow \infty$, the microparticle can neither penetrate inside the barrier nor go beyond the boundary of the well [11].

However, if the depth of well is *finite*, then the *amplitude* of de Broglie wave at the points $x = \pm l/2$ *does not vanish* for any n and has *continuation* beyond the boundary of barrier.

This important result is shown in Fig. 8.39B by solid curves in the *shaded areas*. That is, microparticle in the potential well of ultimate depth U^0 can *penetrate* beyond the boundary of barrier for energies E less than U^0 , which contradicts the law of energy conservation of classical physics and never observed in the “macroworld.”

In addition, Fig. 8.40, which shows the probability density for three different values of n , implies that the *energy spectrum of quantum particle is discrete*, with

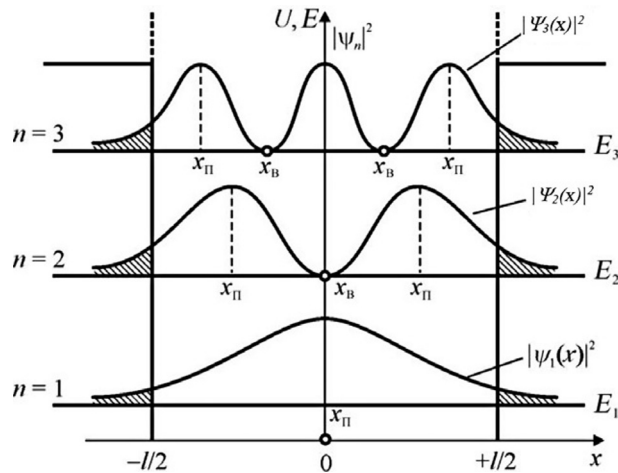


FIG. 8.40

Probability density $|\Psi(x)|^2$ to find microparticle at different points x of potential well of ultimate depth.

its minimal energy not equal to zero ($E_1 > 0$). This energy is called *zero energy* and usually is denoted as E_0 . Zero energy is *peculiar to any quantum systems*: physical vacuum, quarks in hadrons, nucleons in atomic nuclei, electrons in atoms, and atoms in molecules and crystals.

The probability to find microparticle at different points *inside* the well, as seen from Fig. 8.40, greatly varies. There are some points, where the probability of “visiting” by particle shows the maximum: they are called the *parochial* (x_p). There are also points in which the particle is never seen, these are *nodes* (x_n). Such behavior is not a characteristic of macroparticles [10].

It should be noted that the probability of penetration by microparticle beyond the boundary of barrier ($x = \pm 1/2$) is not equal zero, but it only gradually decreases with the increasing distance from boundaries of barrier (shaded areas in Fig. 8.40).

The most important result is that, if the barrier width is not infinite, then probability of its penetration by microparticle beyond boundary of barrier (tunneling effect) is different from zero.

Quantum dimensional effects in the electronic structures are observed when the role of critical length l_{cr} is played by electron de Broglie wavelength (λ_e) when the size of a structure at least in one dimension has the order of λ_e . Thus, the quantum-dimensional effects are conditioned by wave nature of electrons. In the nanosized regions, the behavior of electrons is determined by the reflection of electronic waves from the boundaries of such areas, as well as the interference of electronic waves or the passage of waves through potential barriers. The quantization of the energy of electrons might be *spatially limited* in their displacements, or the passage of electrons through nanometer *dielectric layers* is quantized, or the electrical resistance of *nano-wires* has quantum peculiarities, and others.

De Broglie wavelength for electron that has effective mass m^* and moves in crystal with velocity v has following meanings: $\lambda_e = \hbar/p = \hbar/m^*v$, where $p = m^*v$ is the impulse. Thus, the greater the de Broglie wavelength, the less effective the electron mass. In most metals, it is almost same as in vacuum: $m_{ef} \sim m_e$ so that de Broglie wavelength is relatively small. However, in some semiconductors, the effective mass of electron can vary widely. At room temperatures (~ 300 K), in silicon $m_{ef} = 0.9m_e$ and $\lambda_e = 8$ nm, in gallium arsenide $m_{ef} = 0.07m_e$ and $\lambda_e = 30$ nm. Thus in semiconductors the size of de Broglie wave has a nanoscale, and hence the quantum-dimensional effects are technologically easier to implement in semiconductors.

The density of states $g(E)$ determines the *number of quantum states* of electrons per unit of volume (or area, or length, depending on dimension of object) referenced to single energy interval. According to this definition, the density of states is equal to $dn(E)$, that is, the number of states in the energy ranges from E to $E + dE$:

$$g(E) = dn(E)/dE.$$

The knowledge of density of states $g(E)$ and probability of their filling by electrons $w(E)$ allows determining the distribution of electrons of a system in quantum states and to describe many electrical, optical, and some other properties of a system. The electrons are characterized by a half-integer spin; therefore the probability of their

filling quantum states is determined by statistics of Fermi-Dirac, which is based on Pauli principle.

The energy spectrum of electrons $E(k)$ as the density of their quantum states are most important characteristics of a quantum object, which determine its electronic properties and the response to external influences [11].

The energy spectrum characterizes the dependence of energy of a particle on its impulse: $E(p)$ or, equivalently, the energy dependence on wave vector $E(k)$, as the impulse and wave vector are bound by simple relation $p = \hbar k$ (module of wave vector is determined by inverse wavelength: $k = 2\pi/\lambda$). The energy spectrum represents a set of possible values of particle's energy in given conditions. If energy is quantized, then energy spectrum is called the *discrete* (quantum), but if energy can accept continuous series of values, then the spectrum is called the *continuous*.

The "infinite" crystal quantum properties are characterized by 3D long-range ordering of atoms. From the point of view of nanophysics, even a crystal having 1-micron size already looks "infinite," because in all directions its regular crystalline lattice consists of thousands of atoms being many times larger than the de Broglie wave.

Metals and semiconductor crystals are filled with 3D electronic gas, in which electrons can move freely in any direction. The energy of electron in the 3D gas is:

$$E = \frac{(m_{ef}v)^2}{2} = \frac{p^2}{2m_{ef}} = \frac{(p_x^2 + p_y^2 + p_z^2)}{2m_{ef}} = \frac{\hbar^2(k_x^2 + k_y^2 + k_z^2)}{2m_{ef}},$$

where p and k are quasi-impulse and quasi-vector of electron, respectively, and m_{ef} is electron's effective mass (for simplicity it is considered as isotropic). The spectral dependence of $E(k)$ looks like *quasicontinuous*, as the energy levels are as much as electrons held in the 3D crystal (10^{14} – 10^{22} cm⁻³).

In Fig. 8.41B, the dependence of energy on wave vector components near the bottom of conduction band is given, and in Fig. 8.41C, the density of quantum states of electrons $g(E)$ for unbounded (3D) crystal is described by parabola: $g(E) \sim E^{1/2}$.

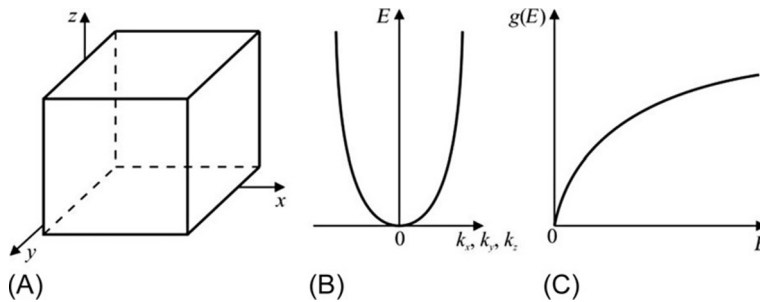


FIG. 8.41

Electronic spectrum of unlimited size crystal: (A) model of sample; (B) dependence of electron energy on quasivector components; (C) dependence of quantum density of states $g(E)$ on electron energy E .

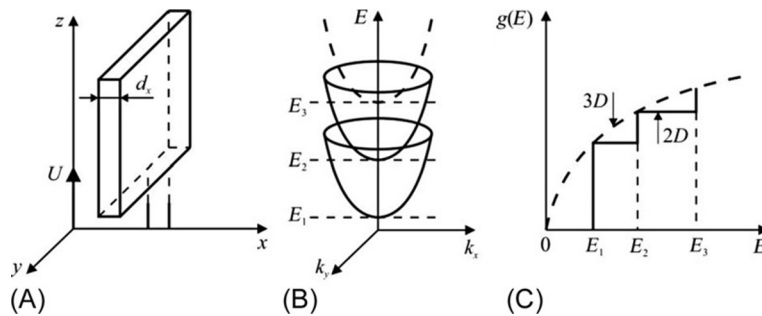


FIG. 8.42

Two-dimensional (2D) nano-object (thin layer): (A) schematic representation of thin layer limited along x -axis; (B) potential holes for electrons in this layer; (C) dependence of state density on energy in case of infinitely deep quantum well.

Near the bottom of the conduction band, the density of states is small, but it gradually increases with the increase in E , reaching saturation. From Fig. 8.41B, it is evident that in limits of the same band the functions $E(k)$ and $g(E)$ are quasicontinuous, and therefore properties of 3D crystals under the influence of fields change continuously.

The quantum well corresponds to two-dimensional (2D) object. Usually this is a thin layer of crystal whose thickness d is commensurate with de Broglie wavelength ($d \sim \lambda$). The system of electrons in such a layer is called the 2D electronic gas. Fragment of such a layer is presented in Fig. 8.42A. The motion of electrons in this layer is limited by the segment d_x in the direction x but is not limited in the directions y and z .

Moving in the x direction, the electron is not able to leave thin layer, as its output work (equal, e.g., in aluminum arsenide ~ 4.5 eV) is much greater than thermal energy (0.026 eV at room temperature). Therefore, the motion in x direction is legitimately regarded as a movement in the 1D deep enough rectangular potential well with the width d_x (Fig. 8.42A). The energy of such motion is quantized and characterized by quantum number $n = 1, 2, 3, \dots$

If the quantum well is infinitely deep, then on its width d_x the integer quantity of half-wave $\lambda_n/2$ will be spaced. This means that only those states of electron motion would be stationary, which will correspond to standing wave formed by de Broglie waves falling and reflected from the walls of a well (dotted curves in Fig. 8.42). The values of E_n are called *quantum-dimensional levels* [9].

The motion energy along the y and z is not quantized and is determined by the same expressions as for free particle (or for volume). Therefore the total electron energy, taking into account that effective mass of electron m_{ef} is the same for movements in all directions, can be represented as

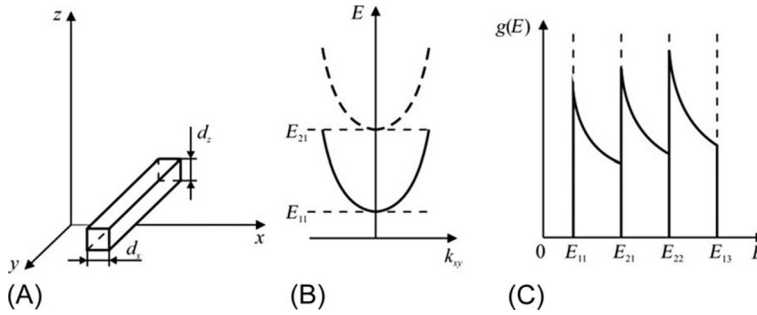


FIG. 8.43

One-dimensional (1D) nano-object (quantum wire): (A) model representation; (B) energy dependence on wave vector; (C) density of states dependence on energy.

$$E = \frac{\hbar^2 (k_y^2 + k_z^2)}{2m_{ef}} + E_n = \frac{\hbar^2 (k_y^2 + k_z^2)}{2m_{ef}} + \frac{\hbar^2 \pi^2 n^2}{2m_{ef} d_x^2},$$

where the quantum number $n = 1, 2, 3, \dots$

Consequently, the energy spectrum of electron in a quantum well of 2D nano-object will be discrete-continuous. Each dimensional level of E_n corresponds to a set of possible energy values due to free motion of electron along the axes y and z . This set of energies E is the 2D sub-bands of dimensional quantization. The dependences $E(k_y, k_z)$ and $g(E)$ are shown in Fig. 8.43B and C.

The graph of $E(k_y, k_z)$ dependences is the paraboloid system; the bottom of n th paraboloid corresponds to the level $E = E_n$. The dependence of $g(E)$ is straightforward. Each dimensional sub-band makes equal contribution $m_{ef}/\pi\hbar^2$ to the density of states. Strictly speaking, the model of infinitely deep rectangular potential well is valid only for the motion of electron in the “isolated” thin film with nanosized thickness d . Such a separate model—flat-parallel film of nanometer thickness—is difficult to implement. In practice, in the nanosized structures such nanolayers are formed *inside* the crystal so that charge carrier movement is limited in one dimension; therefore it is possible to assume that these carriers are in the 1D potential well. An example is the electrons in the nanometer layer of narrowband semiconductor between two layers of broadband semiconductor.

Practical examples of quantum wells with 2D electronic gas may be conductive channels in the unipolar transistors (metallic oxide-semiconductor structures on silicon) and narrowband layers in heterostructures from compounds $A^{III}B^V$ (required for injection lasers). Systems of closely spaced equilibrium quantum wells, which make electronic tunneling possible, form the superlattices that are the heterostructures made of two or more different materials. In these heterostructures, an important role belongs to the transition layer, that is, the boundary between two materials.

All elemental materials of which semiconductor heterostructures are made (Zn, Cd, Hg, Al, Ga, In, Si, Ge, P, As, Sb, S, Se, Te, Ti, Ti) are located in the central part of the

periodic table of elements and belong to groups II–VI. In the middle, there is *silicon*, which, in the technology of electronic materials, occupies the same important place as the steel in manufacture of structural materials. In addition to silicon, electronics often deal with semiconductor $A^{III}B^V$ compounds and their solid solutions, as well as with $A^{II}B^{VI}$ compounds. Of the compounds of the type $A^{III}B^V$, the most often used is GaAs (gallium arsenide) that forms solid solutions $Al_xGa_{1-x}As$. The use of such solid solutions allows creation of heterostructures with a continuous (rather than abrupt) change in the relative content of elements of group III. In such heterostructures, the width of the bandgap also changes continuously.

For the manufacture of heterostructures, it is very important to reconcile the parameters of crystalline lattices of two contacting materials. If two materials that have different spacings of crystal lattices, they are grown on one another, then, with increasing thickness of layers, large mechanical stresses and deformations can occur in the area of the interface that can cause occurrence of microcracks in layers. Mechanical stresses and deformations appear irrespective of whether the transition between two layers is smooth or not. In order to reduce deformations, the permanent spacings of two materials must differ at least. That is why in the application of heterostructures, the solid solutions of AlAs-GaAs system are most often used, because aluminum and gallium arsenides have almost identical lattice parameters. In this case, GaAs in a form of single crystals is the ideal substrate for heterostructure growth. Another natural substrate is the indium phosphide, InP, used in combination with solid solutions of GaAs-InAs, AlAs-AlSb, and other semiconductors of $A^{III}B^V$ type.

The progress in the creation of thin-layer heterostructures is due to the emergence of practical technologies for the growth of thin layers by methods of molecular beam epitaxy (MBE), metal-organic vapor phase epitaxy (MOVPE), and liquid phase epitaxy. These methods provide opportunity to grow heterostructures with a very sharp boundary [3].

The quantum wire (quantum thread) is a 1D nano-object. The motion of the electrons is limited along the axes x and z by sizes d_x and d_z , respectively, but not limited along the axis y (Fig. 8.43). Square cross section of quantum wire is only a comfortable model for calculation but does not violate the generality of reasoning. It is important that potential well for free electrons in a quantum wire (thread) is 2D.

The motion of electron is associated with the axis y but its energy must be quantized in 1D potential wells d_x and d_z . The total electron energy is

$$E = \frac{\hbar^2 k_y^2}{2m_{ef}} + E_{mn} = \frac{\hbar^2 k_y^2}{2m_{ef}} + \frac{\hbar^2 \pi^2 n^2}{2m_{ef} d_x^2} + \frac{\hbar^2 \pi^2 m^2}{2m_{ef} d_z^2},$$

where $m, n = 1, 2, 3, \dots$ and E_{mn} is the energy of dimensional levels. The position of each of them depends on two quantum numbers m and n and on the quantities d_x and d_z while in the direction of wire (y) spectrum is continuous. The conduction band in the quantum wire is divided into 1D sub-bands (Fig. 8.43B). The density of states per unit length $g(E)$ has a series of sharp peaks (Fig. 8.43C) corresponding to

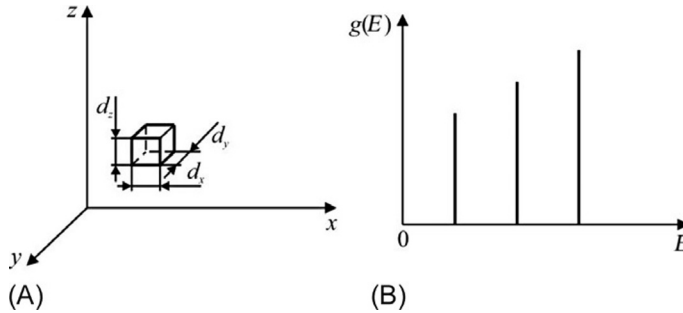


FIG. 8.44

Zero-dimensional (0D) nano-object (quantum dot): (A) model of quantum dot; (B) density of states dependence on energy.

dimensional levels. This means that most electrons in sub-bands have energy near corresponding dimensional level.

Most methods for producing quantum wires are based on the fact that in a system with 2D electronic gas (usually on basis of heterostructure) by one or another technological way the motion of electrons in one of directions is limited. The conductivity of quantum wires can have important features. The most interesting case is *short wire* with a length that is less than the length of free run of electrons (the latter is determined by scattering on impurities and defects of wire). In this case, the electron, having flown from one contact of a quantum wire, reaches to another contact without collisions, similar to a projectile released from a cannon. Such analogy has led to the consideration of structures in which there are no collisions of electrons, which are often referred to as *ballistic* ones.

The quantum dot is zero-dimensional (0D) nano-object, where the motion of electrons is limited in all three dimensions: x , y , z . Fig. 8.44A shows a convenient model for calculations; in fact, the shape of quantum dot is usually different from cubic.

Potential well for a quantum dot is 3D. The energy of free electrons should be quantized for movements in all three dimensions. The energy spectrum of electrons in quantum dot is quite discrete, as in a separate atom. Energy is defined by the expression:

$$E = \frac{\hbar^2 \pi^2 l^2}{2m_{ef} d_x^2} + \frac{\hbar^2 \pi^2 n^2}{2m_{ef} d_y^2} + \frac{\hbar^2 \pi^2 m^2}{2m_{ef} d_z^2},$$

where $l, m, n = 1, 2, 3, \dots$, as well as d_x, d_y, d_z are dimensions of the dot in three directions. The energy spectrum of electrons consists of separate dimensional levels of E_{lmn} , which resembles the spectrum of isolated atom. The energy E_{lmn} depends on three quantum numbers l, m, n and sizes d_x, d_y, d_z . The graph of the density of states $g(E)$ for the quantum dot has a type of delta function: $g(E) = \infty$, if $E = E_{lmn}$

(E coincides with dimensional level) and $g(E)=0$, if $E \neq E_{lmm}$ (i.e., E lies in the gap between dimensional levels).

The nanocrystals can serve as an example of quantum dots, grown on the surface of epitaxial layer of another material. In the Introduction to this book, one quantum dot is shown (Fig. 1.5), the islet from Ge atom is grown on Si surface while its image is obtained by atomic-force microscopy [10].

Historically, first studied quantum dots were microcrystals of cadmium selenide (CdSe). Electrons in such microcrystal are in 3D potential well so that they have several stationary levels of energy with a characteristic distance between them (exact expression for energy levels depends on the shape of the quantum dot). While electrons make transitions between the energy levels of a quantum dot, the photons can emit, similar to transition of electron between the energy levels of an atom. It is possible also to arouse (“throw”) the electron into the highest energy level and obtain the radiation from transition between levels (luminescence). In this case, unlike actual atoms, the frequencies of transitions are easy to control, changing size of a microcrystal. Once again note that observations of CdSe nanoclusters (“microcrystals”) luminescence served as the first opening of quantum dots.

The length of optical wave in CdSe clusters (and fluorescence color) depends on the size of these clusters. It is determined that fluorescence of CdSe nanoclusters corresponds: for size of ~ 400 nm—purple; ~ 450 nm—blue; ~ 500 nm—green; ~ 600 nm—yellow; and ~ 700 nm—red. Observed blue shift of absorption band begins with nanoparticles of 10^{-12} nm in size. In the semiconductor clusters absorption of photons generates electron-hole pairs, while their recombination is accompanied by fluorescence. In other words, the decrease in the size of semiconductor nanoparticles is accompanied by a shift in the absorption band in the high-frequency region.

At present, many experiments are devoted to quantum dots formed in 2D electronic gas. In 2D electronic gas, electron motion perpendicular to the plane is limited; therefore, the region on plane can be isolated by means of gate metallic electrodes superimposed on heterostructure. Quantum dots in the 2D electronic gas can be connected by the tunnel contacts to other regions of the 2D gas, and electrical conductivity through a quantum dot can be studied.

Ballistic conductivity of nanoscale conductors. The conductance (G) of ordinary wire with a circular cross section equals $G=1/R=\sigma S/L$, where $S=\pi r^2$ is cross-sectional area, L is length of wire, r is its radius, and σ is specific electrical conductivity. This formula is valid, if r and L are much greater than the mean free path of electrons λ_e . In these conditions, the motion of electron in the conductor has a diffusive character, and trajectory of its motion is the broken line (Fig. 8.45A).

If $\lambda_B > L$ and $\lambda_e > r$ (Fig. 8.45B), then the electron flies from one contact to another practically without collision with atoms of crystalline lattices. Such a mode of movement is called *ballistic*. The moving electron does not feel resistance to its motion in conductor’s volume [9].

In Fig. 8.46A, ballistic conductor 3 with quantum-sized diameter is placed between two metal contacts 1 and 2. Suppose that temperature is of the order of

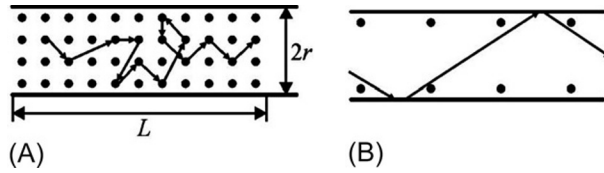


FIG. 8.45

Schematic representation of diffusion-type (A) and ballistic-type (B) electron motion in a conductor.

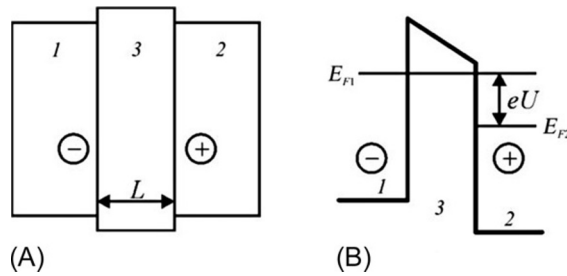


FIG. 8.46

Spatial (A) and energy (B) presentation of ballistic conductor (3) and contacts to it.

several degrees Kelvin and all electrons in contacts on energy diagram of Fig. 8.46B are located below Fermi levels E_{F1} and E_{F2} . If the potential U is applied between the contacts, then energy levels of metal 2 are reduced by a value of eU with respect to levels of metal 1. In this case $E_{F1} - E_{F2} = eU$. The current can be generated only by electrons that have energy in the range from E_{F1} to E_{F2} . It is only these electrons that from contact 1 that have the ability to move to free levels of contact 2.

If conductor (or semiconductor) 3 is characterized by N -dimensional subzones, and it is ballistic (i.e., it has quantum-dimensional section S), then its electrons are located in the dimensional subzones. Only those electrons of subregions can participate in the transfer of a current that are located in the range from E_{F1} to E_{F2} . In this case, the conductance is $G = 2e^2N/h$, while the resistance is $R = h/2e^2N$.

Thus, unlike the classical wire, the resistance of ballistic quantum wire *does not depend on its length* L . The number N is determined by a distance between subzones, and this distance increases when the intersection of wire S decreases. If one gradually reduces the diameter of a wire, then from the interval $E_{F1} - E_{F2}$ will in turn, one after another, the dimensional sub-band deviate. With the departure of each of sub-band, the conductance G shows a jump decrease by $2e^2/h$. When there is not a single sub-band in the interval $E_{F1} - E_{F2}$, the conductivity G will vanish.

The quantity $2e^2/h$ is called the *conduction quantum* while reciprocal value $h/2e^2 = 12.9\text{k}\Omega$ is the *quantum of resistance*. In fact, the quantization of resistance is conditioned by dimensional quantization of energy. To monitor the effect of

quantization of resistance, the sufficiently low temperatures (~ 1 K) are required. At higher temperatures, jumps of conductance G become blurred or disappear because of thermal motion in the contacts it “throws” electrons at the level where $E > E_F$. It should be noted that resistance measured in these conditions is the resistance in contacts. In the ballistic nanowire, there is no scattering of electrons. Consequently, it should not have any electrical resistance.

It should be noted that functioning of many instrumental structures of nanoelectronics is determined by the features described earlier of energy spectra of quantum-dimensional elements. In this case, it is important that quantization of energy is observed only when the size of objects has the order of de Broglie wave (at least in one dimension).

Resonance tunneling. Usual tunneling effect, as already noted earlier, is micro-particle passage through a potential barrier, whose height U^0 is greater than the energy of a moving particle. At the same time, not every time does the particle with this energy pass through the barrier. There is a certain probability of its passage through the barrier, called the *coefficient of transparency*.

The greater the magnitude of transparency coefficient, the smaller the width of the barrier l and the difference between its height U^0 and energy of particle E , that is, the smaller “deficit” of particle energy inside the barrier; $U^0 - E$. In a nanosized structure, this effect can be manifested, for example, in the passage of electrons through thin layers of dielectric.

The tunneling time is very small: $\hbar/(U^0 - E) \approx 10^{-15}$ s, which can be estimated using the uncertainty relation: $\Delta E \cdot \Delta t \geq \hbar/2$. Tunneling effect has essential probability, if the barrier width l is comparable to de Broglie wavelength of electron. It defines the boundaries of functioning elements in the integrated circuits, constructed on the basis of traditional principles. However, if tunneling effect is considered on the basis of principle of device, it can increase its speed to hundreds of terahertz. For example, some single-electron devices work only on this principle.

Important features show the so-called *resonance tunneling effect*, which manifests itself in the two- or multidirectional periodic structure (Fig. 8.47A), and looks like the sharp increase in probability of particle passing through barriers, if its energy coincides with any dimensional energy level of potential well that separates the barrier (Fig 8.47C).

Resonance tunneling through a series of barriers only occurs if the width of wells and barriers is of the order of de Broglie wavelength. In this effect, the time of electron passage of structure includes, in addition to time of tunneling, the time of electron life in the well, that is, the time τ of its life on the resonant level. For example, according to estimations, for double heterostructure consisting of layers $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ (5 nm)- $\text{Ga}_{0.7}\text{As}$ (7 nm)- $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ (5 nm) at the barrier height of 0.2 eV the time is $\tau \sim 10^{-11}$ s. Thus, tunneling time provides the operation of corresponding devices in the terahertz range. It should be noted that the value of τ decreases also with further reduction in structure size [12].

Two-barrier structures are of great interest to electronics, because they can operate at ultrahigh-frequency (microwave) devices in the range of hundreds of gigahertz

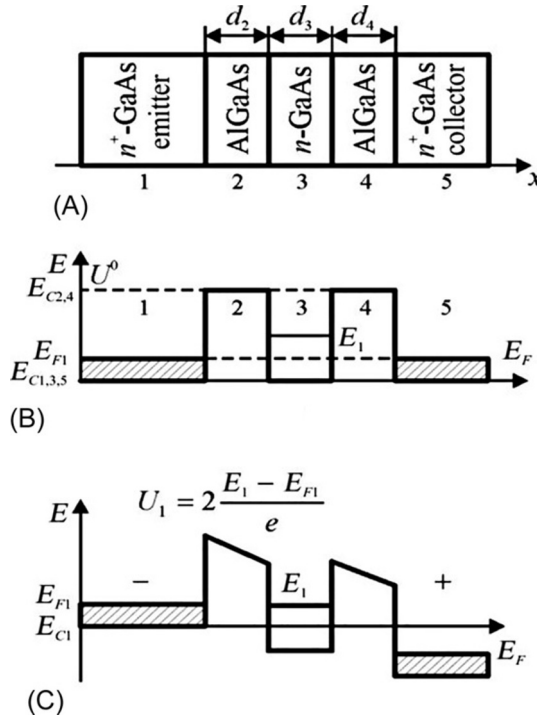


FIG. 8.47

Structure (A) and energy diagrams (B, C) resonance tunnel diode.

(GHz) and switches with a delay time less than 1 ps (picoseconds). These devices are created on the basis of two-barrier structure—resonant-tunnel diode and transistor. Moreover, multibarrier structures, called superlattices, are developed and used widely.

Superlattices. As is known, there are various possibilities for creating p - n junction, but they all were implemented in the same semiconductor, doped in different ways—acceptors and donors. In the light of subsequent presentation, such junction can be called the homojunction, for example, the p Si- n Si.

The following *heterojunctions* occur at the contact of two different chemical components of semiconductors. In this contact not only the bandgap changes, but also other fundamental properties change: band structure, effective masses of charge carriers, their mobility, and physical-chemical and optical properties.

Heterojunctions can be sharp and smooth. In a sharp heterojunction, the change in chemical composition occurs at a distance of the order of lattice constant. In an ideal heterojunction, there are no defects and boundary states at interface. The possibility of obtaining monocrystalline heterojunctions, that is, the close contacts of various chemical compositions of semiconductors carried out in single crystal is associated with development of technological methods of epitaxial growth of semiconductor

crystals, that is, the formation of uniformly oriented layers of crystals of one substance on the surface of another substance.

The heterostructure is the combination of several heterojunctions in a single-crystal structure that forms part of a semiconductor device. In the case of many layers of alternating crystals, the planar periodic structure is formed, which is called the *superlattice*. The characteristic dimensions of the layers in heterostructures and superlattices are nanometers, and therefore corresponding semiconductor devices belong to relatively new generation of electronic devices—nanoelectronics [11].

Superlattices are crystalline structures in which, apart from the periodic potential of crystalline lattices, there is *another periodic potential*, the period of which far exceeds the constant lattice, but corresponds to the nanoscale.

The most widely used are the *semiconductor* superlattices. They consist of layers of two semiconductors, differing in either chemical composition or type of conductivity. Superlattices are made, for example, with the help of MBE technology, which allows build-up layers of any compound and thickness. Period of repetition of layers lies from several nanometers to tens of nanometers (for comparison, lattice constant of crystals Si and GaAs is about 0.5 nm).

Two types of semiconductor superlattices are widely used: the *composite* and the *doped* (it is appropriate to note that there are also superlattices made of metals, superconductors, and dielectrics).

The composite superlattices are heterostructures made of *different* chemical composition layers having different widths of bandgap, but with the close values in the magnitude of lattice parameters. For example, composite superlattices are $\text{Al}_x\text{Ga}_{1-x}\text{As-GaAs}$, $\text{In}_x\text{Ga}_{1-x}\text{As-GaAs}$, $\text{In}_x\text{Ga}_{1-x}\text{As-InP}$; ZnS-ZnSe; and many others. Additional periodic potential in them is created by the periodical changing of the width of bandgaps.

The doped superlattice is a periodic sequence of layers of the *n*- and *p*-type of *same semiconductor*. Donor atoms in *n*-layers supplies electrons, which bind to acceptor atoms in the *p*-layers. Embedded in the crystal lattice charges of ionized acceptors and ionized donors create obviously the need for superlattice *additional periodic potential*.

This potential crucially changes the band structure of semiconductor on the basis of which superlattice was created. Therefore superlattice can be considered as a new, *synthesized* semiconductor that does not exist in nature and has unusual properties. The choice of composition materials for layers can broadly vary the band structure of the superlattice. The combination of methods for obtaining materials with a modified zone structure is at the basis of the so-called *band engineering*.

Energy diagram of superlattice (its potential profile) is given as an example in Fig. 8.48A for the composite superconductor $\text{Al}_x\text{Ga}_{1-x}\text{As-GaAs}$ in the direction perpendicular to layers. Due to periodic change in the bandgap $\Delta E_e = E_{C1,2} - E_{V1,2}$ a sequence of rectangular quantum wells separated by the barriers is created. The wells are formed in a narrowband semiconductor: for electrons, in the conduction band, and for holes, in the valence band. It should be noted that there are superlattices with much more complex profile, for example, in the structures $\text{Ge}_x\text{Si}_{1-x}\text{-Si}$ or GaAs-Ga.

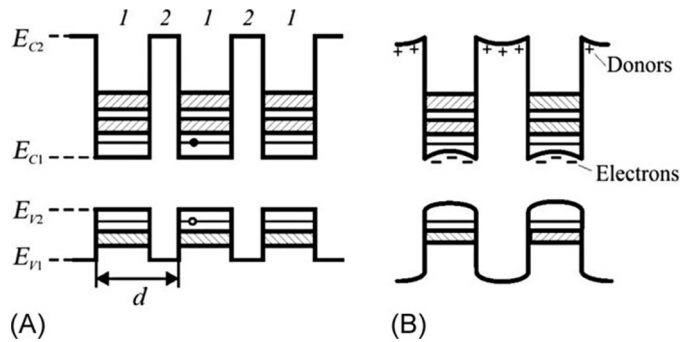


FIG. 8.48

Energy charts of simple composite (A) and modulated-doped (B) superlattices: d —period of superlattice.

In Fig. 8.48B, the potential profile of modulated-doped superlattice is shown. In this case, the donor impurity is implemented only in the wideband material. Electrons from donor levels pass into the quantum wells, spatially separating by ionized donors. The alternation of charges causes periodic bends of the edges of bands. Fig. 8.48A and B shows the minibands on which the valence band and the conduction band are divided [12].

Fig. 8.49 shows potential profile of doped superlattice. The charges of ionized donors and acceptors create a sequence of potential wells for electrons and holes. The electrons and holes are *spatially separated*: the holes are located in the potential wells of valence band of p -layer, while the electrons are in the potential wells of conduction band of n -layer. Bindings show miniband; E_g is the width of bandgap of output semiconductor, and ΔE_{ef} is the effective width of superlattice bandgap. To obtain doped superlattices GaAs is often used.

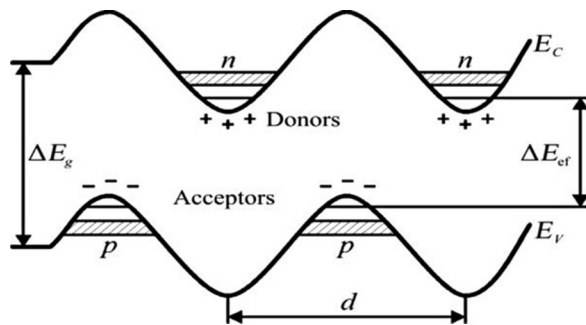


FIG. 8.49

Energy charts of doped superlattice: ΔE_{ef} —effective width of suppressed band gap; d —its period.

To construct energy spectrum of electrons in the superlattice, Schrödinger equation can be solved (as in case of a bulk crystal) by taking into account the additional periodic potential. Using the results of band structure calculations, it is possible to make qualitative conclusions about energy structure of superlattice. Its potential is periodic, so the dimensional levels are *split into bands*. Spectrum has band character—in the band as many levels as the wells in structure. As the period of superlattice is much larger than the usual lattice constant, the superlattice bands represent a finer splitting of energy band of ordinary semiconductor that are called the *minibands*. The splitting of the conduction band and the valence band into minibands is shown by the hatching in Figs. 8.48 and 8.49.

The smaller the width of the well, the greater the distance between minibands and the more effective the bandgap ΔE_{ef} , and the smaller the width of the barrier the wider the miniband. Thus one can rebuild the energy spectrum of superlattice by a simple change in the thickness of layers, which is easy to accomplish in the MBE method. The graph of states density $g(E)$ has a stepped form, like for a quantum well, but with another form of steps.

Electronic gas in superlattices differs in the fact that electrons and holes are spatially separated. Generated by light electron-hole pairs (equilibrium carriers) become also spatially separated, which prevents their recombination and increases their lifetime to $\sim 10^{-3}$ s.

In modulated-doped composite superlattice, a wideband semiconductor (e.g., AlGaAs) is doped with donor impurity. The electrons from donor levels of barrier pass into the wells in the conduction band of the narrowband semiconductor (e.g., GaAs, Fig. 8.48B). Ionized donors (impurity centers) remain in the barriers while 2D electronic gas with high density and mobility of electrons is formed in the wells. High mobility is because the density of electrons in the narrow-gap layer is greater than the density of centers of dispersion, and donor impurity centers are located in wide-gap layers.

The effect of increasing mobility is especially significant at low temperatures, when the main contribution to electron dissipation makes their scattering on impurities. High-mobility electrons allow creating high-speed devices on superlattices, such as transistors with high-conductive channels. The switching time of such transistors can be picoseconds.

Thus semiconductor superlattices are solid-state structures in which, besides 3D periodic potential of crystalline lattices, there is additional 1D potential, the period of which substantially exceeds lattice constant. The presence of such a potential substantially changes energy spectrum so that superlattices have a number of interesting properties that are absent in ordinary semiconductors.

Superlattices represent a unique opportunity to almost randomly modify their band structure. The features of the superlattice luminescence (possibility of rearranging the emitted wavelengths, the excitonic nature of radiation up to room temperature, the strong limitation of impurity capture, the femtosecond kinetics, etc.) are used to create a new generation of light-emitting devices. Acoustic properties of superlattices are characterized by the presence of selective reflection of phonons.

Semiconductor superlattices are characterized by essentially nonlinear transport properties due to the presence of very narrow minibands in their energy spectrum.

8.8 SUMMARY

1. Semiconductors are materials whose conductivity occupies intermediate position between conductors and insulators; semiconductors are different from conductors by strong dependence of conductivity on impurity concentration, temperature, and different types of radiation. The main property of these materials is *exponential increase* in electrical conductivity with temperature rise. Near absolute zero semiconductors are close to insulators.
2. Semiconductors are crystals with a *bandgap* in electronic spectrum, which is in the range of 0.1–2.5 electron-volts. For example, gallium arsenide can be grouped under wide-gap semiconductors, while indium arsenide is a narrow-gap semiconductor. Among semiconductors there are some chemical elements (germanium, silicon, selenium, tellurium, arsenic, etc.), several alloys, and compounds. Almost all inorganic materials of the surrounding world are semiconductors. In nature, the most common semiconductor is silicon that occupies almost 30 percent of the earth's crust.
3. Semiconductors have conducting as well as dielectric properties. In semiconductor crystals, atoms are usually joined by *covalent bonds* (i.e., pair of electrons bounded with two atoms); these electrons require a certain level of internal energy to release from atom that characterize the difference between semiconductors and dielectrics. This energy can be applied by energy fluctuations in crystal (room temperature thermal energy level is 0.026 eV).
4. The analysis of Schrodinger equation for electrons in crystal (Bloch theorem) shows that electronic wave function depends on the wave vector k that module has a dimension of inverse length; by this way, the quasi-impulse $p = \hbar k$ can be introduced in a consideration. This concept is very useful for examining many problems in electronic theory of solids.
5. In crystalline semiconductors, spatial atomic structure has long-range ordering, that is, the position of individual atoms (or groups of atoms) is repeated periodically within volume of crystal. Accordingly, the *potential field* is periodically changed with distance, divisible to period of structure. The compact recording of this condition is: $U(r) = U(r + na)$, where a is period of structure and n is integer. In the simplest case—chain of identical atoms—this period coincides with the distance between atoms.
6. Wave functions of free electrons are also periodical functions of potential field—there are plane waves whose amplitude is modulated with potential period (so-called *Bloch waves*). The minimal possible length of electronic

wave is $2a$, the maximal is $2L$, where a is lattice parameter and L is length of linear chain of atoms. Thus the length of electronic waves is located in the range: $2a \leq \lambda \leq 2L$.

7. The length of electronic waves in a crystal is *discrete*, and it can vary with multiplier of $2a$. Wave clearness means that wave vector $k = 2\pi/\lambda$ and impulse $p = \hbar k = h/\lambda$ are *discrete* quantities. Wave vector is within $-\pi/a \leq k \leq +\pi/a$; at that, signs “-” and “+” take into account that two identical waves can exist that move in opposite directions. If one expands the main range of wave vector by adding left and right sections of π/a , then new values of k do not arise, because all of them are included in the main range. In this regard, the main range of wave vector ($-\pi/a \leq k \leq +\pi/a$) is *coerced zone of wave vectors*.
8. In 3D space, the coerced zone occupies reduced volume, and, unlike actual physical space, the space of impulses has dimension of *inverse length*. This space is called the *reciprocal lattice*. Geometric configuration of coerced zone and all reciprocal lattices is uniquely determined by the direct lattice structure, which defines spatial position of atoms in crystal. The number of wave vector values within the zone is defined by the number of atoms N and their valence, that is, each valence state of each atom contributes one permitted value to the wave vector. For group IV semiconductors (Ge, Si), this number is $\sim 10^{22} \text{ cm}^{-3}$.
9. The energy of electrons in the permitted bands is single-valued function of wave vector and impulse. Taking into account the discontinuity of wave vectors and impulses, as well as a huge number of allowed states, it can be argued that energy of electrons is unambiguous, a quasicontinuous function of impulses, and wave vectors. Electronic energy spectrum of crystals, that is, distribution of electrons on energies in permitted bands, is usually described in the impulse space, that is, in reciprocal lattice. Dispersion law $E(p)$ for free electrons in allowed energy bands is quite different from dependence $E(p)$ for electrons moving in vacuum.
10. Inertial property of electron is characterized by its mass, but accelerating force finding for electrons in the *periodic structure* of crystal is rather difficult task; therefore, the difference is compensated by replacement of real mass m to *effective mass* m^* (effective mass is inertial mass of electron moving in potential field of a crystal).
11. As temperature of semiconductor rises, the number of free electrons and holes increases, and, therefore, resistivity of semiconductor (that contains no impurities) decreases. It is assumed that semiconductors are crystals in which electronic bandgap energy is less than 2–3 eV. The electron-hole conduction mechanism is found in the *intrinsic* semiconductors (i.e., without impurities). This conductivity of semiconductors is also called as intrinsic.

12. The *hole* appears as a space in the electronic shell of atom, when one of bonds between electrons and atomic core becomes broken; this causes the transition of bonding electron from the neighbor atom to the atom with empty bond. To the atom, from which electron goes away, another electron jumps from the neighbor atom, and so on (moving through covalent bonds in lattice). Thus this is the movement of “positive charge” without moving of host atom. Such positive charge is the hole. Usually hole mobility in semiconductors is lower than electron mobility.
13. Some impurities in semiconductors (*donors*) can easily release electrons while other impurities (acceptors) can capture electrons from atoms of semiconductor. Introduction of such impurities, even in small quantities, in a large degree increases the number of free carriers, and, thus, increases the conductivity of semiconductor.
14. In semiconductors under the influence of thermal fluctuations free charge carriers are *generated continuously*, producing electrons and holes that determine conductivity. Also electrons and holes *recombine* continuously. Under the influence of these two processes (generation and recombination), the semiconductor acquires the *equilibrium concentration* of free charge carriers. The value of equilibrium concentration of electrons and holes in semiconductor depends on temperature and on the type and concentration of introduced impurities.
15. Various external influences, such as light illumination that excites photoelectric effect, can essentially increase the concentration of electrons and holes. However, if the impact of foreign influences (light, radiation, etc.) stops, the excess concentration (imbalanced) carriers will decrease rather rapidly, tending to the equilibrium value (due to *recombination*).
16. The rate of recombination of excess charge carriers characterizes the *lifetime* τ of charge carriers. The value of τ can be controlled technologically; at that, lifetime can be reduced 1000 times by imbedding special impurities in the semiconductor that leads to appearance of deep energy levels in the forbidden band. However, these deep levels increase not only the rate of recombination, but the rate of generation of charge carriers as well.
17. Free charge carriers are always in a state of chaotic random movement that occurs with high velocity. Under normal conditions, average velocity of this chaotic traffic is 10^5 – 10^6 m/s. If electrical field is applied to this chaotic movement, superimposed directed drift of charge carriers occurs. The average *drift velocity* in a rather weak electrical field is proportional to the field. In the strong electrical field, drift velocity of charge carriers becomes saturated: drift velocity cannot surpass chaotic velocity of thermal motion.
18. If concentration of free charge carriers in the semiconductor is distributed heterogeneously, the diffusion appears: the flow of charge carriers is from the region with higher concentration to the region of lower concentration.

19. Near the border that divides crystal and environment, a certain force exists (work function) that prevents electrons to leave crystal and fly out. The energy that electron requires to overcome effect of work function is approximately a few electron-volts. In semiconductors, the value of work function depends on the type of semiconductor and on the level of its doping. Also, it depends on the spectrum of surface states, that is, the surface energy levels that always exist on the surface of the semiconductor. The negative surface charge increases the level of work function, the positive surface charge reduces it.
20. The width of potential barriers in a semiconductor depends on the level of doping. For low-doped semiconductors potential barrier equals from hundreds to thousands of atomic layers (tens of micrometers), while in heavily doped semiconductors this width is only several atomic layers (thousandths parts of a micrometer).
21. Between two parts of semiconductor crystal, one of which is doped by donors and another by acceptors, a potential barrier appears: the p - n junction. In the absence of external voltage, the height of this barrier approximately equals to the bandgap E_g of semiconductor. External electrical bias field when “+” is applied to p -region of junction and “-” to n -region (reverse voltage) increases the height of a barrier. In this case through the p - n junction only very small “reverse” current can flow. Therefore, the diode with p - n junction at inverse bias shows very high resistance. In case of direct bias field (“+” of external voltage is applied to n -region while “-” is applied to p -region of diode), the height of barrier is reduced. The density of “direct” current through the diode increases sharply with increasing voltage, and can reach very high values.
22. The *light absorption*—a decrease in intensity of optical radiation during its passing through crystal and interacting with crystal—causes light energy transformation into other forms of energy. Absorption coefficient does not depend on light intensity, but it is different for various wavelengths. Based on exponential law of light intensity attenuation in depth of sample, the indicator of photon absorption can be interpreted by parameter α while value α^{-1} is the *middle free path* of photon in crystal.
23. The *internal photoelectric effect* is due to the fact that in case of semiconductor surface illumination the number of generated free electron-hole pairs greatly increases. Increased concentrations of electrons and holes, in turn, lead to conductivity increase. The *photovoltaic effect* (Dember effect) is the bipolar diffusion of charge carriers that generate the EMF conditioned by different mobility of electrons and holes in case of light absorption.
24. Mechanisms of light absorption vary in different spectral intervals. The *intrinsic* (or fundamental) absorption is due to electronic transitions between allowed energy bands. Absorption by *free charge carriers* is conditioned by electron (or hole) transitions between allowed bands (or sub-bands). The

doping absorption is due to electron (or hole)-allowed transitions between bands and impurity levels in the forbidden band. Light absorption by impurities can be explained as electron (or hole) absorption at their transitions between impurity states in the forbidden band. The absorption by *excitons* is conditioned by the appearance of electron-hole pairs that makes significant contribution *near* fundamental absorption edge, because energy of exciton state is smaller. Absorption of light by crystal lattice in semiconductors can also lead to the *absorption by phonons*. The *plasma absorption* is the light wave energy absorption by electron-hole plasma that results in plasma transition on the higher quantum state.

25. The combined effect of electrical and magnetic fields in the conductors and semiconductors leads to some *galvanomagnetic effects*, at which the difference in electrical potentials occurs, or temperature change in electrical and thermal conductivities is observed.
26. *Hall's effect* is occurrence of transversal potential when placing conductor (or semiconductor) in the crossed electrical and magnetic fields. This effect is widely used for measuring magnetic fields, as well as for determining charge carrier *concentration* in semiconductors and metals.
27. Electrical resistance of conductor or semiconductor changes in magnetic field: this is the *magnetoresistance* (Gauss effect). The physical cause of magnetoresistance is the change of charge carrier velocity in longitudinal direction (along current flow). Based on the magnetoresistance effect, many types of magnetic field sensors are elaborated.
28. In a semiconductor placed in the magnetic field, the following are also observed:
 - *temperature gradient* occurrence in the direction perpendicular to magnetic field B and to current density vector j : this is transverse *galvanothermomagnetic* effect that can be explained by combined influence of Hall magnetic field and Lorentz force influence on fast and slow charge carriers;
 - *longitudinal galvanothermomagnetic* effect caused by a diverse influence of Hall magnetic field and Lorentz force on charge carriers, moving with different velocities.
29. The *pinch effect* is a self-compression of electrical discharge that can be seen in the electroconducting environment, and is conditioned by the action of its own (generated by same current) magnetic field. This effect is the only characteristic of such conductive environments in which mobile charge carriers (electrons and holes in semiconductors) are presented in approximately *equal* quantities.
30. The *thermomagnetic* effects (being conditional by electronic thermoconductivity) are a flow of charge carriers caused by temperature gradient. Magnetic field bends the trajectories of carriers, prompting electrical and thermal gradients.

31. The electrical thermomagnetic effects include:
- *transverse electrical field* in a magnetic field that arises in case of temperature gradient presence (this effect is used for investigation of nondegenerated semiconductors);
 - *longitudinal electrical field* that arises because a magnetic field, deflecting moving electrons, reduces their average velocity and thus reduces the energy transfer in this direction;
 - *transverse temperature gradient* that occurs in a magnetic field in a conductor, in which there is the heat flow; while magnetic field is switched on, the flux diffusion is deflected by Lorentz force at some angle; this effect is a thermal analog to Hall's effect;
 - *longitudinal temperature gradient* that occurs in magnetic field toward existing temperature gradient because charge carriers along diffuse heat flow twist their trajectory on some angle and change charge carrier velocity in the direction of temperature gradient.
32. In semiconductors, illuminating by light, in the *strong magnetic field* some effects occur:
- *photoelectromagnetic effect* that is conditioned by bipolar diffusion from illuminating surface of semiconductor; generated by light diffuse streams of excess electrons and holes are turned by magnetic field to opposite sides of sample (similar to Hall's effect);
 - *interband magnetoabsorption* that occurs due to light intrinsic absorption in magnetic field (these are transitions between Landau's levels in the valence band or in the conduction band);
 - *magnetoabsorption by dopings* that occur during optical transitions of electrons and holes from the ground state to the excited state, as well as during transitions between impurity states and Landau levels in permitted bands;
 - *turning of polarization plane* (Faraday magneto-optical effect) by free charge carriers and interband transitions, determined by a difference in optical paths of two opposite circular polarization of electromagnetic waves;
 - *birefringence*—Voigt effect, related to Faraday effect of free charge carriers; the research of Faraday and Voigt effects directly determines effective mass of charge carriers.
33. If an illuminating semiconductor is placed in constant magnetic field, the interaction of light with crystal at the frequency close to plasma frequency produces various *magnetoplasma phenomena*.
34. In magnetic field the oscillations of different electrical properties can appear in semiconductors: oscillations of magnetic susceptibility, light propagation, ultrasound absorption, and thermoelectromotive power.

35. The *dimensional effects* occur, if the body length at least in one dimension becomes comparable with some critical magnitude l_{cr} . For classical dimensional effects, the l_{cr} is a classical value, such as the diffusion length, the length of the free run of electrons, etc. However, during charge carrier movement in *nanostructure semiconductors* the wave nature of electrons and discreteness of energy spectrum vividly manifest themselves. The *quantum mode* of charge carrier transfer occurs; in contrast to classical regime, in nanostructures free path of electrons is much smaller than the size of a system, and electrons are regarded as classical charged particles.
36. The *quantum-dimensional effects* in the electronic structures are observed when the role of critical length l_{cr} is played by de Broglie wavelength for electrons λ_e , at that, the size of structure at least in one dimension has the order of λ_e . Quantum-dimensional effects are due to *wave nature* of electrons. In the nanosized regions, the behavior of electrons is determined by the reflection of electronic waves from the boundaries of such areas, as well as by interference of electronic waves or by passage of waves through potential barriers. The quantization of electrons energy is *spatially limited* their displacements, as well as the passage of electrons through nanometer dielectric layers and quantization of electrical resistance of nanowires, and others.
37. The *quantum well* corresponds to 2D object. Usually this is a thin layer of crystal, whose thickness d is commensurate with de Broglie wavelength ($d \sim \lambda$). The system of electrons in such a layer is called the 2D electronic gas. The *quantum wire* (quantum thread) is 1D nano-object. Motion of the electrons is limited along two axes, but not limited along one wire axis. The potential well for electrons in the quantum wire (thread) is 2D. The *quantum dot* is 0D nano-object, where motion of electrons is limited in all three dimensions.
38. The *ballistic conductivity* in the nanoscale conductors gives a chance to find the quantity of *quantum of conduction* ($2e^2/h$) and its reciprocal value the *quantum of resistance*: $h/2e^2 = 12.9 \text{ k}\Omega$. The quantization of resistance is conditioned by dimensional quantization of energy, but to monitor the effect of quantization of resistance sufficiently low temperatures ($\sim 1 \text{ K}$) are required.
39. The *resonance tunneling* is characterized by very small time, $\hbar/(U^0 - E) \approx 10^{-15} \text{ s}$, which can be estimated using uncertainty relation: $\Delta E \cdot \Delta t \geq \hbar/2$. Tunneling effect has essential probability, if barrier width l is comparable to *de Broglie wavelength* of electron. It defines the boundaries of functioning of elements of integrated circuits, constructed on the basis of traditional principles. The tunneling effect now is the basic principle of many devices: it can increase their speed to the terahertz region. For example, only on this principle some single electron devices can work.

40. The *superlattices* are crystalline structures in which, apart from the periodic potential of crystalline lattices, there is *another periodic potential*, the period of which far exceeds lattice constant, but corresponds to the nanoscale dimensions.
41. The *semiconductor superlattices* are solid-state structures in which, except for the 3D periodic potential of crystalline lattice, there is additional 1D *potential*, the period of which substantially exceeds lattice constant. The presence of such potential essentially changes energy spectrum so that superlattices show many interesting properties that are absent in the ordinary semiconductors.
42. Superlattices represent the *unique opportunity* to almost randomly *modify their band structure*. The features of superlattices luminescence are used to create a new generation of light-emitting devices. Semiconductor superlattices are characterized by essentially nonlinear transport properties due to the presence of very narrow minibands in their energy spectrum.

REFERENCES

- [1] K.V. Shalimova, *Physics of Semiconductors*, Energoatomizdat, Moscow, 1985.
- [2] V.I. Ilchenko, Y.M. Poplavko, *Physics of Semiconductors*, Avers (Ukraine), Kiev, 2010.
- [3] R. Waser (Ed.), *Nanoelectronics and Information Technology*, Wiley, Weinheim, 2005.
- [4] C. Kittel, *Introduction to Solid State Physics*, fifth ed., John Wiley, New York, 1976.
- [5] P.S. Kireev, *Physics of Semiconductors*, Higher School, Moscow, 1975.
- [6] K. Zeeger, *Semiconductor Physics: An Introduction*, Springer, New York, 1973.
- [7] Y.M. Poplavko, A.V. Borisov, Y.I. Yakimenko, *Microelectronics and Nanoelectronics*, Kiev Polytechnic Inst, Kiev, 2010.
- [8] V.V. Gorbachev, L.G. Spitsyna, *Physics of Semiconductors and Metals*, Metallurgy, Moscow, 1982.
- [9] D.M. Zayachuk, *Nanotechnology and Nanostructure*, Lvivska polytechnika, Lvov, 2009.
- [10] A.F. Kravchenko, V.N. Ovsyuk, *Electronic Processes in Solid-state Systems of Reduced Dimensionality*, Novosibirsk University, Russian Federation, 2000.
- [11] H.S. Nalwa, *Nanostructured Materials and Nanotechnology*, Academic Press, San Diego, 2002.
- [12] Y.M. Poplavko, A.V. Borisov, Y.I. Yakimenko, *Nanophysics, Nanomaterials, Nanoelectronics*, Kiev Polytechnic Inst, Kiev, 2012.

Polar dielectrics in electronics

CONTENTS

9.1 Simplified Description of Primary Effects	511
9.2 Piezoelectric Effect	515
9.3 Inverse Piezoelectric Effect	527
9.4 Electromechanical Coupling in Piezoelectric	533
9.5 Electrostriction	537
9.6 Pyroelectrics and Electrets	544
9.7 High-Permittivity Dielectrics and Paelectrics	557
9.8 Ferroelectrics and Antiferroelectrics	566
9.9 Ferrielectrics and Ferroelastics	575
9.10 Nonlinearity of Ferroelectrics and Paelectrics	580
9.11 Different Effects Interdependence in Polar Crystals	585
9.12 Summary	595
References	599

The term “insulator,” which signifies a substance that practically cannot conduct direct electrical current, has previously been considered as a synonym to term “dielectric.” In present-day electronics, in addition to good insulation, other properties of solid dielectrics have gained importance, namely, those that are used for the conversion of energy or information. Among such *functional dielectrics* are, for example, piezoelectrics that convert mechanical energy into electrical energy and vice versa, which are widely used in electronic devices. Another example is pyroelectrics that convert heat energy into electricity energy, which are used in sensitive radiation detectors, thermal vision devices, and so on [1].

Nonlinear properties of ferroelectrics and paelectrics, external electrical field induced by electrets, high optical activity of liquid crystals enable the application of such active dielectrics for modulation; detection; amplification; registering, storing, and displaying; and other types of electrical and optical conversion of signals carrying information. In view of a possibility to use some dielectrics as the “active” (converting) elements in electronics, one should identify and describe their properties considering not only their exclusive electrical characteristics but also their capability to manifest various electrical, optical, mechanical, and thermal effects. These materials are important also for the miniaturization of microwave and telecommunication equipment. For these reasons, regarding materials science, the electronic industry shows considerable interest in ferroelectrics, paelectrics,

piezoelectrics, and pyroelectrics, precisely because of their new applications in instrumentation engineering and electronics, as well as owing to significant progress in the field of modern microelectronic and nanoelectronic technologies [2].

Polar dielectric materials, applied in the electronic equipment, sometimes are referred to as “smart” or “adaptive” materials in literature, whereas Russian-language literature prefers the term “active” dielectrics. These materials are particularly relevant to modern and future instrumentation based on micromachining. In this trend, group technology of microelectronics is used for a variety of technical fields. Based on modern equipment, micromachining is organically connected with microelectronics and nanoelectronics.

Among contemporary applications of active dielectrics, the following areas of particular relevance should be noted [3]:

- ferroelectric and paraelectric thin films, integrated with semiconductors;
- microsystems that combine sensors, processors, and actuators;
- microwave microelectronics, based on active dielectric components;
- nanodielectrics that have some perspectives for sensors and memories.

Once again, dielectrics are termed as active if they can convert energy or information. Active (or adaptive, or controlled) dielectrics may easily react to the changes in temperature, pressure, mechanical stress, electrical and magnetic fields, light illumination, and even smell. Active dielectrics can be classified as ferroelectrics, piezoelectrics, electrets, quantum electronics components, superionic conductors, and others.

Electromechanical and electrothermal properties have been given priority for the application of active dielectrics. In particular, mechanical properties such as elasticity, which defines the practical use of crystals in piezoelectronics, acoustoelectronics, acousto-optics, and so on, are focused. The discrete structure of crystals can be overlooked with regard to its elastic properties; hence, the crystal is considered as a continuous homogeneous medium (continuum approximation). This approach is justified to frequencies below 10^{12} Hz, which is much greater than the frequency of operation of conventional electronic devices (up to 10^{11} Hz).

The most important scientific and technical fields of piezoelectric effect application are [4] as follows:

- (1) *piezoelectronics* (piezotechnique of bulk acoustic waves) are used in the development of piezoelectric receivers, piezoelectric transformers, and piezoelectric motors, phone cards, adaptors, microphones, piezoelectric resonators, and piezoelectric filters;
- (2) *acoustoelectronics* (piezotechnique of surface waves) are applied in microelectronic data converters: delay lines, filters, sensors of external influences, convolvers, and so on;
- (3) *acousto-optics* involves the interaction of optical waves with acoustic waves and are used in developing deflectors, optical filters, and other optical devices.

The progress in these areas of technology depends on the search for more effective materials on the development of new technologies and on methods of improvement of piezoelectric properties.

9.1 SIMPLIFIED DESCRIPTION OF PRIMARY EFFECTS

The classification of basic physical effects that are manifested in different active dielectrics is given in Table 9.1. With the purpose of simplification and visibility, the “impact-response” method of analysis is used.

The *impact* on a material is realized by externally applying various fields: electrical, mechanical, and thermal. In case of dielectrics, first, the application of electrical field is the most important. However, electrical field in metals and highly doped semiconductors is screened by free carriers and usually equals zero. Therefore, only in wide-gap semiconductors and dielectrics, electric field can make a significant impact. Table 9.1 shows the classification of the main effects that occur as a result of influence on materials under different fields [1,5].

The *responses* of a material are physical effects, induced by the impact.

“Trivial” (or conventional) responses are those of physical nature that corresponds to the impact nature. For example, electrical field determines electrical current (charge transfer) and electrical polarization (charge separation), which are described by conductivity σ and permittivity ϵ , respectively. These responses are shown on the main diagonal of Table 9.1. Such effects can be observed not only in active dielectrics but also in any dielectrics. Dielectrics that are characterized by only the “trivial” effects can be called as “regular”; they play a technically important and well-defined role.

However, when electrical field is applied, the responses might have not only electrical nature (electrical current or polarization) but also *mechanical* nature

Table 9.1 Primary Effects in Active Dielectrics, Classified by the “Impact \Rightarrow Material \Rightarrow Response” Method

Impact	Response		
	Electrical	Mechanical	Thermal
Electrical field E	Polarization $P = \epsilon_0 \chi E$ Induction $D = \epsilon_0 \epsilon E$	Inverse piezoeffect $x = dE$ Electrostriction $x = RE^2$	Electrocaloric effect $P = \xi \delta T$
Mechanical stress X	Direct piezoeffect $E = dX$	Strain $x = sX$	Elastothermal effect $\delta T = \eta X$
Temperature change δT	Pyroelectric effect $P = \gamma \delta T$	Thermal expansion $x = \alpha \delta T$	Heat capacity $\delta Q = C \delta T$

(deformation), *thermal* nature (change of substance temperature), and others. Particular attention is given to materials in which the *nondiagonal* (“cross”) effects are strongly pronounced. These effects are, primarily, the piezoelectric effect that characterizes electromechanical properties of dielectrics that have internal polar directions. The pyroelectric effect is also concerned with the cross effects. If these effects are greatly inhibited, it enables the relevant material to be considered as active dielectrics, which includes pyroelectrics and piezoelectrics, as well as many other corresponding magnetic and optical counterparts. As a rule, crosselectrical and optical effects are most clearly manifested in ferroelectrics.

Impact usually is the *vector* field: electrical, magnetic, and high-frequency electromagnetic field (light), but impact may also have *scalar* nature (heat), and it may be the second-rank *tensor*, for example, mechanical stress. It should also be noted that many phenomena, which represent considerable interest both for physics and technical use of dielectrics (especially in electronics), occur in case of *joint impacts* of several factors such as light and electrical field, mechanical stress and light, and so on. These effects are not shown in Table 9.1.

Usually dielectric and magnetic properties of substances are independent. The exceptions are ferroelectric-ferromagnetics (*ferroics*), in which a subsystem of ordered spins (magnetism) is strongly associated with spontaneously polarized lattice (a subsystem of ordered dipoles). In these substances, different magnetoelectrical effects may be pronounced. For example, the magnetic field can displace ferroelectric Curie point and affect the permittivity, whereas the electrical field can control frequency and quality factor of ferromagnetic resonance. However, presently, these effects are mainly a subject of research laboratories. It should be noted that different magnetomechanical and magnetothermal phenomena (magnetostriction, piezomagnetic, and other effects) are substantial only in those dielectrics that have ferromagnetic or paramagnetic properties simultaneously. For example, the magnetocaloric effect in some paramagnetics is quite significant and it is used in cryogenic technology to obtain ultralow temperatures.

Principal effects that occur in dielectrics under the influence of the electrical field have been discussed in Chapter 7. It should be noted that the strong electrical field polarization (Fig. 9.1A) and conductivity (Fig. 9.1B) become nonlinear, so that simple linear relations, as shown in Table 9.1, become more complicated. Regarding electromechanical responses (Fig. 9.1B), it should be noted that there are two responses: one effect by its nature is odd and linear (piezoelectric effect), whereas the second is even and quadratic (electrostriction). Similarly, electrothermal responses are characterized and shown in Fig. 9.1D: in addition to total quadratic effect of energy loss for all substances, a linear electrocaloric effect exists in the polar crystals.

When the intensity of impact exceeds a certain threshold, dielectrics may pass into the irreversible state such as electrical breakdown, mechanical disruption, melting, sublimation, or a combination of these. It is appropriate to note that Table 9.1, in total, is applicable only to polar (active) dielectrics.

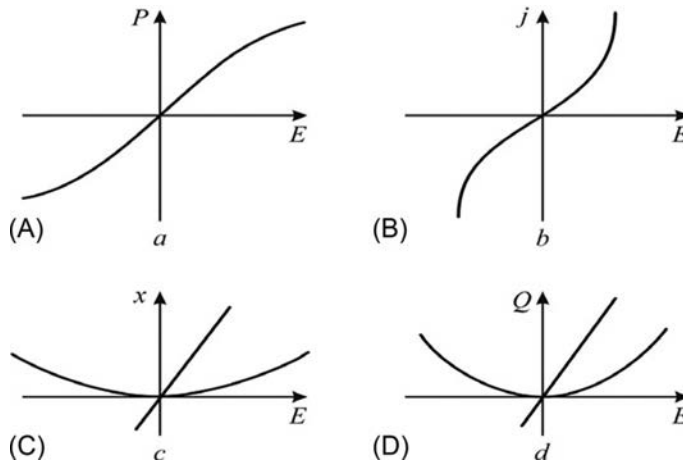


FIG. 9.1

Linear and nonlinear dielectric responses to electrical impact: (A) polarization; (B) conductivity; (C) piezoelectric effect and electrostriction; and (D) quadratic effects of dielectric loss and linear electrocaloric effect [1].

Bonding diagram for elastic, electrical, and thermal effects. Elastic, thermal, and electrical properties of polar crystals are interdependent. This diagram appears as two triangles, connected by their apexes (Fig. 9.2). Nine lines, connecting apexes, represent the *nine linear effects* that may be observed in the polar crystals [6].

Three lines of this diagram, connecting the apexes of inner and outer triangles, represent separately thermal, electrical, and mechanical interaction. The line, which connects the right apexes of triangles, symbolizes the equation $\Delta Q = C\Delta T$, which

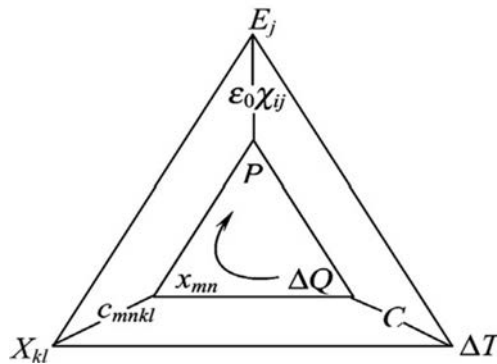


FIG. 9.2

Graph of electrical, mechanical, and thermal properties of bonding in the polar crystal.

After W.O. Cady, *Piezoelectricity*, Amazon com, New York, 1946.

describes the relationship of basic *thermal* parameters of the crystal: ΔQ and ΔT are the changes of heat and temperature, whereas C is the specific heat. The line, connecting the upper apexes of triangles, indicates *electrical* parameters of the crystal in case of electrically induced polarization: $P_i = \epsilon_0 \chi_{ij} E_j$. The line, joined to the left apexes of the diagram, symbolizes *mechanical* properties of the crystal (Hooke's law): $X_{kl} = c_{klmn} x_{mn}$, where X_{kl} and x_{mn} are stress and strain tensors, whereas c_{klmn} is the elastic stiffness tensor.

The six edges of two triangles in a diagram represent linear effects, thus reflecting the connection of thermal, elastic, and electrical properties in the polar crystal. In particular, the lower (horizontal) lines indicate *thermoelastic* effects, for instance, $x_{mn} = \lambda_{mn} \Delta T$, where λ_{mn} is the thermoelastic tensor of mechanically clamped crystal.

In case of mechanically free crystal, any type of stress is absent; hence, $x_{mn} = \alpha_{mn} \Delta T$, where α_{mn} is the tensor of thermal expansion coefficient. Depending on the implementation of this process—adiabatically ($\Delta Q = 0$) or isothermally ($\Delta T = 0$)—as well as depending on the mechanical conditions of the crystal, which might be mechanically free ($X_{kl} = 0$, which means allowing strains) or mechanically clamped ($x_{mn} = 0$, which means banned strains), thermoelastic effects can be described by different linear equations. Moreover, two opposite directions of these effects are possible: the primary influence might be thermal (change in heat or temperature), whereas the response is mechanical (change in strain or stress). Alternatively, the primary influence may be the mechanical impact on the crystal, whereas reaction is the change in temperature or heat. For example, during stretching, cooling occurs in the crystal, whereas during compression, heat is produced in the crystal. As a result, *eight* linear equations might describe all thermoelastic effects—interrelations between ΔQ , ΔT , X_{kl} , and x_{mn} .

The left side of the diagram (Fig. 9.2) corresponds to *linear electromechanical phenomena*. If the initial perturbation of the equilibrium state is mechanical deformation, x_{mn} (the crystal is free), or mechanical stress, X_{kl} (the crystal is clamped), the open-circuit electrical response to this influence will be the electrical field: $E_i = h_{imn} x_{mn}$ or $E_i = g_{ikl} X_{kl}$, respectively. In case of closed-circuit conditions, electrical response is due to the occurrence of polarization: $P_j = e_{jmn} x_{mn}$ or $P_j = d_{jkl} X_{kl}$. Thus depending on boundary conditions, a direct piezoelectric effect is described by *four* linear relationships. Inverse piezoelectric effect meets a similar situation for two mechanical boundary conditions (free or clamped crystal), as well as two electrical conditions (open circuit and closed circuit). The piezoelectric effect might be not only direct but also inverse. Thus, the left side of the diagram symbolizes *eight* linear equations that describe linear electromechanical effects: interrelations of E_i and P_j on one hand and between X_{kl} and x_{mn} on the other hand.

The right side of the diagram (Fig. 9.2) describes *eight electrothermal effects* in the polar crystal. The pyroelectric effect occurs when the disturbance factor is thermal influence, whereas the response is of electrical nature. Depending on thermal conditions (adiabatic with $\Delta Q = 0$ or isothermal with $\Delta T = 0$) and electrical conditions (open circuit and closed circuit), four possible equations describe the pyroelectric effect: $P_i = \gamma_i \Delta T$, $P_i = \gamma'_i \Delta Q$, $E_j = \gamma''_j \Delta T$, and $E_j = \gamma'''_j \Delta Q$, where different

pyroelectric coefficients correspond to various boundary conditions. The electrocaloric effect is the inverse of the pyroelectric effect, and it may also be described through four different linear relationships—depending on the boundary conditions.

One important consequence of the relationship between electrical, thermal, and elastic effects in the polar crystal is the appearance of *secondary effects*. The path of one of these effects, denoted by an arrow, is seen in the diagram. In this example, it is possible to observe the *secondary pyroelectric* effect that occurs under certain boundary conditions: due to thermally induced expansion of the free crystal, electrical polarization appears through the piezoelectric effect.

Another consequence of the relationship shown is the dependence of thermal, electrical, or mechanical processes, present in the polar crystals, on the boundary conditions. For example, in the *open-circuit* pyroelectric, specific heat C^E differs from specific heat C^P obtained in the *closed-circuit* crystal. In the same manner, there is a difference in specific heat of mechanically free (C^X) and mechanically clamped (C^Y) crystals. Similarly, the elastic stiffness, according to Hooke's law, for polar crystals depends on electrical conditions: the elastic stiffness of open circuit (c_{klmn}^P) differs from that of closed circuit (c_{klmn}^E). Although Hooke's law is studied in the polar crystal, its elastic stiffness depends on the isothermal (c_{klmn}^T) or adiabatic (c_{klmn}^S) conditions.

9.2 PIEZOELECTRIC EFFECT

Piezoelectric effect was discovered by Pierre and Jacques Curie in 1880. The first technical application of piezoelectrics became well known in the year 1920 when P. Lanzheven created ultrasonic transducer for transmitting and receiving signals in water, which became the prototype of modern ultrasonic transducers used currently for navigation in submarines as well as to detect shoals of fish and for other purposes. Sometime later, B. Cady developed piezoelectric filters for use in telecommunication applications [6].

The area of practical application of instruments and devices that use the piezoelectric effect in their designs is constantly expanding. Some products such as watches, cameras, mobile phones, televisions, computers, and piezoliters have become the objects of everyday life. Many electronic devices are not possible without piezoelectric elements. There are radiators and antennas of sonar; frequency stabilizers in computers; electronic devices for reference time; power line filters and delay lines in radio and telephone communications; sensors to measure acceleration, vibration, and acoustic emission nondestructive testing; piezotransformers and piezomotors; medical ultrasound imaging and medical instruments for various purposes; and so on [7].

Piezoelectric materials include bulk ceramics, ceramic thin films, multilayer ceramics, single crystals, polymers, and ceramic-polymer composites. In recent years, many types of piezoelectric films have been developed and tested for different microsystems and microelectronic components. Film and bulk piezoelectrics can

also be used in microwave MEMS devices. New relaxor-ferroelectric ceramics and crystals exhibit extremely high efficiency of piezoelectric energy conversion, which is of interest, in particular, for medical imaging devices and for other applications such as special drives for industrial nondestructive testing.

In dielectrics, although the electrical field is applied, different electromechanical effects occur: the “free” crystal under the influence of field is deformed, whereas in the “clamped” crystal, elastic stress occurs. The physical causes of electromechanical effects are the *microscopic displacement of electrical charges* in the applied electrical field because electrical polarization is obviously accompanied by the mechanical effect. The dependence of electrically induced mechanical strain on the electrical field is determined by the symmetry of dielectric structure.

In dielectrics with a centrosymmetric structure, sign of their deformation in the applied electrical field (compression or tension) is independent of field polarity. This effect is called *electrostriction*, which occurs in all dielectrics without exception. In most dielectrics, mechanical stretching is observed in the direction of the applied field; however, this effect of electrostriction is usually very small.

In dielectrics with a noncentrosymmetric structure, a more pronounced effect is observed: the *piezoelectricity*. It is assumed that the reason for this effect is the *intrinsic electrical moment* existing owing to structural peculiarities: it is the internal interaction of electronic shells of ions or molecules, which results in their shift. In the event of the piezoelectric effect, if the electrical polarity of externally applied electrical field is changed, the sign of electrically induced mechanical deformation reverses. Moreover, in the noncentrosymmetric dielectrics, an opposite effect is observed: external mechanical stress causes electrical polarization.

Thus the piezoelectric material is capable of converting mechanical energy into electrical energy, or, conversely, electrical energy into mechanical energy. The first of these effects is the direct piezoelectric effect, whereas the second is the inverse effect [8].

In case of the direct piezoelectric effect, under the influence of mechanical stress X (or the elastic deformation of x caused by mechanical stresses), noncentrosymmetric dielectrics (piezoelectric) generate electrical polarization, as shown in Fig. 9.1B and C. Because the electrical conductivity of the piezoelectric material (which is usually a good insulator) is very small, its polarization is expressed in the form of induced electrical charges that appear on the surface of the deformed piezoelectric material. The density of these charges is described by the polarization P , and the direction of the polarization vector is selected from the mark “−” to the mark “+,” as shown in Fig. 9.1B and C. Polarization P is proportional to electrical induction D , as shown in Fig. 9.1G.

If mechanical stress is not applied ($X=0$), no free charges exist on the surface of the piezoelectric material. Therefore it is electrically neutral, as shown in Fig. 9.3A. The piezoelectric material becomes polarized as a result of positive (stretching) deformation when $x>0$ or negative (compression) deformation when $x<0$. When mechanical stress changes its sign (such as when compression changes to stretching, as shown in Fig. 9.3B and C), the sign of the mechanically induced electrical

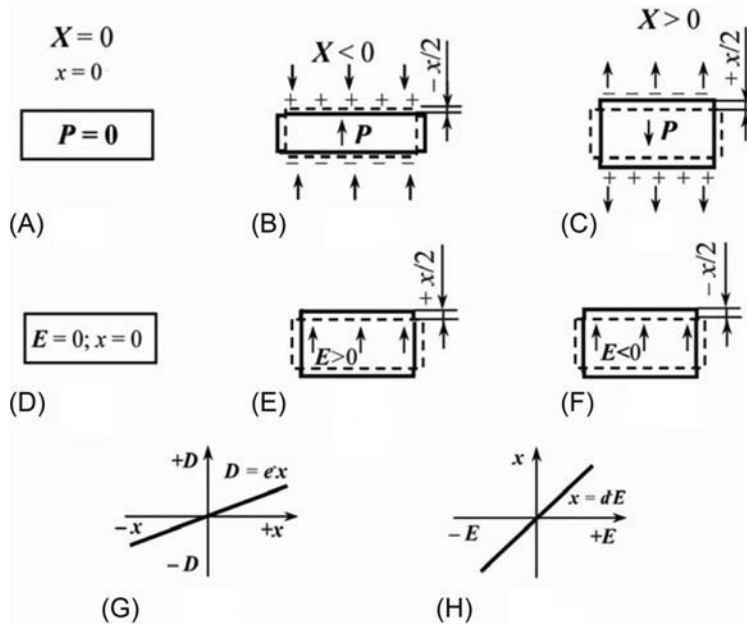


FIG. 9.3

Explanation of direct (A, B, C, H) and inverse (D, G, H, E) piezoelectric effects.

polarization P supersedes. In case of the direct piezoelectric effect, polarization is directly proportional to strain:

$$P = \epsilon x,$$

where ϵ is the piezoelectric strain constant.

Inverse piezoelectric effect occurs when the electrical field deforms the noncentrosymmetric crystal structure, as shown in Fig. 9.3E and F. The sign of the electrically induced strain varies with the sign of electrical influence, as shown in Fig. 9.3H. Further, deformation (strain) varies linearly with electrical field:

$$x = dE,$$

where d is the *piezoelectric modulus*.

The most simplified explanation of the direct piezoelectric effect in the α -quartz (SiO_2) is presented in Fig. 9.4. The generally accepted model of hexagonal quartz structure is the hexagon with positive silicon ions and negative oxygen ions that form a noncentrosymmetric structure. Some forms of hexagon deformation can produce electrical polarization. If deformation is absent, no polarization is observed, as shown in Fig. 9.4A [6,9]. The stretching of a model cell in the horizontal direction induces charges and electrical field, as shown in Fig. 9.4B; this is the *direct longitudinal piezoelectric effect*. Further, the “-” charge dominates on the left side of a cell, whereas the “+” charge appears on the right side of a cell. The upper and lower

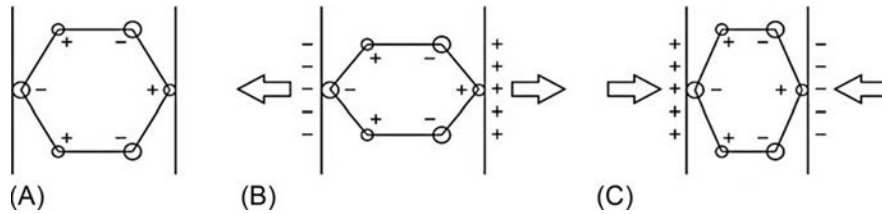


FIG. 9.4

A simplified model of the piezoelectric effect in quartz: (A) deformation is absent; (B) stretching of a model cell in the horizontal direction; (C) compression of a model cell in the horizontal direction.

parts of the concerned cell have no generated charges: they remain neutral (no transverse effect).

A similar result might be obtained by the compression of the concerned cell in the horizontal direction, as shown in Fig. 9.4B: in the vertical direction, no longitudinal effect is seen compared to that seen in the upper and lower parts of the cell. However, due to horizontal compression, the *transverse* piezoelectric effect occurs again on the left and the right sides of the cell. The point is that the horizontal direction of the selected cell is *polar*, whereas the vertical direction is nonpolar. Fig. 9.4C demonstrates that a change in the sign of mechanical impact gives rise to the change in piezoelectric polarity, as it should be in the case of *linear effects*.

Therefore a simple model, shown in Fig. 9.4, describes the longitudinal piezoelectric modulus, when electrical response has a direction same as that of mechanical influence. In this case, the highest value of piezoelectric modulus (d_{\max}) is determined. In various directions of the quartz crystal, modulus d has another value, whereas the piezoelectric response distribution might be rather complicated, as shown in Fig. 9.5.

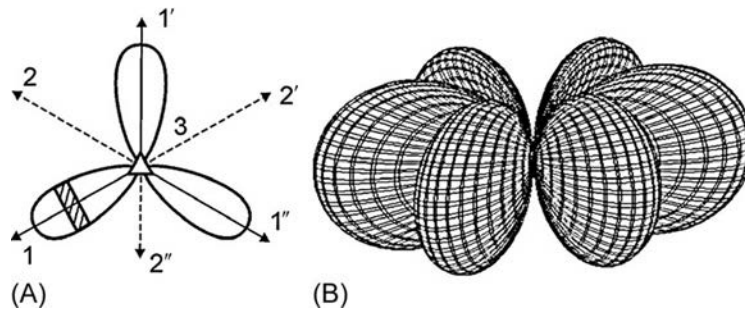


FIG. 9.5

Longitudinal piezoelectric modulus of quartz: (A) planar distribution, Curie cut is shown by strokes [6], (B) spatial distribution of modulus (guide surface—indicatrix [10]).

From Fig. 9.5, one can see that the piezoelectric effect in quartz is absent in the vertical axis of a cell; in Fig. 9.5A, it is denoted as axis 3 (this three-order axis of quartz is nonpolar). Similarly, the piezoelectric effect cannot be observed along other three *nonpolar* axes that are indicated as $2, 2'$, and $2''$). The highest possible piezoelectric effect is seen along the three *polar* axes $1, 1'$, and $1''$. The cut of a quartz crystal, made perpendicular to this direction, is the *Curie cut*. There is decreased piezoelectric activity slanting to Curie cut planes, and its distribution in a plane is described as $d = d_{\max} \cos 3\varphi$, where φ is the plane angle.

Spatial distribution of piezoelectric modulus in the polar coordinates is described as $d = d_{\max} \sin^3 \theta \cos 3\varphi$, where θ is azimuth angle. This spatial pattern, as shown in Fig. 9.5B, appears as almond grains: six surfaces joined in the center. In the Z-axis, as well as in the three Y-axes, the piezoelectric effect in quartz does not occur. Through the radius vector directed from the center of the figure as a certain angle, the size of piezoelectric modulus can be determined in any cut of quartz. It is obvious that the highest possible effect occurs along any of the three X-axes.

The main feature of the piezoelectric effect is its *linearity*, and this important fact enables to distinguish inverse piezoelectric effect from *electrostriction*. In any dielectric, external electrical field produces the deformation that is characterized by its quadratic dependence on the field:

$$x = RE^2,$$

where R is the constant of electrostriction, as shown in Fig. 9.6A. It is seen that strain in case of electrostriction does not change its sign with change in electrical field polarity.

Except quadratic type of $x(E)$ dependence, electrostriction is different from the piezoelectric effect¹ also by a fact that electrostriction has no retroactive mechano-electrical effect. Regarding the linear piezoelectric effect, one can see the direct and inverse effects.

In compliance with such an *electrically induced* (artificial) piezoelectric effect (possible in any solid dielectric), one can suppose that the usual piezoelectric effect also would be explained as the “linearized electrostriction,” as shown in Fig. 9.6B. In case of piezoelectric effect, no external field is applied, but the internal (spontaneous) distribution of electrical charges can be roughly characterized as “bias effective

¹*Comments.* Linear electromechanical effect (which is a peculiar property of polar structures) can sometimes be interpreted as *linearized electrostriction*, as shown in Fig. 9.6B. Suppose the external direct electrical field (bias field E_b) is applied to *usual* centrosymmetric crystal (which is nonpiezoelectric at $E = 0$), the applied electrical field changes the original symmetry of the crystal due to electrical polarization: conditioned by electrostriction, strain x_b occurs that corresponds to the bias field E_b . In this way, under external direct voltage, the structure of the crystal turns into the *polar structure* (becomes noncentrosymmetric). In that case, the *imitation of linear* electromechanical response can be observed: on a wing of electrostriction parabola: in the presence of bias field, alternating (sinusoidal) electrical field E' generates practically linear mechanical response: $x' \approx d'E'$, where d' is the *electrically induced piezoelectric module*. Calculations show that $d' \approx 2Q\epsilon_0^2\epsilon^2E'$, where ϵ is the permittivity and Q is the electrostriction.

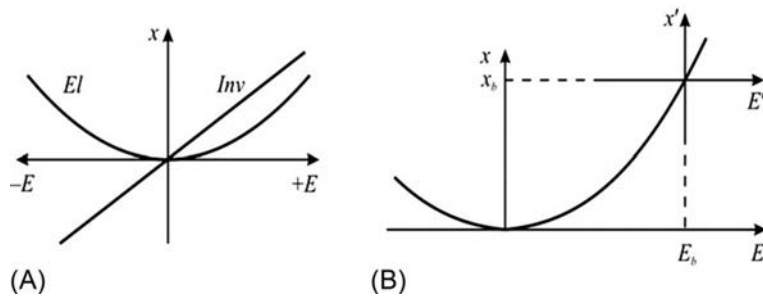


FIG. 9.6

Comparison of inverse piezoelectric effect (Inv) and electrostriction (El): (A) parabolic and linear field dependences of strain; (B) in bias field E_b the quasilinear dependence $x'(E)$ imitates the piezoelectric effect.

internal crystal field, E_b .” However, in the noncentrosymmetric crystal, this “imaginary field” looks large, whereas the externally applied voltage can only give rise to a slight change in “spontaneous mutual displacement” of ions but cannot change the overall direction of internal polar arrangement of a crystal. This assumption might be advanced according to the conception that the fundamental reason for intrinsic polarity of the crystal is the asymmetry in electronic density distribution along polar bonds between ions that have quite different electronegativity.

Electrostriction differs from the piezoelectric effect by a fact that it has no opposite effect, that is, this effect is exceptionally *electromechanical*, but not *mechano-electrical*. In case of the direct piezoelectric effect, induced electrical moment (polarization) occurs due to charged particle displacement in the noncentrosymmetric dielectric. However, in the centrosymmetric dielectric, any displacement of charged particles under the influence of mechanical force does not result in the polarized state of dielectric just due to the presence of a symmetry center in its structure: any electrical moments, created by the displacement of positively and negatively charged particles, are compensated. Therefore, electrostriction has no opposite effect.

Although the symptom of piezoelectric properties is necessarily due to the presence of noncentrosymmetric structure, the observation of electrostriction has no restrictions in symmetry, and it is manifested *in all dielectrics*.

In most cases, electrostriction effect is quite small that it cannot be considered—not only in technical application but also in dielectric research (relative deformation due to electrostriction rarely exceeds 10^{-8}). However, recently, such active dielectrics have been discovered (relaxor ferroelectrics), which demonstrate the “*giant*” *electrostriction* when the strain in the external field reaches 10^{-4} – 10^{-3} , that is, surpass even best piezoelectrics. Such electrostrictive materials are very important in technical applications because they do not show hysteresis in their strain-field characteristics.

Typically, in solid-state physics, mechanical and electrical properties are studied as independent. However, in piezoelectrics, owing to their special structure, electrical and mechanical properties are mutually conditioned. Piezoelectricity and electrostriction are the electromechanical effects that are close to each other by their physical nature. The piezoelectric effect refers to special electromechanical properties of certain dielectrics that have a polar structure (maximal effect is seen just in polar directions).

Electromechanical parameters of crystals include piezoelectric modules, electromechanical coupling coefficients, and piezoelectric Q -factors (mechanical and electrical), which indicate energy loss in the piezoelectric transducers. In addition to the aforementioned parameters, in accordance with one or another technical application of piezoelectric materials, other “quality factors” are applied to compare different piezoelectric materials to select them for certain practical applications.

Therefore, as already noted, internal polarity of noncentrosymmetric crystals enables the conversion of mechanical energy into electrical energy (direct piezoelectric effect) or, the converse, electrical energy into mechanical energy (inverse piezoelectric effect). All these effects are described by different *linear relationships* with dependence on the combination of various boundary conditions, under which polar crystals are used or studied.

Next, the idealized electrical and mechanical boundary conditions, under which the polar crystal might be applied or studied, are considered.

Electrical boundary conditions:

When $E = 0$, the polar crystal is *electrically free*, which means that the entire surface of the crystal is equipotential. If the electrical induction is $D = \epsilon_0 E + P$, then in the electrically free crystal, $D = P$. In case of the *static* procedure of piezoelectricity study (or application), the condition $E = 0$ can be realized by entirely metallized crystal. In practice, this condition is performed by a shorting of electrodes deposited on the piezoelectric. Under *dynamic* testing, when mechanically or thermally induced polarization is variable in time, the condition $E = 0$ leads to electrical current, that is, the crystal is the source of current.

When $D = 0$, the polar crystal is *electrically disconnected*, $D = \epsilon_0 E + P = 0$. In the *static* case, implementation of this condition in research requires extremely low conductivity of the piezoelectric: only in this case, piezoelectric polarization P is totally compensated by mechanically induced electric field: $\epsilon_0 E = -P$. In case of *dynamic* excitation of the polar crystal, the condition $D = 0$ is true, for example, for acoustic waves that have *longitudinal* polarization.

Mechanical boundary conditions:

When $X = 0$, the polar crystal is in a *mechanically free* state, in which all components of stress tensor are equal to zero. In the *static* studies, this condition can be realized by providing a total freedom for crystal deformation. Under the *dynamics* condition, $X = 0$ can be realized with the same caution and, in addition, the polar crystal should be explored at frequencies *below its mechanical resonances*.

When $\mathbf{x} = \mathbf{0}$, the polar crystal is *mechanically clamped*. Theoretically, to provide this condition in the *static* experiment, the crystal must be surrounded by an “infinitely rigid” shell that is “rigidly stuck” to the crystal. Such studies are either impossible or impractical. During experiments, mechanical clamping is realized by introducing the *dynamic* method using high-frequency range studies: frequencies higher than those of all electromechanical resonances of the crystal. In this case, the inertia of the crystal by itself prevents deformation; hence, the condition $x = 0$ is satisfied.

These are only idealized boundary conditions, and any approach to their implementation can be special when setting research goals: the study of electromechanical properties of the crystal. In practice, piezoelectric crystals are used in the *intermediate conditions*: they are partially clamped—partially free, not entirely short circuit nor entirely open circuit but are loaded into a certain impedance value. However, the listed idealized boundary conditions should be assumed as a basis for piezoelectric effect study [11].

For short circuit and mechanically clamped crystal description of the direct piezoelectric effect, the equation is

$$P_i = d_{ijk} X_{jk}, \quad (9.1)$$

where P_i is the component of polarization, X_{jk} is the component of mechanical stress tensor (second-rank tensor), and d_{ijk} is the component of piezoelectric module (third-rank tensor). From [relation \(9.1\)](#), it follows that the dimension of piezoelectric module is $[d] = [P]/[X]$. Considering that $[P] = \text{C/m}^2$ and $[X] = \text{N/m}^2$, for dimension of piezoelectric stress coefficient, we have: $[d] = \text{C/N}$. The module d_{ijk} is the component of the third-rank tensor; hence, the indices i, j , and k in the expression mean summation. In the expanded record of Eq. (9.1) for the crystals of lowest symmetry tensor, d_{ijk} could have 27 components (listed in [Table 9.2A](#)). In fact, due to the symmetry of elastic stress tensor ($X_{jk} = X_{kj}$), the tensor of the piezoelectric module is symmetric as the last two indices: $d_{ijk} = d_{ikj}$, whereby the number of independent components is reduced to 18, as it can be seen in [Table 9.2B](#).

For a more convenient abbreviated *matrix representation* of the third-rank tensor, it is better to use the same form of the matrix applied in [Chapter 2](#) in case of fourth-rank tensors of elastic stiffness and compliance. The first subscript for d_{ijk} has the values $i = 1, 2$, and 3, but two others indexes j and k should be changed into the indexes $n = 1, 2, \dots, 6$ in compliance with rules given in [Table 2.1A](#). Appropriate new designations of piezoelectric stress coefficients are given in [Table 9.2B](#). In the matrix notation, the equations of the direct piezoelectric effect are

$$P_i = d_{in} X_n. \quad (9.2)$$

The right side of these three equations has not nine but six terms. It is obvious that the number of independent piezoelectric coefficients for low-symmetry crystals really is not 27, but 18.

The higher the symmetry, the smaller is the number of nonzero components in the matrix. For example, in [Table 9.2C](#), a matrix of quartz piezoelectric coefficients is shown, whereas in [Table 9.2D](#), components of barium titanate crystal piezoelectric

Table 9.2 Various Records for Piezoelectric Modulus Components

(A)									
	X_{11}	X_{12}	X_{13}	X_{21}	X_{22}	X_{23}	X_{31}	X_{32}	X_{33}
P_1	d_{111}	d_{112}	d_{113}	d_{121}	d_{122}	d_{123}	d_{131}	d_{132}	d_{133}
P_2	d_{211}	d_{212}	d_{213}	d_{221}	d_{222}	d_{223}	d_{231}	d_{232}	d_{233}
P_3	d_{311}	d_{312}	d_{313}	d_{321}	d_{322}	d_{323}	d_{331}	d_{332}	d_{333}

(B)						
	X_1	X_2	X_3	X_4	X_5	X_6
P_1	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}	d_{16}
P_2	d_{21}	d_{22}	d_{23}	d_{24}	d_{25}	d_{26}
P_3	d_{31}	d_{32}	d_{33}	d_{34}	d_3	d_{36}

(C) Quartz Piezoelectric Modulus Components						
$d_{1n} = \begin{bmatrix} d_{11} & d_{12} & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{25} & d_{26} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \text{ where: } \begin{cases} d_{11} = -d_{12} \\ d_{25} = -d_{14} \\ d_{26} = 2d_{11} \end{cases}$						

(D) BaTiO ₃ Piezoelectric Modulus Components						
$d_{1n} = \begin{bmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{24} & 0 & 0 \\ d_{31} & d_{32} & d_{33} & 0 & 0 & 0 \end{bmatrix}, \text{ where: } \begin{cases} d_{24} = d_{15} \\ d_{32} = d_{31} \end{cases}$						

module are given as examples. Each of those components is a proportionality factor between the components of polarization P_i and the components of stress X_n . Piezoelectric modules of different crystals and textures can vary significantly in their magnitude and sign; for example, the main components of the piezoelectric modulus of ADP (ammonium hydrogen phosphate) are: $d_{14} = -1.34 \times 10^{-12} \text{C/N}$ and $d_{36} = 20 \times 10^{-12} \text{C/N}$ (all other components of matrix are zero).

It is seen that the *unit of piezoelectric module* in system SI is very small. Therefore in practice, the more convenient unit pC/N (picocoulomb) is used, where $1 \text{ pC} = 10^{-12} \text{ coulomb}$. In such units in barium titanate, $d_{33} = 150 \text{ pC/N}$, $d_{31} = 70 \text{ pC/N}$, and $d_{15} = 250 \text{ pC/N}$, whereas $d_{24} = d_{25}$ and $d_{32} = d_{31}$, as shown in Table 9.2D. In the quartz crystal, the components of piezoelectric module are much less: $d_{11} = 2.3 \text{ pC/N}$, $d_{12} = -d_{11}$, $d_{14} = 2.7 \text{ pC/N}$, $d_{25} = -d_{14}$, and $d_{26} = -2d_{11}$.

The upper part of Fig. 9.7 shows physical interpretation of the main quartz piezoelectric coefficients. For example, from Eq. (9.2), along axis l , one component can be selected

$$P_1 = d_{11} X_1 + d_{12} X_2 + d_{13} X_3 + d_{14} X_4 + d_{15} X_5 + d_{16} X_6. \quad (9.3)$$

As it follows from Table 9.2, for quartz, piezoelectric module components $d_{13} = d_{15} = d_{16} = 0$, so that Eq. (9.3) can be simplified:

$$P_1 = d_{11} X_1 + d_{12} X_2 + d_{14} X_4.$$

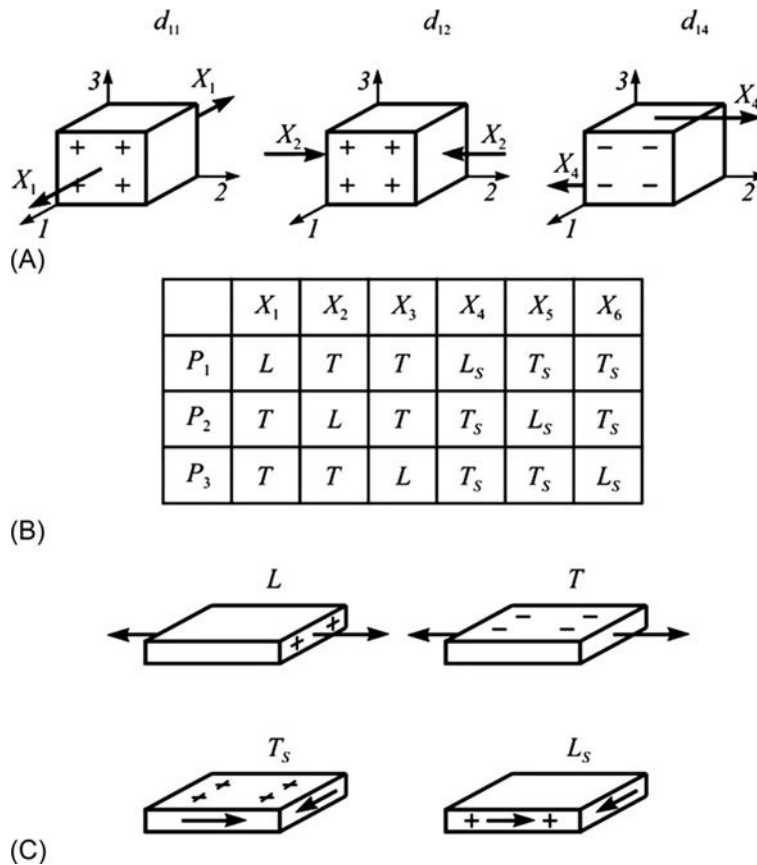


FIG. 9.7

Geometric patterns that explain the longitudinal, transverse, and shear piezoelectric effect in quartz: (A) physical interpretation of the main quartz piezoelectric coefficients; (B) total matrix of piezoelectric module components; (C) two types of shear piezoelectric effects (longitudinal shear L_S and transverse shear T_S).

Stress tensor component X_1 characterizes compressive or stretching stress along axis l . Therefore, piezoelectric modulus component d_{11} corresponds to the *longitudinal* piezoelectric effect because polarization occurs along the same direction in which mechanical stress is applied. Longitudinal effect is sometimes referred to as the *L*-effect.

The same physical meaning have components d_{22} and d_{33} (Table 9.2B): they characterize the longitudinal piezoelectric effect along axes 2 and 3, respectively. If the indices in piezoelectric matrix $|d|$ are the same, these components describe one of the three *longitudinal* piezoelectric effects. However, in the quartz crystal, the *L*-effect occurs only along axis 1, whereas in the barium titanate, this effect exists only along axis 3 (Table 9.2C and D).

Piezoelectric coefficient d_{12} corresponds to *transverse* piezoelectric effect. In fact, elastic tension is applied along axis 2, but piezoelectric effect is observed along axis 1 that is perpendicular to axis 2. As it is seen from Table 9.2B, the components of transverse matrix are also piezoelectric coefficients d_{12} , d_{21} , d_{13} , d_{31} , d_{23} , and d_{32} . They describe the appearance of polarization along one of the three axes (1, 2, or 3) at the influence of stretching-compression stresses along the axis perpendicular to the axis of response.

In Fig. 9.7B, the total matrix of piezoelectric module components is shown. Further, due to the L -effects, the T -effects occupy the *left half* of the matrix. For example, piezoelectric coefficient d_{14} (shown in the upper part of Fig. 9.7A) describes piezoelectric polarization that arises under the influence of *shear stress*. Quartz crystal, as seen from matrix of its piezoelectric coefficients (Table 9.2C), has three nonzero *shear components*: d_{14} , d_{25} , and d_{26} . The physical meaning of d_{14} is illustrated in Fig. 9.7A: the pair of forces, applied along axis 2, induces polarization along axis 1. From Table 9.2, it is clear that in quartz, $d_{14} = d_{123} + d_{132}$ ($d_{123} = d_{132}$), because stress matrix is symmetrical, and the shear stress components are $X_{23} = X_{32}$.

In the barium titanate, the nonzero shear piezoelectric coefficients are d_{15} and d_{24} (Table 9.2D), whereas, for example, in potassium dihydrogen phosphate (KDP) crystals, only one nonzero shear coefficient exists: d_{31} . The method to distinguish between the two types of shear piezoelectric effects (longitudinal shear L_S and transverse shear T_S) is shown in Fig. 9.7C. The coefficients d_{14} , d_{25} , and d_{36} correspond to L_S -components, which are characterized by polarization vector that is parallel to the shear axis and perpendicular to the shear plane. The coefficients d_{15} , d_{16} , d_{24} , d_{26} , and d_{34} и d_{35} correspond to the transverse shear effect (T_S). In such case, polarization vector is perpendicular to the shear axis and lies in the plane of shear [11].

Expressions (9.2) are only one of the four possible definitions of the direct piezoelectric effect peculiar to the *electrically free* ($E=0$) and *mechanically clamped* crystal ($x=0$): $P_i = d_{in}X_n$. A combination of other idealized boundary conditions leads to other three piezoelectric equations:

$$\begin{aligned} P_i &= e_{im}x_m \\ E_j &= -g_{jm}x_m \\ E_j &= -h_{jn}X_n \end{aligned} \quad (9.4)$$

Here and subsequently exactly, *matrix* notation of the third-rank tensor is used: e_{im} , g_{jm} , and h_{jn} . All these piezoelectric coefficients, as d_{in} , characterize the properties of noncentrosymmetric crystals and textures. The units of different piezoelectric coefficients can be determined from Eq. (9.4):

$$[d] = [\text{C/N}]; \quad [g] = [\text{Vm/N}];$$

$$[e] = [\text{C/m}^2]; \quad [h] = [\text{V/m}].$$

According to the discussed boundary conditions, when d_{in} is measured independently, the component acting on the piezoelectric stress tensor X_n is determined, as well as the component of the resulting polarization P_i . This method corresponds

to the static measurement of piezoelectric coefficient; just by this way, the piezoelectric effect was first discovered. In the event that the piezoelectric coefficient e_{im} should be determined directly, piezoelectric crystal will be as electrically free ($E=0$); hence mechanically free ($X=0$), that is, the deformation x and the density of induced electrical charge are measured.

Each of the four piezoelectric coefficients: d_{in} , e_{im} , g_{in} , and h_{jm} can be calculated from any other factor, if elastic parameters (c_{mn} or s_{mn}), dielectric parameters (ϵ_{ij} and β_{ij} —dielectric “tightness,” inverse tensor to ϵ_{ij}), and piezoelectric parameters are known. For example, from Eq. (9.3), it follows that $P_i = d_{in}X_n$, where $X_n = c_{nr}x_m$ and $P_i = d_{in}c_{nr}x_m$. Comparing with expressions (9.4), it is possible to get one of constraint equations for piezoelectric coefficients:

$$e_m = d_{in}c_{mn}. \quad (9.5)$$

In these and others, similar relationship conditions at which components c_{mn} and s_{mn} were obtained cannot be ignored: for the short-circuit ($E=0$) or the open-circuit ($D=0$) cases, inasmuch as $c_{mn}^E \neq c_{mn}^D$ and $s_{mn}^E \neq s_{mn}^D$. In others, relationships between piezoelectric coefficients and the tensor components ϵ_{ij} and β_{ij} are included, which are different in mechanically free ($\epsilon_{ij}^X, \beta_{ij}^X$, i.e., $X=0$) and mechanically clamped ($\epsilon_{ij}^x, \beta_{ij}^x$, i.e., $x=0$) crystals or textures.

While studying the direct piezoelectric effect, considering electrical freedom of crystal, elastic stiffness should be included in Eq. (9.5) with the upper index E . This means that it is determined at $E=0$. Hence, the given ratio should be written as:

$$e_{im} = d_{in}c_{mn}^E. \quad (9.6)$$

While determining the e_{im} from the direct piezoelectric effect, piezoelectric is electrically free ($E=0$), so that another equation of elastic compliance is indicated:

$$d_{in} = e_{im}s_{nm}^E. \quad (9.7)$$

The remaining relationships that tie together piezoelectric coefficients, including formulas (9.5)–(9.7), are contained in following, more comprehensive, value, which considers conditions of dielectric and elastic parameters determining:

$$d_{in} = \epsilon_0 \epsilon^X_{ij} g_{jn} = e_{im} s_{nm}^E = \epsilon_0 \epsilon^x_{ij} h_{jm} s_{nm}^E;$$

$$e_{im} = \epsilon_0 \epsilon^x_{ij} h_{jm} = d_{in} c_{nm}^E = \epsilon_0 \epsilon^X_{ij} g_{jn} c_{nm}^E;$$

$$g_{in} = (\beta^X_{ij}/\epsilon_0) d_{im} = h_{im} s_{nm}^D = (\beta^x_{ij}/\epsilon_0) e_{jn} s_{nm}^D;$$

$$h_{in} = (\beta^x_{ij}/\epsilon_0) e_{jn} = g_{im} c_{nm}^D = (\beta^X_{ij}/\epsilon_0) d_{jm} c_{nm}^E.$$

This section concludes with the examples of piezoelectric coefficient matrixes that are most studied and widely used in engineering piezoelectrics; quartz and barium titanate piezoelectric coefficients are shown previously in Table 9.2B.

For Rochelle salt crystal at $t = (-18 \dots +24)^\circ\text{C}$:

$$d_{in} = \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{25} & d_{26} \\ 0 & 0 & 0 & 0 & d_{35} & d_{36} \end{bmatrix}.$$

For KDP crystal above 150K:

$$d_{in} = \begin{bmatrix} 0 & 0 & 0 & d_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{25} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \text{ where } d_{25} = d_{14}.$$

9.3 INVERSE PIEZOELECTRIC EFFECT

The voltage, applied to any dielectric, always results in its deformation because during electrical polarization, all charged particles are displaced. Therefore in all dielectrics, the quadratic electromechanical effect (or electrostriction) is seen to occur, but this effect usually is very small. However, in some solid dielectrics, namely, in crystals having a noncentrosymmetric structure, usually a much larger electromechanical linear effect is observed that is the *inverse piezoelectric effect*:

$$x_m = d_{mj}E_j, \quad (9.8)$$

where $m = 1, 2, \dots, 6$; $j = 1, 2, 3$ according to matrix notation. Eq. (9.8) includes the same piezoelectric modules as in the case of direct piezoelectric effect (in Eq. 9.3) and with the same components that were previously listed.

Direct and inverse piezoelectric effects occur only in 20 of possible 32 classes of crystals, and each is characterized by a peculiar symmetry group. These groups consist of a set of symmetry elements, allowed only in certain combinations. Crystals, which have a center of symmetry, cannot exhibit piezoelectricity. There are 11 such classes (of the possible 32). Thus, nonpiezoelectric classes of crystals are 11; from the remaining 21 classes of crystals, 20 are piezoelectric ones. All of them belong to the noncentrosymmetric crystal.

It is obvious that there is one class, which, being noncentrosymmetric, does not show the piezoelectric effect (as it turns out, in the linear case).

Note. This “mysterious” noncentrosymmetric but nonpiezoelectric class of crystal, in reality, becomes piezoelectric but in the *strong electrical field*. The description of the inverse piezoelectric effect by a simple linear relation $x = dE$ in the strong electrical field should be considered only as the first term of odd series: $x = dE + d'E^3 + d''E^5 + \dots$. Really, in a weak electrical field for noncentrosymmetric class 4.2, the first term of this series is $d = 0$, but already the *second term* of this series, namely, the d' is nonzero. Thus in a strong electrical field, a nonlinear (but odd) piezoelectric effect in this “mysterious” class actually exists.

Table 9.3 lists all piezoelectric classes [11]; symbols of these classes are given in the international classification (reflecting the main elements of symmetry) and specify the number of nonzero tensor components of elastic stiffness (or compliance) in

Table 9.3 The Number of Components of Basic “Material” Tensors of Piezoelectric Crystal Classes

Crystal Symmetry	Syngony of Lattice	Number of Nonzero Components ϵ_{ij}	Number of Independent Components ϵ_{ij}	Number of Nonzero Components C_{mn}	Number of Independent Components C_{mn}	Number of Nonzero Components d_{in}	Number of Independent Components d_{in}
1	Triclinic	9	6	36	21	18	18
2	Monoclinic	5	4	20	13	8	8
<i>m</i>		5	4	20	13	10	10
222	Orthorhombic	3	3	12	9	3	3
<i>mm</i> 2		3	3	12	9	5	5
4	Tetragonal	3	2	16	7	7	4
422		3	2	12	6	2	1
4		3	2	16	7	7	4
4 <i>mm</i>		3	2	12	6	5	3
42 <i>m</i>		3	2	12	6	3	2
3	Trigonal (rhombohedral)	3	2	24	7	13	6
32		3	2	18	6	5	2
3 <i>m</i>		3	2	18	6	8	4
6	Hexagonal	3	2	12	5	7	4
6		3	2	12	5	6	2
622		3	2	12	5	2	1
6 <i>mm</i>		3	2	12	5	5	3
6 <i>m</i> 2		3	2	12	5	3	1
23	Cubic	3	1	12	3	3	1
43 <i>m</i>		3	1	12	3	3	1
∞m	Polarized ceramics	3	2	16	7	5	3

all classes, as well as the number of nonzero tensor components of piezoelectric coefficients (according to the matrix shown in Table 9.2B).

It is seen that with increase in the number of elements of crystal symmetry, the independent components of tensors become smaller. The number of nonzero components of piezoelectric coefficients is gradually reduced in the last classes of piezoelectric crystals, finally, with only one independent component. Such piezoelectrics are the easiest objects to study.

It is pertinent to recall that matrix of elastic stiffness c_{mn} (and inverse matrix of elastic compliance s_{mn}) is symmetrical as to the diagonal components of matrix with 6×6 elements; therefore, in general, the number of independent components in it is 21, which is typical for triclinic crystals. When the number of symmetry elements of crystal increases, the number of zero elements of matrix also increases while the number of independent component reduces; for high symmetrical cubic crystals in the matrix of elastic constants, there are only three independent components of 21 nonzero components [11].

For example, the presence of axis 4 leads to the following relationships:

$$c_{11} = c_{22}, c_{13} = c_{23}, c_{44} = c_{55}, c_{16} = -c_{26},$$

$$c_{14} = c_{15} = c_{24} = c_{25} = c_{34} = c_{35} = c_{36} = c_{45} = c_{46} = c_{56} = 0.$$

Thus, the crystals of point groups 4 and $\bar{4}$ have 7 independent elastic constants (from 16 nonzero components).

As for the centrosymmetric crystals (these classes of symmetry in Table 9.3 are not shown), all 18 components of piezoelectric coefficients equal zero, that is, they have no electromechanical linear effect (piezoelectricity) but they have quadratic effect—electrostriction.

Polarized piezoelectric *ceramics* are widely used in modern technology—this is a texture, characterized by the axis of symmetry of infinite order (∞) and by plane of symmetry m passing through this axis. The polar axis of symmetry is turned along the direction of electrical field that has been applied from outside to create artificial polarization in ceramics. Designation ∞m of polarized texture corresponds to polar vector symmetry and consistent with “symmetry of cone.” After turning off the polarizing field, the induced polarized structure is maintained for a long time, and it has a set of elastic constants and piezoelectric coefficients that correspond to the tetragonal crystal of class $4mm$ (Table 9.3).

Normally, nonpolarized (isotropic) ferroelectric ceramics are synthesized with the “symmetry of a ball” that is the highest possible symmetry of solids. As it was mentioned, this symmetry is converted into the noncentrosymmetric texture with “symmetry of cone” by the mentioned “electrical polarization.” In this technology, which involves a strong electrical field at increased temperature, ferroelectric domains present in ceramics usually are oriented randomly and acquire preferential orientation along the applied field, thus forming stable unipolar texture of domains.

Eq. (9.8) describes the inverse piezoelectric effect for mechanically free ($x=0$) and open-circuit (electrically free, $D=0$) sample and is only one of the possible

descriptions of linear electromechanical interaction in the noncentrosymmetric crystals and textures. Under various boundary conditions listed earlier, the inverse piezoelectric effect can be described by four equations:

$$\begin{aligned} x_m &= d_{mj}E_j; & X_n &= e_{nj}E_j; \\ x_m &= g_{mi}P_i; & X_n &= h_{ni}P_i, \end{aligned} \quad (9.9)$$

where d_{mj} , g_{mi} , e_{nj} , and h_{ni} are the same piezoelectric coefficients that are used previously to describe the direct piezoelectric effect. The ratios between these parameters are given in the previous section.

The inverse and direct piezoelectric effects enable the determination of piezoelectric coefficients experimentally. To measure piezoelectric module component d_{im} , for example, it is necessary, first, to measure the electrical field applied to the piezoelectric: $E = U/l$ (where U is the electrical voltage supplied from low-impedance power and l is the thickness of the sample). Second, using dilatometer, induced mechanical deformation Δl is determined, from which dimensionless relative deformation $x = \Delta l/l$ can be found.

In accordance with Eq. (9.9), using the inverse piezoelectric effect, g_{mi} and h_{ni} can also be experimentally determined. In these experiments, the piezoelectric should be polarized with electrical current as the source, thus having a very high internal electrical resistance, which provides for piezoelectric $E \approx 0$. In case of the coefficient g_{mi} , the determining experiment must be provided at free piezoelectric strain, whereas the determination of h_{ni} deformation should be excluded; the measured parameter is the stress.

The necessity of piezoelectric effect description by *four* different coefficients is justified by different cases of technical applications. For example, when there is a need for selecting the piezoelectric material for *ultrasound emitter* that is generally used in the sonar and echo sounder, it is necessary to generate mechanical deformation under the influence of electrical voltage. In this case, to evaluate effectiveness of various piezoelectric materials, it is necessary to compare them accordingly with the largest piezoelectric module used in equation: $x = dE$.

In case of the *ultrasound receiver* that uses the direct piezoelectric effect, requirements for piezoelectric are different: maximal voltage from mechanically clamped piezoelectric, that is, the “force sensor”: $e = gX$. Hence, the best piezoelectric is that in which the factor $g = d/X$ is large. In the other case, for instance, in piezoelectric adapters, the important coefficients are h and e .

All the listed equations of the piezoelectric effect, which characterize different connections between mechanical parameters x and X and electrical parameters P and E , may be represented by the diagram—*piezoelectric square*—shown in Fig. 9.8, in the corners of which parameters x , X , P , and E are located.

In the left vertexes of the square, the *mechanical* parameters (stress and strain) are placed and their linear relationship is represented by a straight line, symbolically characterizing different views of Hooke’s law: $x = sX$ or $X = cx$. The right vertexes in the square (Fig. 9.8) represent the electrical parameters (electrical field E and

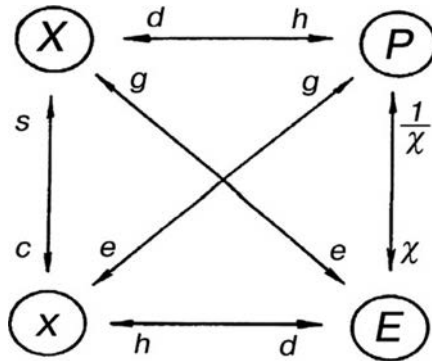

FIG. 9.8

Diagram explaining different piezoelectric descriptions depending on electrical and mechanical boundary conditions.

polarization P), whereas the connecting line characterizes purely *electrical* effects: $P = \epsilon_0 \chi E$ and $E = (\epsilon_0 \chi)^{-1} P$.

In the horizontal lines of the diagram, as well as the diagonals of the square, the several linear equations describe all direct and inverse piezoelectric effects. Near the straight lines of these “connections,” equivalent piezoelectric coefficients are shown. The ratio, located near the arrow on the link, must be multiplied by the parameter closest to it. For example, the top line of the piezoelectric square represents piezoelectric equations $P = dE$ and $X = hP$, whereas the bottom line corresponds to equations $x = dE$ and $E = hx$.

Piezoelectric contribution to permittivity (ϵ_{EM}) induced by electrical field polarization and electrical induction are described by the equation $D = \epsilon_0 \epsilon E = \epsilon_0 E + P$, where ϵ_0 is electrical constant and ϵ is relative permittivity. To describe polarization of *mechanically free* piezoelectric, it is necessary to also consider the direct piezoelectric effect: $P = -ex$, where e is the piezoelectric constant and x is the mechanical deformation:

$$D = \epsilon_0 \epsilon E + ex. \quad (9.10)$$

Electroelastic contribution to stress X is characterized by the same piezoelectric constant e : according to Hooke’s law ($X = cx$, where c is the elastic stiffness), the inverse piezoelectric effect gives additional stress:

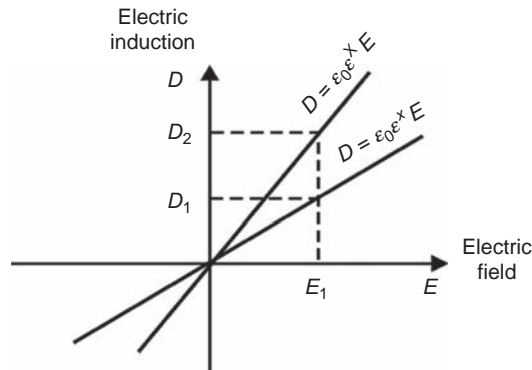
$$X = cx - eE.$$

Under this condition, any mechanical stress in piezoelectric is absent ($X = 0$); hence from the previous equation, it follows:

$$x = eE/c.$$

By substituting this expression in Eq. (9.10), the following can be obtained:

$$D = \epsilon_0 \epsilon E + (e^2/c)E = (\epsilon_0 \epsilon + e^2/c)E. \quad (9.11)$$

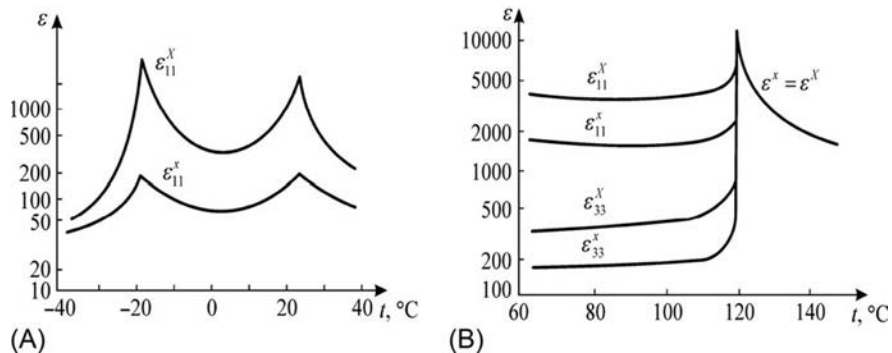

FIG. 9.9

Electric induction dependence D on the electrical field E for free ($X=0$) and clamped ($x=0$) piezoelectric.

From expression (9.11) and Fig. 9.9, it follows that mechanical response to the applied electrical field enhances the dielectric induction. In the clamped piezoelectric, in which deformation is impossible ($x=0$), external field E_1 induces $D_1 = \epsilon_0 \epsilon^x E$. In the free piezoelectric ($X=0$) at the same field E_1 , the induction is bigger: $D_2 = \epsilon_0 \epsilon^X E$. The piezoelectric reaction appears to be an additional mechanism of polarization because it mimics corresponding contribution to dielectric constant. If piezoelectric is free, its permittivity equals to $\epsilon = \epsilon^X$, whereas in mechanically clamped piezoelectric, $\epsilon = \epsilon^x$. From Eq. (9.11), the next relationship between ϵ^x and ϵ^X follows:

$$\epsilon^X = \epsilon^x + e^2 / \epsilon_0. \quad (9.12)$$

Comparison of permittivity in the free and clamped piezoelectrics is shown in Fig. 9.10. For example, well-studied ferroelectrics are selected. The piezoelectric crystal Rochelle salt is studied over a wide temperature range; permittivity of free


FIG. 9.10

Temperature dependence of the dielectric permittivity for free (at frequency of 10^3 Hz) and clamped (at 10^{10} Hz) crystals: (A) Rochelle salt, (B) barium titanate.

and clamped crystals everywhere varies greatly. In the vicinity of ferroelectric, Curie points that the clamping effect is particularly large: $\varepsilon^X/\varepsilon^x \approx 50$.

Barium titanate above its Curie point is found in cubic (centrosymmetric) phase that is not piezoelectric; hence $\varepsilon^X = \varepsilon^x = \varepsilon$. However, below the Curie point ($T_C \approx 120^\circ\text{C}$) in single-domain crystal of BaTiO_3 near room temperature, the ratio $\varepsilon^X/\varepsilon^x$ is around 2. In the polarized ferroelectric ceramic BaTiO_3 , the ratio $\varepsilon^X/\varepsilon^x$ is slightly < 2 .

At lower frequency, ε^X is measured because the piezoelectric reaction of the studied sample has enough time to be set freely ($X=0$), and it contributes to the dielectric constant. At ultrahigh frequency that is much higher than frequencies of all piezoelectric resonances, the *own mechanical inertia* of the test sample makes impossible its piezoelectric deformation in the external field ($x=0$), and the ε^x is determined. Transition from increased permittivity ε^X obtained for free crystal (at low frequencies) to the reduced permittivity ε^x inherent to clamped crystal (at very high frequencies) is accompanied by many electromechanical resonances. As an outstanding example, the KH_2PO_4 (KDP) crystal might be selected because in this case the record ratio $\varepsilon^X/\varepsilon^x \approx 100$ is observed.

9.4 ELECTROMECHANICAL COUPLING IN PIEZOELECTRIC

Piezoelectric is a converter of energy: during the direct effect, mechanical (elastic) energy is converted into electrical energy, and during the inverse effect, electrical energy is converted into mechanical energy. Therefore, elastic and electrical properties of the piezoelectric should be considered together because any change in the electrical conditions of the polar crystal leads to the changes in its mechanical state, and vice versa [8].

Relationship between electrical and mechanical properties of noncentrosymmetric crystals and textures (that exhibit piezoelectricity) is characterized by the electromechanical *coupling coefficient* K_{coup} . This is one of the most important parameters not only for piezoelectric materials but also for piezoelectric devices.

Electromechanical coupling coefficient can be defined as follows: the square of electromechanical coupling factor shows how much energy supplied to the piezoelectric (W_{br}) is converted into another type of energy (W_{conv}):

$$K_{\text{coup}}^2 = \frac{W_{\text{conv}}}{W_{\text{br}}}.$$

This definition of K_{coup} resembles the definition of efficiency coefficient; however, energy losses are not included in this expression: electrical conductivity, mechanical damping, and dielectric losses are neglected when determining K_{coup} .

In case of the direct piezoelectric effect, the crystal receives mechanical energy that is spent not only on the elastic deformation (leading to accumulation of elastic energy W_{elas}) but also on the creation of electrical polarization, which causes the accumulation of electrical energy W_{elec} :

$$K_{\text{coup}}^2 = \frac{W_{\text{elec}}}{W_{\text{br}}} = \frac{W_{\text{elec}}}{W_{\text{elas}} + W_{\text{elec}}}. \quad (9.13)$$

In case of the inverse piezoelectric effect (as well as for electrostriction), the formula for coupling coefficient varies:

$$K_{\text{coup}}^2 = \frac{W_{\text{elas}}}{W_{\text{br}}} = \frac{W_{\text{elas}}}{W_{\text{elas}} + W_{\text{elec}}}. \quad (9.14)$$

In this case, the electrical energy given to a crystal is spent not only on the electrical polarization but also on the elastic deformation of the piezoelectric. The difference between these coerced relations does not mean that K_{coup} is the same for the direct and inverse piezoelectric effects: during calculation of energies, different boundary conditions should be considered (the crystal can be free or clamped and short circuit or open circuit).

The elastic energy can be defined as a quadratic form of the strain x or mechanical stress X , according to various choices parameters characterizing the elastic process:

$$W_{\text{elas}} = \frac{1}{2}xX = \frac{1}{2}cx^2 = \frac{1}{2}sX^2, \quad (9.15)$$

where c is the elastic stiffness and s is the elastic compliance (tensor, inverse to c).

Accordingly, the energy of electrical polarization in the field E is expressed through induction D and dielectric constant ϵ , or through tensor β inverse to tensor ϵ :

$$W_{\text{elec}} = \frac{1}{2}ED = \frac{1}{2}\epsilon_0\epsilon E^2 = \frac{1}{2}\left(\frac{\epsilon_0}{\beta}\right)D^2. \quad (9.16)$$

For mixed elastic-to-electrical processes, the contribution to energies W_{elas} and W_{elec} can also be expressed in other relationships. In addition to these relationships, electromechanical coupling coefficient is defined as the ratio of elastic energy density to densities of elastic and electrical energies:

$$K_{\text{coup}}^2 = \frac{W_{\text{em}}^2}{W_{\text{elas}}W_{\text{elec}}}.$$

In case of mechanically clamped (not deformed) piezoelectric, the density of electromechanical energy is $W_{\text{em}} = d \cdot X \cdot E$, and for mechanically free piezoelectric, it is $W_{\text{em}} = e \cdot h \cdot E$, where d and e are corresponding piezoelectric modules: $d = P/X$ and $e = P/x$ (tensor indices at components of modules are neglected here to simplify records of relationships).

The developers and researchers of piezoelectric and electrostrictive materials, as well as designers of piezoelectric devices (surface acoustic wave devices, delay lines, filters, convolver-type signal converters, etc.), to determine K_{coup} sometimes use the change in velocity of elastic waves in piezoelectric:

$$K_{\text{coup}}^2 = \frac{2\Delta v}{v_0} + \left(\frac{\Delta v}{v_0}\right)^2,$$

where v_0 is the elastic wave velocity without the piezoelectric effect, and Δv is the change in velocity, obtained due to electromechanical coupling [4].

Although K_{coup} is a scalar parameter, this coefficient depends on the direction of external influences and on other causes. For example, the polarized ferroelectric ceramics (i.e., the texture of $\infty \cdot m$ symmetry) can be identified by 12 different coupling coefficients, depending on the system of boundary conditions (for many forms of piezoelectric sample), as well as on the manner of clamping and fixing. Numerical values of K_{coup} are defined by piezoelectric material properties. Most crystals, ceramics, and textures that are used in practice have $K_{\text{coup}} = 0.1 \dots 0.5$, although in some crystals in their particular orientation parameter, K_{coup} reaches a value of $0.8 \dots 0.95$.

Electromechanical coupling becomes apparent, especially during elastic wave excitation by the electrical field. Two types of elastic bulk waves can exist in the homogeneous elastic medium: the longitudinal waves, in which the displacement of particles occurs in the direction of wave propagation, and the transverse waves in which particle displacement takes place in the plane perpendicular to the direction of wave propagation.

Longitudinal and transverse waves are three-dimensional oscillations of the elastic medium. Bulk waves are used in various piezoelectronic devices: in the mode of *standing waves* in the resonant devices and in the mode of *traveling waves* in case of surface acoustic waves. Some examples of bulk wave application are shown in Section 2, Fig. 2.6, for simple piezoelectric elements made from polarized ferroelectric ceramics.

It is possible to give some examples of K_{coup} depending on piezoelectric sample orientation. As shown in Fig. 2.6A, thickness-polarized piezoelectric plate, transversal electromechanical coupling coefficient is

$$k_{31} = d_{31}(\epsilon_{33}s_{11})^{-1/2}.$$

In case of shear strain, Fig. 2.6B, for the same piezoelectric plate coupling coefficient is

$$k_{15} = d_{15} \left(\frac{\epsilon_{11}}{s_{44}} \right)^{-1/2}.$$

Thickness-polarized piezoelectric disk with radial deformations, as shown in Fig. 2.6C, has coupling coefficient

$$k_p = k_{31} \left(\frac{2}{1-\sigma} \right)^{1/2},$$

where σ is the Poisson's ratio: $\sigma = -s_{12}/s_{11}$. For thickness vibrations of the same disk

$$k_t = h_{33} \left(\frac{\epsilon_{33}}{c_{11}} \right)^{-1/2}.$$

Obtained relations enable the determination of the ratio between K_{coup} and permittivity of free and clamped crystals:

$$\varepsilon^X = \varepsilon^x + K_{\text{coup}}^2 \varepsilon^x; \quad \varepsilon^X / \varepsilon^x = 1 + K_{\text{coup}}^2. \quad (9.17)$$

At low frequency, when piezoelectric response has adequate time to settle and give contribution to permittivity, the ε^X is measured. At high frequency (much higher than the frequency of electromechanic resonances), the ε^x is determined.

Mechanical properties dependence on the electrical state of piezoelectric. From the earlier equations and Fig. 9.11, it is shown that the elastic compliance of the open-circuit piezoelectric crystal s^D should exceed that of the short-circuit piezoelectric crystal s^E .

Indeed, piezoelectricity does not affect the elastic compliance s^E in the short-circuit crystal ($E=0$). But in the open-circuit case ($D=0$), compliance s^D of the piezoelectric plate essentially decreases, as shown in Fig. 9.11. In accordance with these designations, deformation x can be presented as follows:

$$x = \left(s^E - \frac{d^2}{\varepsilon_0 \varepsilon} \right) X = \left(s^E - K_{\text{coup}}^2 s^E \right) X = s^D X;$$

$$s^D = s^E (1 - K_{\text{coup}}^2). \quad (9.18)$$

Experimental evidence of this effect is shown in Fig. 9.12A. In the polar phase of barium titanate (below its ferroelectric phase transition), one can see differences between elastic compliance of open-circuit and short-circuit crystal. Above ferroelectric phase transition, barium titanate has a cubic centrosymmetric structure; therefore the piezoelectric effect is absent; that is why, in the paraelectric phase $s^E = s^D$.

However, below phase transition, barium titanate enters into the tetragonal polar phase that is characterized by rather strong piezoelectric effect; hence, elastic compliance for various electrical conditions becomes quite different: $s^E > s^D$. The most critically elastic compliance s^D reduces at the Curie point [2].

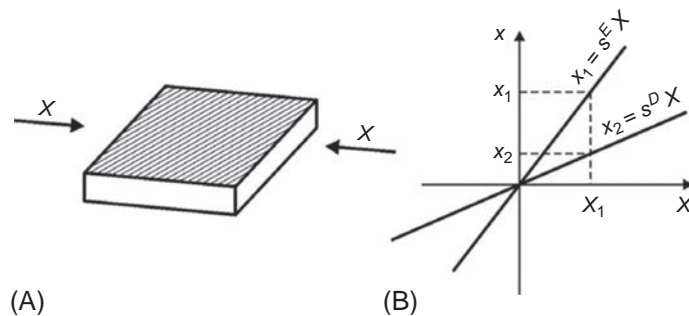
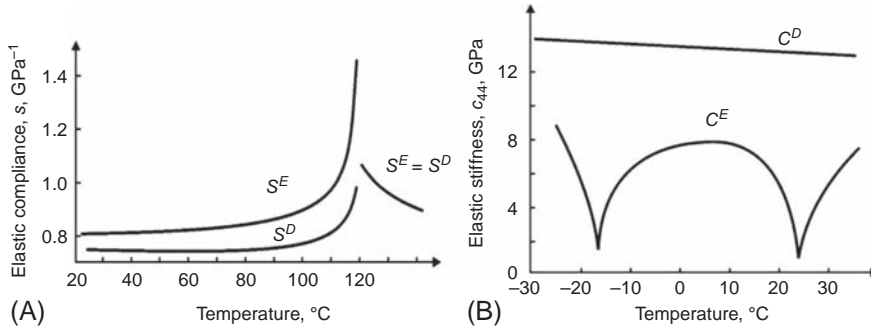


FIG. 9.11

Difference in elastic compliance of piezoelectric: (A) mechanical action on the piezoelectric plate; (B) Hooke's law in the open-circuit (s^D) and short-circuit (s^E) crystals.


FIG. 9.12

Effect of electrical conditions on piezoelectric elastic property: (A) barium titanate elastic compliance, s^D —open-circuit and s^E —short-circuit crystal. (B) Rochelle salt elastic stiffness: c^D —open-circuit and c^E —short-circuit crystal.

In the same way, experiments with Rochelle salt (Fig. 9.12B) show essential difference in the elastic stiffness c_{mn} (i.e., tensor, inverse to the elastic compliance tensor s_{mn}). Elastic stiffness c_{mn} , where $m, n = 1, 2, \dots, 6$, is different for open-circuit and short-circuit cases:

$$\frac{c_{mn}^D - c_{mn}^E}{c_{mn}^D} = K_{\text{coup}}^2. \quad (9.19)$$

The most impressive difference one can see in both Curie points of Rochelle salt. These data are obtained by sound velocity measurements, and in the Curie point, sound velocity slows down by eight times.

As elastic module together with density ρ of the piezoelectric define velocity of acoustic waves

$$v_{\text{sound}} = (c/\rho)^{1/2}. \quad (9.20)$$

Difference in the elastic module of electrically opened and shorted crystals $\Delta c = c^D - c^E$ leads to the difference in sound velocity. As can be seen from Fig. 9.12B, sound velocity in Rochelle salt (corresponding to stiffness component c_{44}) reduces by several times at the Curie points as compared with that at room temperature.

9.5 ELECTROSTRICTION

In contrast to the inverse piezoelectric effect, which is characterized by linear (“odd”) strain dependence on the electrical field, the electrostriction shows a quadratic (“even”) effect. The sign of a strain at electrostriction is not dependent on the direction of the applied electrical field, and in majority of the solid dielectrics, the extension of dielectric ($x > 0$) is observed along the applied field. The value of

electrostriction usually is small, and as a rule, electrostriction is 100–1000 times smaller than piezoelectric effect. Only in the very large electrical field, the deformation obtained by electrostriction can be compared with piezoelectric deformation (in quartz, this is observed at field strength of 35 kV/cm).

Limits by boundary conditions in case of electrostriction are the same as those for the piezoelectric effect. Depending on whether the dielectric is electrically and mechanically clamped or free, dependence of x or X on E or P can be described by four equations. However, only the electrostriction in *mechanically free* dielectrics will be considered here. In this case, for $E = 0$ and $D = 0$, respectively, it is possible to obtain

$$x_{ij} = Q_{ijkl}P_kP_l + Q'_{ijklgh}P_kP_lP_gP_h + \dots;$$

$$x_{ij} = R_{ijkl}E_kE_l + R'_{ijklgh}E_kE_lE_gE_h + \dots; \quad (9.21)$$

In these series, it is sufficient to consider only the first terms of the expansion for polarization and for the electrical field. Only in a special case of giant electrostriction (observed in ferroelectrics with diffuse phase transition) in the given series, three terms of expansion might be considered:

$$x(E) = RE^2 + R'E^4 + R''E^6. \quad (9.22)$$

Coefficients of electrostriction Q_{ijkl} and R_{ijkl} ($i, j, k, l = 1, 2, 3$) are fourth-rank tensors. However, because of strain tensor x symmetry, the Q_{ijkl} and R_{ijkl} tensors have not 81 but 36 independent components. In practice, fourth-rank tensors can be presented as *matrix components*: Q_{mn} and R_{mn} , where $m, n = 1, 2, \dots, 6$.

The fourth-rank material tensors (elastic stiffness c and elastic compliance s) already have been presented in a symmetrical matrix form, so that low-symmetry crystals can have a maximum of 21 of c and s independent components. However, tensors of electrostriction in case of low symmetry can have all $6 \times 6 = 36$ independent components. Nevertheless, in practice, this difficult case does not occur: the majority of these components are usually equal to zero. With increase in symmetry, the number of tensor components Q_{mn} and R_{mn} is significantly reduced, but never occurs (as in piezoelectric for d_{in}) when *all* accessed electrostriction tensor components turn to zero.

In the highest symmetry material (isotropic medium), only two components of electrostriction tensor are seen: Q_{11} and Q_{12} (correspondingly, R_{11} and R_{12}), characterizing longitudinal expansion and transverse contraction of dielectric in the electrical field, respectively. Just this case is usually used in practice because giant electrostriction is a characteristic of nonpolarized relaxor ferroelectrics with diffuse phase transition [1].

The electrostriction tensor Q_{mn} is more consistent for analyses, as its components are only slightly different in various solids, and they are weakly dependent on the change in external conditions. Even in ferroelectrics components, Q_{mn} only slightly changes with temperature and frequency, and they are almost independent on the electric field.

Therefore, it is assumed that only the Q_{mn} is a *fundamental* characteristic of the electromechanical coupling of atoms, ions, or molecules in a given structure of dielectric. By contrast, the components of the tensor R_{mn} are strongly dependent on permittivity, and hence they are dependent on temperature and frequency of the applied field (all these conditions substantially influences on permittivity). That is why for ferroelectrics, in which permittivity can exceed 10^4 , electrostriction may reach a magnitude of deformation same as that of the piezoelectric effect: it becomes giant for the coefficient R_{mn} (but not for the coefficient Q_{mn}). Giant electrostriction is used to control hysteresis-free micromoving (in actuator) and enables to realize electrical control by the electromechanical effect.

Electrical control of device parameters including piezoelectric effect electrical controlling has some scientific and technological interest. Changing piezoelectric properties by influence of the electrical (controlling) field is used in electrically controllable delay lines, in elements based on surface acoustic waves (SAW convolvers), and in electrically tunable piezoelectric filters.

Physical mechanisms of piezoelectric parameter control by electrical field are different in paraelectrics and piezoelectrics, although both of them are used to control sound velocity owing to the change of elastic compliance (elastic stiffness) in the electric field. Fig. 9.13 shows some typical cases of sound velocity change by the electrical bias field in different dielectrics.

The lithium niobate crystal is piezoelectric (being “hard” ferroelectric) of symmetry class $3m$. The change in sound velocity v_{sound} in the LiNbO_3 crystal in the electrical field equals only few hundredths of percent, but this is sufficient for effective management of SAW devices. Linearity of $\Delta v/v_0(E)$ dependence indicates that in the polar crystal LiNbO_3 , the strength of intracrystalline bonding (that determines piezoelectric properties) much exceeds the applied external controlling field.

More effective electrical control of sound velocity is available in usual ferroelectrics: for example, the nonpolarized PZT ceramic characteristic is shown in Fig. 9.13B. The strength of intracrystalline bonding in PZT is comparable with that of the controlling electrical field, which by the reorientation of ferroelectric domains significantly affects v_{sound} . The mechanism of sound velocity change in the applied electrical field in this case is rather complicated. First, electrical field influences the elasticity, which *increases* because of orientation of domains. Second, controlling electrical field by electrostriction increases *internal stresses* between crystallites of ceramics (that changes the dielectric constant of ferroelectric ceramics in the electrical field). The tendency of ferroelectric domains to save their orientation leads to hysteresis (aftereffect) in the characterization of controlling, which is highly undesirable in practical devices.

In first two examples of the piezoelectric effect electrical controlling, as shown in Fig. 9.13A and B, *changes* in piezoelectric properties are seen in noncentrosymmetric crystal and in ferroelectric ceramics. However, the piezoelectric effect can be also *induced* by electrical field in the nonpolar (centrosymmetric) dielectric. In Fig. 9.13C, the dependence of transverse piezoelectric coefficient d_{31} on electrical *bias field* is shown: three different centrosymmetric dielectrics are selected. In the

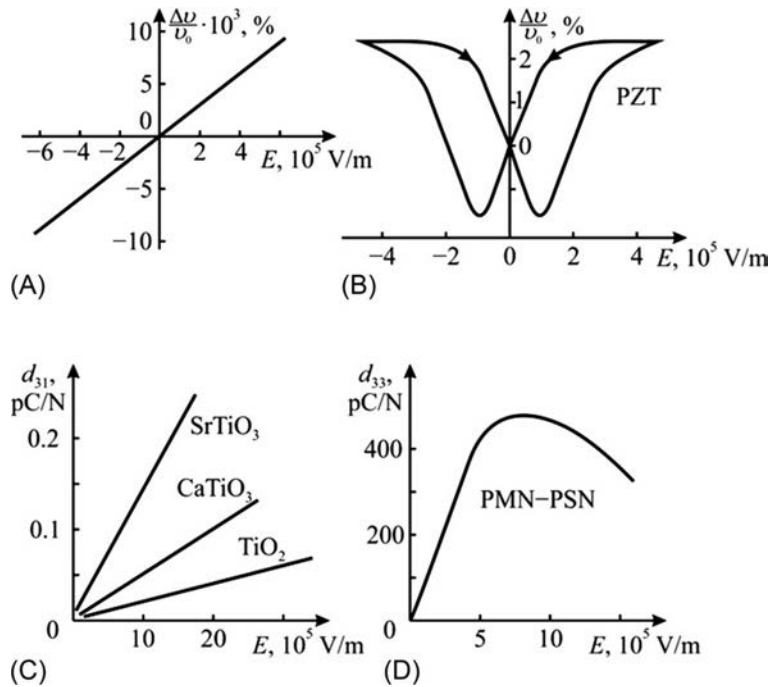


FIG. 9.13

Electrically tunable and induced piezoelectric effect: (A) lithium niobate crystal used in controlled delay lines; (B) nonpolarized PZT ceramics; (C) paraelectric ceramics; (D) diffused phase-transition ceramics [7].

absence of electrical field, piezoelectric effect in such structures is impossible. However, as it can be seen from Fig. 9.13C and D, “piezoelectric effect” in these dielectrics can be electrically induced (not only electrically controlled).

The point is that owing to electrostriction, electrical field transforms the structure of any isotropic dielectric into the noncentrosymmetric structure, and induces in it *electromechanical coupling*—piezoactivity. In the dielectrics with low permittivity, this effect is negligible and even difficult to be observed (electrostriction is very small). However, in the dielectrics with increased permittivity, the electrically induced piezoelectric effect is quite noticeable, although artificial piezoelectric module is relatively small. A comparison of characteristics in titanium oxide ceramics with $\varepsilon \sim 100$ (TiO_2 , rutile), in calcium titanate ceramics with $\varepsilon \sim 150$ (perovskite, CaTiO_3) and in strontium titanate ceramics with $\varepsilon \sim 300$ (paraelectric SrTiO_3) is shown in Fig. 9.13C.

Piezoelectric modules for *electrically induced piezoelectric effect* can be calculated from the equation of electrostriction (9.21), omitting, for simplicity, the indices of tensor components as in formula (9.22). Induced deformation is the even function of polarization that can be described by rapidly convergent series

$$x = QP^2 + Q'P^4 + \dots$$

In relatively small electric fields, only the first term of this series can be considered: $x = QP^2$. The electrostriction x_{\sim} in the *alternating* electrical field becomes linearized and can be presented as “piezoelectric effect”:

$$x_{\sim} = 2QP_y P_{\sim} = d E_{\sim},$$

where parameter d acts as “piezoelectric module” caused by electrostriction Q :

$$d \cong 2Q\epsilon_0^2 \epsilon^2 E, \quad (9.23)$$

where $E = E_{\text{bias}}$ is electrical *bias field* that induces piezoelectric properties.

Clearly, that electrically induced piezoelectric module is directly proportional to the controlling field and to the *square of permittivity*. Therefore, it is not surprising that such artificial “piezoelectric effect” the higher the bigger permittivity. If in dielectrics with increased permittivity ($\epsilon = 100\text{--}300$) induced piezoelectric module reaches only $\sim 0.3\text{ pC/N}$, in specially designed electrostriction ceramics $\text{PbMg}_{1/3}\text{Nb}_{2/3}\text{O}_3\text{--PbSc}_{1/2}\text{Nb}_{1/2}\text{O}_3$ (PSN-PMN), where $\epsilon = 30,000$, piezoelectric module induced by not very large electrical field ($E_{\text{bias}} = 10^6\text{ V/m}$) is close in its magnitude to piezoelectric coefficients of commonly used in technique piezoelectric PZT ceramics ($d_{31} \cong 500\text{ pC/N}$).

In addition to important technical applications (in actuators, tunable piezoresonators and electrically tunable filters with resonance frequency and bandwidth management), electrically induced piezoelectric effect is interesting in a sense that it can explain the microscopic nature of electromechanical coupling. Correspondent physical models and the thermodynamic theory of piezoelectric effect will be discussed in the next sections.

Relaxor ferroelectrics are characterized by high permittivity ($\epsilon \sim 20,000\text{--}40,000$), and consequently, they have very high induced polarization. Comparison of induced polarization of relaxor ferroelectric $\text{Pb}(\text{Mg}_{1/3}\text{Nb}_{2/3})\text{O}_3 = \text{PMN}$ with a similar polarization of the paraelectric material $\text{Ba}(\text{Ti}_{0.6}\text{Sr}_{0.4})\text{O}_3 = \text{BST}$ (that also has rather high $\epsilon \sim 10,000$) is shown in Fig. 9.14. It is seen that induced polarization in PMN many times exceeds one of the BST.

Moreover, in the relaxor ferroelectric–induced polarization P_i depends on temperature (similar to the P_S of ferroelectrics), as it can be seen in Fig. 9.15. Therefore under electrical bias field, relaxor also exhibits electrically induced *pyroelectric effect*; the magnitude of *artificial pyroelectric coefficient* in the relaxor ferroelectric might exceed the pyroelectric coefficient of the usual pyroelectric. Moreover, in comparison with widely used artificial pyroelectric effect in the BST type paraelectrics, electrically induced pyroelectric effect in PMN should be much more thermally stable because of the less sloping dependence of $P_i(T)$.

Artificial piezoelectric effect in relaxors is the *linearized electrostriction*. In ferroelectric materials (paraelectrics and, especially, relaxors), electrostriction is large enough for practical applications. Explanation of this fact is shown in Fig. 9.16: electrical bias field E_b produces some constant internal strain x_0 at the parabolic

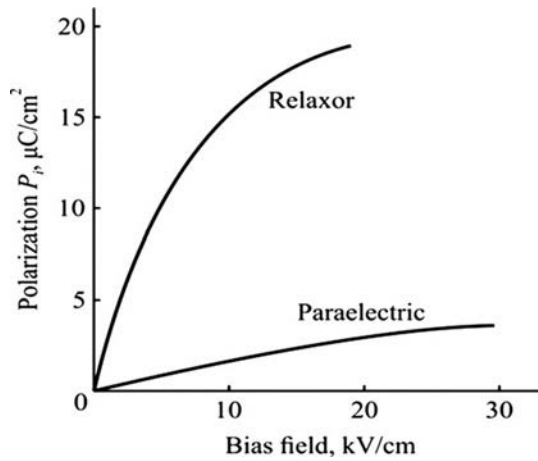


FIG. 9.14

Relaxor ferroelectric characteristics: comparison of electrically induced polarization P_i in PMN relaxor and P_i in paraelectric BST.

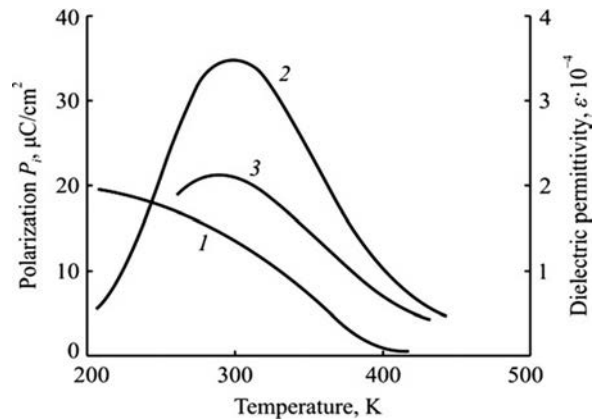


FIG. 9.15

PMN characteristics: 1—induced by $E_b=10\text{ kV}/\text{cm}$ polarization obtained by pyroelectric measurements; 2—permittivity without bias field; 3—permittivity at bias field.

dependency of strain x on field E . In addition to steady and relatively big bias field E_b , smaller alternating electrical field E' can be applied to a given dielectric material. As a result, the *pseudolinear "piezoelectric effect"* appears, which is shown in a new scale: $x' - E'$.

It is obvious that electrically induced piezoelectric effect is a particular case of electrostriction. As it is seen, relaxor ferroelectric shows much greater electrically induced strain x , even in comparison with paraelectric material, in which electrically

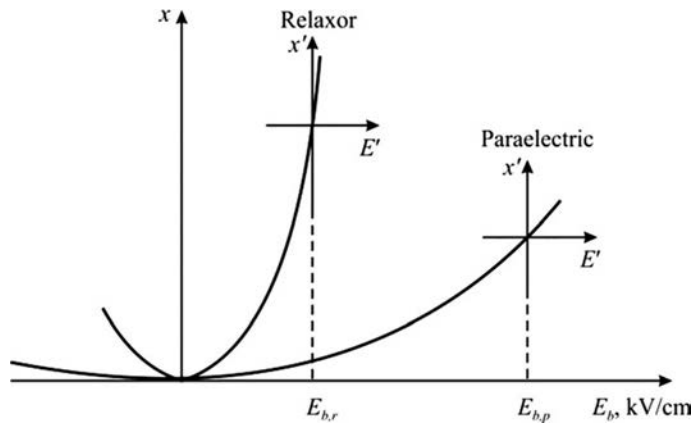


FIG. 9.16

Explanation of linearized electrostriction: induced piezoelectric effect in high- ϵ noncentrosymmetric dielectrics [3].

induced piezoelectric effect is also large. Previously, in Fig. 9.13C, some experimental tests that are commonly used in electronic ceramic materials are shown using relatively high dielectric permittivity ceramics (from $\epsilon \sim 100$ in TiO_2 to $\epsilon \sim 300$ in SrTiO_3). Any piezoelectric properties in this ceramics are absent without external bias field: only due to the bias field, these materials became artificial piezoelectric. Piezoelectric effect appears *instantly* after the bias field is applied, and it disappears *immediately* after the bias field is switched off. Correspondent response time is $< 10^{-9}$ s but in the mentioned ceramics, the electrically induced piezoelectric effect is small: $d \leq 0.3$ pC/N. In comparison with these materials in the relaxor ferroelectrics, electrically induced piezoelectricity is large, but relaxation time is longer.

Theoretical calculation for the electrically induced piezoelectric effect is given for the artificial module $d = 2Q\epsilon_0^2\epsilon^2E$. Here Q is the electrostriction coefficient, ϵ_0 is the electrodynamic electrical constant, ϵ is the material permittivity, and E is the bias field. It is obvious that electrically induced piezoelectricity is substantial only in dielectrics with very high permittivity. However, in the paraelectric material-induced piezoelectricity appears and disappears without inertia. At the same time, in some PMN-PSN relaxor with $\epsilon \geq 40,000$, piezoelectric modulus reaches $d = 2000$ pC/N (more than as in the best piezoelectric ceramics of the PZT type) (Fig. 9.17).

For comparison, it might be noted that paraelectrics show the transverse component d_{31} of electrically induced piezoelectric module two orders of magnitude smaller than the magnitude of the electrically induced module d_{31} in relaxor ferroelectric. Record high strain in electrical field shows crystalline structure of $\text{PbZn}_{1/3}\text{Nb}_{2/3}\text{O}_3$ -4.5% PbTiO_3 (PZN-4.5% PT). Its electrically tunable deformation is 10 times higher than the deformation in widely used piezoelectric ceramic PZT-8 and, unlike piezoelectric, allows large controllability without hysteresis.

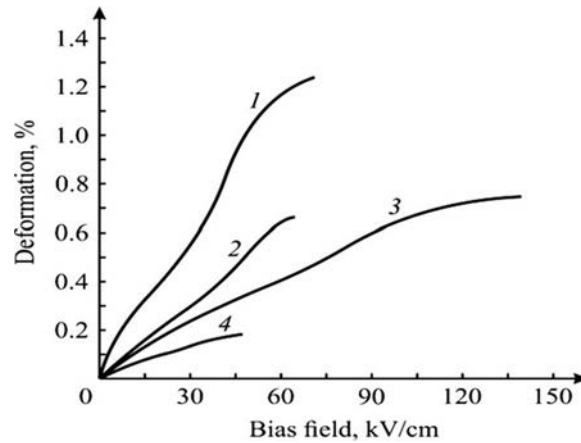


FIG. 9.17

Electromechanical effect in relaxors: 1—crystal PZN-4.5% PT; 2—crystal PZN; 3—crystal PMN-24% PT; 4—piezoceramics PZT-8.

Therefore, electrically induced piezoelectric effect occurs while bias field is applied and disappears immediately after field shutdown. The operation speed of such control depends on the inertia of relaxor ferroelectric polarization, whereas its performance is limited by the permittivity dispersion that appears at frequencies near hundreds of kilohertz.

9.6 PYROELECTRICS AND ELECTRETS

Some of dielectrics behave in the absence of external electric field as if they are polarized. This peculiar internal electrical state in the dielectrics can be *energetically advantageous* (being associated with a particular crystal structure—the presence of polar axis), hence *metastable* (which can be disturbed by external electrical influences). In the former case, internally polarized electrical state is characterized by the *spontaneous polarization* (pyroelectricity), whereas in the second case, it is characterized by *residual polarization* (electrets).

Pyroelectric effect traditionally is considered as a property of spontaneously polarized crystal, but it is preferable to assume that the structure of the pyroelectric crystal is such that it demonstrates electrical response when its temperature changes. Thermal energy in the pyroelectric can be converted directly into electrical energy owing to electrically active intrinsic structure of such crystal. Therefore, pyroelectric and piezoelectric presents solid-state energy converter; however, piezoelectric is the electromechanical (or vice versa mechano-electrical) power converter, whereas pyroelectric is the thermoelectric (or vice versa electrothermal) power converter.

In principle, energy transformation in solid dielectric is only possible if the crystal (texture or polymer) is polarized. However, in the absence of external influences, the existence of intrinsic polarization is not evident. The point is that intrinsic polarization at constant temperature should be completely compensated by the electrical charges, which are precipitated on the surface of polar dielectric or on its electrodes.

The so-called spontaneous polarization manifests itself only when external conditions *change dynamically*. As it was noted in the previous section, it might be the *change in stress* that results in the piezoelectric polarization of polar dielectrics. The pyroelectric polarization in polar dielectric also occurs only *due to temperature changes*.

Temperature increase or decrease alters the intensity of the thermal motion of particles in the polar dielectric, and therefore changes as the orientation of polar complex (molecules); hence, the distance between them leads to a change in spontaneous polarization. Consequently, on the surface of the polar dielectric, the uncompensated electrical charges appear. If a pyroelectric element with electrodes is connected to an amplifier, then the pyroelectric *current* flows through this device and then amplified. In the case of open-circuit crystal, the pyroelectric *voltage* appears on the crystal. Over time, however, if the temperature of the pyroelectric remains invariable, pyroelectric current (or pyroelectric potential) decreases gradually to zero.

Pyroelectric effect has been described (but not understood) in the ancient sources, mentioned around 2000 years ago by the Greek philosopher Pliny. This effect was observed in the semiprecious mineral tourmaline (with time, this crystal was called as “electrical” crystal). The term “pyro-” comes from the Greek word meaning “fire” because the effect is elicited during tourmaline heating in fire. When heating, the electric charges are generated, accompanied by cracking sound—electrical discharges. Moreover, heated tourmaline attracts objects with light. Recent measurements have shown that in a rather thin (around 1 mm) plate of tourmaline, electrical potential of around 1 kV is generated when temperature changes only by 10 degrees. However, it should be noted that tourmaline is still relatively weak pyroelectric.

Regarding the *electrical* phenomenon, the pyroelectric effect was qualified around 200 years ago by F. Aepinus. However, the main aspects of symmetry and physical mechanism of the pyroelectric effect have been described only in the early twentieth century by W. Voigt. Among minerals and among artificially synthesized crystals, pyroelectrics are relatively rare materials. The natural pyroelectric mineral is tourmaline: $\text{NaMg}[\text{Al}_3\text{B}_3\cdot\text{Si}_6(\text{OOH})_{30}]$ with different impurities, whereas synthetic pyroelectrics are lithium sulfate ($\text{LiSO}_4\cdot\text{H}_2\text{O}$), lithium niobate (LiNbO_3), potassium tartrate ($\text{K}_4\text{C}_8\text{O}_{12}\cdot\text{H}_2\text{O}$), and many others. Semiconductors of the $\text{A}^{\text{IV}}\text{B}^{\text{VI}}$ group (CdS , ZnO , etc.) by their symmetry also relate to pyroelectrics, but the pyroelectric effect among them is small. Interestingly, pyroelectric crystals contain sugar ($\text{C}_{12}\text{H}_{12}\text{O}_{11}$); this is the reason why these crystals are being used in homeopathic medicine.

All *ferroelectric* materials, potentially, are pyroelectrics because they are spontaneously polarized. However, to use ferroelectric as a pyroelectric element, it must

have a *single-domain* structure. Otherwise, the pyroelectric effect, which occurs in many different ways in oriented ferroelectric domains, is mutually compensated. Monodomainization of ferroelectrics (converting them into a homogeneously polarized structure) can be realized in many ways including temperature polarization: heating ferroelectric in the externally applied electrical field. Currently, to obtain single-domain ferroelectric-pyroelectric crystal, several methods for growing crystals are developed, which permits already in the process of crystal growth to get practically a single-domain structure.

Pyroelectrics are applied in electronics as uniquely sensitive uncooled temperature sensors and as detectors of radiation. Compared with semiconductor temperature sensors, pyroelectric sensors have several advantages: they do not require cooling (can be used at room temperature) and show the wide spectrum range of sensitivity. Numerous technical applications stimulate rapid development of the physics of pyroelectric. Presently, dozens of new pyroelectrics are synthesized and investigated, and many of them have already found technical application.

Thus, it is supposed that the pyroelectric effect is caused by the temperature change during spontaneous polarization in the *polar crystals*; however, a similar effect can be artificially induced in any solid dielectric, if an external electrical field is applied to them. Without application of an external electrical field, and in the absence of mechanical influences, change in polarization with temperature is possible only in those crystals in which a peculiar structure is traditionally described as spontaneous polarization.

Simplified model of pyroelectric effect. Pyroelectrics are closely connected to piezoelectrics. In all pyroelectrics, piezoelectric effects (direct and inverse) are also observed, but only in the pyroelectric, it is possible to observe the *volumetric piezoelectric effect*, that is, the electrical response to hydrostatic pressure, and this is very important for practical applications.

The pyroelectric effect can be explained by the simplest model of one-dimensional polar crystal, as shown in Fig. 9.18. This model of pyroelectricity involves a simple molecule consisting of a pair of ions separated by distance a , which is larger than distance b of the neighboring polar unit cell. This asymmetry is explained by a large difference in the electronegativity of positive and negative ions.

The hidden (or latent) internal polarity of pyroelectrics is no more than the *ability to provide electrical (vectorial) response to any nonelectrical scalar impacts*, in a given case—when temperature changes. To describe this ability, it is assumed that the polar crystal has intrinsic electrical moment P_S , which summarizes many elemental moments p_0 ; that is why each unit cell is marked in Fig. 9.18B–D as simple dipoles. A possible example of such structure is spontaneously polarized (at low temperatures) pyroelectric crystal HCl; wherein Fig. 9.18A, is intended to remind the wide-gap semiconductor of CdS ($A^{II}B^{IV}$ type crystal) belonging to pyroelectrics of $6\text{-}mm$ point symmetry class.

In this one-dimensional model, one can observe not only the pyroelectric effect but also the piezoelectric effect, which contributes to the pyroelectric response. Indeed, mechanical stretching or compression of shown dipole chain results in the

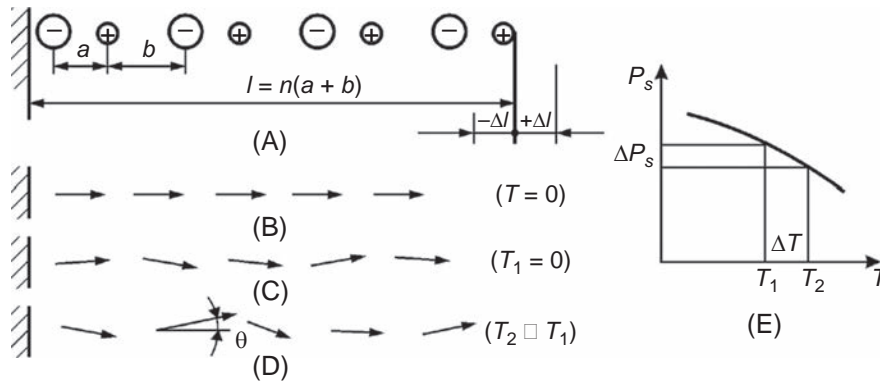


FIG. 9.18

One-dimensional model of the pyroelectric crystal: (A) polar file of two-ion molecules; (B) dipole moment orientation without thermal motion ($T=0$); (C and D) different degrees of thermal disordering and thermal expansion of a dipole chain; (E) spontaneous polarization dependence on temperature.

change in specific electrical moment: $P \sim \Delta l/l$. Thus not only from general considerations but also from simple model, it follows that any pyroelectric should have piezoelectric properties (but opposite conclusion is not fair).

Thus, as shown in Fig. 9.18B, polar molecules are replaced by the arrows that present single dipole moments. In the idealized state (when absolute temperature $T=0$), all dipoles are strictly oriented. As temperature increases, the thermal chaotic motion, first, results in a partial disordering of dipoles, and, second, leads to thermal expansion of the crystal. In mechanically *free crystal*, both these mechanisms give rise to spontaneous polarization P_s decrease with increasing temperature, as shown in Fig. 9.18E. The first mechanism (polar unit disordering) is *always* found in any polar crystal (clamped or free), but the second mechanism (crystal thermal expansion) is possible to observe only in *mechanically free crystal*.

Temperature increase changes P_s in the *linear* (“hard”) pyroelectric such as tourmaline or lithium sulfate crystals mostly due to their thermal expansion or compression. This type of pyroelectricity is produced by the *piezoelectric conversion of thermal strain* that is referred as *secondary pyroelectric effect*. The temperature change of P_s of the *nonlinear* (“soft”) pyroelectrics (which include most of the ferroelectrics) is caused mainly by *thermal disordering* of the dipole structure. Dipole orientation alteration results in the *primary pyroelectric effect*.

Owing to the fast decrease in spontaneous polarization with temperature increase in ferroelectrics near the Curie point (abrupt change in dependency: dP_s/dT), exactly, these materials are mostly used as the pyroelectric sensors. In the model, as shown in Fig. 9.18, the elementary electrical dipole moment change is $dp = p_0(1 - \cos \theta)$. As the angle θ is small, it can be considered as being proportional to the intensity of thermal motion: $\theta \sim k_B T$. Therefore the change in polarization is $\Delta P = \gamma^{(1)} \Delta T$, where $\gamma^{(1)}$ is the *primary pyroelectric coefficient*.

For secondary pyroelectric effect, change in the proportionality of ΔP and ΔT is, first, the result of thermal expansion linear law: $\Delta l = \alpha \Delta T$, where Δl is the linear deformation, whereas α is the thermal expansion coefficient. Second, the linear direct piezoelectric effect is described as $\Delta P = e \Delta l / l$, where e is the piezoelectric strain constant. From these two formulas, it is possible to obtain a linear equation for secondary pyroelectric effect: $\Delta P = \gamma^{(2)} \Delta T$, where $\gamma^{(2)}$ is the *secondary pyroelectric coefficient*.

Consequently, considering the deposits from both mechanisms of pyroelectricity, it is possible to obtain an equation for thermally induced polarization

$$\Delta P = (\gamma^{(1)} + \gamma^{(2)}) \Delta T,$$

where temperature T is the scalar value, but polarization is the vector value, and pyroelectric coefficient γ is the vector. However, it is a peculiar (“*material*” vector) that differs fundamentally from the “force-type” vectors (such as vectors E , D , or P).

A first-rank tensor material describes the spatial distribution of pyroelectric response in a crystal; appropriate *indicatory surface* (indicatrix) is shown in Fig. 9.19, being represented by two spheres. They are located above and below of the symmetry plane m and can be characterized by the equation $\gamma(\varphi) = \gamma_{\max} \cos \varphi$. It is evident that the spatial distribution of pyroelectric coefficient corresponds to spontaneous polarization orientation in the polar crystal: $P = P_{\max} \cos \theta$. The upper sphere is an indicatory surface for the upper orientation of P_s , whereas the bottom sphere means only a change in sign of γ_i , if spontaneous polarization has the opposite direction.

Material vector γ shows maxim in that direction of ordinate, which coincides with spontaneous polarization direction. Hence, the γ_{\max} might be measured in a cut of the crystal that is perpendicular to the polar axis. The angle φ between the ordinate and the vertical to slanting cut of the crystal determines the magnitude of the pyroelectric

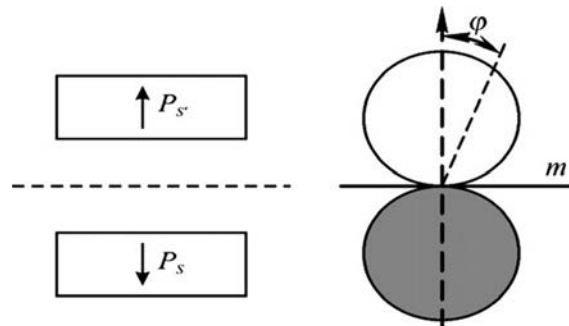


FIG. 9.19

Orientation of spontaneous polarization in crystal and corresponding guide surface (indicatrix) for pyroelectric coefficient; the *black area* shows the negative part of pyroelectric coefficient.

effect. By the radius vector drawn from the center of a figure (shown in Fig. 9.19), it is possible to determine pyroelectric effectiveness in any *slanting cut* of a crystal. It is obvious that perpendicular to ordinate, any pyroelectric effect is absent.

Pyroelectricity, as piezoelectricity, is determined by crystal symmetry. However, if in event that piezoelectric properties, the necessary condition is the *absence of the center of symmetry* in the crystal, the pyroelectric effect is possible only in the crystals that have a special element of symmetry—the *peculiar polar axis*. It is this axis that provides polar acentricity of a crystal, so that any pyroelectric should also have the piezoelectric properties (but not vice versa).

Of the 20 classes of piezoelectric crystals (described in the previous section in Table 9.4), only 10 classes are the pyroelectric ones; their designations are as follows: 1, 2, 3, 4, 6, m , $2m$, $3m$, $4m$, and $6m$. As it was noted earlier, the number indicates the order of polar axis, whereas the letter m means the plane of symmetry, which passes through the polar axis [11].

In addition to polar crystals, the polarized ferroelectric ceramics also have pyroelectric properties: under conditions of increased temperature and under the influence of externally applied electric field, most of the ferroelectric domains (spontaneously polarized microregions) become oriented. Thus after cooling down to normal temperature and turning off the field, the pyroelectric *texture* occurs with a group of polar symmetry $\infty \cdot m$ (∞ is the order of symmetry axis). Because of mechanical strength and high chemical resistance, the polarized ferroelectric ceramics, as well as polar crystals, are used in the pyrometry, although pyrosensitivity in the ceramics is less than that in ferroelectric crystals.

As in the case of piezoelectrics, the *boundary conditions* are very important for the *electrothermal* effects (pyroelectric and electrocaloric). All these effects can be described by eight different linear relationships—depending on the combination of various boundary conditions, under which polar crystals are studied or explored.

The *electrical boundary conditions* for pyroelectrics are similar to those of the piezoelectrics. The first condition corresponds to *electrically free* pyroelectric when $E=0$, which means that the entire surface of the crystal is equipotential. As the inductance $D=\epsilon_0 E+P$, then in this case, $D=P$. This electrical condition is performed by the shorting of electrodes deposited on the pyroelectric; in practice, a pyroelectric element is loaded on the input resistance of an amplifier (which has a resistance thousand times less than the resistance of pyroelectric). The condition $E=0$ results in pyroelectric *current*, that is, the crystal is the source of current.

Second idealized electrical boundary condition ($D=0$) means that the polar crystal is electrically disconnected, $D=\epsilon_0 E+P=0$. The implementation of this condition means very low conductivity of the pyroelectric; in this case, pyropolarization $\delta P=\gamma\delta T$ that occurs is compensated by the electrical field: $\epsilon_0 E=-\delta P$, that is, the crystal is the source of voltage. Originated by pyroelectric effect, voltage $V=El$ can be measured by high-impedance (static) voltmeter (l is the thickness of pyroelectric element).

The *thermal boundary conditions* are two idealized cases:

$T = \text{const}$, *isothermal condition* means the invariable temperature, when the crystal during its measurements (or exploitation) has enough time for energy exchange with the environment. As a rule, the thermodynamic theory uses exactly this approximation as a comfortable case for theoretical calculation (only isothermal condition is supposed in theories of ferroelectric phase transition). In practice, correspondent experiments should be provided in the quasistationary conditions.

$S = \text{const}$, *adiabatic condition* means permanent entropy, when there is no energy exchange with the environment during measurement or exploitation. This is the usual case at comparatively high frequency.

Similar to the electromechanical coupling coefficient K_{EM} in piezoelectrics, analog power conversion factor—the coefficient of *thermoelectric coupling* K_{TE} —might be introduced for pyroelectrics (it plays the role of efficiency). In basic pyroelectric materials, this ratio is seen in the range of 1%–4%. Such relatively low efficiency of thermoelectric power conversion K_{TE} is due to the physical nature of this phenomenon in crystals, which tend to be “electrically hard” relative to external influences.

To describe the efficiency of pyroelectric sensors that convert infrared radiation into electrical energy, not only pyroelectric coefficient γ but also several *quality parameters* are evaluated: γ/C_V , $\gamma/(C_V\epsilon_0\epsilon)$ and $\gamma/[C_V(\epsilon_0\epsilon tg\delta)^{1/2}]$, where C_V is the volumetric heat capacity. There are a number of options for pyroelectrics; these quality parameters define the *current sensitivity* ($S_J = \gamma/C_V$) and the *voltage sensitivity* $S_V = \gamma/(C_V\epsilon_0\epsilon)$.

The *first* group of pyroelectrics are *nonlinear* pyroelectrics-ferroelectrics: triglycine sulfate and crystals isomorphic to it (they are grown with special additives for the purpose to obtain single-domain samples); lithium niobate and lithium tantalate (polarized by the current bias during crystal growth); thin films of potassium nitrate in the ferroelectric phase; lead titanate and lead zirconate-titanate polarized ceramics with different impurities. Pyroelectric properties of ferroelectrics are mostly due to the *primary* pyroelectric effect. Near Curie point (T_C), when spontaneous polarization change with temperature is expressed very strongly, pyroelectric coefficient reaches maximum; hence the pyroelectric effect can be used with maximal efficiency.

The *second* important group of pyroelectrics is the *linear* pyroelectric crystals. In contrast to ferroelectrics (which are usually divided into domains with arbitrary direction of P_S), the linear pyroelectrics P_S has the same direction throughout the crystal. Furthermore, this direction cannot be changed by an external electrical field. The value of P_S in the linear pyroelectrics, varying with the temperature, never decreases to zero (as in ferroelectrics). These crystals may belong to pyroelectrics of the CdS-type ($A^{II}B^{VI}$ crystals with wurtzitic structure), as well as lithium sulfate, lithium tetraborate, and others. It is important to note that in such pyroelectrics the

contribution from *secondary* pyroelectric effect dominates, exceeding the contribution of the primary pyroelectric effect.

The *third* group of pyroelectric materials is *polar polymers* such as PVDF film. Owing to special processing, which involves film stretching 3–5 times with the next temperature polarization (in field nearly 1 MV/cm at 130°C), the polymeric film acquires pyroelectric properties. Despite the fact that the pyroelectric coefficient of polymeric materials is lower than those in single crystals and pyroelectric ceramics, technical application of pyropolymers is very promising owing to their excellent mechanical properties (thin and elastic films).

According to Curie principle, any linear effects in crystals must have the *opposite effect*. For example, opposite to the direct piezoelectric effect is the inverse piezoelectric effect. Similarly, inverse to the pyroelectric effect is the *electrocaloric* effect. This effect can be applied for electrically controlled reduction in temperature (e.g., to achieve better cooling). Thus, pyroelectric effect not only can convert thermal energy into electrical energy but also vice versa. Controlled by voltage, the electrocaloric cooling (or heating) depends on the polarity of the applied electrical field. The alternation voltage can generate extended temperature wave.

In polar crystals, the electrocaloric effect influences the value of permittivity. When thermal equilibrium in the studied sample is entirely established at the time of electrical field application (possible at very low frequency), the pyroelectric crystal completely absorbs electrical energy applied to a crystal and converts it into thermal energy. This is the isothermal process of pyroelectric polarization, which is characterized by *isothermal permittivity* ϵ^T . On the contrary, in case of rather fast-changing electrical field, the energy process is the adiabatic one (thermal equilibrium has no time to be set). It appears that there is a decrease in the capacitance of pyroelectric element. Therefore at higher frequency, the *adiabatic permittivity* $\epsilon^S < \epsilon^T$ can be determined.

Residual (quasipermanent) polarization of electrets. Sometimes, relaxation time of the polarization process can be extremely large and it exceeds many months. In such dielectric usually *polarized state exists*, although this state is thermodynamically unstable: this is the *electret* type of materials. Therefore in addition to spontaneous polarization of pyroelectrics, some solid dielectrics may have another kind of constant polarization that exists without application of an external electrical field to the dielectric. The structure of electrets must be inhomogeneous, so polarization, originally initiated by an electrical field (or by other influence), is stored in the electret structure for a long time, although remaining metastable.

Typically, electrical polarization is induced by an external electric field, but after this, when the field is switched off the polarization quickly disappears, and this returns the dielectric to its equilibrium state that usually corresponds to the nonpolarized state. However, in some cases, the polarized state remains for a long time after the external field is switched off: this is the *residual* polarization. It is typical for inhomogeneous dielectrics and it can be caused, for example, by “freezing” of some mechanisms of thermally activated or migratory polarization, when relaxation time for one or the other reason can be sharply increased. This quasipermanent

polarization can also be created by the electrical charges, embedded in the dielectric and fixed in the surface or in the volumetric “traps.” If residual polarization is not shielded by the metallic electrodes, such dielectric creates *electrostatic field* in the surrounding space, similar to how a permanent magnet generates magnetic field.

Dielectrics that maintain electrical charges in their volume or on surface for a long time and create in the surrounding space constant electric field are called the electrets.

There are different methods for producing electrets. The usual way is to create residual polarization by a strong electric field and use *additional activation impact* on dielectric. Additional influence accelerates the process of polarization: orientation of dipoles, polar complexes, radicals, and domains. It might be also a process of electrification: migration of electrons or ions and their attaching by traps. Depending on the nature of the activation effect (heating, lighting, radiation, magnetic field, and mechanical stretching), the electrets are classified as thermal, photo, radio, magnetic, mechanical, and other electrets. Although electrets are in the metastable state, which is formed from very low leakage materials, they can retain excess charge or polarization for many years.

Quasipermanent (residual) polarization can be configured by two main mechanisms, as shown in Fig. 9.20:

- *heterocharge* formation, in which the sign is opposite to the charges on forming electrodes.
- *homocharge* formation, captured in dielectric, which has a sign same as that of the charges on electrodes.

Important conditions for long-time existence of formed electrets is low electrical conductivity of the original dielectric ($\sigma < 10^{-11}$ S/m) and large energy barriers for dipole relaxation or trapped charge carriers.

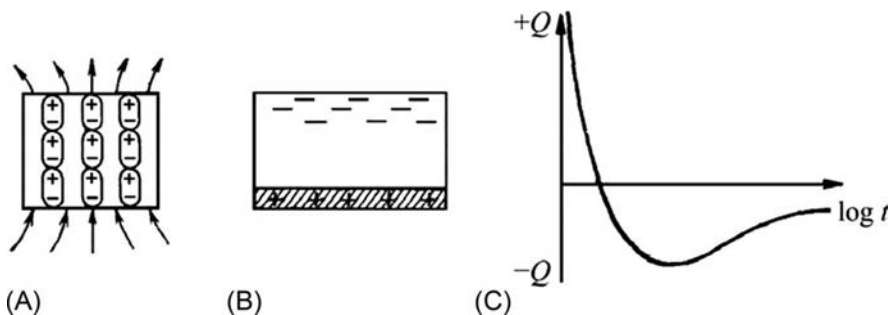


FIG. 9.20

The main mechanisms of residual polarization while formation of electrets: (A) macrodipole polarization; (B) implantation of electrical charge (electrons); (C) surface charge variation in time.

The *thermoelectrets* can be obtained by heating of heterogeneous dielectric in the strong electrical field. By this method, homocharge and heterocharge can be formed. The homocharge can be obtained, for example, owing to thermostimulated injection of electrons from cathode (or holes from anode). The heterocharge appears through various “impurity-type” polarization mechanisms: thermally activated polarization (electrons, ions, or dipoles), migration polarization (space charge, according to Maxwell-Wagner mechanism).

Under electrical field influence, the *absorption current* passes through a heated dielectric, thus indicating that spatially separated space charges are really accumulated in a dielectric and resulting in macroscopic electrical moment—polarization. After such forced “temperature polarization,” the dielectric is cooled without electrical field turning off; hence, its polarized state becomes “frozen.”

At the time of thermoelectret formation, namely, during the cooling period, a tremendous increase in the relaxation time occurs. As usual, $\tau \sim \exp(U/k_B T)$, but for a substantial increase in τ , that is, to increase the duration of electret state existence, not only exponential dependence $\tau(T)$ is required but also the possibility of an abrupt increase in τ at the expense of *potential barrier* U changing. Therefore the waxy dielectrics, during thermoelectret preparation, can be polarized at temperatures higher than their melting point; in the same way, polymers can be polarized at a temperature higher than that of their glass transition, whereas ferroelectrics can be polarized at a temperature higher than their Curie point.

Because of fundamental instability of electrets, their properties vary with time; however, the main changes occur immediately after their manufacture. The overpatching is seen similar to those in the heterocharge and homocharge, and the net charge of electrets may subsequently change the sign.

The thermoelectret *materials* include [12] the following:

- Mixtures of amorphous materials (wax, bitumen, and tar); only from these materials, the first electrets were prepared; currently, such electrets have no technical applications because they are nonthermally stable and have low mechanical strength.
- Ceramic-based materials (calcium and strontium titanates, rutile, etc.) as well as glasses, pyrocerams, and various single crystals (oxides, fluorides, chalcogenides). These electrets have sufficiently stable properties but have not enough manufacturability.
- Ferroelectric ceramics (mostly, based on lead zirconate-titanate), for which temperature polarization is provided by domain field orientation. Ferroelectric electrets differ from polarized piezoelectric elements by lack of electrodes.
- Polymeric materials manufactured by thermoelectret technology (for instance, polyvinylidene fluoride).

Residual polarization can be obtained in almost any solid dielectric; if thermally activated, charge carriers in the presence of polarizing field are captured by traps—structural defects. Then, while heating, thermally activated current of depolarization

enables to determine the energy structure of dielectric, as well as the concentration of defects in crystals or polymers.

In addition to thermoelectrets, many other mechanisms of residual polarization conservation are known, as shown in Fig. 9.21.

The **photoelectrets** are composed of dielectrics that have low dark conductivity ($\sigma < 10^{-12}$ S/m), but increased photosensitivity. Formation of such electrets depends on the magnitude of the electrical field and light brightness. At illuminated places of photosensitive dielectric, because of the photoelectric effect, charge carriers occur and then they drift to the shadow places, thus settling on traps and forming the homocharge.

As a result, after electrical field and light turning off, an “electrical image” appears on the surface of photoelectrets. It can be read by electronic beam reflection or by using pigment powder, whose particles are attracted to the charged places by electrostatic forces. The effect is reversible: photoelectrets image can be eliminated (“erased”) by the strong electrical field and by the light striking of the photosensitive layer.

The main materials used for photoelectrets are photosensitive dielectrics, which are convenient for large plate manufacturing (selenium, zinc oxide, zinc and cadmium sulfides, selenides, and others). The photoelectret state can be also obtained in many crystals and polycrystals, for example, sulfur, anthracene, naphthalene, silikossilinit ($\text{Bi}_{12}\text{SiO}_{20}$), and others. Photoelectrets are widely used in the xerography (“dry pictures”) and photocopier devices.

Typically, the process of xerography somewhat differs from photoelectret polarization because it uses the light depolarization of the electret layer. The electrophoto-sensitive plate is the substrate coated with a thin layer of photoelectrets, which is charged in advance by the corona-discharge in the dark. Then the image is projected onto the plate for reproduction. Owing to low dark conductivity, the homocharge of electrets persists for adequate time. In the alight places, charge relaxes,

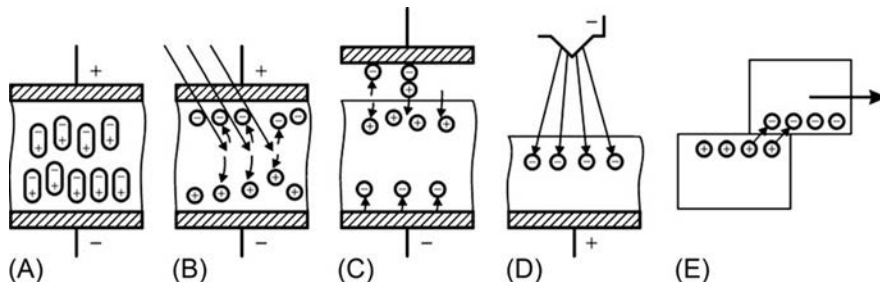


FIG. 9.21

Various types of electrets: (A) thermoelectrets; (B) photoelectrets; (C) electroelectrets; (D) radioelectrets; (E) triboelectrets.

After B.M. Tareev, *Physics of Dielectric Materials*, Energia, Moscow, 1973.

thus creating latent electrical image; then by using pigment powder, the image is transferred to the paper.

Electroelectrets. In some cases, for residual polarization formation, only sufficiently strong electric field is used (without an additional activating effect). Such electroelectrets, preferably, are made from the polymeric films. The homocharge of electroelectret is formed on the upper surface of a film by its electrification (bottom surface of the film is precoated using a metallic electrode).

Different methods can be used for electrification, for instance, the injection of space charge in the polymer might be applied. The corona type of electrical discharge is often used over the surface of the electrified film, whereas bombarding electrons are fixed on the surface “traps.” Some of these electrons diffuse into the dielectric, filling “traps.” Compensating charge is formed on the bottom electrode.

In some cases of electrets formation, the spark discharge in the gas located above the surface of the polymer is used. Injection of electrons in this case is more intense, but the spatial distribution of charge is less homogeneous. To improve the homogeneity of homocharge, contact methods of electrification are used when the electrical field is applied to the polymeric film through a thin layer of liquid dielectric.

Electronic beams are also employed for homocharge formation. The promising method to obtain highly stable electrets is also the method of ionic implantation (which is usually used when semiconductor materials are doping). These technologies enable to control the amount of injected charge and the energy of electrons (or ions), bombarding the polymer, as well as set the depth of their penetration into the polymer. Thus, it is possible to control the distribution of homocharge in the electrets.

Electroelectrets can be made from almost any polymeric dielectric. One of the best materials for such electrets is the polytetrafluoroethylene (PTFE) and its various copolymers on the basis of $[-CF_2-CF_2-]_n$, $[CF_2-CH_2]_n$, and so on. These elastic materials are thermally and moisture resistive, thus possessing high mechanical and electrical strength as well as extremely low conductivity: $\sigma < 10^{-14}$ S/m. These parameters determine the long-term stability of polymeric electret properties.

Radioelectrets are obtained by the irradiation of dielectrics by fast particles or by hard radiation. Regardless of the nature of initial particles under the influence of illumination, the activated electrons occur and then they are captured by structural defects to form a space charge. In radioelectrets, the division of positive and negative charges takes place, which leads to the formation of residual polarization (electrets state). This division can be obtained because of the application of an external electrical field and also without it. In latter case, electrical charges are separated because of uneven absorption of hard radiation in the volume of dielectric: further, electrical field gradient is induced, which distributes electrons and holes in the traps. However, this method of electrification is not widely applied in the electrets but in the radiometers to measure the doses of radiation.

The state of electrets in dielectrics can be obtained by some other methods. For example, to create polarized state in the *magnetolectrets*, the electrical and magnetic fields are used simultaneously. The *mechanolectrets* are obtained by means

of mechanical deformation of some dielectrics, usually by stretching of polymeric films. Polar polymers under mechanical tension acquire noncentrosymmetric structure and become polarized (polyvinylidene fluoride).

Triboelectrets are obtained by the friction of two different dielectrics. For this purpose, dielectrics with different electronic work function are used: the electrons from the dielectric with a low work function move into the dielectric with a high work function. It is interesting to note that historically first electrical phenomena were observed with the friction of amber on fabric (Greek name of amber is “electron”), and exactly, this phenomenon is used to name such scientific and technical fields as “electrical technique” and “electronics.”

Thermally stimulated depolarization. The *thermally stimulated depolarization* of charged electrets is widely used to analyze energy levels, characteristic times, and other *activation parameters* of molecular motion in the dielectrics and semiconductors, to study their structural and physical transformations, and to investigate mechanisms of accumulation and relaxation. During this process, the studied dielectric generates electrical current, which changes with time when heating the sample. The experimental method of the *thermally stimulated current (TSC) spectroscopy* is used to study **energy levels** of traps in dielectrics and **semiconductors** and can be applied also to study many electrophysical phenomena in these materials; particularly, the TSC method is widely used to study the physics and chemistry of polymers.

The experimental procedure of the TSC method comprises the following stages:

- application of a direct electrical field E_0 (or other activation factors) to the studied sample at primordial temperature T_1 ;
- sample cooling under the action of field to lower temperature T_0 (optical or electrical injection of charge carries may be used);
- switching off (or change) polarized field (or other activation factor);
- gradual heating of the sample at a *constant rate* and recording emergent electrical current as a function of time/temperature.

After the electrical field, which produced polarization, is switched off, temperature-stimulated transition gradually occurs from the nonequilibrium state (created by field and additional influence) to the unpolarized equilibrium state. In this case, if the bonded charges in a sample liberate themselves, temperature dependence of the measured current shows some peaks of *thermally depolarization current*. Theoretical description of TSC spectra uses relaxation equation of dielectric depolarization with dependence on temperature relaxation time. Linear law of sample heating enables change in these equations from temporary-variable to temperature-variable. Comparison of TSC spectra with theoretical models provides information on the energy state of impurities.

Many dielectrics and high-resistance semiconductors after their excitation by external radiation (e.g., strong electrical field or radiation) are able to maintain their excited state for a long time. As described in the previous section, this electret state can be formed, for example, by the macrodipole orientation, by the appearance of

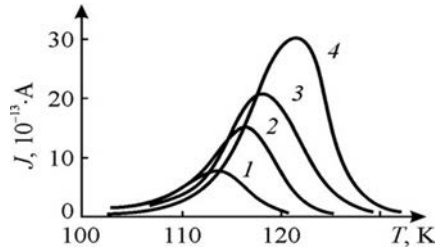


FIG. 9.22

Thermally stimulated current from polarized AgCl at different heating rates: $v_1 < v_2 < v_3 < v_4$.

thermoexcited electrons and then captured by traps, as well as by the arising polarons that form space charge.

Therefore if a polarized dielectric (that has electrodes) is subjected to gradual heating, then in the external circuit (capacitor closed to resistor) the thermally stimulated current appears and changes over time. This current is caused by partial or complete disorientation of dipoles, by liberation of charge carriers from traps, and so on. Thermally stimulated current is a function of temperature, and it has a characteristic shape—the highs $j(T)$ whose magnitude and position depend on the type of dielectric, conditions of excitation, and rate of temperature increase, as shown in Fig. 9.22. If the residual (quasipermanent) polarization P_{res} has been established in the field E_0 by oriented particles of one type with elementary electrical moments p_0 and bulk density N , then $P_{\text{res}} = N(p_0^2/3k_B T)E_0$. Under normal conditions of constant temperature, $P_{\text{res}}(T)$ decreases with time according to the exponential law, and depolarization current density is defined as

$$j(T) = P_{\text{res}}(t)/dt = P_{\text{res}}(t)/\tau.$$

During the TSC study, the temperature increase over a time is controlled by the linear law from the initial value T_0 : $T = T_0 + bt$, where $b = dT/dt = \text{const}$. Depolarization current occurs because the relaxation time of polarized polar particles (or complexes) varies as $\tau = \tau_0 e^{U/k_B T}$, where τ_0^{-1} is the relaxation frequency of particles and U is the activation energy. As temperature increases, depolarization current initially increases (due to τ temperature dependence) and then decreases because the concentration of excited particles gradually decreases, as shown in Fig. 9.22. The activation energy can be determined from a plot of $j(T)$ at the initial section of the chart, if the scale is: $\ln j(T) = \text{const} - U/k_B T$.

9.7 HIGH-PERMITTIVITY DIELECTRICS AND PARAELECTRICS

Different dielectrics with a given value of ϵ and $TC\epsilon$ (temperature coefficient of permittivity) are widely used in the ceramic capacitor production. Owing to the fact that in most cases the capacitors are used at high frequencies, only fast polarization mechanisms can play a role in them. One of the requirements for the properties of capacitors is their thermostability.

Concerning, it is necessary to mention that *electronic* polarizability somewhat increases the value of ϵ but stimulates *negative* $TC\epsilon$. Conversely, if the predominant mechanism is *ionic* polarization, $TC\epsilon$ becomes *positive*. Electronic devices, in particular electrical capacitors, need dielectrics with different ϵ values and various sign $TC\epsilon$. Capacitors with various $TC\epsilon$ are applied for temperature compensation of other elements of electronic apparatus. In many cases, it is necessary to have $TC\epsilon \approx 0$ in a broad temperature range. Importance of this problem is also attributed to the fact that the relative number of ceramic capacitors in equipment reaches 50%.

Contemporary technology of polycrystalline solid solutions enables to obtain different ϵ and $TC\epsilon$ values. This is especially important at high frequencies (microwave ϵ_{mic}). During technical dielectric designing, it is considered that change in magnitude and sign of $TC\epsilon$ depends on what dominates the polarization of the dielectric—electronic or ionic. The lattice (ionic) polarization also depends on electronic shell displacement, but frequency characteristic of this mechanism of polarization is given by the elastic forces and masses of ions. Therefore, this frequency is much less than optical frequency but quite adequate for electronic applications.

The role of electronic shell displacements is especially important for high-permittivity dielectrics. The fact that short-range repulsion force between ions depends mainly on the interaction between shells of adjacent ions, this interaction is smaller for “soft” shells. If the crystal structure is such that the elastic force of ion repulsion is reduced, then ϵ_{ir} (due to “contribution” of infrared polarization) can be very large. Thus, electronic polarization of ionic crystals affects the frequency of infrared vibrations. Accounting for this effect leads to reduction in frequency of transverse optical phonons ω_{TO} . As the frequency of longitudinal lattice oscillations ω_{LO} is practically unchanged with temperature, then, from the Lideyn-Sachse-Teller relation, it follows that only the value of frequency ω_{TO} determines ϵ_{mic} .

If it is considered that electronic (optical) polarization affects ionic polarization (infrared), then instead of the simple expression $P = N\alpha F$ (where α is the polarizability, N is the concentration of polarizable particles, and F is the acting Lorentz force), it is possible to obtain the equation

$$P = \left[a + \frac{nq^2}{m} \frac{1}{(\omega_0^2 - \omega^2)} \right] \left(E + \frac{1}{3\epsilon_0} P \right), \quad (9.24)$$

where $1/(3\epsilon_0)$ is the Lorentz factor and a is a parameter, considering the effect of optical polarization.

After some transformations, for frequencies of transverse and longitudinal optical phonons as well as for dielectric contribution from infrared polarization, the following equations can be obtained [13]:

$$\begin{aligned} \omega_{TO}^2 &= \frac{c}{m} - \frac{nq^2}{3\epsilon_0 m} \frac{\epsilon(\infty) + 2}{3}; \\ \omega_{LO}^2 &= \frac{c}{m} + \frac{2nq^2}{3\epsilon_0 m} \frac{\epsilon(\infty) + 2}{3}; \\ \epsilon(0) - \epsilon(\infty) &= \frac{nq^2}{\epsilon_0 m \omega_{TO}^2} \frac{(\epsilon(\infty) + 2)^2}{9}. \end{aligned} \quad (9.25)$$

These formulas might be reduced to a simplistic form, if the influence of electronic-shell polarization would be neglected, that is, make substitution $\epsilon(\infty)+2=3$. Note that in formula (9.25), there is multiplier $(\epsilon+2)/3$, which is the ratio between Lorentz local field F and external electric field E .

Expression for frequency ω_{LO} is the sum; therefore, inclusion of electronic polarization almost does not change the frequency ω_{LO} (as opposed to the value of ω_{TO}). From Eq. (9.25), it is seen that the lower the frequency ω_{TO} , the greater is the value of ϵ_{mic} (measured at microwave frequencies). Comparison shows that calculations by formula (9.25) are very close to the experimental data for cubic alkali halide crystals. For example, these parameters for NaCl and TlBr crystals can be compared: in NaCl, $\epsilon(\infty)=2.25$, $\omega_{TO}=3.1 \times 10^{13}$ Hz, and $\epsilon(0)=5.6$ ($\epsilon_{\text{exper.}}=5.6$); crystal TlBr has higher optical refractive index, respectively, it has $\epsilon(\infty)=5.1$, $\omega_{TO}=1.1 \cdot 10^{13}$ Hz, and $\epsilon(0)=30$ ($\epsilon_{\text{exper.}}=31$).

Formula (9.25) enables to describe temperature dependence of ϵ in the ionic crystals. In the majority of the ionic crystals with $\epsilon=4-8$, the temperature coefficient is positive ($TC\epsilon > 0$), which distinguishes them from dielectrics that are characterized mainly by electronic polarization and have $TC\epsilon < 0$. In those ionic crystals, where $\epsilon > 10$, the infrared polarization is significantly affected by electronic shell polarization, and therefore $TC\epsilon$ might be negative. This interesting (and important for technical applications) result follows from the analysis of formula (9.25).

Formula for frequency ω_{TO} shows that with increasing temperature, because of thermal expansion, on one hand, the term c/m reduces, but, on other hand, dependence on $\epsilon(\infty)$ subtrahend decreases. The value of $\epsilon_{mic}(T)$ depends on which of these effects will be overwhelming.

In crystals with high electronic polarizability, the effect of reducing $\epsilon(\infty)$ with increasing temperature predominates, which leads to ω_{TO} increase with temperature and to temperature decrease of $\epsilon_{mic} \sim 1/\omega_{TO}^2$, respectively. Thus crystals with big ϵ , unlike other ionic crystals, are characterized by $TC\epsilon < 0$. In connection with these examples, it should be noted that in NaCl crystals ($\epsilon_{mic}=5.6$), temperature coefficient of ϵ is positive: $TC\epsilon = +4 \times 10^{-5} \text{ K}^{-1}$, whereas in TlBr crystals ($\epsilon_{mic}=31$), it is negative: $TC\epsilon = -2 \times 10^{-3} \text{ K}^{-1}$.

Paraelectrics of displacement type. Nonpolar ionic crystals with $\epsilon \sim 100$ (and much above) occupy a special place among dielectrics with high permittivity. Typical representatives of such dielectrics are rutile (TiO_2) and perovskite (CaTiO_3). It should be noted that these crystals are characterized by increased electronic (optical) polarization: $\epsilon_{opt} > 5$. Moreover, microwave permittivity of rutile and perovskite is strongly temperature dependent, with a negative $TC\epsilon$ value. These dielectrics are related to paraelectrics.

Paraelectric crystals belong to dielectrics with particular temperature dependence of their permittivity, described by the Curie-Weiss law:

$$\epsilon(T) = \epsilon_1 + \frac{C}{T - \theta}, \quad (9.26)$$

where θ is the Curie-Weiss temperature, C is the Curie-Weiss constant, and ϵ_1 is a part of permittivity practically independent on temperature. This equation is in good agreement with experimental data, as shown in Fig. 9.23A.

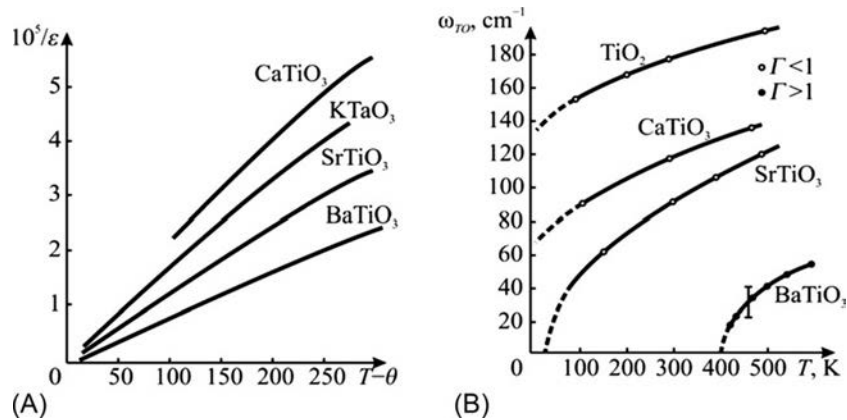


FIG. 9.23

Temperature dependence of inverse permittivity (A) and frequency ω_{TO} (B) for paraelectrics TiO_2 , CaTiO_3 , SrTiO_3 , and ferroelectric BaTiO_3 (in its paraelectric phase); Γ is the correspondent oscillator damping factor, $1 \text{ cm}^{-1} = 30 \text{ GHz}$ [13].

For example, dependence $\epsilon(T)$ in CaTiO_3 (perovskite) can be described in a broad temperature range, if in formula (9.26), one puts $\epsilon_1 = 60$, $C = 4 \times 10^4 \text{ K}$, and $\theta = -90 \text{ K}$. Calcium titanate can be considered as a material *related* to paraelectrics. However, in *typical* paraelectrics, which include, for example, SrTiO_3 or KTaO_3 , Curie-Weiss temperature is positive ($\theta > 0 \text{ K}$), whereas CaTiO_3 is distinguished by a fact that it has negative characteristic temperature ($\theta < 0 \text{ K}$).

In most paraelectrics, at temperature $T = T_c \approx \theta$ (i.e., *critical temperature*) phase transition occurs to ferroelectric (or antiferroelectric) phase. That is the reason why paraelectricity usually is associated with ferroelectricity. Similar to the ferroelectrics, the paraelectrics can be divided into two basic classes.

The paraelectrics of the “*order-disorder*” type are polar crystals (containing dipoles), in which, as temperature decreases, dipole-to-dipole interaction gives rise to a gradual ordering of dipole orientations, until, finally, when temperature becomes $T_c \approx \theta$, spontaneous polarization arises, at which most of dipoles are steadily oriented. Such paraelectrics near their *second-order* phase transition are characterized by rather sharp maximum of $\epsilon(T)$ that decreases with temperature increase. This fast decrease in $\epsilon(T)$ signifies *rather small* Curie-Weiss constant ($C \approx 10^3 \text{ K}$).

The paraelectrics of *displacement type* are ionic (not dipole) crystals, in which, however, ionic-covalent bonds between atoms are very significant. They usually show *first-order* phase transition and relatively *flatter* $\epsilon(T)$ dependence in their paraelectric phase that is characterized by a *big* Curie-Weiss constant: ($C \approx 10^5 \text{ K}$).

These properties of displacement-type paraelectrics can be explained by the dynamic theory of lattice vibrations. The central point in this theory is the fact that phase transition arises because the crystal lattice loses its stability owing to relatively

one of the *transverse* optical vibrations. This is reflected in the decrease in frequency ω_{TO} with decreasing temperature, and it is assumed that, at temperature $T \rightarrow \theta$, the frequency $\omega_{TO} \rightarrow 0$. The relationship between frequency and temperature is described by the Cochran law:

$$\omega_{TO} = A\sqrt{T - \theta}, \quad (9.27)$$

where A is the temperature coefficient of frequency. This dependence is confirmed by many experiments, as shown in Fig. 9.23B. Direct relationship between permittivity and temperature dependence on frequency of transverse optical lattice mode in paraelectrics, that is, the correlation of $\epsilon(T)$ and $\omega_{TO}^2(T)$ is seen, whereas the two parts of Fig. 9.23 are compared.

Explanation of the decrease in ω_{TO} when temperature decreases and high polarization in crystals of *perovskite* structure is as follows. Each ion is in equilibrium position under the action of long-range electrical forces of attraction and short-range forces of repulsion. High polarization means that the application of even a weak electrical field leads to unusually large displacements of ions from the equilibrium position (or, according to the model of “soft” ions, to large deformation of electronic shell of ions). This also means that the elastic force of repulsion of ions is small (i.e., corresponds to lower oscillation frequency).

It is logical to assume that in perovskite-type structures, such conditions are created for some ions, when compensation of short-range repulsion forces and long-range attraction forces occurs. Thus the effects, caused by interaction of electronic shells of ions, lead to a very high value of permittivity along with anomalous temperature dependence.

The frequency of transverse optical vibration mode in ionic lattice, which tends to zero when temperature $T \rightarrow \theta$, is the “*soft*” *vibration mode*. Using the Liddelyn-Sachs-Teller relationship: $\epsilon(0)/\epsilon(\infty) = [\omega_{LO}/\omega_{TO}]^2$, it is possible to show that the Cochran law (9.27) gives rise to the Curie-Weiss law: $\epsilon(T) \approx C/(T - \theta)$. To explain the possibility of decrease in frequency ω_{TO} in the perovskites, one needs to use the *polarizable* ion model, in which the interaction of electronic shells has a significant effect on the repulsive force of ions. This interaction can cause a condition at which the force, conditioned by polarization of ions, decreases; this allows to assume $\omega_{TO}(T) \rightarrow 0$ and to $\epsilon(T) \rightarrow \infty$. In this sense, it might lead to the understanding of the expression “wasting of crystal stability”: when small external perturbation (i.e., external electrical field) leads to a great response, the polarization is and ϵ is also high.

The latter assumption can be illustrated on the basis of the formula (9.25). The parameters c and q , if one uses the “shell model,” depend on the characteristics of the crystal structure: c characterizes short-range force that returns displacement of ion, and q is the effective charge. Note that the longitudinal frequency ω_{LO} is not critical to changes in model parameters, as it is determined by the sum of two terms. Conversely, the frequency ω_{TO} is strongly dependent on their difference.

In alkali-haloid crystals, the minuend in Eq. (9.25) is approximately two times greater than that of the subtrahend; hence, frequency ω_{TO} is only slightly dependent

on temperature. However, in the perovskite-structured crystals, the minuend in Eq. (9.25) is very close to that of the subtrahend; hence ω_{TO} becomes very low (much $< 10^{13}$ Hz). Moreover, even small changes in external conditions such as temperature, pressure, or voltage substantially affect the ω_{TO} . As a result, in the vicinity of “paraelectric-to-ferroelectric” phase transition, not only temperature but also electrical field or pressure greatly changes dielectric properties of such a crystal.

To analyze the possibilities of how it is possible to reduce temperature dependence of ϵ_{mic} , it is desirable to express Curie-Weiss constant through parameters of the discussed model. The approximate form of Curie-Weiss law $\epsilon(T) \approx C/(T - \theta)$ can be obtained, if the following formula is substituted in Eq. (9.25):

$$\frac{m\omega_{TO}^2}{c} = 1 - \frac{nq^2(\epsilon(\infty) + 2)^2}{9c\epsilon_0} = \gamma(T - \theta).$$

By substituting this value in the expression for permittivity in formula (9.25), it is possible to obtain.

$$\epsilon(0) - \epsilon(\infty) = \frac{nq^2(\epsilon(\infty) + 2)^2}{9c\gamma\epsilon_0} \frac{1}{T - \theta}; \quad A = \sqrt{\frac{c\gamma}{m}}; \quad C = \frac{nq^2(\epsilon(\infty) + 2)^2}{9c\gamma\epsilon_0}.$$

Displacement-type paraelectrics appears as a very attractive material for use in microwave microelectronics that need high ϵ_{mic} and low loss materials. The main obstacle to use paraelectrics at microwaves is ϵ_{mic} temperature instability. As the nature of this instability is the electronic subsystem of a crystal, the ways to overcome this instability should be sought in the methods of impact on this subsystem. The main contribution to ϵ_{mic} is given by far infrared (lattice) polarization. Formally, it is mentioned as “ionic” polarization, but in fact, it is associated with the susceptibility of ion electronic orbitals. It is found that most part of ϵ_{mic} in the rutile and perovskite is caused by highly polarizable oxygen octahedrons TiO_6 connected at their vertices. In this case, electronic clouds that link ions in the system of octahedrons provide enough freedom for easy polarization and leads to the $\epsilon_{mic} > 100$.

It is possible that there are possibilities for the rutile structure to have an impact on crystal properties by allocation of various ions between the octahedrons. That is why it is possible to govern the freedom of electronic shell interaction and their reciprocal displacement, when different ions are implemented to the structure. By this way, it is possible to significantly weaken critical $\epsilon(T)$ dependence (described by the Curie-Weiss law) but maintain high permittivity.

Thermal stable microwave dielectrics. Some examples of such elaborations are shown in Table 9.5, where most important microwave dielectrics are listed. As the loss factor has a tendency to linearly increase with frequency, to compare different microwave dielectrics, the *special quality factor* is used: $K = \nu / \tan \delta$, where ν is the frequency in gigahertz [13].

All high-permittivity microwave dielectrics mentioned in Table 9.4 are associated with perovskite-structure paraelectrics or ferroelectrics; however, none of listed thermostable microwave dielectrics in Table 9.4 exactly match typical paraelectrics: rather, they can be defined as materials “related to paraelectrics.”

Table 9.4 Permittivity and Quality Factor for Thermally Stable ($TC\epsilon < 5 \text{ ppm K}^{-1}$) Microwave Dielectrics

Ceramics	ϵ_{mic}	$K \cdot 1000$	Ceramics	ϵ_{mic}	$K \cdot 1000$
(Mg,Ca)TiO ₃	20	50	BaO ₄ ·TiO ₂	37	30
Ba(Sn,Mg,Ta)O ₃	40	200	BaO _{4,5} ·TiO ₂	40	40
Ba(Zr, Zn,Ta)O ₃	30	150	CaTiO ₃ ·LaAlO ₃	40	50
(Zr, Sn)O ₂ ·TiO ₂	38	50	BaO·Ln ₂ O ₃ ·TiO ₂	90...120	3 ÷ 7

Note. Ln-La, Gd, Nd, Sm, Eu.

Table 9.5 Location of Electron Spins in the Orbitals and Parameters of REE

	4f	5s	5p	5d	6s	Magnetic Moment, μ_B	r_{ion} , nm
Ba	□ □ □ □ □ □ □	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □	↑↓	0	0.149
La	□ □ □ □ □ □ □	↑↓	↑↓ ↑↓ ↑↓	↑ □ □ □ □ □	↑↓	0	0.117
Ce	↑ ↑ □ □ □ □ □	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □ □	↑↓	2.56	0.115
Pr	↑ ↑ ↑ □ □ □ □	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □ □	↑↓	3.62	0.113
Nd	↑ ↑ ↑ ↑ □ □ □	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □ □	↑↓	3.68	0.112
Pm	↑ ↑ ↑ ↑ ↑ □ □	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □ □	↑↓	2.83	0.111
Sm	↑ ↑ ↑ ↑ ↑ □ □	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □ □	↑↓	1.60	0.110
Eu	↑ ↑ ↑ ↑ ↑ ↑ □	↑↓	↑↓ ↑↓ ↑↓	□ □ □ □ □ □	↑↓	3.45	0.109
Gd	↑ ↑ ↑ ↑ ↑ ↑ ↑	↑↓	↑↓ ↑↓ ↑↓	↑ □ □ □ □ □	↑↓	7.94	0.108

The main reason of large ϵ in paraelectrics (or ferroelectrics) is the strong correlation between high-polarizable octahedrons. This correlation gives birth to ferroelectricity but at the same time leads to nonthermostability. To control octahedron correlation, some additional ions should be placed between BO₆-octahedrons, and this is main sense of microwave dielectric compositions.

The outer electronic shell of barium ion ($5s^2p^6$ hybrid) is very remote from the Ba-core owing to empty more deep electronic state 4f (Table 9.6); that is why the $5s^2p^6$ shell is rather compliant to interaction with neighborhood. Taking root between TiO₆-octahedrons, the large ion Ba²⁺ might have a strong influence on the outer electronic shells of the surrounding O²⁻ ions. This is one possible explanation of thermal stability.

The microwave ceramics with $\epsilon = 80-120$. Ceramics Ln_{2,3-x}M_{3x}TiO₃ with $\epsilon \approx 80$ can be obtained as solid solutions of perovskite structure, where Ln is a rare-earth element and M is the alkali metal ions that partially substitute rare-earth ions. Among these components, some have the positive $TC\epsilon$, whereas others have negative $TC\epsilon$. For example, La_{0,5}Li_{0,5}TiO₃ has perovskite structure with $TC\epsilon > 0$; however, if M=Na, K materials are characterized by $TC\epsilon < 0$. That is why it is

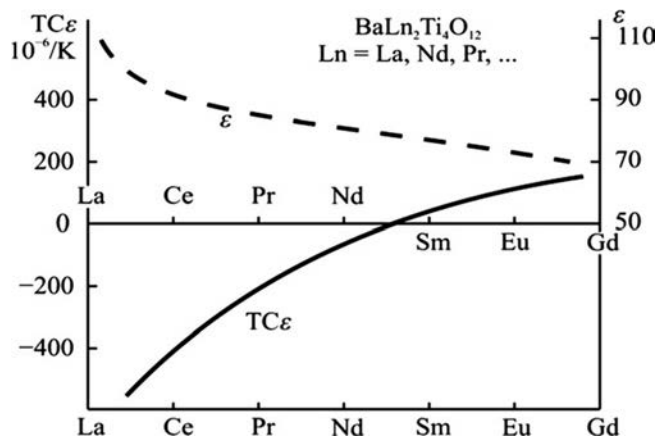
Table 9.6 BLTs Permittivity and $TC\epsilon$ at 300 K According to Measurements at a Frequency of 9.4 GHz

Parameter	Lanthanoid								
	⁰ La	¹ Ce	² Pr	³ Nd	⁵ Sm	⁶ Eu	⁷ Gd	TiO ₂	CaTiO ₃
ϵ_{mic}	110	90	85	83	80	75	65	100	150
$TC\epsilon, 10^{-6}K^{-1}$	-700	-400	-250	-80	+60	+100	+160	-900	-1600

Note. Upper index before RE shows the filling of the 4f-shell; for comparison of rutile and perovskite parameters are given as well.

possible to find compositions, in which permittivity temperature dependence would be compensated. At that, rare-earth ions are simultaneously substituted by alkali metal ions with large radius (Na, K) and small radius (Li); these compositions might have high temperature stability of dielectric parameters in microwave range.

The composition of $BaLn_2Ti_4O_{12}$ (BLTs) is a paraelectric-like material with imbedded rare-earth ions. These polycrystalline dielectrics are allocated among many other microwave dielectrics: their permittivity ϵ is several times higher than those in other types of microwave ceramics. As it was shown with the example of rutile, there are some possibilities that have an impact on crystal properties by allocation of various ions between octahedrons. In other words, it is possible to govern dielectric properties by the degree of electronic shell interaction and their reciprocal displacement by means of different ion introduction to the structure. By this way, it is possible to significantly weaken critical $\epsilon(T)$ dependence (according to Curie-Weiss law) and maintain a high polarizability. This method is used in $BaO \cdot 4TiO_2$ ceramics. Fig. 9.24 shows another way of microwave high-permittivity dielectric elaboration

**FIG. 9.24**

BLT permittivity and temperature coefficient for rare-earth cerium row.

(with $\epsilon_{\text{mic}} \approx 80$), using the rare-earth ions doping of TiO_2 : Ln_2TiO_5 , $\text{Ln}_2\text{Ti}_2\text{O}_7$, or $\text{Ln}_2\text{Ti}_3\text{O}_{11}$ ceramics, where Ln means cerium row: Ce, Pr, Nd, Pm, Sm, Eu, and Gd. The lanthanum ion has the electronic configuration $5s^25f^6$ (similar to that of Ba^{2+}), but La^{3+} is much smaller than barium ion (~ 0.1 nm).

It is also important that the La^{3+} ion has an empty $4f$ shell; therefore, it is diamagnetic (as barium ion). However, while barium introduction to rutile more than twice decreases its permittivity, in TiO_2 - La_2O_3 ceramics, microwave properties remain practically similar to those of TiO_2 , including $\epsilon_{\text{mic}} \approx 110$ and $TC\epsilon \approx -700$, as shown in Table 9.6. Increased value of ϵ_{mic} is explained by the relaxation process [13].

In the TiO_2 - Ln_2O_3 system, paramagnetic lanthanide leads to a noticeable effect on microwave properties, especially by increasing thermal stability. Lanthanide electronic configuration is $4f^{(1 \dots 7)}5s^25p^6$; hence they are paramagnetics with a gradual increase in magnetic moment: from one to seven of Bohr magnetons. Holding rather big dielectric permittivity ($\epsilon_{\text{mic}} \approx 80$ that exceeds twice ϵ_{mic} in the TiO_2 - BaO system), the TiO_2 - Ln_2O_3 compositions can change $TC\epsilon$ from the negative sign to the positive one, as shown in Table 9.6.

As described in Table 9.6, the compositions appear as those of paraelectrics, but they are simultaneously the paramagnetics; in the $\epsilon(T)$ characteristic, the impact of paramagnetism predominates over paraelectricity.

In TiO_2 - Ln_2O_3 composition, paraelectric Curie-Weiss law becomes gradually suppressed: from Ce to Gd (as paramagnetism becomes stronger), $TC\epsilon$ changes greatly, as shown in Fig. 9.24. It is obvious that using specially selected solid solution, for example, $(\text{Nd—Sm})\text{La}_2\text{TiO}_5$ composition, it is possible to reach zero thermal coefficient $TC\epsilon$.

It is necessary to note that relatively simple compositions of the $n\text{TiO}_2$ - Ln_2O_3 type (as Ln_2TiO_5 or $\text{Ln}_2\text{Ti}_2\text{O}_7$ ceramics) in which we can get $TC\epsilon \approx 0$ are nonstable in processing, whereas more stable in technology composition; $\text{Ln}_2\text{Ti}_3\text{O}_{11}$ is characterized by $TC\epsilon < 0$ (as in TiO_2 -rutile ceramics). Efficient solution to the problem of the high ϵ and low $TC\epsilon$ combination is to use the “structural stabilizer” Ba^{2+} together with Ln^{3+} for TiO_2 - Ln_2O_3 ceramics. As a result, most prospective compositions are elaborated in a condition when both lanthanides and barium are used together to obtain thermostable ceramics (that has $\epsilon \geq 100$ and $TC\epsilon \approx 0$).

Perovskite-like structure of barium-lanthanum tetratitanates $\text{BaLn}_2\text{Ti}_4\text{O}_{12}$ is usually titled as BLT. The Monophase perovskite-like structure BLTs exist only for cerium group of lanthanide row: Ln = La, Ce, Pr, Nd, Sm, and Eu. The valence of Ln^{3+} is quite different from that of Ba^{2+} ; hence, BLTs are vacant (nonusual) perovskites. For technical application, such dielectrics are synthesized as complex monophase systems, in which thermal stability could be controlled by change in their composition, as shown in Fig. 9.24. It is seen that the case $TC\epsilon = 0$ is located between neodymium and samarium. Some more recent elaborations using BLT-doped system enable achieving $\epsilon \geq 140$.

9.8 FERROELECTRICS AND ANTIFERROELECTRICS

Traditional and comfortable modeling of ferroelectrics is the assumption of its *spontaneous polarization* P_S , the direction of which can be switched by an externally applied electrical field. Currently, there is opinion as to another nature of ferroelectricity: it might be assumed that the structure of ferroelectric is able to demonstrate such nonlinear polar response to externally applied field, as if the *switching of polarization* occurs in it.

However, in this tutorial, it is better adhered to the *traditional theory* of ferroelectrics as crystals with reversible spontaneous polarization.

As stated earlier, other than in ferroelectrics, internal polarization is peculiar to the electrets and pyroelectrics. However, unlike residual polarization of electrets, cold spontaneous polarization represents the *stable thermodynamic state* of polar dielectrics. Indeed, residual polarization in electrets disappears during their heating or irradiation, whereas spontaneous polarization appears a structural feature of polar crystal. Really, its value can be changed under external influences, but then completely restores when initial conditions return.

It should be recalled that pyroelectricity is one of the possible manifestations of the peculiar structure of polar crystals. However, an applied electrical field can change the direction of spontaneous polarization in the *linear* pyroelectric, which persists up to the melting of a crystal. Being a *nonlinear* pyroelectric, ferroelectric not only switches its P_S under the electrical field but also shows significant change in $P_S(T)$ dependence until P_S completely disappears well before the melting of crystal.

Thus ferroelectrics are a subclass of pyroelectrics in which the polarized state is not stable enough, but it is quite labile. This polarized state can be changed by many external influences: electrical field, temperature, and pressure.

An important feature of ferroelectrics that suggests them as the electrical analog of ferromagnetics is their spontaneous division into a plurality of domains. Within each domain, the spontaneous polarization P_S is the same, but in various domains, P_S has a different orientation. The subdivision of the ferroelectric structure into a great number of domains is energetically advantageous, as the single-domain crystal would create an external electrical field (as in the case of electrets) in the environment. Obviously, energy of this field decreases with diminution in the size of domains.

Externally applied electrical field causes, at first, junction of randomly oriented ferroelectric domains into one domain; next, its polarization reaches saturation. As it can be seen from Fig. 9.25A, after switching off external field, polarization tends to maintain its constant direction. If polarity of externally applied field would be changed, polarization, without changing its absolute value, will change its direction abruptly [8].

For such “forced” change in the direction of P_S , that is, for ferroelectric polarization reversal, it is necessary to apply electrical field of a certain value, which is the *coercive field* E_C (see Fig. 9.25A). Sometimes, the value of this field reaches very

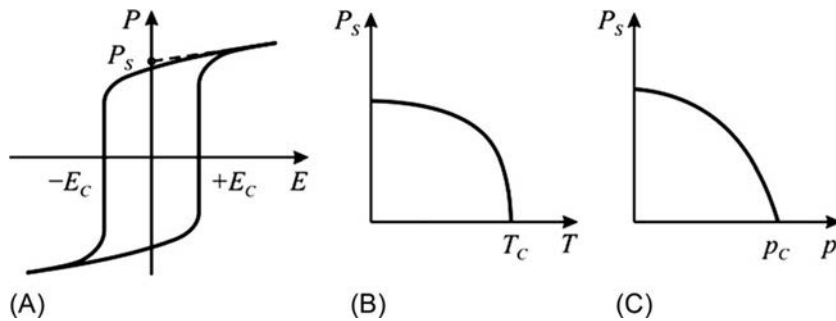


FIG. 9.25

Ferroelectric polarization dependence on the electrical field (A) and spontaneous polarization dependence on temperature (B) and pressure (C).

high values, and then, the ferroelectric cannot be repolarized and behaves like a pyroelectric. However, during such “hard” ferroelectric heating, as it approaches the temperature of Curie point T_C , its coercive field E_C must decrease; therefore close to the Curie point, it becomes possible to observe hysteresis. The coercive field E_C and P_S in ferroelectric becomes zero, if $T = T_C$. The pyroelectric, however, has no Curie point, and until electrical breakdown, its internal polarization does not change direction—such crystal rather can be destroyed than changing the direction of polarization.

It is therefore believed that availability of dielectric hysteresis is *necessary with adequate property* of the ferroelectric state. If temperature exceeds the critical value T_C , then hysteresis loop and ferroelectric state disappears. In the same way in ferroelectric, P_S affects the increase in hydrostatic pressure, as shown in Fig. 9.25C. By contrast, linear pyroelectric does not change its polarized state under a pressure up to being destroyed.

Summarizing, it might be concluded that ferroelectric is a special, nonlinear pyroelectric. Ferroelectrics are significantly different from linear pyroelectrics of tourmaline or lithium sulfate types.

In case of *active dielectric applications*, first, ferroelectrics or dielectrics close to them have gained the largest interest. In fact, exactly in ferroelectrics, their “transforming functions” are most pronounced. For example, the greatest value of piezoelectric module is observed in Rochelle salt crystals and, in the ferroelectric, antimony sulfiodide (SbSI). The highest values of pyroelectric coefficients are also seen in ferroelectric crystals (three glycine-sulfate). Therefore for thermal infrared receiver manufacture that uses pyroelectric effect, only the ferroelectrics (nonlinear pyroelectrics) are applied. The most applied piezoelectrics are also ferroelectrics, in particular, the ferroelectric ceramics of the PZT type ($\text{Pb}(\text{Zr},\text{Ti})\text{O}_3$). In optical detectors (that use the photopolarization effect), some of ferroelectric crystals are also applied, whereas the ferroelectric crystals strontium-barium niobate and lithium niobate are used for recording optical holograms.

Model conception of ferroelectricity. It is important to establish the main cause of ferroelectricity appearance in the ionic crystal. It can be shown that the great importance for ferroelectricity constitutes the anharmonicity in ion movement. This means a substantial nonlinearity in the law of reciprocal displacement of neighboring ions in the crystal lattice. This peculiar property occurs in some special structures [7].

For the simplest analysis, the one-dimensional chain of ions is investigated. The energy of linear chain of ions can be expanded in a row in power of dynamic displacement x :

$$U(x) = \frac{1}{2}cx^2 + \frac{1}{4}bx^4 + \dots \quad (9.28)$$

When considering polarization of ordinary (“linear”) dielectric, sufficient approximation is to consider only the first term of this expansion: $U(x) = \frac{1}{2}cx^2$, where c is the coefficient of elasticity. To determine the role of anharmonicity, it is enough to consider the next (anharmonic) term $\frac{1}{4}bx^4$ with the coefficient of anharmonicity $b > 0$. The lattice concerned can be stable only with assumption that the coefficient of anharmonicity is positive: $b > 0$. Exactly, this guarantees the stability of a given lattice in case of large fluctuations. As to the coefficient of elasticity, it might be positive ($c > 0$) or negative ($c < 0$).

Eq. (9.28) corresponds to the fact that ferroelectric is found *above* the Curie point (T_C), that is, in the nonpolar (paraelectric) phase that has the centrosymmetric structure. Below T_C , this 1D crystal passes onto the noncentrosymmetric (ferroelectric) phase. In case of transition to spontaneously polarized state (below T_C), to expansion series (9.28), the energy of ion spontaneous displacement $F \cdot x$ should be added:

$$U(x) = \frac{1}{2}cx^2 + \frac{1}{4}bx^4 - Fx, \quad (9.29)$$

where x is the ion deviation at the equilibrium state and F is the internal (spontaneous) electrical field. Fig. 9.26 shows the function $U(x)$ for both cases: when $c > 0$ (on

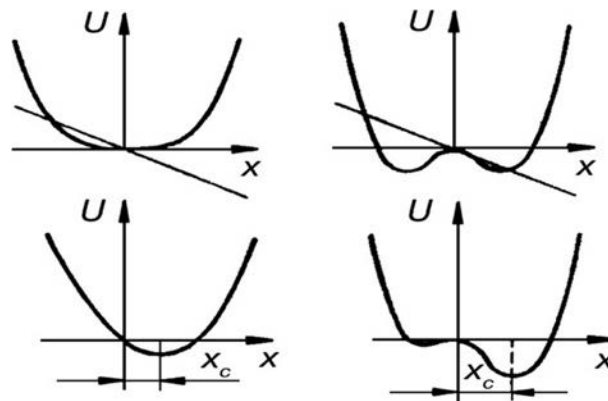


FIG. 9.26

Anharmonicity and ferroelectricity: energy U dependence on “ferroactive” ion deviation from equilibrium state in a lattice.

the left) and when $c < 0$ (on the right), it is seen that below the Curie point, the spontaneous deformation x_S arises, at which the energy $U(x)$ becomes minimal.

As the polarized state at $x = x_S$ now is the equilibrium state, a total force acting on the system of charges in this state equals zero: $\partial U(x)/\partial x = 0$, which means

$$cx + bxc^3 - F = 0. \quad (9.30)$$

With spontaneous polarization, the electrical field is associated, which is called the coercive field: $F_C = \beta P_S$, where β is the Lorentz factor. In case of one-dimensional model of ferroelectric represented by simple linear chain of ion polarization $P_S = np_S = nqx_S$, where n is ion concentration and q is ion charge. By substituting these data to Eq. (9.30), a cubic equation can be obtained

$$cx_S + bx_S^3 - nq^2\beta x_S = 0. \quad (9.31)$$

where the term cx_S describes the “elasticity,” whereas member bx_S^3 characterizes the anharmonicity. This equation has three roots:

$$x_1 = 0; x_{2,3} = \pm [(nq^2\beta - c)/b]^{1/2}. \quad (9.32)$$

As far as the only *spontaneously polarized phase* (with spontaneous deformation $x_S \neq 0$) is considered, the first solution $x_1 = 0$ of Eq. (9.32) is a side effect and will not be implemented here. The analysis of the other two obtained solutions provides an opportunity to make the following conclusions.

First, the sign « \pm » means two equivalent possible directions of the spontaneous polarization, which corresponds to two equal in magnitude but opposite in direction ion displacements: $\pm x_S$. This corresponds to two opposite values of P_S . Indeed, spontaneous polarization of ferroelectric material in some parts of a crystal can be directed in one direction, but in other parts—in the opposite direction (these areas of P_S in the opposite direction are called domains).

Second, in crystals with very small anharmonicity (when $b \cong 0$), the spontaneous displacement of ions is impossible. Therefore, the anharmonicity of ionic displacements is one of the *defining properties* of ferroelectric crystals.

Third, Eq. (9.32) has real roots $x_{2,3}$ only under the condition when $nq^2\beta > c$ (because parameter $b > 0$). To clarify the physical meaning of this important inequality (i.e., essential conditions for spontaneous polarization arising), it is necessary to multiply the left and the right sides of the expression $nq^2\beta > c$ by deformation x :

$$nq^2\beta x > cx. \quad (9.33)$$

The right-hand side of Eq. (9.33) corresponds to the elastic force that counteracts ferroelectric spontaneous displacement x_S . The nature of electronic shell interaction is such that it seeks to *return* the nonpolar state. Obviously, the left side of inequality (9.33), namely, the $nq^2\beta x$, has also the dimensions of force, which is the *leading interaction* (i.e., leads to ferroelectricity). Therefore spontaneous polarization occurs in such crystals, where *the leading interaction exceeds the returning interaction*.

In further analysis of Eq. (9.33), there is an opportunity to make the conclusion of what should be atomic parameters that contribute to the emergence of the

ferroelectric state in the ionic crystals. The first factor is the *high density* of a crystal (in this simple model, it is represented by parameter n). As a second factor, the *big electric charge* q of shifting ions can be considered: q^2 in inequality (9.33). The third factor is the *increased Lorentz factor* β .

Comparing these qualitative results with those of the experimental data, their correctness is observed. Indeed, among a large number of well-studied alkali halide crystals (such as NaCl), no ferroelectrics exist: ions in these crystals have single charge (Na^{1+} and Cl^{1-}), whereas the Lorentz factor is small: $\beta = 1/3\epsilon_0$ (ϵ_0 is the permittivity of free space). At the same time, in the barium titanate, for example (BaTiO_3 is the best-known ferroelectric), titanium ion (Ti^{4+}) has valence of +4 (i.e., q^2 in Ti^{4+} is 16 times higher than the q^2 of alkali halides). The Lorentz factor in barium titanate is also five times higher than its usual value in the simple cubic ionic crystals owing to peculiarities of perovskite structure (this term comes from mineral $\text{CaTiO}_3 = \text{perovskite}$).

In the perovskites that have a general formula ABO_3 , the small-sized cation B^{4+} is surrounded by the octahedron formed of six oxygen ions O^{2-} (Fig. 9.27). The displacement of particular “ferroactive” ion B^{4+} provides great contribution to the dipole moment of unit cell, in which spontaneous polarization occurs. The significant shift in small-sized tetravalent cation in the octahedron is conditioned by the fact that surrounding the very small ion B^{4+} , large anions O^{2-} leave considerable space inside the octahedron for easy displacement of the ion B^{4+} . This effect results in the appearance of spontaneous polarization in the perovskites.

Barium titanate is one of the many *ferroelectrics with a perovskite structure*. At the Curie point of barium titanate, its permittivity shows step increase and gradual decrease in the paraelectric phase. As temperature increases, spontaneous polarization first gradually declines and then abruptly falls to zero at phase transition. Heat capacity shows typical for phase transition maximum (Fig. 9.28).

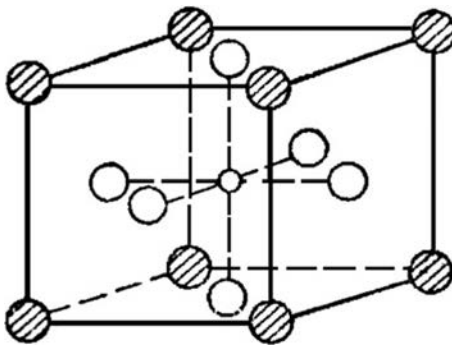
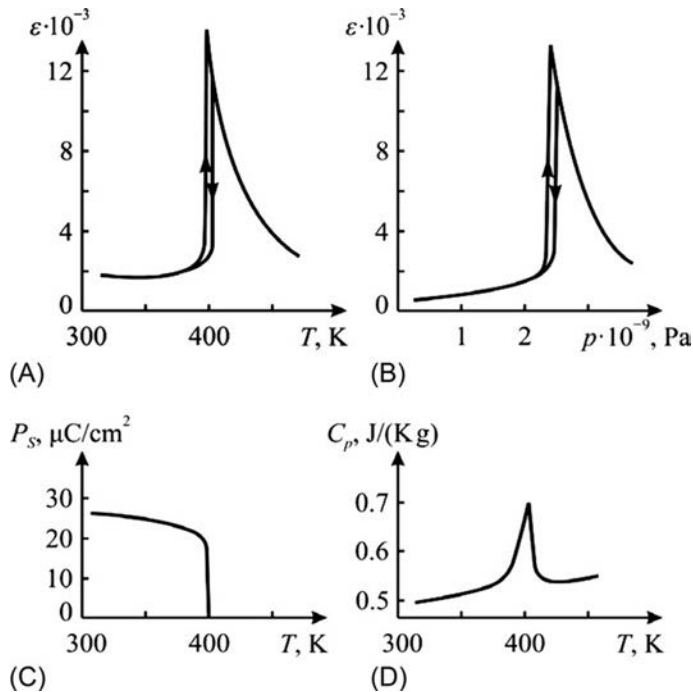


FIG. 9.27

Crystal structure of perovskite ABO_3 : “ferroactive” ions; $\text{O} = \text{B}$ is surrounded by octahedron of oxygen O, whereas crosshatched A ions are located between octahedra.

**FIG. 9.28**

The main characteristics of barium titanate: (A, B) permittivity dependence on temperature and pressure; (C, D) spontaneous polarization and specific heat dependencies on temperature.

Another class of ferroelectrics that have phase transition of order-disorder type (Table 9.7) is quite different from barium titanate (mostly, these are ferroelectrics with hydrogen bonds). First, the Curie-Weiss parameter C in them is less by two orders of magnitude; second, their phase transition temperature T_C is very close to Curie-Weiss temperature θ . Phase transition in these crystals is well described by the second-order transition theory (see Section 10.1).

The main experimental characteristics of ferroelectrics of the order-disorder type ferroelectrics are shown in Fig. 9.29 and listed in Table 9.7. Temperature dependence of dielectric constant, spontaneous polarization, and specific heat corresponds to the thermodynamic theory of phase transitions of the second type. Dynamic properties of these crystals differ from the properties of ferroelectrics with displacement-type transition. A special and interesting property of crystals with order-disorder transition is the isotopic effect—the displacement of Curie point in case of hydrogen replacement by deuterium (Table 9.7). This peculiarity demonstrates the importance of hydrogen bonds for majority of these types of ferroelectrics.

Table 9.7 Main Classes of Ferroelectric Crystals and Their Properties According to Author Microwave Studies [2]

Ferroelectric/ Parameter	P_S , $\mu\text{C cm}^{-2}$	T_C , K	θ , K	C , 10^{-4} , K	E_g , eV	$A/2\pi$, $\text{GHz K}^{-1/2}$
<i>Complex oxides of perovskite type</i>						
CaTiO ₃	–	–	–90	4.5	3.2	170
SrTiO ₃	–	–	35	8.4	3.2	180
BaTiO ₃	30	400	388	12	3.3	75
PbTiO ₃	80	780	730	15	3.1	90
KNbO ₃	30	685	625	18	3.4	95
LiNbO ₃	70	1500	–	–	3.6	–
<i>Crystals with hydrogen bonds:</i>						
Rochelle salt	0.25	297	291	0.17	–	–
	–	255	257	0.14	–	–
Deuterized Rochelle salt	–0.35	308	300	–	–	–
TGS	2.8	323	322	0.28	–	8.1
DTGS	3.2	328	327	0.27	–	10
KDP	4.7	123	118	0.28	–	180
DKDP	4.8	216	208	0.31	–	37
<i>Chalcogenides</i>						
SbSI	50	295	285	23	1.9	–
SbSBr	10	95	82	12	2.2	–
PbTe	–	–	–	14	0.2	–

There are also ferroelectric crystals that do not contain oxygen. Mostly, they are represented by the *chalcogenides* (Table 9.7). These ferroelectrics have a narrow band gap in their electronic spectrum, that is, such crystals belong to the ferroelectric semiconductors.

Properties of ferroelectrics are considerably dependent on their *domain structure*. The origin of multidomain structure in ferroelectric crystal below phase transition is energetically favorable. Single-domain crystal (Fig. 9.30A) creates an electrical field in the surrounding space (such as electrets), to which some energy W_1 is spent.

As shown in Fig. 9.30B, the energy of an external field in a two-domain crystal is smaller than that in single-domain crystal. Thus in case of multiple-domain structure, the total energy of the crystal must be reduced. This reduction in energy is limited by the growth of energy W_2 , expended on the formation of domain walls that separate regions with different directions of P_S , Fig. 9.30C. The average size of domains (whose sum $W_1 + W_2$ is minimal) depends on temperature, structural defects, and

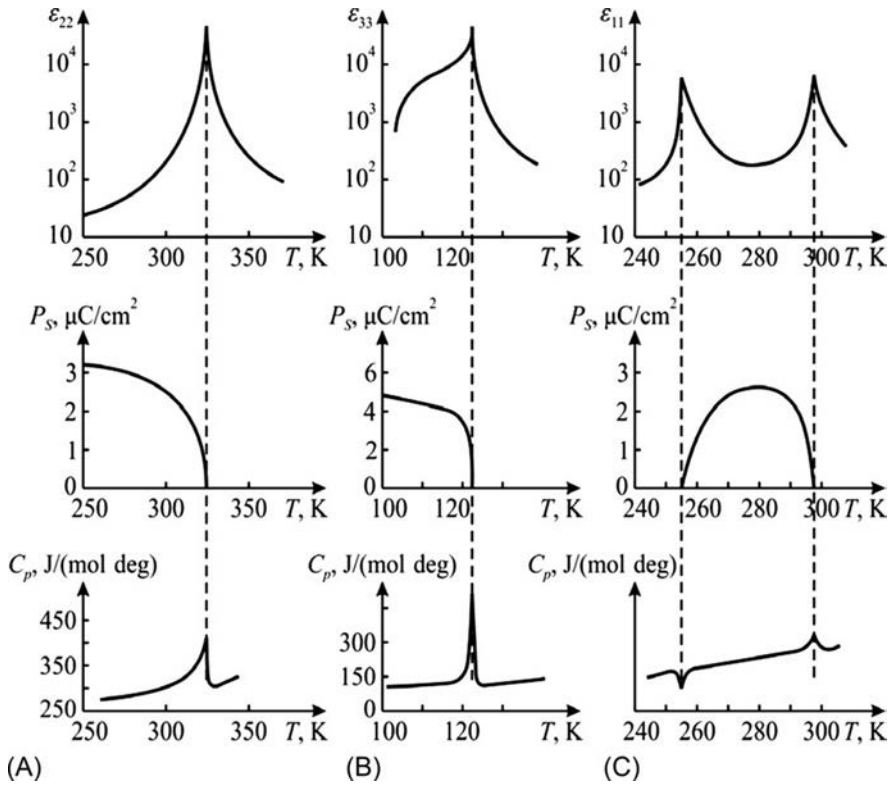


FIG. 9.29 Temperature dependence of relative dielectric constant, spontaneous polarization ($\mu\text{C}/\text{cm}^2$), and specific heat (J/(mol deg)) crystals, close to model of order-disorder phase transition: (A) TGS = triglycinesulfate ($\text{NH}_2\text{CH}_2\cdot(\text{COOH})_3\text{H}_2\text{SO}_4$); (B) KDP = potassium dihydrogen phosphate (KH_2PO_4); (C) Rochelle salt (RS) $\text{KNaC}_4\text{H}_4\text{O}_6\cdot 4\text{H}_2\text{O}$.

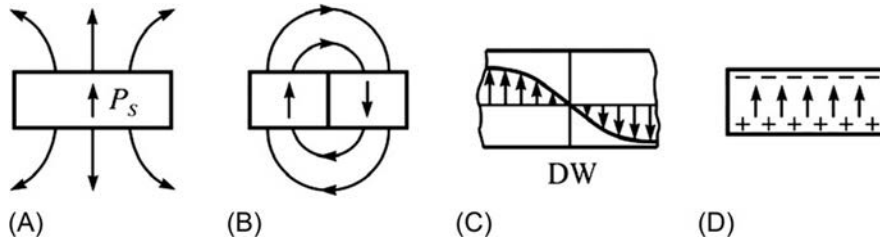


FIG. 9.30 Domain structure of ferroelectrics: (A) single-domain crystal creates in the surrounding area depolarizing electrical field; (B) in two-domain crystal, the depolarizing field is reduced; (C) domain wall structure in the vicinity of which P_s gradually changes its direction to opposite; (D) comparison with pyroelectric, wherein field is shielded by charges on the surface.

electrical conductance of dielectric, as well as on environment properties. Multidomain structure in ferroelectrics is relatively stable; the equilibrium state of ferroelectric domains usually corresponds to domain size from a few hundredths of millimeter to several millimeters.

Linear pyroelectrics that are characterized by the “hard” orientation of spontaneous polarization never split up into domains. However, they usually do not induce electrical field in the environment, as their spontaneous polarization is shielded by electrical charges accumulated on the surface (Fig. 9.30D). With change in ambient temperature, the alteration in polarization has not adequate time to be compensated by conductivity and a pyroelectric effect appears. Over time, spontaneous polarization in crystal again remains compensated.

The possibility of spontaneous splitting into the domains is caused by the changeableness of “soft” ferroelectric state in comparison with “hard” pyroelectric state. In this regard, sometimes ferroelectric is defined as *pyroelectric that divides into domains*.

However, in technical applications, sometimes just single-domain ferroelectric crystals should be used. This is needed, for example, in the pyroelectric temperature sensors, particularly, in heat television tube (vidicon) that converts invisible infrared image of objects into optical image seen on a screen. Single-domain structure in ferroelectrics can be created by various methods: thermal electrical polarization, radiation exposure of crystals with applied electrical field or by the introducing specific impurities that impede formation and movement of domain walls.

Application of ferroelectric ceramics also needs to create polarized structure: inasmuch as ceramic sample would consist of a plurality of domains and crystallites, oriented in different directions, and piezoelectric effect will not occur. The polarized piezoceramics often are obtained as thermoelectrets—by heating and subsequent cooling in the strong direct electrical field. This method uses temperature dependence of coercive field E_C that is significantly reduced when temperature increases (in Curie point $E_C=0$). In heated ceramics, ferroelectric domains can be easily oriented by electrical field; later, when temperature decreases, most of the domains remain in the polarized state. Thus the piezoelectric ceramics is the *texture of oriented ferroelectric domains*. That is why, when operation with polarized piezoelectric ceramics, their overheating should be avoided because it may result in domain depolarization and, consequently, in the loss of piezoelectric properties as domains become disordered in high temperature.

Frequency and optical nonlinear properties of ferroelectrics can be also determined by the motion of domain walls under the influence of electrical field. Low-frequency nonlinearity is characterized by the hysteresis. In its first cycle, the hysteresis loop (Fig. 9.25A) is due to forced orientation of domains; then, they partially maintain their polarization until the field of opposite direction (E_C) makes domains switching. Reversal polarization in ferroelectrics specifies domain contribution to the dielectric constant: $\epsilon_{\text{dom}} \sim dP/dE$. This contribution depends on the electrical field; this nonlinear dependence $\epsilon(E)$ sometimes is applied in technique. However, at microwave frequencies switching of domains, as a rule, does not have

time to occur. Therefore, the use of nonlinearity conditioned by domains motion is limited by the radiofrequency range.

Dielectric hysteresis loop characterizes two different states of polarization of ferroelectric crystal. This bistability is clearly manifested, if coercive field of ferroelectric is big enough. Exactly, the bistable polarized state can be used in the memory devices of computers and in other devices of modern electronics. It should be noted that in the bulk ferroelectrics, domain switching is possible only at relatively low frequencies (typically, $<10^6$ Hz). However, in the thin ferroelectric films ($<1 \mu\text{m}$), switching time can be reduced to 10^{-8} s.

9.9 FERRIELECTRICS AND FERROELASTICS

The antiferroelectric is close to the ferroelectric by its physical nature, structure, and chemical composition. However, in the antiferroelectric, $P_S=0$, as spontaneous polarization that occurs during phase transition is totally compensated within a single unit cell. Inasmuch as the energy of antipolar state is not very different from the energy of polar phase, external influence can turn antiferroelectric into ferroelectric. For example, phase transition from antipolar to polar states can be induced by a strong electrical field ($E > E_{cr}$); in this case, the double hysteresis loop is observed, as shown in Fig. 9.31F.

Phase transition between antiferroelectric and ferroelectric can occur not only under the influence of electrical field but also sometimes as a result of temperature change. This situation is observed, for example, in sodium niobate (NaNbO_3 , Fig. 9.31B). Antiferroelectric phase in this crystal exists between temperatures of 630 and 80 K. Below temperature of 80 K, the NaNbO_3 crystal enters into the ferroelectric phase, when ferroelectric and antiferroelectric states coexist.

However, most often, antiferroelectric phase occurs upon cooling from the paraelectric phase, usually with the “multiplication” of crystal unit cell. Therefore below the Curie point, the size of antipolar phase unit cell is 2, 4, or 8 times bigger than the unit cell in the paraelectric phase. Spontaneous polarization in this case is compensated by the displacement of opposite charges within new enlarged unit cell [14].

Thus in antiferroelectrics, owing to unit cell “multiplication” in comparison with the original (nonpolar) phase, polar shifts of ions during phase transition are compensated at the elementary level; hence, total spontaneous polarization is absent ($P_S=0$). In this connection, it is necessary to note that during phase transition from paraelectric to ferroelectric states, multiplication of elementary unit cell usually is not observed: each unit cell below the Curie point becomes polarized in the same way, and this effect is condensed in the crystal, forming $P_S > 0$.

This means that in antiferroelectrics, the critical reduction in the frequency of vibrational soft mode occurs not in the center of Brillouin zone (as in ferroelectrics) but on the boundary of Brillouin zone, and therefore the size of antiferroelectrics Brillouin zone decreases as a result of crystal symmetry lowering below phase transition (a more detailed description will be given in Chapter 10).

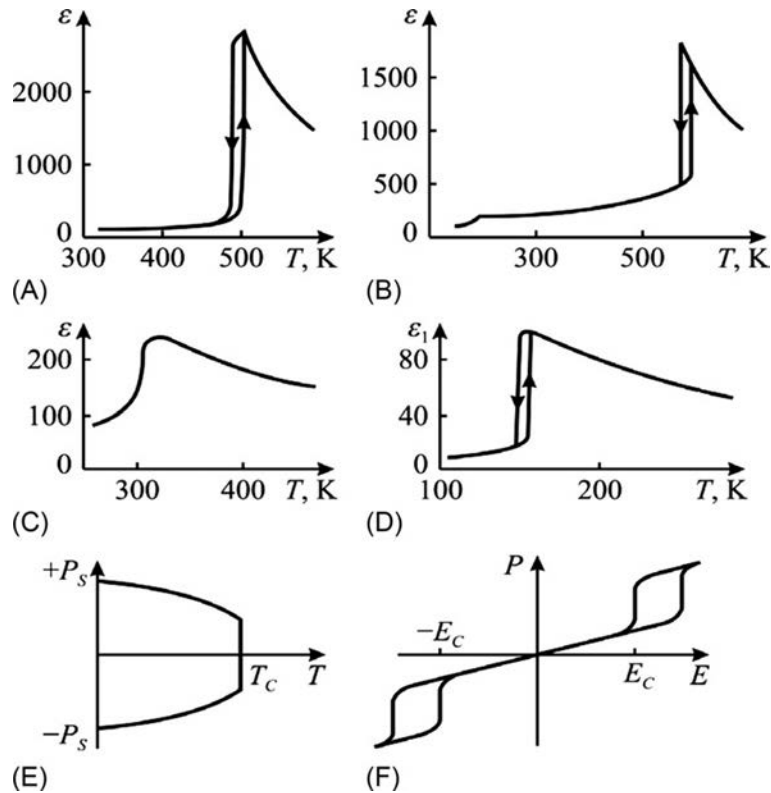


FIG. 9.31

Temperature dependence of permittivity in antiferroelectrics PbZrO_3 (A), in NaNbO_3 (B), PbMgWO_6 (C), and NH_4PO_4 (D), as well as spontaneous polarization compensation in unit cell of antiferroelectric (E) and double hysteresis loop in unit cell of antiferroelectric (F).

Ferrielectrics are crystals in which spontaneous polarization P_S is compensated only partially—by analogy with ferrimagnetics that are characterized by partial compensation of their spontaneous magnetization. Therefore ferrielectric is not entirely compensated antiferroelectric. Compounds such as sodium niobate (NaNbO_3 below 80 K), tungsten oxide WO_3 , $\text{PbCd}_{1/2}\text{W}_{1/2}\text{O}_3$, and some other isostructural compounds are related to ferrielectrics.

Ferroelastics are crystals that in their structural properties are close to the ferroelectrics. Phase transitions in them are accompanied by critical changes in the *elastic constants* of these crystals. Thus at temperature below the critical T_C , a *spontaneously deformed state* occurs, similar to spontaneous magnetization that occurs in the ferromagnetic or spontaneous polarization that occurs in the ferroelectric. By analogy with ferroelectrics, these materials are called the *ferroelastics*.

In the ferroelastic, phase transition from one direction of spontaneous deformation to another direction can be realized by applying external *mechanical stress*. Similar to ferrielectrics, ferroelastics below temperature T_C are divided into mechanical domains, where spontaneous strain has different directions. Similar to analog process of ferroelectric domain orientation (under electrical field influence), the uniform mechanical stress can make a single-domain ferroelastic.

Thus ferroelastics are spontaneously deformed crystals, in which deformation can be reoriented by the external mechanical influences.

If the sign of strain X is changed, the sign of spontaneous deformation x_S also changes. Mechanical rigidity of ordinary crystals is large enough, so that their deformation is very small and *linearly* depends on mechanical stress (according to Hooke's law, Fig. 9.32A). Conversely, the mechanical stiffness of ferroelastic *in one peculiar direction* of crystal is small, and therefore deformation increases sharply in case of increasing corresponding stress X , Fig. 9.32B). However, after achieving certain coercive force X_C , the stiffness increases again, and a hysteresis characteristic is observed with the nonlinear saturation of $x(X)$ characteristics.

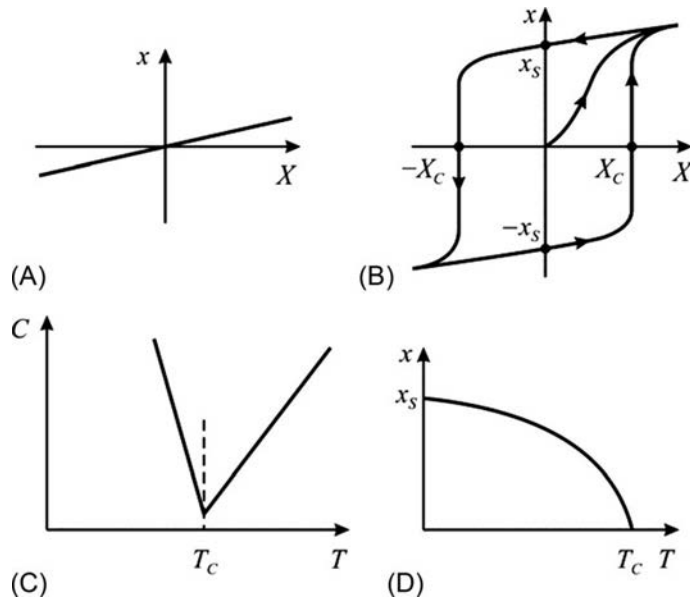


FIG. 9.32

Key features of ferroelastics: (A) linear dependence of $x(X)$ in ordinary crystals; (B) mechanical hysteresis in ferroelastics; (C) temperature dependence of the critical component of elastic stiffness in the vicinity of phase transition; (D) spontaneous strain temperature dependence.

After removing external stress, unlike conventional crystals, immediate restoration of the undeformed state is not observed in the ferroelastics; for some time, they can save their spontaneous deformation $+x_S$. To change strain direction (from $+x_S$ to $-x_S$), it is necessary to apply to a crystal the opposite sign of mechanical force that exceeds the coercive strain X_C . Therefore ferroelastics, for a long time, can exist in one of the two spontaneously deformed states: in the state of tension ($+x_S$) or in the state of compression ($-x_S$). Over time, however, the ferroelastic domains occurs, and total (macroscopic) spontaneous deformation gradually relaxes to an average value of $x=0$, approaching to the original (zero) point of mechanical hysteresis, as shown in Fig. 9.32B.

The *ordering parameter* of the ferroelastic phase transition is one of the components of crystal mechanical deformation. Therefore, the ferroelastics in the vicinity of phase transition may not have either dielectric or magnetic anomalies. However, owing to the symmetry change at phase transition, particularly, because of appearance (or change) of the piezoelectric effect, small anomaly in the permittivity in the vicinity of ferroelastic phase transitions can be observed, as shown in Fig. 9.33B. For example, such change in ϵ takes place in the ferroelastic lead orthophosphate ($\text{Pb}_3(\text{PO}_4)_2$).

In ferroelastic gadolinium molybdate ($\text{Gd}_2(\text{MoO}_4)_3$, sometimes used in optoelectronics, the main parameter of phase transition is mechanical deformation, but as a consequence of ferroelastic transition, spontaneous polarization also occurs, as shown in Fig. 9.33A. It is interesting to note that, unlike conventional ferroelectric crystal, in this case, spontaneous polarization P_S increases while cooling, which is not according to Landau dependence $P_S \sim (T_C - T)^{1/2}$ (as in ferroelectric or ferromagnetic) but linearly: $P_S \sim (T_C - T)$.

Ferroelectric phase, which occurs in this case as “by-effect” of ferroelastic transition, is the *improper ferroelectric*. In case of improper ferroelectric transition, the ordering parameter is not vector quantity (polarization) or the tensor physical quantity (component of deformation), but it is the *mixed* parameter. Thus the improper ferroelectric $\text{Gd}_2(\text{MoO}_4)_3$ is the ferroelastic and the ferroelectric simultaneously.

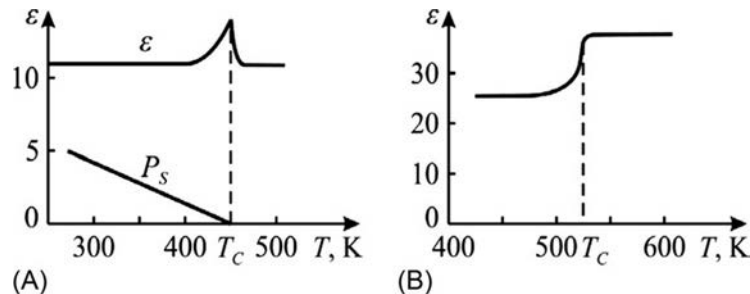


FIG. 9.33

Temperature dependence of permittivity and spontaneous polarization in gadolinium molybdate (A); $\epsilon(T)$ dependence for lead orthophosphate (B).

A small peak in permittivity near phase transition (Fig. 9.33A) is observed in this crystal only at low frequencies, when the crystal is free to deform under the electrical field—as electromechanical contribution to permittivity ϵ_{EM} from the piezoelectric effect. In microwave frequency range, where piezodeformation in crystal has no time to occur, the dependence $\epsilon(T)$ for $\text{Gd}_2(\text{MoO}_4)_3$ has no significant anomalies near its phase transition.

In the ferroelastic, spontaneous deformation changes with temperature according to law: $x_S \sim (T_K - T)^{1/2}$, as shown in Fig. 9.32D. Therefore, the strain component x_S in the ordered phase of ferroelastics changes by a law similar as that for spontaneous magnetization M_S in ferromagnetics or ferroelectric spontaneous polarization P_S .

The main parameter that is critically changed at the Curie point (similar to the permittivity of ferroelectric or magnetic permeability of ferromagnetic) in case of ferroelastic is a component of elastic compliance tensor s_{ijkl} that causes spontaneous deformation in crystal at temperature below T_C . Temperature dependence of elastic stiffness tensor critical component c_{ijkl} (inverse to elastic compliance s_{ijkl}) is shown in Fig. 9.32C, for one of the ferroelastics. As the $1/\epsilon$ of ferroelectrics or the $1/\mu$ of ferromagnetic, this option ($1/s$) of ferroelastics at Curie point tends to be zero.

Critical reduction of elastic stiffness components in the *paraelastic* phase determines the reduction in sound velocity (v_{sound}) in a certain direction of the crystal. In some cases, in the vicinity of ferroelastic phase transition, sound velocity decreases to 300–400 m/s (in this connection, it is appropriate to remind that in the ordinary crystals, $v_{\text{sound}} \sim 4000$ m/s, whereas in water, $v_{\text{sound}} = 1500$ m/s).

After transition from paraelastic to ferroelastic phases, the velocity of sound increases again but still remains much lower than that in most of the dielectric crystals. All these show that peculiar mechanical properties of ferroelastics are caused by the soft mode in the *acoustic* lattice vibrations, whose frequency is critically reduced in the vicinity of phase transition—as well as in ferroelectrics, spontaneous polarization is a result of a decrease in the frequency of transverse *optical* lattice vibrations mode.

Ferroelastic transitions can be both the second and the first order. In ferroelastic lead orthophosphate ($\text{Pb}_3(\text{PO}_4)_2$), phase transition of first-order type (PT-1) takes place at 450 K with a jump in deformation. However, in another ferroelastic BiVO_4 , its transition at 530 K is the second-order transition (PT-2). As mentioned, an additional ferroelectric phase can occur below the transition temperature in the ferroelastic.

The low velocity of sound indicates significant sensitivity of ferroelastic to external influences. That is why they can be used in the optics as *light deflectors*, designed for spatial scanning of light beam. Elastic waves can excite the crystal through a piezoelectric effect (usually in the frequency range of 30–300 MHz), and they form a kind of *optical diffraction grating*, whose pitch depends on the frequency of controlling electrical field. By varying the frequency of ultrasound, it is possible to control the angle of light beam deflection passing through the crystal. In addition to deflectors, ferroelastics and paraelastics can be applied as sensors of pressure and strain, as well as in other measuring devices.

9.10 NONLINEARITY OF FERROELECTRICS AND PARAELECTRICS

Ferroelectric materials are widely used in electronics. If we consider, for example, modern devices such as mobile phone, it will appear that ferroelectric materials are used in many of its components. First, in the mobile, the *microphone* is present (which uses the direct piezoelectric effect in ferroelectric) and so is the *speaker* (working on the principle of the inverse piezoelectric effect). Furthermore, piezoelectric filters are applied in the radiofrequency *filters* implemented with surface acoustic waves created with ferroelectric films. The microwave signal filter uses materials with high dielectric constant related to ferroelectrics. In addition, in the mobile phone, the subminiature film ceramic *capacitors* are applied, also made of ferroelectric materials.

However, one of the main characteristics of ferroelectrics, which provide their practical application, is the reversal of spontaneous polarization, in other words—the effect of *polarization switching*. At a certain value of coercive field, this effect is expressed stronger in case of more rectangular hysteresis loop (Fig. 9.34A). Two stable states of polarization can be interpreted as either 0 or 1 in the binary system that is the mathematical basis of modern computing.

Two opposing values of spontaneous polarization ($+P_S$ and $-P_S$) in the ferroelectric crystals of films can be stored for a long term, thereby providing computer memory recording and save information in the binary code. For example, the pulse of positive polarity orients ferroelectric domains that result in the residual polarization (that approximately equals to $+P_S$). Further, positive pulse (“readings” information) in this case does not lead to polarization reversal, and current through ferroelectric “storage element” is negligible. If originally the negative pulse was “written,” then reading positive pulse occurs repolarization—jump of polarization from $-P_S$ to $+P_S$, thus resulting in significant boost of current through ferroelectric.

Similarly, ferroelectric film may control the current in field-effect transistors (FETs), when gate is the ferroelectric film deposited during the manufacture of integrated circuits. This is an example of the practical application of *thin film integrated with semiconductor* in the memory devices: they are used as a matrix of logic elements. The principal feature of such devices is based on the polarization switching; their advantages are the ability to obtain a high integration density (up to 10^{12} bit/cm²), low power consumption, high radiation resistance, and others.

In addition to computing, nonlinear dielectric enables to convert electrical signals (modulation, amplification, detection, and others). In conventional dielectrics, the electrical induction is strongly proportional to electrical field: $D \sim \epsilon E$ or, what is the same, induced by electrical field polarization *linearly* relates to the field: $P \sim \chi E$, whereas $\epsilon = 1 + \chi$. However, in ferroelectrics (and paraelectrics), these linear relationships are not fulfilled, as permittivity itself depends on the field strength, $\epsilon = \epsilon(E)$ or $\chi = \chi(E)$.

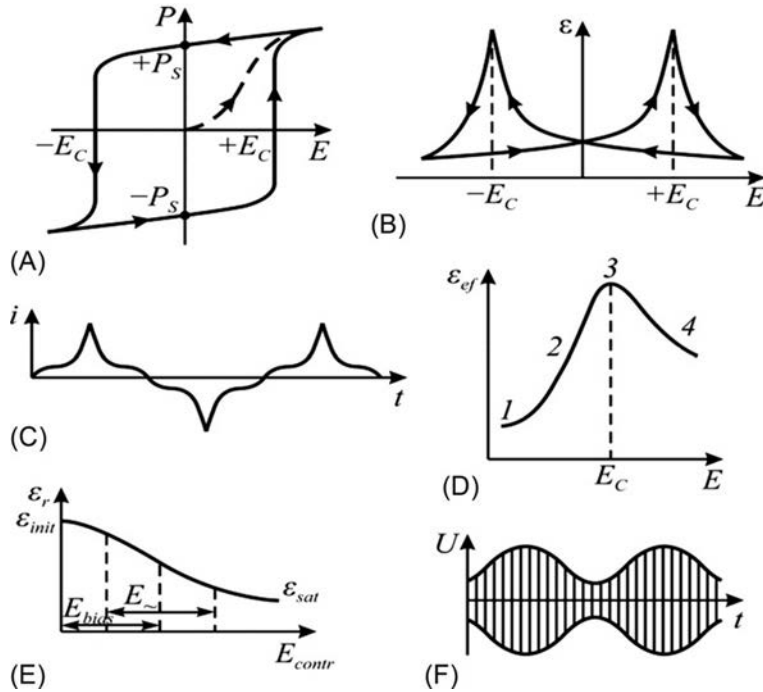


FIG. 9.34

Basic nonlinear properties of ferroelectrics: (A) dielectric hysteresis; (B) dynamic nonlinearity; (C) nonlinear current in ferroelectric variable capacitor; (D) effective nonlinearity; (E) reversing nonlinearity; (F) amplitude modulation using variable capacitor [9].

Permittivity is characterized by the ratio of polarization to electrical field: $\epsilon = 1 + P/\epsilon_0 E$. In ferroelectrics, $\epsilon \gg 1$; hence $\epsilon \approx P/\epsilon_0 E$, where $P = P_{\text{ind}} + P_{\text{or}}$; P_{ind} is the induced polarization, whereas P_{or} is the orientation (domain) polarization. As seen in Fig. 9.34B, the $\epsilon(E)$ dependence owing to polarization switching has sharp maxims that are seen near E_C . Thus during a period of sinusoidal voltage, “instantaneous” capacitance of ferroelectric capacitor passes twice through peak values, whereby current that flows through *nonlinear capacitor* has pronounced non-sinusoidal shape, as shown in Fig. 9.34C.

By averaging of permittivity for a period, it is possible to get *effective* dielectric constant ϵ_{ef} , whose dependence on the electrical field is characterized by a curve with maxim, as shown in Fig. 9.34D (that reminds the maxims of instantaneous values of ϵ in Fig. 9.32B).

In area 1 of $\epsilon_{\text{ef}}(E)$ dependence (in a relatively small electrical field), the nonlinearity is practically absent because field strength is not adequate to cause domain orientation, and permittivity (ϵ_{start}) is defined only by P_{ind} . In area 2 of $\epsilon_{\text{ef}}(E)$ dependence, permittivity increases rapidly because of domain orientation: polarization P_{or}

is added to P_{ind} . In the strong field, when almost all domains participate in polarization, dependence $\epsilon_{\text{ef}}(E)$ shows maxim (ϵ_{max} in area 3). In case of further increase in field, in the area 4, effective permittivity reduces because the contribution of P_{or} no longer increases, although field strength E continues to increase.

In practice, domain nonlinearity is described by coefficient $K_{\text{ef}} = \epsilon_{\text{max}}/\epsilon_{\text{start}}$. In nonlinear (variable) ceramic capacitors, *coefficient of effective nonlinearity* reaches 20 (in crystals, K_{ef} reaches 100). Effective nonlinearity of variable capacitors can be applied for frequency multiplication (higher harmonics generation) to improve contrast of image in photoluminescent screen and as the limiters of voltage (because capacity increases sharply with voltage, resistance to alternating current decreases).

In variable capacitors, in addition to effective nonlinearity, the *reversible nonlinearity* is seen (and also can find application in electronics). In this case, permittivity in alternating field ϵ_{\sim} can be controlled by simultaneously applied bias voltage, as shown in Fig. 9.34F. The bias field “hardly” orients domains and excludes them from repolarization that results in the ϵ_{\sim} reduction and saturation ϵ_{sat} . Nonlinearity coefficient of such dielectric “varactor,” that is, the ratio of primary permittivity ϵ_{start} to permittivity of saturation regime $K_p = \epsilon_{\text{start}}/\epsilon_{\text{sat}}$ usually reaches 10.

Reversible nonlinearity can be applied for capacitor value controlling by bias field. Effectiveness of such control is increased, if capacitor is included into resonant circuit. That is why dependence $\epsilon_{\sim}(E_{\text{bias}})$ may be used, for example, for *amplitude modulation* of the RF signal (Fig. 9.34F). If the modulation characteristic should be linear, except for controlling variable field, the application of direct electric bias E_b is necessary, as shown in Fig. 9.34E. Reversible nonlinearity, unlike effective nonlinearity, can be used in a very wide frequency range, if the voltage of high-frequency signal is small. In this case, the controlling field results in the ϵ -anisotropy and polarization saturation. In case of nonpolar (paraelectric) phase, reversible nonlinearity can be used in the microwave range.

Nonlinearity of paraelectrics manifests itself as a dependence of permittivity on electrical bias field, that is, $\epsilon(E_b)$. It is appropriate to remind that paraelectrics have high permittivity that depends on temperature according to the Curie-Weiss law: $\epsilon(T) \approx C/(T - \theta)$.

It is pertinent to recall that paraelectrics are ferroelectrics above the Curie point, as well as similar to the crystals (SrTiO_3 , KTaO_3), which at low temperatures do not enter into polar phase because of quantum effects. Paraelectrics have no domains; hence, the high-frequency dielectric losses (caused in ferroelectrics by domains) in the paraelectrics are relatively small. Therefore, a paraelectric can be used as a *tunable and nonlinear* dielectric up to a frequency range of 10^9 – 10^{11} Hz.

Permittivity of paraelectrics depends on temperature T and electrical field E [1]:

$$\epsilon(T) \approx \frac{C}{T - \theta}; \quad \epsilon(E) \approx \epsilon(T) \sqrt[3]{\frac{1}{1 + 3\beta\epsilon_0^3\epsilon^3(T)E^2}}, \quad (9.34)$$

where C is the Curie-Weiss constant; θ is the Curie-Weiss temperature, β is the coefficient at term P^4 in Landau expansion (see Chapter 10).

For further analysis, Eq. (9.34) should be presented in the form of rapidly convergent series:

$$\epsilon(T, E) \approx \frac{C}{T-\theta} - \frac{K}{3} \frac{C^4 E^2}{(T-\theta)^4} + \frac{2}{9} K^2 \frac{C^7 E^4}{(T-\theta)^7},$$

where $K = 3\beta\epsilon_0^3$. From this expression, the temperature coefficient $TC\epsilon$ and dielectric nonlinearity N can be determined:

$$TC\epsilon = \frac{1}{\epsilon} \frac{\partial \epsilon}{\partial T} \approx -\frac{1}{T-\theta} + \frac{KC^3}{(T-\theta)^4} E^2 - \frac{K^2 C^6}{(T-\theta)^7} E^4; \tag{9.35}$$

$$N = \frac{1}{\epsilon} \frac{\partial \epsilon}{\partial E} \approx -\frac{2}{3} + \frac{KC^3}{(T-\theta)^3} E + \frac{2}{3} \frac{K^2 C^6}{(T-\theta)^6} E^3. \tag{9.36}$$

Eqs. (9.35), (9.36) are illustrated graphically in Fig. 9.35.

The value of $TC\epsilon$ depends on the electrical field, whereas nonlinearity N depends on temperature. At certain temperature, a maxim of nonlinearity is observed, which is of interest in connection with the selection of operating modes of microwave

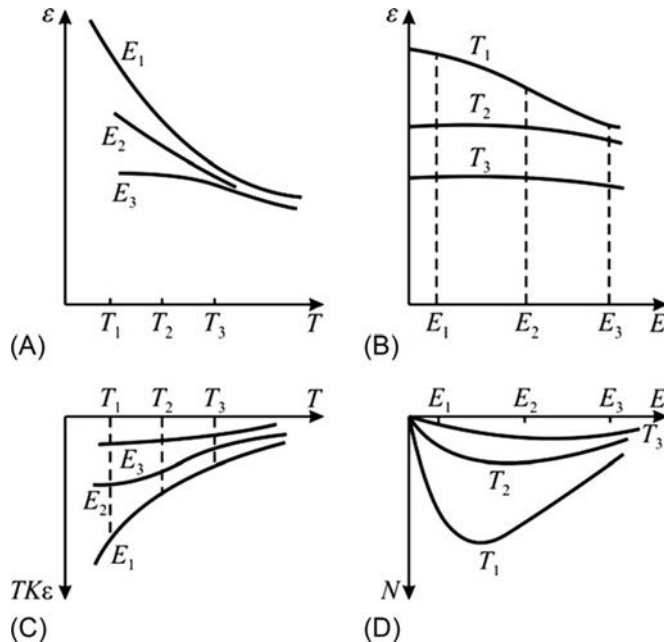


FIG. 9.35

Nonlinear properties of paraelectric: (A, B) ϵ depending on temperature and bias field simultaneously; (C) temperature dependence $TC\epsilon$; (D) nonlinear dependence on temperature and bias field [4].

variable capacitors for use in the technical devices. From Eq. (9.36), it is possible to find field strength E_{\max} at which nonlinearity is the largest:

$$E_{\max}^2 = \frac{(T - \theta)^3}{3KC^3}.$$

This parameter characterizes *paraelectric "rigidity"* and determines the range of operating voltage. In practice, this voltage needs to be reduced, which is caused by requirement to improve device reliability (operation voltage must be significantly lower than electrical breakdown voltage) and to reduce the power of microwave signal (which results in overheating of nonlinear dielectric).

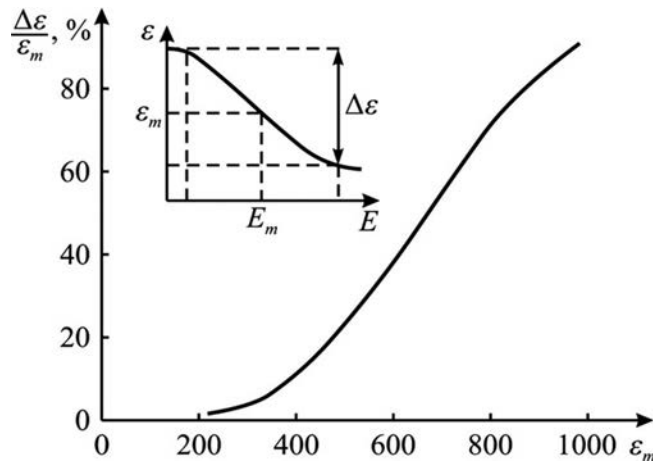
Owing to low inertia of controlling mode and high resistance to radiation, *microwave variable capacitors* are quite promising for use as parametric amplifiers, modulators, microwave signals, phase shifters, and so on. At cryogenic (helium) temperatures, most promising for application is potassium tantalate because, at helium temperatures, it has quite low level of dielectric losses, whereas at liquid nitrogen temperatures, the advisable material for variable capacitors is strontium titanate. At ambient temperatures, the suitable variable capacitors can be prepared from ferroelectric solid solutions (Sr,Ba)TiO₃ and (Sr,Pb)TiO₃ in their paraelectric phase.

When ferroelectric materials are used for applications at microwave frequencies, it is important that the value of permittivity and its temperature dependence would be minimal with maximal nonlinearity. However, as it is seen from Eq. (9.34), nonlinearity is proportional to permittivity in *power 3*, whereas a field, in which nonlinearity has a maxim, increases with deviation from the Curie-Weiss temperature as $(T - \theta)^{3/2}$. Therefore, various technical requirements to be applied in microwave paraelectric varactors (high thermal stability, high nonlinearity, and small losses) are difficult to reconcile; however, compromise is still possible, if paraelectrics is used as a thin film.

Nonlinear paraelectric films. Paraelectric microwave elements are usually made as thin films, deposited on dielectric substrate having high thermal conductivity (BeO or MgO). Ensuring proper heat dissipation prevents overheating of nonlinear film, in which density of energy can reach large sizes. Different thermal expansion coefficients of substrate and film give rise to mechanical stresses in a film, thus resulting in more diffuse dependence of $\epsilon(T)$, whereas film permittivity and losses decrease. In addition, the operating voltage in case of film reduces, whereas breakdown voltage increases.

For application in microwave technology, only ferroelectrics of the *displacement type* of phase transition are promising materials, and they must be regular solid solutions using *above* the Curie point. That is, films should be in the paraelectric phase because at microwaves in the ferroelectric phase dielectric losses show a sharp increase because of the multidomain structure of ferroelectrics.

Deposited on dielectric substrate with $\epsilon_{\text{sub}} \approx 10$, paraelectric film of 0.1–1- μm thickness and permittivity $\epsilon = 300\text{--}1000$ can be applied practically without inertia of controlling. Pectinated interdigital electrodes should be deposited on only one side of a film. Typical characteristics of film designed for microwave devices are shown

**FIG. 9.36**

Parameter $\Delta\epsilon/\epsilon_m$ for paraelectric BST films in dependence on bias field; permittivity electrical control is shown in a window.

in Fig. 9.36. The manageability of paraelectric by bias electrical field corresponds to the parameter $\Delta\epsilon/\epsilon_m$, that is, the bigger the $\Delta\epsilon/\epsilon_m$, the higher is the permittivity. Non-linearity N can be significantly increased with permittivity: $N = \epsilon^{-1}d\epsilon/dE \sim \epsilon^3$.

Generally, the thinner the paraelectric film, the lower is the permittivity and dielectric losses, but ratio of $\Delta\epsilon/\epsilon_m$ reduces. Dielectric losses of films are much smaller than losses of bulk samples made of the same nonlinear composition. Studies have shown that solid solution $(\text{Sr,Pb})\text{TiO}_3 = \text{SPT}$ appears most suitable to use in the microwave range. Both components (SrTiO_3 and PbTiO_3) are characterized by relatively low damage of soft phonons and therefore give relatively small contribution to microwave losses.

9.11 DIFFERENT EFFECTS INTERDEPENDENCE IN POLAR CRYSTALS

Most active dielectrics belong either to *polar crystals* (pyroelectrics, ferroelectrics, and piezoelectrics) or to partially ordered *polar systems* (liquid crystals, electrets, polar polymers, and so on). That is why any physical hypothesis relevant to the nature of *intrinsic polarity* (which is not caused by an external electrical field) deserves particular attention. This “hidden polarity” is the only phenomenon that can cause the *electrical* (vector) response onto nonelectrical impact on a matter. Impact can be scalar (i.e., zero-rank tensor) or it can be vector (first-rank tensor) or even second-rank tensors. For all these reasons, this section is devoted to the conception of spontaneous polarization [15–18].

The **cause of noncentrosymmetry in crystals**. If the simplest structures formed by *polar molecules* (of HCl type) and *polar radicals* (of NH_4^+ type) would be ignored, but only atomic and ionic dielectrics will be considered, the polar structures are inherent to the crystals with *hybrid ionic-covalent bonds*. Exactly, this specificity causes the *symmetry reduction* in the crystals; hence, most of them belong to the non-centrosymmetric classes of symmetry.

Crystals with purely ionic bonds and those with purely covalent bonds are *electrically nonsensitive* to nonelectric impacts (they are nonpolar crystals). Mostly, they belong to the centrosymmetric classes of symmetry. As a rule, in *pure ionic* crystals, the highest symmetry exists, and they have no special orientations in their atomic connections. In the same way, in *simple covalent* crystals, each atom provides one unpaired electron, but the formal charge of atom remains unchanged; hence atomic bonds have a balanced electronic pair. That is why they also have, as a rule, the centrosymmetric structure.

It might be supposed that the primary cause of internal polarity existence is the *asymmetry in distribution of electronic density along atomic bonds*. Such asymmetry is caused by considerable distinction in *electronegativity* of their atoms. It is appropriate to recall that electronegativity is a physical **property** that describes the tendency of an **atom** to attract **electrons**; it depends on the **atomic number** and the size and structure of *outer electronic orbitals*. In the structure of atom, an attraction that outer electrons feel from particular nucleus depends on the number of protons in a nucleus, the distance of outward **orbital** from nucleus, and the amount of screening from the inner electrons. The higher the atomic electronegativity, the stronger the atoms attract electrons toward themselves.

For the reasons listed, in various atoms, the difference in electronegativity might be very substantial. Hereby, the atom with higher electronegativity *attracts* conjunctive electrons more strongly, and therefore its *true charge* becomes more negative. Conversely, the atom with lower electronegativity acquires increased positive charge. Together, these atoms can create the *polar bond* and contribute to the non-centrosymmetric structure.

In the *molecules*, perceptible difference in the electronegativities of atoms gives rise to their interior polarity, for example, the *dipole moments* are created (such as in HCl or H_2O). However, in the *crystals*, a significant difference in the electronegativities of atoms results in *directional polar structural motives* and produces noncentrosymmetric structures.

Recall that out of 32 classes of crystals, 11 are centrosymmetric classes and 21 are noncentrosymmetric classes. However, only in the 20 of noncentral classes, *odd electromechanical* effect (piezoelectricity) becomes apparent: $x = dE$ (strain x is proportional to the applied field E , whereas d is piezoelectric module). Furthermore, out of 20 piezoelectric classes of crystals, 10 classes belong to the pyroelectrics (they have *peculiar polar axis*), whereas other 10 noncentral classes would be referred to as “*true*” (or “*actual*”) piezoelectrics; one can come to the conclusion that they have the *polar-neutral* axes. This means that along any of such axis, polarity can be convincingly manifested, but on the whole (considering the joint action of all axes), the crystal turns out to be polar neutral.

Next, these crystals will be considered, as well as the question: what can be the microscopic explanation of piezoelectricity in case of noncentrosymmetric but non-pyroelectric crystal?

Intrinsic polarity modeling. Generally, the unique properties of some polar crystals are described by the conception of spontaneous polarization, P_S . If there really is a spontaneous polarization, it would be accompanied by the “internal field.” In this connection, contemporary theoretical works doubt in uniqueness of the concept P_S conception [19] because experimentally only the *change* in polarization can be measured: by the pyroelectric effect and by means of “polarization switching” in ferroelectrics.

The model discussed later, based on the asymmetry in the distribution of electronic density along atomic bonds, is free from assumption of internal field existence. Asymmetric atomic bond does not result in any internal field in a crystal but can provide specific response to external impact. That is why it is possible to interpret the “ P_S ” as responsiveness of a crystal for electrical reaction on *uniform* thermal, mechanical, or irradiation impact (which is impossible for centrosymmetric crystals).

For example, in case of uniform but *directed* mechanical influence onto a polar (or polar-neutral) crystal, the electrical response arises (i.e., piezoelectric effect). The point is that reciprocal displacement of atoms compresses or stretches their asymmetric bonds, and by this way, it induces associated electrical charges on crystal surface (i.e., produces piezoelectric polarization). On the contrary, the *hydrostatic* (nondirected) pressure cannot produce the piezoelectric effect in the polar-neutral (“actual”) piezoelectric crystal, but in the polar (pyroelectric) crystal, the volumetric piezoelectric effect arises [16].

If atomic bonds in crystal are symmetric (nonpolar), then no electrical response is possible to any *uniform* mechanical perturbation. It should be noted that this is not in regard to *heterogeneous* mechanical impact, when electrical response occurs in *any* crystalline dielectric—*flexoelectricity*. By the same way, the *heterogeneous* temperature influence induces in *any* piezoelectric (even in polar-neutral crystal) something such as “pyroelectric effect”—*actinoelectricity*. The point is that any nonuniform impact changes the structure of atomic bonds in a crystal, thus converting crystal into an *asymmetric state*. For example, electrical field induces polar axis in *any* solid dielectric: atomic bonds forcedly acquire the asymmetry so that any dielectric demonstrates *electrically induced* piezoelectric and pyroelectric properties (magnitude of these effects is strongly dependent on dielectric permittivity).

The noncentrosymmetric allocation of electrical charges in polar crystals can be described by different *structural polar motives* such as the imaginary dipole (i.e., unidirectional atomic bond), the imaginary sextuple (i.e., two-dimensional formation of asymmetric bonds in a plane), or the imaginary octuple (i.e., three-dimensional asymmetric atomic connections in a volume).

Such modeling of electrical charge allocation is quite common in the case of description on *molecule properties*, for instance, to describe the Van der Waals bonding by the *fluctuating polar structures* in electronic shells (dipoles, sextuples, and

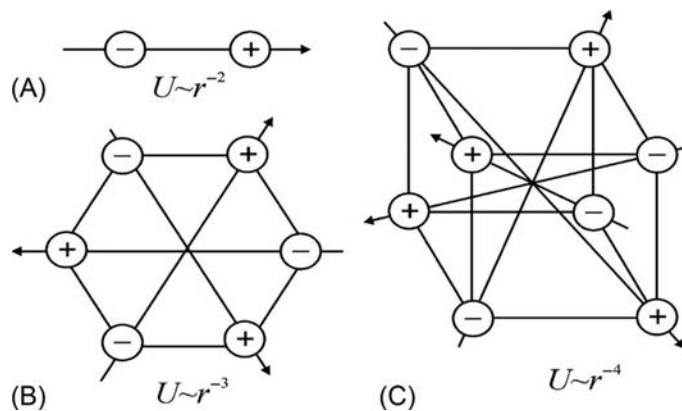


FIG. 9.37

Simplified representation of intrinsic polarity in noncentrosymmetric crystals: (A) dipole moment; (B) sextuple moment, (C) octuple moment [17].

octuples). In case of polar and polar-neutral crystals, such modeling of their properties can be applied to the crystals with asymmetric hybrid ionic-covalent bonds in the one-, two-, and three-dimensional intrinsic polarities of crystals possessing multidirectional *structural motives*, and this will be used further to describe various properties of active dielectrics.

The simplest case is the *dipole-type structural polar motive* (Fig. 9.37A) that corresponds to the so-called spontaneous polarization in crystal. This generally known conception of P_S might be used only as a way to quantify the response of polar crystal onto *homogeneous dynamic* influence: changing in time *uniform heating* (or cooling) that leads to the pyroelectric effect; changing in time *uniform compression* (or stretching) of crystal (including hydrostatic influence) that gives rise to the piezoelectric effect.

Physical property that usually is expressed by the conception of spontaneous polarization might be represented in the macroscopic way by *polar vector*, which is considered as being “built-in crystal structure” and corresponds to polar first-rank tensor. Exactly this “*intrinsic polarization*” results in the *vector-type* reaction to any *external scalar* impact on a crystal. For example, it leads to *volumetric piezoelectric effect* in case of all-round (hydrostatic) pressure dynamic influence and to pyroelectric effect in case of uniform dynamic heating or cooling. Applying rather strong alternating electrical field might result in polarization switching (only in ferroelectrics). Crystals, possessing the so-called spontaneous polarization (i.e., dipole-type hidden intrinsic polarity), belong to 10 pyroelectric classes of symmetry (ferroelectrics are the sub-class of pyroelectrics).

However, many other polar crystals are not pyroelectrics, but they belong to the so-called actual piezoelectrics. In this case, asymmetric allocation of bonded electrical charges needs more complicated modeling than discussed simple dipole-type

structural motive. In such “actual” piezoelectrics, intrinsic polarity can be represented by imaginary *sextuple* (in plane) or by imaginary *octuple* (in volume): they correspond to second- and third-rank tensors, respectively. As seen from Fig. 9.37B and C, these high-rank intrinsic electrical moments are *totally compensated* (in contrast to dipole structural motive).

In both these cases, any arbitrary *scalar* action (e.g., hydrostatic pressure or uniform heating) cannot “awake” in such crystals any vector response. Only the vector or tensor types outside actions, such as mechanical stress X_{ij} (second-rank tensor) or temperature gradient (vector *grad T*), are capable of provoking this hidden intrinsic polarity to produce vector response (i.e., to induce connected (polarization) charges on the surface of the crystal or to generate electrical current, if crystal has electrodes closed onto resistor).

The modeling of intrinsic (or “latent,” or “hidden”) polarity shown later signifies the ability of low-symmetric crystal to produce specific responses to a variety of external influences, but they obviously must be *variable in time*. If the external influence (mechanical, thermal, and optical) is not changing in time, no electrical response is possible to observe. This property of polarization is generally only when the affecting factors (field, pressure, heat, light, and so on) are *changing in time*. If the impact, after being applied (or changed), remains constant, polarization does not manifest itself—unlike conductivity that exists all the time of external factor duration (fields, illumination, radiation, and heat gradient).

Experimental illustrations. There is much experimental evidence of spontaneous polarization ($P_S = M_i$) manifestations in pyroelectrics and ferroelectrics; hence, this is a matter of common knowledge. However, in the *piezoelectrics* that are *not pyroelectrics*, the existence of hidden polarity (capability to produce electrical response onto homogeneous nonelectric impact) needs some experimental grounds [18].

As a rule, experimental possibilities are limited only by the electrical measurements that can be seen as the vector responses (voltage or current). Not surprisingly, *compensated* intrinsic polarity was not detected. However, there is some qualitative evidence of the latent polarity of noncentral crystals.

For instance, the *etching* of quartz crystal occurs more rapidly on the “positive” side of polar *X*-axis, whereas rate of etching is very slow on its “negative” side. Therefore, the etch figures for quartz samples are very different for “+” and “−” surfaces [6]. In only the same way in the cubic but polar GaAs crystal, a considerable distinction in the *chemical properties* between two surfaces of the (1 1 1)-cut plates is seen, which should be considered during electrode deposition.

Next, during the *crystallization* process, the density of piezoelectric crystal decreases in comparison to its melt. For example, growing GaAs crystal can swim in its melt (as ice in water). It may be deduced that polar bond formation expands the noncentrosymmetric crystal. Moreover, temperature dependence of *thermal expansion* coefficient α of any piezoelectric passes through zero with a temperature interval of 10–100 K instead of showing classic dependence $\alpha \sim T^3$. It can be attributed to the assumption that polar bonds in a piezoelectric at low temperatures become more ordered.

Moreover, *heat diffusion* resistance R in all noncentrosymmetric crystals far exceeds R of the centrosymmetric crystals. Increased thermal resistance in the piezoelectrics is conditioned by the peculiarities of the phonon dissipation process in these crystals (binding of acoustical and optical phonons). In addition, the fundamental (lattice) *microwave absorption* of polar crystals is vastly superior to the absorption of centrosymmetric crystals. Dielectric losses by microwaves show an additional maxim of the quasi-Debye type absorption owing to interaction between optical and acoustical phonons [4].

Moreover, the *affinity* pyroelectricity and piezoelectricity is demonstrated by the *polymorphism* of $43m$ (piezoelectric) and $6mm$ (pyroelectric) structures. The zinc sulfide crystal is the best example of piezoelectric and pyroelectric affinity: interatomic interaction provides some relatively stable configuration of the ZnS crystal (zinc blend), in which both structures (sphalerite and wurtzite) can coexist.

In case of sphalerite, the piezoelectric *intrinsic* polar structure of ZnS can be described by the octuple electrical moment, and this simulation might be successfully represented by four threefold polar axes of $[111]$ -type direction that are crossed at an angle of 109.5° , as shown in Fig. 9.38A. Such intrinsic three-dimensional polar structure is absolutely self-compensated.

However, the second principal structure of ZnS is the pyroelectric wurtzite, which includes not only octuple-type polar axes but also a dipole component, as shown in Fig. 9.38B. In spite of difference in symmetries, distinction between atomic displacements in two principal forms of zinc blend is quite small that these structures can *alternate each other* in one crystal. Therefore, there is no big qualitative difference between pyroelectric and piezoelectric intrinsic polarities.

Of the 10 *pyroelectric* classes of crystals, the dominating structural motive is the intrinsic *dipole* motive. The first-rank material tensor (i.e., material vector) describes

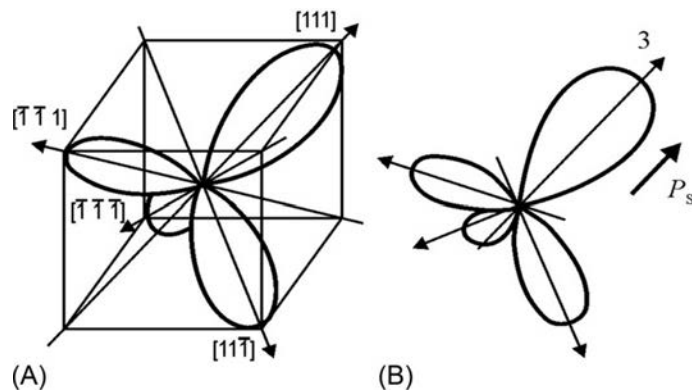


FIG. 9.38

Spatial distribution of 3D octuple-type polarity (only a positive direction of axes is shown): (A) total compensation of polarity; (B) appearance of dipole component that is equivalent to spontaneous polarization P_s [16].

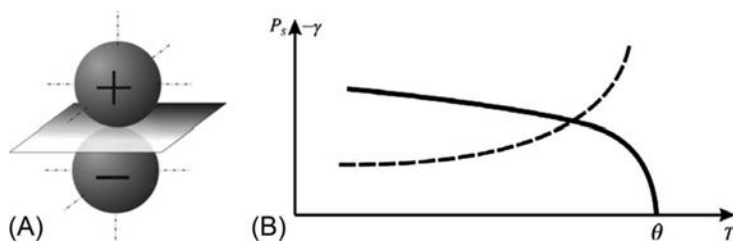


FIG. 9.39

Latent polarity for 1D model: (A) pyroelectric coefficient indicatory surface; (B) internal polarity (*firm line*) and pyrocoefficient (*dotted line*) temperature dependence.

the distribution of pyroelectric response in a crystal. Appropriate indicatory surface (indicatrix) consists of two spheres, as shown in Fig. 9.39A.

These spheres are located above and below the symmetry plane m , being characterized by the equation $\gamma(\varphi) = \gamma_{\max} \cos \varphi$. It is evident that spatial distribution of pyroelectric coefficient corresponds to 1D intrinsic polarity (which is traditionally called as spontaneous polarization: $P = P_{\max} \cos \theta$). The upper sphere is the indicatory surface for upper orientation of P_S , whereas the bottom sphere only means the change in sign of the pyrocoefficient γ_i , if spontaneous polarization would have the opposite direction.

Material vector γ reaches its maxim in the direction of ordinate that coincides with internal polarity. Hence, the γ_{\max} might be measured in the cut of crystal made perpendicular to *peculiar polar axis*. Angle φ is the angle between the ordinate axis and the slanting crystal's cut where the pyroelectric effect is studied.

It should be also mentioned that the energy of dipole-to-dipole interaction comparatively *weakly decreases with distance*: namely, as $\sim r^{-2}$. Such a far range has an influence on the intrinsic polarity temperature dependence (when chaotic movement aspires to destroy internal ordering). It is notable that 1D ordering is relatively stable; hence, it is capable of *strongly withstanding* the inescapable disorientation of thermal fluctuations. That is why common pyroelectric can maintain its internal polarity up to crystal melting. However, if the 3D thermal disordering can overcome the internal ordering steadiness in polar 1D system (this is the case of ferroelectricity, as shown in Fig. 9.39B), its collapse happens very fast (critically) and this gives rise to phase transition into the nonpolar phase, when polarization temperature dependence follows Landau law $P(T) = P_{\max}(\Theta - T)^{1/2}$, that is, with *critical index* "0.5."

Some *piezoelectrics* (nonpyroelectrics!), for instance, quartz, are characterized by *in-plane situated polar structural motives* of intrinsic polarity. Crystals of quartz-symmetry include berlinite (AlPO_4), cinnabar (HgS), tellurium (Te), and so on.

Spatial distribution of charges that characterize internal polarity in quartz-type crystals in the polar coordinates can be described as: $\gamma(\theta, \varphi) = \gamma_{\max} \sin^3 \theta \cos 3\varphi$, as shown in Fig. 9.40A, where θ is the azimuth angle and φ is the plane angle.

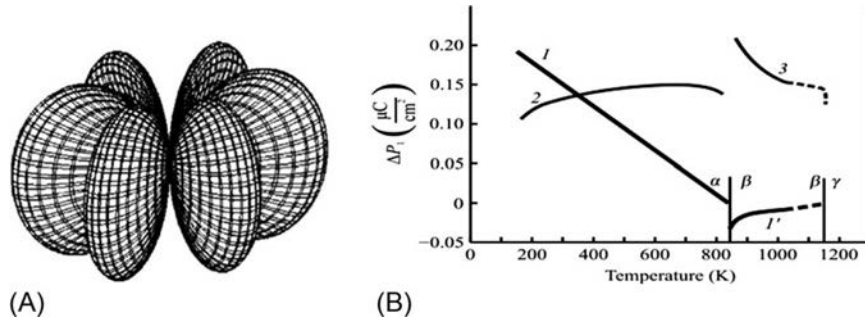


FIG. 9.40

Intrinsic polarity in quartz: (A) indicatory surface; (B) temperature dependence: I — M_1 found in $[100]$ -cut plate in α -phase; I' — M_1 found in $[110]$ rod in high-temperature β -phase; 2 —artificial pyroelectric coefficient of $[100]$ -thin cut, 3 —piezomodule d_{14} [10].

Through the radius vector, handled from the center of the figure shown, one can determine the magnitude of piezosensitivity in *any cut* of quartz-type crystals. It is obvious that maxims of this effect are seen along any of three X -type axes. No piezoelectric effect is possible in the directions of Y and Z axes.

In case of 2D intrinsic polarity correlation, the interaction energy of sextuple moments decreases with distance much *faster* than in case of 1D interaction, namely, as $\sim r^{-3}$, as shown in Fig. 9.37B. That is why such kind of intrinsic polarity (described by polar moment M_{ij} being second-rank tensor) can be destroyed by 3D thermal fluctuations more easily than in case of 1D correlation. It is remarkable that 2D intrinsic polarity correlation also ceases at quite definite critical temperature; hence, the crystals with 2D-polarity show phase transition into nonpolar state, as it is seen from Fig. 9.40B, with the example of quartz.

The most interesting is that “effective pyrocoefficient γ_1 ” disappears in the vicinity of $\alpha \rightarrow \beta$ phase transition of quartz at the temperature of $\theta_1 = 846$ K. It is also remarkable that quite interesting temperature dependence is observed for any of the *components* of intrinsic polarity $P_{[100]} = \Delta P_1$. Calculated from $\gamma_1(T)$ dependence, the intrinsic polarity $\Delta P_1 = \int \gamma_1 dT$ decreases with temperature *linearly*: $\Delta P_1 \sim (\theta - T)$, that is, with Landau *critical index* “1.” Previously, such linear temperature dependence of spontaneous polarization was observed only in the improper ferroelectrics.

It should be noted that other piezoelectrics of quartz symmetry (SiO_2 and AlPO_4) also show additional to $\alpha \rightarrow \beta$ phase transition second high-temperature transition, namely, $\beta \rightarrow \gamma$. In their β phase, quartz and berlinite are also piezoelectrics, but with intrinsic polarity described by the *octuple*. The highest-temperature γ -phase in these crystals is nonpolar.

Other *piezoelectrics* (also called nonpyroelectrics), for instance, gallium arsenide, or KH_2PO_4 (KDP in its *paraelectric phase*), are characterized by the dominating *volume-situated polar structural motive* of intrinsic polarity; typical for octuple (3D), the electrical moment is shown in Fig. 9.37C, which corresponds to the third-

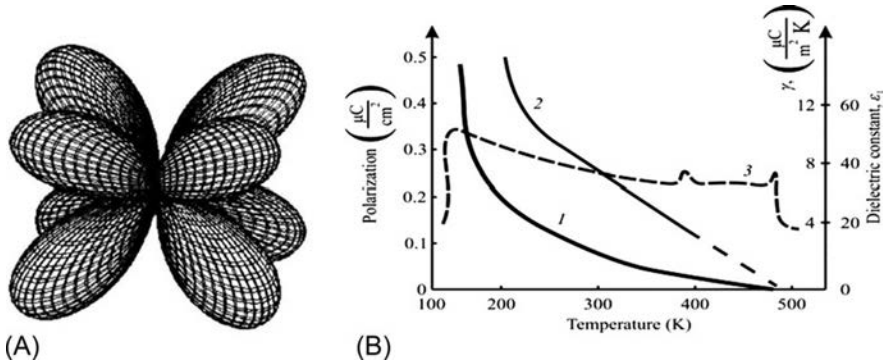


FIG. 9.41

Intrinsic polarity in the paraelectric phase of the KDP crystal: (A) indicatory surface; (B) partially clamped crystal temperature investigation: 1—octuple moment $|M| \sim (\theta - T)^2$; 2—effective pyroelectric coefficient; 3—permittivity ϵ_1 at microwaves [15].

rank material tensor. Spatial distribution of octuple moment is seen in Fig. 9.41A, and can be described as $P_{111} = P_{\max} \sin \theta \sin 2\theta \cos 2\varphi$.

Well-known ferroelectrics of the KDP type (as well as antiferroelectrics of the ADP type) in their *paraelectric phase* belong to the polar-neutral crystals, and at room temperature, they are only “actual” piezoelectrics. The intrinsic (hidden and compensated) polarity is arranged like the electrical moment of the 3D octuple type. It decreases with distance *very fast* (as r^{-4}), and this testifies the weak stability of the corresponding polar bonding. Fig. 9.41B shows some properties of the KDP crystal *above* its ferroelectric phase transition, where KDP is only a piezoelectric of 422 class of symmetry (below 150K, it becomes ferroelectric). It is seen that internal polarity quite gradually disappears by the law $P \sim (\theta - T)^2$, that is, with *critical index* “2.”

It is found that crystals of KDP type in their paraelectric phase also have (similar to quartz) high-temperature phase transition, where their polar-neutral 422 crystal symmetry group changes its symmetry to nonpolar group. Microwave measurements show a *slow decrease* in polar moment with temperature; the permittivity ϵ_1 drops at temperature of $\theta = 483$ K, above which the high-temperature phase transition is seen.

The conception of intrinsic octuple moment can be further applied for many other “actual” piezoelectrics. For instance, in the crystals of 23 cubic class of symmetry main polar directions correspond to four threefold axes. Fig. 9.42 shows temperature dependence of intrinsic polarity $P_{[111]}$ for GaAs and GaP in comparison with component $P_{[100]}$ of quartz intrinsic polarity.

In all studied piezoelectrics, their intrinsic polarity decreases while temperature increases. In the quartz crystal, this polarity disappears with temperature during $\alpha \rightarrow \beta$ transition. It is supposed that in the $A^{III}B^V$ crystals, their P_{111} intrinsic polarity also disappears, but at the melting temperature. It is obvious that in the melt, no firm polar bonds can be settled. However, during crystallization, the $A^{III}B^V$ crystals are

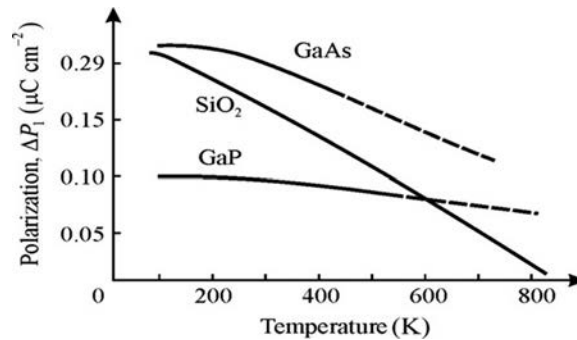


FIG. 9.42

Temperature dependencies of internal polarity component ΔP_{111} in GaAs and GaP crystals, in comparison to polarity component ΔP_{100} of SiO₂ (α -quartz) crystal.

widened because of polar bond formation; hence, the density of crystal becomes less than the density of a melt. While forming, the polar bonds swell up the crystal structure.

It is interesting to notice that effective (artificial) pyrocoefficient γ_{111} (obtained in *partially clamped crystal*) is rather big in case of the sillenite type crystals. Some of them such as Bi₁₂GeO₂₀ and Bi₁₂SiO₂₀ are widely used in electronics industry. Among them, it is possible to find effective pyroelectric coefficients up to 100 $\mu\text{C/m}^2\cdot\text{K}$.

Therefore, the physical nature of mechanoelectrical and thermoelectrical responses of noncentrosymmetric crystals (active dielectrics) is related. Their peculiar properties might be described by a substantial difference in the electronegativity of their atoms, which results in a complicated spatial distribution of their sensitivity to external influences on the crystal.

The pure ionic and pure covalent crystals are defined by their centrosymmetric structure, and they do not show piezoelectric or pyroelectric effects. On the contrary, noncentrosymmetric structures are the demonstration of *mixed* ionic-covalent bonds of their atoms. These bonds are exactly directional and therefore such structures result in different manifestations of asymmetry and complexity of the polar crystal structure. Although polar crystal formation from charged ions (initially, from liquid or vapor state of the material), noncentrosymmetric structures are formed spontaneously. This peculiarity can be described by a combination of electrically active construction of ions that leads to electrical moments, describable by tensors of different ranks.

Temperature steadiness of intracrystalline polar moment can be described by the critical law $M \sim (\theta - T)^n$: it means that a phase transition temperature θ exists, when electrical moment M disappears with temperature increase. The critical parameter is $n = 1$, if all components of intrinsic polarity are arranged in a plane (two-dimensional case, 2D). In the event of spatial (3D) arrangement of the latent polarity, exponent is $n = 2$. This differs essentially from 1D “spontaneous polarization” of ferroelectrics, which shows $P_S(T)$ critical law with $n = 0.5$ (Landau’s critical index).

9.12 SUMMARY

1. Polar (or *active*, or functional, or controlled, or smart) dielectrics may efficiently react to the changes in temperature, pressure, mechanical stress, electrical and magnetic fields, and light illumination. Polar dielectrics are classified as pyroelectrics, ferroelectrics, piezoelectrics, electrets, paraelectrics, optically active materials, and others. Polar dielectrics can be defined as materials that enable to directly convert energy and transform information. Indeed, a piezoelectric material converts electrical energy to mechanical energy (and vice versa); a pyroelectric material is a thermoelectric (and therefore electrothermal) power converter; nonlinear dielectrics enable to transform the frequency of signal and to perform modulation and detection (i.e., conversion of information). These transformative functions are due to the peculiar physical structure and chemical composition of certain dielectrics.
2. *Piezoelectric* can convert mechanical energy into electrical energy or vice versa (electrical energy into mechanical energy). Originally, the mechanoelectric effect was discovered first, and for this reason, it is called as “*direct*” piezoelectric effect; afterwards, the “*inverse*” (electromechanical) effect was detected. Piezoelectric effect is the odd (linear) effect, at which *mechanically induced* polarization is directly proportional to strain (and, vice versa, *induced by electrical field* strain is directly proportional to electrical field). Piezoelectric effect is observed only in the noncentrosymmetric crystals and textures.
3. Piezoelectric effect can be described by *piezoelectric module*, which is the *third-rank tensor*. Mathematical relationships that define piezoelectric effect depend on a combination of various boundary (limiting) conditions, under which piezoelectricity is used or studied. Mechanical boundary conditions mean the possibility of deformation (if the crystal is *free*) or deformation inability (when the crystal is *clamped*). Boundary electrical conditions are as follows: crystals can be *short circuit* or *open circuit*.
4. Mechanical property (elasticity) and electrical property (polarization) of piezoelectric crystals are *interrelated*, and therefore they can be considered together. Their relationship is characterized by the *electromechanical coupling coefficient* K_{EM} —one of the most important parameters of piezoelectric materials and devices. In the case of direct piezoelectric effect imparted to a piezoelectric, mechanical energy not only is spent on its elastic deformation but also creates electrical polarization that causes electrical energy accumulation. Conversely, supplied to piezoelectric, electrical energy (in case of the inverse piezoelectric effect) is spent not only for its polarization but also for its elastic deformation and elastic energy accumulation.
5. The square of electromechanical coupling coefficient K_{EM}^2 shows what part of energy, attached to a piezoelectric, is converted into the energy of other kind. However, this parameter is not a performance factor: first, because the losses of electrical or mechanical power are not considered, and second, the actual conversion efficiency of piezoelectric depends not only on the K_{EM} but also largely on the shape, orientation, and other peculiarities of the piezoelectric element.

6. Manifestation and parameters of piezoelectric effect *can be controlled* by bias electrical field: for example, piezoelectric filter characteristics can be changed, as well as surface acoustic wave parameters. Electrical controlling by piezoelectric properties is most manifested in dielectrics with high permittivity.
7. *Electrostriction* is observed in all dielectrics as the *even effect*, when deformation of dielectric caused by electrical field is proportional to the square of field strength. Thus the sign of strain during electrostriction does not change with field sign alteration. From piezoelectric effect, electrostriction differs by its unidirectionality, that is, the effect of electrostriction is exceptionally an electromechanical, but not “mechanoelectrical,” effect.
8. Piezoelectric effect can be *induced by electrical field* in any solid dielectric (in the form of “linearized electrostriction”); the efficiency of electrically induced piezoelectric effect in diffused phase transition ferroelectrics may even exceed the efficiency of the best piezoelectric material.
9. *Pyroelectric* effect and piezoelectric effect can be considered as solid-state energy converter. However, if piezoelectricity is associated with electromechanical conversion, the pyroelectricity appears to be *thermoelectric* power conversion. Such transformation of energy by a solid is possible, if dielectric (crystal, polycrystal, or polymer) is polarized; as a rule, if it has spontaneous polarization, thermoelectric power conversion is defined as the pyroelectric effect, whereas inverse electrothermal energy conversion is the *electrocaloric effect*.
10. Pyroelectricity is characterized by *primary and secondary* effects. Primary effect is caused by temperature change in spontaneous polarization of polar crystals. Secondary effect is piezotransformed thermal deformation of crystal. It should be noted that similar effects can be artificially induced in all solid dielectrics, if they are to undergo a strong electrical field. Hidden (or latent) intrinsic polarity in pyroelectric is only the ability to provide electrical (vector) response to any nonelectrical scalar dynamic impact—uniform change of temperature or hydrostatic pressure.
11. Research and application of dielectrics is commonly held under the *adiabatic* conditions, when at the time of voltage change the *thermal equilibrium* between dielectric and surrounding environment does not succeed in establishing itself; hence, the entropy is not changed: $\delta S = 0$. Therefore from experiments, the adiabatic dielectric permittivity ϵ^S is measured. In dielectrics whose polarization depends on temperature (ferroelectrics, pyroelectrics, and others), another—*isothermal*—the process of polarization might be important, when $\delta T = 0$ and dielectric permittivity is isothermal: ϵ^T . Analytical determination of ϵ^T and ϵ^S relationship might be important to explain ϵ frequency dependence in the range of subsonic frequencies and in some theoretical calculations. Isothermal dielectric permittivity is always greater than adiabatic: $\epsilon^T > \epsilon^S$. As a rule, this difference is small and can be neglected. However, in pyroelectrics, and, especially, in the vicinity of ferroelectric phase transition, the difference between ϵ^T and ϵ^S can reach 10%–50%; hence, it should be considered.

12. The *electrets* are dielectrics that maintain for a long time their polarized state owing to the *residual polarization*. Electrets are able to create direct electrical field in the surrounding space. Existence of quasipermanent polarization (electret state) is described by two types of localized charges in the insulator—heterocharge and homocharge.
13. The *heterocharge* creates electrical polarization in the volume of dielectric due to the “frozen” *migratory polarization* (orientation of macro-dipoles, or space charge accumulation, and so on). In this case, negative charge is concentrated near the anode, whereas positive charge is located at the cathode, so that residual electrical field has an *opposite* direction to the polarization field.
14. The *homocharge* in the electrets is the result of injection of charge carriers into a dielectric or due to irradiation of dielectric by electronic beam. Then this charge localizes in the traps with corresponding energy levels located in the band gap of dielectric. In case of bonded negative charge, it is located near the cathode, whereas near the anode, the related positive charge is settled; the resulting field in electrets has a direction same as that of the polarization field.
15. Generally, to receive *quasipermanent polarization* in dielectric, it should be placed in a strong electrical field and subjected to *additional physical treatments* that reduce the relaxation time of dipoles (or accelerates the migration of charged particles). However, electret state may be created without application to dielectric electrical field, for example, due to mechanical deformation (*mechanoelectrets*) by charging of dielectric in corona discharge, or under discharge plasma (*electroelectrets*), as well as owing to electrification by friction (*triboelectrets*), and so on.
16. Method of *thermally stimulated depolarization*, also called as thermoelectrets analysis or thermally stimulated current (TSC-*spectroscopy*), is based on the detection of weak currents arising at the process of thermal destruction of created polarization, previously induced in dielectric by a strong direct electrical field. This method utilizes the property of dielectrics to store and maintain for a long time *residual polarized state* that exists in electrets. During TSC study, the dependence of depolarization current against temperature is usually recorded. The investigation of temperature-stimulated electrical currents is one of the widely used methods of determination of electronic local energy levels in solids.
17. *Paraelectrics* are crystals with peculiar temperature dependence of dielectric constant that is described by the Curie-Weiss law: $\epsilon(T) = \epsilon_1 + C/(T - \theta)$. In most of these crystals at a certain temperature called the *critical temperature* ($T = T_C$), the phase transition to ferroelectric or antiferroelectric state occurs. Thus, the properties of paraelectrics are closely related to those of ferroelectrics.
18. Similar to ferroelectrics, paraelectrics can be divided into two basic classes. Paraelectrics of the *order-disorder type* are polar (dipole-type) crystals. With temperature decrease, owing to dipole-dipole interaction, a gradual ordering of dipole orientations takes place, until, finally, below the Curie

temperature T_C , spontaneous polarization arises, at which most dipoles are already ordered. Order-disorder paraelectrics above their phase transition are characterized by rather steep dependence of $\epsilon(T)$: their Curie-Weiss constant is relatively small ($C \approx 10^3$ K), whereas paraelectric-ferroelectric transition usually has a second-order type.

19. The other type (*displacement type*) of paraelectrics are those that are ionic crystals, in which the *covalent bonding* of atoms is very important. In this case, temperature dependence of permittivity appears more flat: $C \approx 10^5$ K, and phase transition to ferroelectric (or antiferroelectric) state occurs at temperature $T_C > \theta$, having a pattern of the first-order type transition. To describe the properties of these paraelectrics, a dynamic theory of lattice vibrations and the model of “*soft phonon mode*” should be applied.
20. *Thermally stable high-permittivity dielectrics* are important components of microwave electronics based on such displacement-type paraelectrics, which never acquire a polar structure. The reason for the high permittivity is low-frequency “soft lattice mode,” whereas increased thermal stability is obtained due to suppression $\epsilon(T)$ critical change by paramagnetism. Low microwave losses can be obtained only in the monophasic composition.
21. The *dominating mechanism of losses* in high- ϵ microwave dielectrics occurs due to the existence of *polar phase*. In case of polar bonds between atoms, their atomic potential manifests pronounced anharmonicity, which is the main microscopic channel to transfer electrical energy into heat. In the dynamic theory, anharmonic potential is described by a coupling between optical phonons and acoustical phonons. With increase in temperature and, consequently, with increase in amplitudes of atom vibration, manifestation of anharmonicity becomes more noticeable, and losses increase. Consequently, during elaboration of microwave dielectrics, any polar (noncentrosymmetric) component in their composition should be avoided.
22. *Ferroelectrics* are characterized by spontaneous polarization, whose direction can be changed by an externally applied electrical field. Except ferroelectrics, the steady polarization is inherent to electrets as well as to pyroelectric crystals. But in contrast to nonequilibrium *residual* polarization of electrets, in pyroelectrics, internal polarization is in a thermodynamically stable state. Ferroelectric differs from pyroelectric by its ability to repolarize: switching of spontaneous polarization in the external electrical field (dielectric hysteresis). Additionally, ferroelectrics may be defined as ferroelectric is a pyroelectric that divides on domains.
23. Ferroelectrics are characterized by strongly expressed *nonlinear properties*. At audio and radio frequencies, nonlinear ferroelectric capacitors (*variable capacitors*) allow the modulation of signals and generation of higher harmonics. At microwave frequencies, a thin film of displacement-type ferroelectrics in their paraelectric phase can be used for microwave phase modulation and for adjustable filters. Optical nonlinearity of ferroelectrics is applied in the electro-optics and nonlinear optics.
24. By analogy with magnetism (where ferromagnetics, antiferromagnetics, and ferrimagnetics can exist), not only the ferroelectrics but also

antiferroelectrics and *ferrielectrics* are known. Spontaneous polarization of antiferroelectric is compensated already in a crystal unit cell, whereas in the ferrielectrics, their antipolarization is not totally compensated. Mechanical analogs of ferromagnetics and ferroelectrics are the *ferroelastics*, in which phase transition occurs owing to *spontaneous deformation*. Coexistence of spontaneous deformation and spontaneous polarization is inherent to the *improper* ferroelectrics.

25. The *nonlinearity* of ferroelectric is expressed as hysteresis loop with large dependence of permittivity on electrical field. Nonlinearity of the *effective permittivity* occurs in alternating field, when effective permittivity $\epsilon_{\text{ef}}(E)$ changes up to 50 times. At *reversible* nonlinearity (i.e., ϵ -controlling by bias electrical field), change in ϵ can reach up to 10 times. Reversible nonlinearity can be applied in various microwave devices based on thin paraelectric films.
26. Spontaneously polarized crystals (pyroelectrics and ferroelectrics) differ from other piezoelectrics by the one-dimensional ordering (1D), conditional with their “*dipole-type structural motive*.” Similarly (and with the same level of approximation), in the “actual” piezoelectrics (nonpyroelectrics), the existence of intrinsic (hidden) polarity might be assumed. Investigation of pyroelectric response in the piezoelectrics with partially limited thermal deformations enables to experimentally decompensate components of “*sextuple (2D) structural motive*” and “*octuple (3D) intrinsic polarity*” in the piezoelectric crystals.

REFERENCES

- [1] Y.M. Poplavko, L.P. Pereverzeva, I.P. Raevskiy, *Physics of Active Dielectrics*, South Federal University of Russia, Rostov-na-Donu, 2009.
- [2] J.C. Burfoot, G.W. Taylor, *Polar Dielectrics and Their Application*, Macmillan Press, New Jersey, 1979.
- [3] Y.M. Poplavko, Y.I. Yakimenko, *Piezoelectrics*, Polytechnic Institute, Kiev, 2013.
- [4] I.S. Rez, Y.M. Poplavko, *Dielectrics: Main Properties and Electronics Applications*, Radio e Svyaz, Moscow, 1989.
- [5] K. Uchino, *Ferroelectric Devices*, Marcel Dekker, New York, 2000.
- [6] W.O. Cady, *Piezoelectricity*, Amazon com, New York, 1946.
- [7] Y.M. Poplavko, L.P. Pereverzeva, A.S. Voronov, Y.I. Yakimenko, *Material Sciences, Part II, Dielectrics*, Polytechnic, Kiev, 2007.
- [8] M.E. Lines, A.M. Glass, *Principles and Application of Ferroelectrics and Related Materials*, Clarendon Press, Oxford, 1977.
- [9] Y.M. Poplavko, *Physics of Dielectrics*, Vischa Skola, Kiev, 1980.
- [10] Y.M. Poplavko, L. Pereverzeva, N.-I. Cho, Y.S. You, Feasibility of microelectronic quartz temperature and pressure sensors, *Jpn. J. Appl. Phys.* 37 (1998) 4041.
- [11] Y.I. Sirotin, M.I. Shaskolskaya, *Basic of Crystal Physics*, Nauka, Moscow, 1979.
- [12] B.M. Tareev, *Physics of Dielectric Materials*, Energia, Moscow, 1973.
- [13] Y.M. Poplavko, Y.V. Didenko, D.D. Tatarchuk, Temperature dependences of losses in high frequency dielectrics, *Electron. Commun.* 19 (4(81)) (2014) 9.

- [14] I.S. Jeludev, *Basics of Ferroelectricity*, Atomizdat, Moscow, 1973.
- [15] Y.M. Poplavko, L.P. Pereverzeva, Pyroelectricity of partially clamped piezoelectrics, *Ferroelectrics (USA)* 130 (1992) 361.
- [16] Y.M. Poplavko, N.-I. Cho, Clamping influence on ferroelectric integrated film microwave properties, *Semicond. Sci. Technol.* 14 (1999) 961.
- [17] L.P. Pereverzeva, Y.M. Poplavko, Pyroelectricity in non-central crystals, *Acta Phys. Pol. A* 84 (2) (1993) 287.
- [18] Y.M. Poplavko, L.P. Pereverzeva, Pyroelectric response of piezoelectrics, *Ferroelectrics (USA)* 134 (1992) 207.
- [19] K.M. Rabe, C.H. Ahn, J.-M. Triscons (Eds.), *Physics of Ferroelectrics: A Modern Perspectives*, Springer, Berlin, 2007.

Phase transitions in solids 10

CONTENTS

10.1 Phase Transitions of First and Second Orders	602
10.2 Physical Meaning of Ordering Parameter	606
10.3 Phase Transitions With Anomalies in Dielectric Properties	609
10.4 Phase Transitions With Conductivity Anomalies	628
10.5 Phase Transitions in Liquid Crystals	644
10.6 Large Parameters and Giant Effects in Electronic Materials	649
10.7 Summary	659
References	662

From a phenomenological point of view, *phase* is defined as a physically and chemically uniform state of matter that has a specific set of properties. After transition to another phase, the substance remains uniform, but acquires other properties. In chemistry, well-known phase transitions (PTs) are the aggregate conversions: “*gas* \Leftrightarrow *liquid* \Leftrightarrow *solid*” phases.

However, in many technical applications of *solid materials*, primary importance is given to another type of PTs that occur *inside* the solid state or *inside* the liquid state without changing the aggregate state. These conversions take place because of the interactions of matter’s structural elements: electrons, ions, dipoles, and molecules (or their complexes).

The question is that near PT the structure of a matter is *extremely sensitive* to any external influences (thermal, electrical, magnetic, or mechanical); even in case of minor changes of T , E , H , or X , electrical, magnetic, optical, and other properties of substance demonstrate considerable variations. This unusually high sensitivity near the PT even to weak external influences is used in many devices and equipment of electronics and instrumentation [1].

Usually, in the event of a change in external conditions, main electrical properties of crystals (conductivity, permittivity, permeability, optical activity, and others) change *smoothly*. However, near the PT these parameters vary essentially, even by a *jump*, sometimes, in several orders of magnitude. This means significant modification in the crystal properties [2].

Provoked by external influence, PTs in the magnets, metals, dielectrics, and semiconductors are accompanied by substantial restructuring of material subsystems: electronic, phonon, dipole, magnetic, or others. This restructuring is obviously accompanied by a change in crystal symmetry in the vicinity of PT. For instance,

in the polymers different PTs are determined by the peculiarities of their intermolecular interactions: vitrifying, orientation melting, crystallization, and so on. Another example is the liquid crystals (LCs), where PTs are conditioned by the appearance (or change) of linear, planar, or helical ordering of anisometric molecules.

When microscopic characteristics of any PT are discussed, a *regulating (ordering) parameter* η should be considered. In crystals, it is the measure of structural deviation from the state of highest symmetry. Depending on what kind of microscopic interactions causes PT and what changes of structure take place, the ordering parameter η acquires different physical meanings. For example, in ferroelectrics the ordering parameter may correspond to the degree of electrical dipole regularity, in the ferromagnetics parameter η describes the ordering in the system of magnetic moments (spins), and so on. The ordering parameter may also have broader content; for example, in case of PT with the aggregate conversion parameter η characterizes the degree of regularity in a mutual arrangement of atoms or molecules [3].

10.1 PHASE TRANSITIONS OF FIRST AND SECOND ORDERS

Despite the wide variety of PTs in different materials, it is possible to introduce some *general criteria* for their classification. According to P. Ehrenfest's theory, this classification is based on thermodynamic function peculiarities and on their derivatives, which demonstrate parameters changing in the vicinity of PT. The type of PT is determined by the *lowest nonzero partial derivative* from the thermodynamic potential (Φ). This potential is a specific function of volume V , temperature T , pressure p , entropy S , and other macroscopic parameters that describe a state of thermodynamic system and represents substance exposed to PT [4].

The classical examples of the *first-order* PTs (PT-I) are melting (or crystallization) and vaporization (or condensation). If, during PT from one phase to another, the *stepped changing* of the entropy S takes place, in which connection heat release (or absorption) is observed, such a transition, following Ehrenfest's classification, belongs to the PT-I.

However, in solid-state physics, the *second-order* PTs (PT-II) are more typical, when the *second derivatives* of thermodynamic potentials show sharp changes. As a rule, although these transitions occur within one (usually crystalline) state of aggregation, they might have some features of first-order transitions, yet being closer in their nature to the PT-II.

A comparison of thermodynamic potential in two phases—phase I (Φ_1) and phase II (Φ_2)—is shown in Fig. 10.1, simultaneously in the temperature and pressure intervals. At lower temperature, phase II is more stable, because its energy is smaller, while at higher temperature, for the same reasons phase I stability is higher than that of phase II.

PT occurs when the energies of phases I and II are equalized—this happens at temperature T_0 . Therefore when temperature increases, the energy of a system varies along

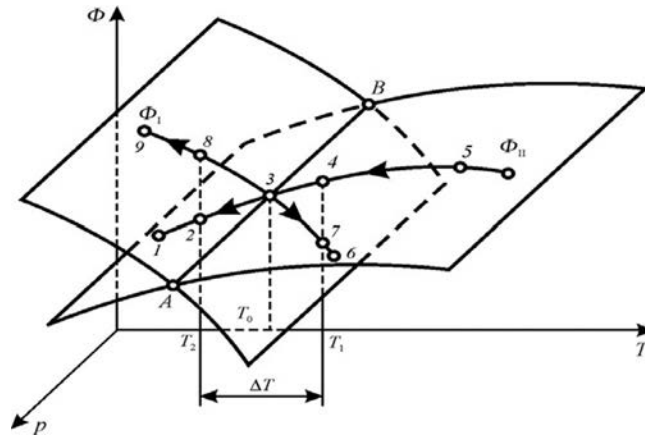


FIG. 10.1

Thermodynamic potential of crystal passing through phase transition; ΔT is temperature hysteresis in case of PT-I.

the line 1–2–3–4–5 while at reduced temperature it follows the line 6–7–3–8–9; in fact, transition can occur in point 3 that corresponds to temperature T_0 . The dependence of potential Φ on the pressure p is also shown in Fig. 10.1. Line segment AB denotes the intersection of two surfaces that meet the same potential of both phases. To the left of this line, less energy is seen for phase II, because this phase is more stable here. To the right of line AB , by contrast, more stable (energetically favorable) is phase I.

First-order phase transition (PT-I) suggests the possibilities of *overheating* and *overcooling*. In case of heating transition occurs at temperature T_1 (line 1–2–3–4–7), while in case of cooling transition occurs at temperature T_2 (following line 6–7–3–8–2). Thus PT-I is characterized by *thermal hysteresis*, namely, during cooling, phase I is converted into phase II not at temperature $T = T_0$, but at lower temperature T_2 , that is, the overcooling of phase I is seen. In the same manner, while heating, phase II conversion occurs at temperature $T = T_1$, which means the overheating of phase II. The temperature range of hysteresis $\Delta T = T_1 - T_2$ depends on many factors, particularly, on the speed of temperature change, as well as on the structural imperfections and on the purity of a substance.

In the vicinity of PT-I, the *entropy* shows a ΔS jump that characterizes the change in latent heat. In the same manner, a sudden change of the *volume* ΔV is observed, as shown in Fig. 10.2A. In the event that PT nature is close to PT-II, the size of these jumps depends on different crystal properties. The smaller the jumps of ΔS and ΔV , the closer the nature of transition to second-order transformation (when $\Delta S \rightarrow 0$ and $\Delta V \rightarrow 0$).

In the event that PT-I occurs, due to jumps in thermodynamic functions, their derivatives are infinite. The first derivative of entropy is the specific heat C_p , while

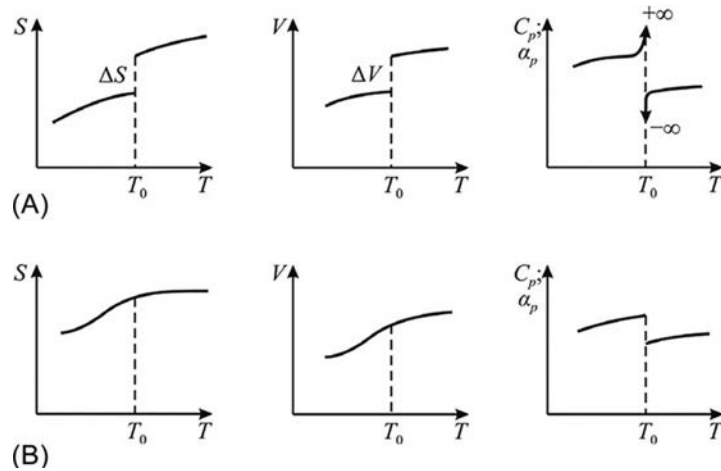


FIG. 10.2

Temperature change of thermodynamic parameters in the vicinity of PT-I (a) and PT-II (b).

the first derivative of volume is the thermal expansion coefficient α_p (both parameters are determined under constant pressure p):

$$C_p = T (\partial S / \partial T)_p; \alpha_p = V^{-1} (\partial V / \partial T)_p. \quad (10.1)$$

Experimental studies of specific heat $C_p(T)$ temperature dependence and thermal expansion coefficient $\alpha_p(T)$ indicate that near PT-I these parameters have narrow but large extremes.

PT-II is characterized by such changes in crystal properties, at which thermodynamic potential, entropy, and volume show uninterrupted change, but their derivatives, in particular C_p and α_p , have finite jump (Fig. 10.2B). In the experiments, in the vicinity of PT-II at critical point (transition point), maximums in temperature dependences of $C_p(T)$ and $\alpha_p(T)$ are observed, but they are not so large, as in the case of PT-I.

PT division to PT-I and PT-II is confirmed by many experiments, devoted to study of thermal, electrical, magnetic, and mechanical properties of different substances. It turns out that a very large jump of entropy, typical for PT-I, is really observed only during transitions, caused by the aggregation state change ($gas \Leftrightarrow liquid \Leftrightarrow crystal$).

If phase transformations are observed in experiments occurring with one of *condensed matter*, that is, within only crystalline or only a liquid substance, these PTs, usually, are only close but not entirely adequate to the mentioned classification. As a rule, transitions of second order dominate that, however, might have *some signs* of first-order transition. During PT-II only the *symmetry changes abruptly* (as a rule, symmetry lowers *below* critical temperature T_0), but thermodynamic potentials change almost continuously (Fig. 10.2B).

Second-order transitions are gradual and smooth, and they do not show temperature hysteresis and they are not accompanied by discontinuous jump in energy or in volume of a crystal. However, as a result of this transition, the new physical property appears as follows: crystal becomes ferroelectric, ferromagnetic, ferroelastic, superconductive, and so on [5].

During the investigation of PTs, as in crystals so in LCs, more complex regularities are observed, differing from the ones described in the idealized limiting cases of PT-I and PT-II. For this reason, Ehrenfest theory offers satisfactory explanation only for some *qualitative features* of material properties while a phase transitions.

Thermodynamic theory that *quantitatively* describes the changes in matter properties in the vicinity of PT-II (and in the vicinity of some PT-I occurring with low jump of entropy, i.e., close to PT-II) is the theory of Landau. This theory can be applied exceptionally to PTs with a change in symmetry, but only for the liquid or crystalline state of a matter.

According to *Landau's theory*, a variable with the temperature *ordering parameter* η is introduced. Depending on the symmetry of a crystal, in one phase (usually—in the higher temperature phase) the studied system is not ordered and $\langle \eta \rangle = 0$. The correspondingly ordered phase (at lower temperature) shows $\langle \eta \rangle \neq 0$. According to Landau, the thermodynamic potential of the system $\Phi(T, \eta)$ can be represented as series in powers of the parameter on order parameter:

$$\Phi(T, \eta) = \Phi_0(T) + A\eta + B\eta^2 + C\eta^3 + D\eta^4 + E\eta^5 + \dots \quad (10.2)$$

where $\Phi_0(T)$ is the part of free energy that does not depend on η , while coefficients A, B, \dots may depend on temperature (for simplicity, pressure is assumed as being constant) [2].

As any phase exists at the thermodynamic equilibrium, then a minimum of free energy should exist at the condition: $\partial\Phi/\partial\eta = 0$. Hence, as evidence shows, the coefficient at first degree of η must be equal to zero: $\partial\Phi/\partial\eta = A = 0$. In addition, it is supposed that the disordered (higher temperature) phase has *center of symmetry*; that is why minimum $\Phi(\eta)$ in vicinity $\eta = 0$ should be also symmetric, and series $\Phi(T, \eta)$ should be characterized only by *even numbers*. Therefore, all coefficients at odd powers of η in [expression \(10.2\)](#) must be zero. This allows to rewrite [expression \(10.2\)](#) for thermodynamic potential in a form:

$$\Phi(T, \eta) = \Phi_0(T) + \frac{1}{2}\alpha\eta^2 + \frac{1}{4}\beta\eta^4 + \frac{1}{6}\gamma\eta^6 + \dots, \quad (10.3)$$

where α, β , and γ are coefficients of terms of series of thermodynamic potential $\Phi(\eta)$. Regarding these, the new designations (α, β, γ instead of B, D, \dots) are used, while coefficients $\frac{1}{2}, \frac{1}{4}, \dots$ are introduced solely to facilitate expressions, derived from differentiation of thermodynamic potential.

Analyzing the expansion in series [\(10.3\)](#), it is possible to arrive at the *critical dependence of α on temperature*. In case of transition PT-II terms β and γ temperature dependence is not critical; therefore, in a simple case, their temperature dependence might be neglected [2].

The critical temperature dependence of α could be justified by the next argumentation. Phase transformation occurs at $T = \theta$; then, from the condition of thermodynamic potential *minimum* ($\partial^2 \Phi / \partial \eta^2 = 0$) it follows that at lower temperatures (below θ), that is, in the ordered phase, where $\eta \neq 0$, the value of α should be *negative*: $\alpha < 0$. Similarly, it might be shown that at higher temperature (in the disordered phase), where the average ordering parameter $\eta = 0$, the value of first coefficient should be *positive*: $\alpha > 0$. Hence, parameter α is dependent on temperature, and in the very point of transition $\alpha = 0$.

Using Taylor series for $\alpha(T)$ by expansion in powers of $(T - \theta)$ in the vicinity of phase transition and paying attention to *small deviation* from transition, it is possible to limit by the first member of this series:

$$\alpha(T) = \alpha_0(T - \theta), \quad (10.4)$$

where α_0 is independent of temperature coefficient.

According to Landau's theory, the type of phase transition (PT-I or PT-II) is determined by the sign of coefficient β at fourth degree of ordering parameter.

If $\beta > 0$, it is the PT-II. In this case, there is no need to consider coefficient γ at η^6 (as well as higher degrees of ordering parameter) because the sustainability of system is entirely guaranteed by the member $\frac{1}{4}\beta\eta^4$ in expansion (10.3). Thus thermodynamic description of PT-II by expansion $\Phi(T, \eta)$ takes *polynomial form*:

$$\Phi(T, \eta) = \Phi_0(T) + (\alpha/2)\eta^2 + (\beta/4)\eta^4, \quad (10.5)$$

where $\alpha = \alpha_0(T - \theta)$ and $\beta > 0$.

If $\beta < 0$, it is the PT-I. At the condition $\beta < 0$, the sustainability of a system must be provided by taking into account the *next term* $\frac{1}{6}\gamma\eta^6$ in expansion (10.3) by the assumption $\gamma > 0$. The expansion of thermodynamic potential in this case takes also polynomial form.

$$\Phi(T, \eta) = \Phi_0(T) + (\alpha/2)\eta^2 + (\beta/4)\eta^4 + (\gamma/6)\eta^6, \quad (10.6)$$

where $\alpha = \alpha_0(T - \theta)$. Coefficient β *also can vary with temperature*, but the last coefficient should be positive ($\gamma > 0$) because this ensures stability of phases.

Owing to *polynomial form* of free energy presentation, Landau's theory allows not only quantitative description of changes of crystal properties near PT, but also prediction of many physical characteristics. The particular shape of functions (10.5) and (10.6) depends on physical meaning of the ordering parameter.

10.2 PHYSICAL MEANING OF ORDERING PARAMETER

Similar to other phenomenological parameters, the ordering parameter η may be regarded as the *macroscopic value*. It is important to note that the value of this parameter can fluctuate and change both in space (from point to point) and in time. The value of η can be represented as generalized consideration of such physical quantity that shows an abnormal change due to PT (magnetization, polarization, deformation, and so on).

Thermal motion in matter significantly affects the degree of its ordering. In case of temperature reduction, the degree of ordering increases, but the rate of temporary setting of process (its relaxation) at low temperature slows down (relaxation lies in thermodynamic equilibrium establishing in a system). Conversely, the higher the temperature, the faster the setting of any degree of ordering. However, the rate of ordering with a time is the *slowest* near the point of PT—this feature is *critical slowing down* of a process. Relaxation of ordering in this area is very slow, thanks to the convergence of free energy of both phases.

The system, in which PT happens, can have different degrees of freedom. However, the only *critical* degree of freedom is associated with the ordering parameter that abnormally reveals itself at PT. Other (noncritical) degrees of freedom of system remain fast enough when equilibrium establishes with a time.

Reflecting the *anisotropy* of crystal properties, ordering parameter might be the tensor of appropriate rank. For example, in ferromagnetic the parameter η is the spontaneous magnetization *vector* \mathbf{M} (first-rank tensor); ordering parameter in ferroelectrics is the spontaneous *polarization vector* \mathbf{P} , while in ferroelastics η is one of components of *second-rank tensor* of mechanical deformation x_{ij} . In the simplest cases, ordering parameter can also be *scalar* (zero-rank tensor).

As an example of scalar η case, the self-ordered alloy AB will be considered. In the simplest case it consists of two components: A and B, in which the number of atoms is same; the well-ordered phase corresponds to structure ABAB... To each of the components its own sublattice can be attributed. Denote the probability of A-type atom detection in the A sublattice as P_{AA} , while in the sublattice B it is P_{AB} . As the ordering parameter, the value $\eta = P_{AA} - P_{AB}$ should be selected.

In more complex cases, related to the transitions in crystals, the parameter η becomes multicomponent (corresponding values and their relationship with crystal are defined as the Lifshits groups). Then consider such changes in a structure, which take place during PT-II, limited to simplest models (Strukov). Consider the flat model of diatomic crystal (Fig. 10.3A) [3]. The point group of symmetry of this structure contains following elements of symmetry: $\bar{1}$ (center of symmetry), 4, $\bar{2}$, 2 (axis of symmetry), m_1, m_2, m_3, m_4 (planes of symmetry).

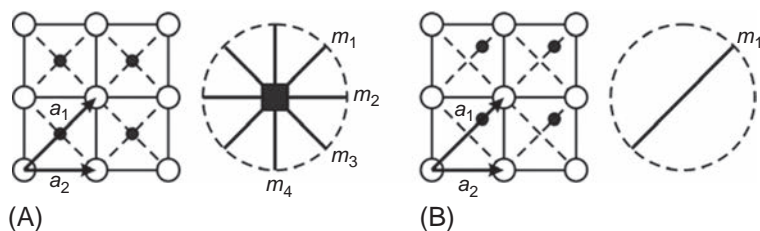


FIG. 10.3

Flat diatomic crystal model and elements of point symmetry (A); symmetry reduction as a result of relative displacement of sublattice elements of point symmetry (B).

Suppose that at certain temperature one of sublattice shifts relatively to another in the direction \mathbf{a}_1 , as shown in Fig. 10.3B. After such small displacement, the symmetry group of this model will have only two elements of symmetry: $\bar{1}$ and m . The initial phase, shown in Fig. 10.3A, is a symmetric one, but the new phase with lower symmetry (Fig. 10.3B) is asymmetric.

As a result of the shift, the coordinates of lattice atoms in crystallographic system are changed as follows:

The coordinates of atoms before shifting:

$$\begin{aligned} n_1\mathbf{a}_1, n_2\mathbf{a}_2 &\text{—for atoms O,} \\ (n_1 + \frac{1}{2})\mathbf{a}_1, n_2\mathbf{a}_2 &\text{—for atoms } \bullet. \end{aligned}$$

The coordinates of the atoms after shifting:

$$\begin{aligned} n_1\mathbf{a}_1, n_2\mathbf{a}_2 &\text{—for atoms O,} \\ (n_1 + \frac{1}{2}\eta)\mathbf{a}_1, n_2\mathbf{a}_2 &\text{—for atoms } \bullet, \end{aligned}$$

where η is part of \mathbf{a}_1 , on which atoms of “internal” sublattice are shifted. Exactly this specific parameter of PT is called the ordering parameter, and it is seen that $\eta = 0$ for the symmetric phase and $\eta \neq 0$ for the asymmetric phase.

The ordering parameter may vary with temperature rise, as shown in Fig. 10.4. Continuous change in the ordering parameter with temperature in the ordered phase is a characteristic feature of PT-II (Fig. 10.4A). On the contrary, the ordering parameter can be changed by a jump (Fig. 10.4C); this is an ideal case and it corresponds to PT-I.

In many experimental situations, $\eta(T)$ dependence very often changes by one of “intermediate” ways, as shown in Fig. 10.4B. In this case, PT is *transition of first order, close to transition of second order*. Here the ordering parameter η first changes with temperature gradually, but then abruptly falls down to the disordered phase [3].

PT of second order can be accompanied by multiplication of the size of crystal unit cell. Then the volume of unit cell in the low-symmetry phase (more ordered) increases in 2, 4, and 8 times, as well as the translational symmetry of unit cell also changes.

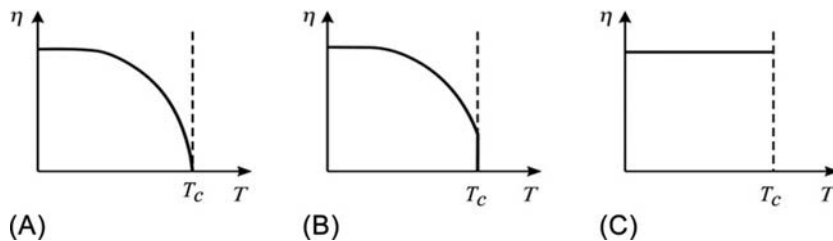


FIG. 10.4

Temperature dependence of order parameter for PT-II (A); PT-I close to PT-II (B), and for PT-I (C).

Based on the microscopic changes in a structure, PTs are divided into the *displacement type* and the *order-disorder type*. As shown in Fig. 10.3, the model of PT-II is characterized by a fact that in point of PT the displacements of atoms occur, changing the symmetry of a crystal. Therefore, these transitions are called the PTs of displacement type.

The regulating parameter for the PT of order-disorder type is described likewise. The symmetry of crystal can be changed as a result of *probability redistribution* for atom location in different states of equilibrium in the unit cell. Suppose that the number of these states is two, and they are “settled” according to concentrations of atoms N_1 and N_2 (numbers of atoms) and their energy positions are 1 and 2, respectively. At sufficiently high temperatures they are settled evenly; therefore, the probability of the presence of atom in each position is same: $N_1 = N_2$. As the average (in time) stay of atoms remains at the center of a cell, the symmetry of structure does not change.

When temperature decreases, PT to ordered phase appears described by the *difference of probability* of particles to stay in different positions, that is, as difference between the number of particles in these positions. Below the point of PT parameter

$$\eta = (N_1 - N_2) / (N_1 + N_2)$$

acquires nonzero value. In this case, more natural, parameter η can be called as the order parameter, than in case of the displacement-type PT. Therefore such structural transition, at which the symmetry changes as a result of redistribution of particles between earlier equivalent positions, is called the transition of order-disorder type.

Nevertheless, a clear border between displacement and order-disorder types of PT cannot be determined. In terms of symmetry, there are no differences between them: considering the structure, always the average position of atoms is taken into consideration. Therefore, it does not matter how this averaging is performed: either by the discrete way or by the continuous way. Regarding some other properties, especially, dynamics of PT, the displacement type of transition and order-disorder type of transitions vary considerably.

10.3 PHASE TRANSITIONS WITH ANOMALIES IN DIELECTRIC PROPERTIES

In case of ferroelectric crystal, higher temperature nonpolar phase crystal becomes spontaneously polarized lower temperature phase (without application of electrical field). Thus ferroelectric transitions can be described by PT models of both the second and the first order: in various ferroelectric crystals PTs might be quite different. Currently, the number of known ferroelectrics and related crystals (antiferroelectrics, ferroelastics, virtual ferroelectrics, improper ferroelectrics) exceeds 600, and taking into account their solid solutions—5000. Therefore, there is a sense to limit this consideration only by most common phenomena, caused by PT in the ferroelectrics.

According to phenomenological classification, PTs in ferroelectrics are divided into two major classes: displacement type and order-disorder type. From the

microscopic point of view, when *displacement-type* PT occurs, above the transition temperature (in paraelectric phase) crystal shows structural *instability* relatively to one of lattice vibrations modes, called the *soft mode*. As temperature of paraelectric phase decreases and approaches to T_C , the frequency of this mode reduces, and in the extreme case goes to nearly zero. As a result, in the point of PT the shift of crystal's sublattices occurs, restoring crystal's dynamic stability. Owing to this shift, the ferroelectric phase with the spontaneous polarization P_S appears (in antiferroelectrics mutual shift of sublattices cannot give total polarization) [5].

The feature of *order-disorder-type* ferroelectrics is peculiar structural elements of crystal (molecule, radical, group of ions), which can be in two or more equilibrium positions that can be described by the orientation of dipole moments. In the nonpolar (high-temperature paraelectric phase), the energy of thermal motion disordering exceeds the energy dipole-dipole interaction, so that the orientation of dipoles is chaotic and total polarization is absent ($P_S = 0$). As temperature decreases, due to dipole-dipole interactions in the vicinity of PT, the self-ordering of polar structural elements occurs and, as a result, the spontaneous polarization appears ($P_S > 0$). In the antiferroelectrics ordered dipoles are directed *antiparallel*, and, despite their streamlining, total polarization does not occur ($P_S = 0$). However, not very far from PT, the external electrical field may forcibly change antiparallel orientation into the parallel with the advent of $P_S > 0$.

Using Landau theory from the general considerations the Curie-Weiss law can be obtained that characterizes permittivity temperature changing, as well as theoretically describing temperature dependence of P_S , to clarify dielectric hysteresis loop appearance and to explain nonlinear properties of ferroelectrics.

In the thermodynamic potential expansions in series of Eqs. (10.5), (10.6) for ferroelectric PT, it is common to assume the polarization P as the ordering parameter. In fact, *above* Curie point T_C ordering parameter $\eta = 0$ and polarization $P = 0$, while *below* T_C , where $\eta > 0$, spontaneous polarization appears: $P > 0$ (Fig. 10.4).

Ferroelectrics with PT-II. Considering that electric field can be defined as derivative $\partial\Phi/\partial P$, expression (10.5) can be rewritten as

$$E = \alpha P + \beta P^3. \quad (10.7)$$

The inverse dielectric susceptibility $\partial P/\partial E$ can be found as

$$\chi^{-1} = \frac{\partial E}{\partial P} = \frac{\partial^2 \Phi}{\partial P^2} = \alpha + 3\beta P^2. \quad (10.8)$$

It is possible to consider that $\chi \approx \varepsilon$, since $\varepsilon = 1 + \chi$ and permittivity of ferroelectrics is very large: $\varepsilon \gg 1$.

First, consider the *nonpolar (paraelectric) phase*, that is, the temperature range, in which the first term in Landau polynomial $\alpha > 0$, and the considered temperature interval is located *above* critical temperature, $T > T_C$ (Fig. 10.5). The conditions of phase sustainability are

$$\frac{\partial \Phi}{\partial P} = 0, \quad \frac{\partial^2 \Phi}{\partial P^2} > 0.$$

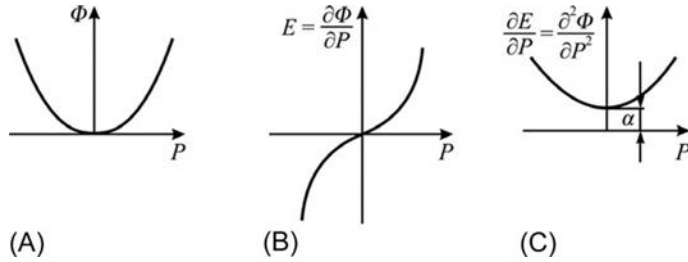


FIG. 10.5

Dependences of thermodynamic potential (A) and its derivatives (B, C) on ordering parameter P above T_C .

The first of these expressions can be reduced to cubic equations $\alpha P + \beta P^3 = 0$. This equation can have *only one* valid root $P_1 = 0$: polarization is the ordering parameter and above T_C the high temperature phase is disordered. Naturally, above the Curie point, spontaneous polarization is absent. Roots $P_{2,3} = \pm \sqrt{-\alpha/\beta}$ are imaginary (since for PT-II parameter $\beta > 0$, and above T_C also $\alpha > 0$).

Next consider temperature dependence of the *permittivity* above the Curie point (Fig. 10.6B). Peculiar to the nonpolar phase, the *Curie-Weiss law* follows from formulas (10.5) and (10.8):

$$\frac{1}{\varepsilon} = \alpha_0(T - \theta), \quad \varepsilon = \frac{C}{(T - \theta)}.$$

PT occurs at Curie-Weiss temperature, when parameter α changes its sign. Therefore, in the nonpolar phase, that is, above the Curie point permittivity depends not only on temperature but also on strength of electrical field. From expressions (10.7) and (10.8), it follows that above T_C , that is, in the paraelectric phase significant *dielectric nonlinearity* should be observed (Fig. 10.6C). Dependence $P(E)$ is characterized by saturation area. This means that permittivity in paraelectrics *decreases* in a strong electrical field, because $\varepsilon \sim \partial P / \partial E$. The general formula that takes into account both ε -nonlinearity and temperature ε -dependence in the paraelectric phase is

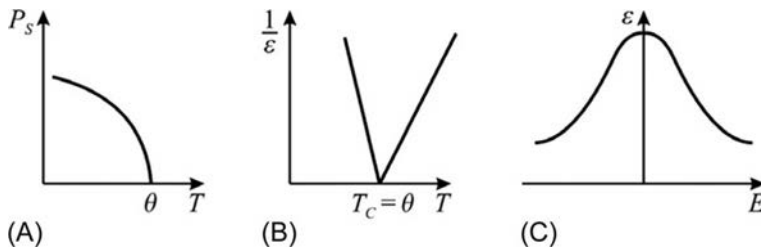


FIG. 10.6

Temperature dependence of spontaneous polarization (A), inverse permittivity (B), and dielectric nonlinearity in nonpolar phase (C).

$$\varepsilon(T, E) = \frac{C}{(T - \theta)} \left[1 + 3\beta\varepsilon_0^3 E^2 \frac{C^3}{(T - \theta)^3} \right]^{-\frac{1}{3}}. \quad (10.9)$$

It can be seen that higher the nonlinearity in the nonpolar phase, the closer the temperature to the PT point [6].

In the **polar phase, below the Curie point** spontaneous polarization appears; therefore, all roots of cubic equations $\partial\Phi/\partial P = \alpha P + \beta P^3 = 0$ are valid. However, since $\alpha < 0$ (see formula 10.3), root $P_1 = 0$ now corresponds to the *maximum* of $\Phi(P)$. However, by definition, the polar phase is stable if the value $\Phi(P)$ is minimal, that is, at $P_{2,3} = \pm\sqrt{-\alpha/\beta}$. Substituting $\alpha = \alpha_0(T - \theta)$ in this expression, it is possible to find temperature dependence of spontaneous polarization (Fig. 10.6):

$$P_c^2 = \frac{\alpha_0(\theta - T)}{\beta}. \quad (10.10)$$

Then, temperature dependence of permittivity can be determined from Eqs. (10.10), (10.8):

$$\frac{1}{\varepsilon} = 2 \frac{(\theta - T)}{C}; \quad \varepsilon = \frac{C}{2(\theta - T)}. \quad (10.11)$$

Thus, thermodynamic theory predicts that below the Curie point, at the same distance from T_C , the value of permittivity is twice smaller than in the paraelectric phase at $T > T_C$ (Fig. 10.6B). This prediction of theory is well confirmed by the experiments.

Further, the *nonlinear properties* of ferroelectrics in the polar phase will be considered. The relevant characteristics of this phase—thermodynamic potential and its derivatives—are shown in Fig. 10.7. In $\Phi(P)$ dependence points of extremes are marked: at these points the function $E(P)$ crosses the axis P . Dependence of $E(P)$ is characterized by the unstable area, the extent of which is shown by dashed lines. In this range, permittivity ($\varepsilon \sim \partial P/\partial E$) would be negative (but that is impossible).

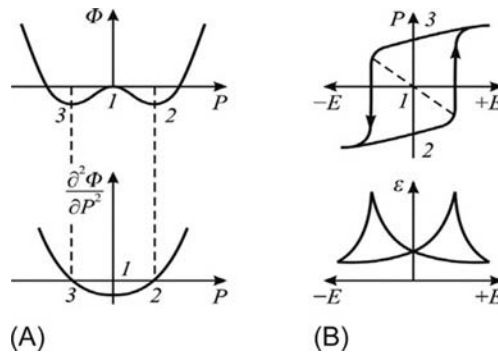


FIG. 10.7

Thermodynamic potential (A) and its derivatives (B) for ferroelectrics with PT-II: points 1, 2, 3—roots of equation.

The result is $P(E)$ hysteresis, which is the main feature of a ferroelectric. The *differential permittivity* also shows “butterfly wings”-type hysteresis (Fig. 10.7B) that corresponds to instantaneous value of ε when it passes through two maxims when the electrical field changes.

Ferroelectrics with PT-I close to PT-II also can be analyzed by Landau thermodynamic theory that allows explaining their properties in detail. Correspondent expression for thermodynamic potential has a form

$$\Phi(P) = \Phi_0(T) + \frac{\alpha}{2}P^2 + \frac{\beta}{4}P^4 + \frac{\gamma}{6}P^6, \quad (10.12)$$

where $\alpha = \alpha_0(T - \theta)$; $\beta < 0$ i $\gamma > 0$.

When this relationship is studied in detail, some special points for $\Phi(P)$ function and for its derivatives can be found. If instead of the image of the function $E(P)$ the more convenient coordinates $P(E)$ would be used, polar phase existence will be explained by the region of instability that corresponds to dielectric hysteresis loop similar to the illustration in Fig. 10.7B. Accordingly, permittivity depends on the field strength. Thus the main characteristics of ferroelectrics in their *polar phase* (hysteresis loop and $\varepsilon(E)$ dependence) do not depend on the type of PT that takes place in the Curie point.

It should be noted that first-order PT (with condition $\gamma > 0$) occurs not at temperature $T = \theta$ (when $\alpha = 0$) but at value $\alpha_K = \frac{3\beta^2}{16\gamma}$. That is why, in case of PT-I, transition temperature T_C is *greater* than Curie-Weiss temperature θ . Therefore, spontaneous polarization arises at $T = T_C$ by a jump (unlike PT-II), and the size of this jump equals $\Delta P_S = \frac{3\beta}{4\gamma}$.

Permittivity also shows a jump at temperature T_C , and its graded change is $\Delta\varepsilon = \frac{4\gamma}{3\beta^2}$ (correspondingly, in the Curie point permittivity is not expected to be infinite, as in case of PT-II) but its maxim equals $\frac{16\gamma}{3\beta^2}$ (Fig. 10.8).

The most studied ferroelectric that PT nature is close to PT-I is the barium titanate. This ferroelectric demonstrates all features of ferroelectric transition of first order. Temperature maximum of permittivity in pure BaTiO₃ occurs at $T_C = 400$ K, while Curie-Weiss temperature ($\theta = 388$ K) is below at 12 K. Curie-Weiss constant in

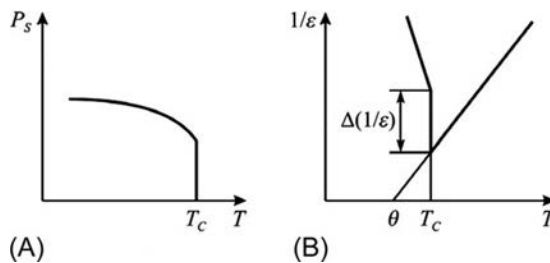


FIG. 10.8

Temperature dependence of spontaneous polarization (A) and inverse permittivity (B) in ferroelectrics with PT-I.

barium titanate equals $C = 1.2 \cdot 10^5$ K, while temperature maximum of permittivity is $\epsilon_{\max} = 10^5$. This maximum of permittivity and transition to nonpolar phase can be achieved not only by barium titanate *heating* but also by the *increase in pressure*. This is also consistent with thermodynamic theory, but in previous analyses, for simplification, pressure is considered as permanent [6].

Moreover, the thermodynamic potential, formula (10.12), which describes main characteristics of barium titanate, can characterize also two low-temperature subsequent PTs from one ferroelectric phase to another. Fig. 10.9 shows the change of BaTiO₃ unit cell in temperature range below 400K. First, there is a transition from high-temperature nonpolar cubic phase (with lattice parameters $a = a = a$) to the tetragonal polar phase (with lattice parameters $a = a \neq c$). Second, at temperature of about 300K, transition from the tetragonal ferroelectric phase to another ferroelectric phase (orthorhombic) occurs; finally, by cooling to temperature ~ 200 K, the most low-temperature (rhombohedral) ferroelectric phase appears [2].

Experimental data show that changes in BaTiO₃ lattice parameters in the vicinity of these transitions everywhere demonstrate *temperature hysteresis*, which clearly indicates that all these transitions are close in their nature to PT-I. Spontaneous polarization in BaTiO₃ occurs abruptly, as foreseen by PT-I theory; the permittivity at the Curie point also decreases by a jump.

Thus most important changes in dielectric properties of ferroelectrics that occur as a result of first- or second-order PT can be successfully explained by thermodynamic theory.

Selecting order parameter from phenomenological Landau theory is based on the choice of the most important property of a crystal. In case of ferroelectrics, the *polarization* is selected as the ordering parameter that can explain not only temperature dependence of P_S , but also large maxim of permittivity in the vicinity of PT. However, in some other cases, the spontaneous polarization arises without

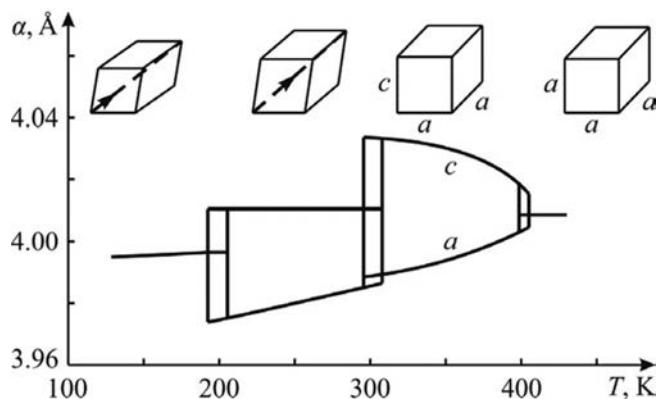


FIG. 10.9

Temperature change of crystal lattice parameters and symmetry of BaTiO₃ in series of its phase transition.

noticeable maximum in $\epsilon(T)$ dependence, for example, in the gadolinium molybdate ($\text{Gd}_2(\text{MoO}_4)_3$).

Ferroelectrics in which temperature dependence of permittivity does not have clearly expressed maximum $\epsilon(T)$ are the *improper ferroelectrics*. The polarization is not their ordering parameter. In case of $\text{Gd}_2(\text{MoO}_4)_3$, for example, PT is caused by *mechanical strain*, while spontaneous polarization originates during PT as one of the properties of the ordered low-temperature phase [2].

However, $\epsilon(T)$ maximum appearance is not obviously caused by the fluctuations of polar phase. Some examples of $\epsilon(T)$ dependences with maximum are given in Chapter 9 for crystals called *antiferroelectrics*. In these crystals, the $\epsilon(T)$ maximum at transition point may be also large (as, e.g., in lead zirconate PbZrO_3), or may be absent (as in lead magnesium-tungstate $\text{PbMg}_{1/2}\text{W}_{1/2}\text{O}_3$). PT into the *antipolar* phase from a nonpolar phase might be close as to PT-I so to PT-II. It is necessary to mention that among antiferroelectrics there are perovskite-type oxides; therefore there are some crystals comprising hydrogen (ammonium dihydrogen phosphate ($\text{NH}_4\text{H}_2\text{PO}_4$)).

Permittivity frequency dispersion near the phase transition. A wide-range frequency study of ferroelectrics is required not only because of their applications, but also because of important physical properties of these materials. Lattice dynamics theory of crystals predicts strong anomalies in dielectric properties of ferroelectric at microwaves. The question is that in the vicinity of ferroelectric PT one of lattice vibration modes decreases down to microwaves. Furthermore, most of ferroelectrics have a multidomain structure, and domain wall resonant (or relaxation) frequency is also located in the microwave range. For this reason, frequency study can support investigation of many fundamental characteristics of ferroelectrics [7].

In the polar (ferroelectric) phase, the dispersion of permittivity is observed thanks to reversible movement of domain walls. In the nonpolar (paraelectric) phase, the permittivity dispersion (usually, relaxation) is due to the contribution of lattice vibration mechanism to a permittivity that varies with temperature, following Curie-Weiss law: $\epsilon(T) = C/(T - \theta)$.

In ordinary (nonferroelectric) dielectrics possessing thermally induced type of polarization their relaxation time shows exponential dependence on temperature: $\tau \sim \exp(U/k_B T)$, where U is activation energy and k_B is Boltzmann constant. Dielectric permittivity dispersion that explains thermal orientation of dipoles in ordinary dielectrics leads to shift of $\epsilon_{\text{max}}(T)$ toward higher frequencies at temperature increase (see Fig. 7.30 in Section 7.8). This model predicts linear dependence of $\log(\omega_{\text{max}}/\Omega)$ on $1/T$, and, using the slope of this line, the activation energy U can be found (relaxation frequency Ω is determined by the relaxation time: $\Omega = 1/\tau$).

However, in the *order-disorder-type ferroelectrics* relaxation type of ϵ -dispersion does not follow traditional temperature dependence: it is disrupted in the vicinity of PT. Relaxation time is characterized by the *critical* temperature dependence: $\tau = \tau_0/(T - \theta)$ that corresponds to Curie-Weiss law: $\epsilon(T) = C/(T - \theta)$. This law is conditioned by significant change with temperature such parameters that in ordinary relaxation-type dielectrics remain practically constant.

Thus in the order-disorder-type ferroelectrics, PT assumes the presence of groups of atoms with individual dipole moments. In the paraelectric (high-temperature) phase they look like *disordered* dipoles, that is, crystal structure is characterized by several equally probable orientations of the dipole moments of polar groups, which, under the influence of thermal vibrations of crystal lattice, are randomly distributed along these directions. As temperature decreases, the *interaction* of dipoles leads to spontaneous ordering of these polar groups so that PT occurs in the ferroelectric (spontaneously polarized) phase. Therefore, in lower-temperature (polar) phase, many polar groups are still not oriented [7].

Among many *order-disorder-type ferroelectrics* there are Rochelle salt (whose dielectric dispersion was shown previously in Section 7.8, Fig. 7.32) and the triglycine sulfate (TGS). This type of ferroelectrics has two main frequency intervals of dielectric permittivity dispersion: domain wall relaxation in polar phase and dipole relaxations in all phases. The most interesting for physics of PT in the ferroelectrics is permittivity frequency/temperature dependences, as shown in Fig. 10.10 for TGS.

In contrast to Rochelle salt, the TGS crystal is not piezoelectric in its paraelectric phase. Therefore electromechanical deposit to ε_2 in paraelectric phase is absent; however, below the Curie point in ferroelectric phase, electromechanical contribution to ε_2 exists (shown in Fig. 10.10 by stroke region).

It is noteworthy that in the Curie point dependence $\varepsilon'(T)$ at microwaves demonstrates not the maximum *but the minimum*. It can be explained, considering that family of $\varepsilon^*(\omega, T)$ characteristics can be described by the modified Debye equation

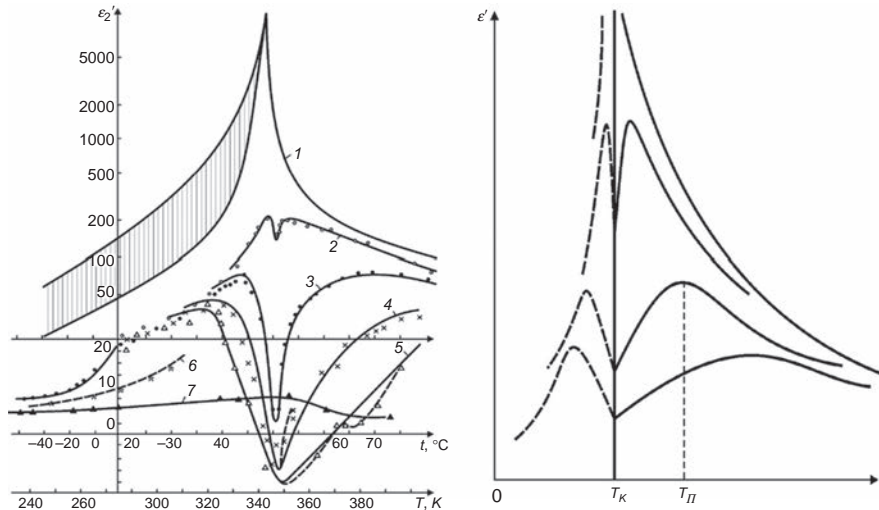


FIG. 10.10

TGS crystal microwave study: (A) ε_2' temperature dependence at frequencies: 1—1 kHz, 2—GHz, 3—16 GHz, 4—26 GHz, 5—37 GHz, 6—80 GHz, 7—250 GHz; (B) theoretical explanation of $\varepsilon'(T)$ dependence.

$$\varepsilon^*(\omega, T) = \varepsilon_{\text{IR}} + \frac{C}{T - \theta + i\omega\tau_0}$$

In the paraelectric phase, TGS dynamic properties can be described by parameters: $C = 3200\text{K}$, $\theta = 321\text{K}$, and $\tau_0 = 2 \cdot 10^{-10}\text{ s/K}$ [7].

Crystals of TGS type are one-dimensional ferroelectrics. In full microwave range and in a wide temperature interval near PT they show large absorption ($>30\text{ db/mm}$), but only in the direction of polar axis. In two other crystallographic axes, these crystals are similar to customary ionic crystals, and they are well transparent at microwaves. For this reason, thin oriented plates of these crystals might be used as *polarizers* in full millimetric and even submillimetric range of waves.

Generalized dielectric spectrum of order-disorder ferroelectrics [7]. Typical frequency-temperature dependence of permittivity in ferroelectrics with PT of order-disorder type is shown in Fig. 10.11. As frequency increases from infra-low to optical range, the ε -dispersion occurs in several well-separated frequency intervals. Consequently, permittivity of ferroelectric crystals can be considered as the sum of several *dielectric contributions* $\Delta\varepsilon_j$ conditioned by different mechanisms of polarization. Above and below PT ε temperature, these polarization mechanisms might be different.

In the *polar phase* (below the Curie point), it is possible to distinguish six polarization mechanisms. Primarily, this is optical dielectric contribution $\Delta\varepsilon_1$ from electronic elastic polarization; in the ferroelectrics this contribution is small, and it is characterized by negative $TC\varepsilon$. In ferroelectrics of order-disorder-type contribution $\Delta\varepsilon_2$ from crystal lattice (far infrared) polarization is also small, but its temperature

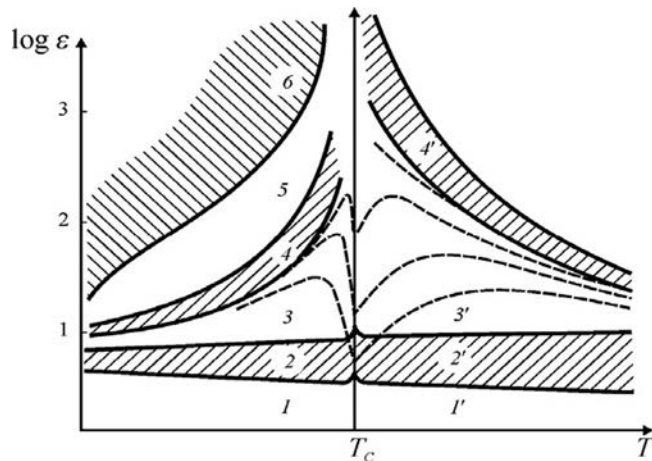


FIG. 10.11

Generalized temperature dependences of different mechanism dielectric contributions for order-disorder-type ferroelectrics: above T_C —data are close to parameters of KDP crystal; below T_C —data are close to the parameters of TGS crystal.

dependence is characterized by positive $TC\epsilon$ ($\Delta\epsilon_2$ differs only a little from usual ionic crystals). In the vicinity of the Curie point, as a rule, $\Delta\epsilon_2$ shows only a small anomaly (in contrast to displacement-type ferroelectric).

Other contributions to dielectric constant (plots 3...6 below T_C), shown in Fig. 10.11, are characterized by relatively *low-frequency* polarization mechanisms, caused by peculiarities of ferroelectric crystal. Common to them is the decrease in these contributions as distance from PT point increases. Contribution $\Delta\epsilon_3$ is given by the relaxation of *disordered part of polar groups*. These atomic groups appear and disappear (“flicker”) in a crystal with frequency depending on distance from PT. Therefore, the variance of $\Delta\epsilon_3$ at temperatures below T_C is similar to the $\Delta\epsilon_3$ dispersion in the paraelectric phase.

Electromechanical contribution to permittivity $\Delta\epsilon_4$ in some crystals, for example, in Rochelle salt, is large; in other ferroelectrics (such as TGS), it is relatively small. Dispersion of $\Delta\epsilon_4$ occurs in a form of one or more dielectric resonances at radio frequencies, when “inertial clamping” of ferroelectric crystal happens, usually in the frequency range of 10^4 – 10^6 Hz.

Dispersion of $\Delta\epsilon_5$ in ferroelectrics is studied in detail. This dielectric contribution is associated with high *polarizability of domain walls*, the cause of which some authors consider as relaxation while other authors suppose resonant mechanism of domain wall polarization. The plot of $\Delta\epsilon_6$ in Fig. 10.11 corresponds to very low-frequency contribution to permittivity. This dispersion clearly depends on the domain structure of crystal, but, unlike $\Delta\epsilon_5$, it occurs in the range of subsonic frequencies. The $\Delta\epsilon_6$ value strongly depends on aging processes and on the presence of defects in a crystal.

In the nonpolar phase (above the Curie point), dielectric dispersion of order-disorder-type ferroelectrics is characterized by the dynamic properties of dipoles and *polar groups ordering* in a crystal in the vicinity of PT. Dotted lines in area 3' (Fig. 10.11) show the $\Delta\epsilon_3(T)$ change at different frequencies. Near PT these dependencies are characterized by pronounced minimum.

Displacement-type ferroelectric frequency characteristics. From a microscopic point of view, in case of displacement-type ferroelectric, above the transition point T_C , there is crystal lattice instability due to one of optical lattice vibrations that is called as the “*soft mode*.” As temperature decreases approaching T_C , the frequency of this mode diminishes *critically*, and in the limit tends to zero. As a result, displacement-type phase ferroelectric transition occurs in a form of *spontaneous shift of the sublattices* of a crystal, restoring its dynamic stability. Then, spontaneous shift of sublattices causes spontaneous polarization P_S .

The example of dielectric dispersion in the displacement-type ferroelectrics is shown in Fig. 10.12. The domain-type dielectric contribution shows dispersion mainly at microwaves; this dispersion has relaxation character. In the paraelectric phase (above PT temperature) in most of displacement-type ferroelectrics, the microwave ϵ -dispersion is absent.

It is important to determine how Curie-Weiss temperature affects the presence or absence of microwave ϵ -dispersion. Investigations show that in lead titanate (PbTiO_3), in which Curie-Weiss temperature ($\theta = 740$ K) is rather high, microwave

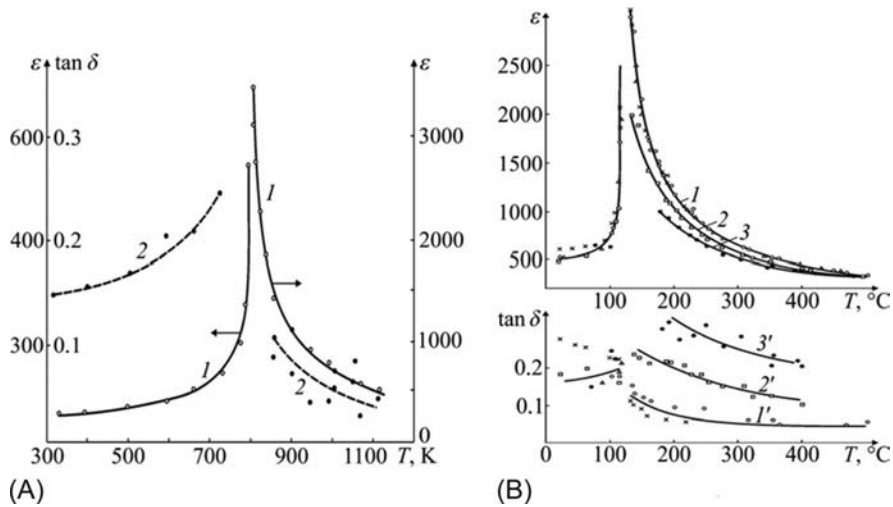


FIG. 10.12

Temperature dependence of ϵ and dielectric losses $\tan \delta$: (A) PbTiO_3 at 37 GHz, (B) BaTiO_3 at frequency range of 10.4... 37 GHz (1), at 46 GHz (2), at 75 GHz (3).

ϵ -dispersion is absent. In the PbTiO_3 even at millimeter waves, the $\epsilon(T)$ characteristic above the Curie point coincides with this dependence at a frequency of 1 MHz (Fig. 10.12A, curve 1).

However, in the ferroelectric phase of PbTiO_3 microwave losses (curve 2) are large, which indicates strong ϵ -dispersion *below* the Curie point, in the ferroelectric phase. Thus PbTiO_3 investigation suggests the assumption that microwave dispersion in the paraelectric phase is absent; therefore ϵ -dispersion does not depend on PT temperature. In practice, these studies recommend using exactly paraelectric phase of displacement-type ferroelectrics for nonlinear microwave components with small microwave losses, creating *solid solutions*, such as $(\text{Sr,Pb})\text{TiO}_3$ [6].

In contrast, in other ferroelectrics, which are considered only as “close to PT of displacement type” (for instance, in BaTiO_3) the microwave dispersion in the paraelectric phase appears quite strongly in the range of millimeter waves that brings to increased microwave losses (Fig. 10.12B). The reason for this microwave dispersion in paraelectric phase is that PT in BaTiO_3 has some indications of order-disorder-type transition. Investigations have shown that this dispersion, which starts at millimeter waves and continues in the submillimeter waves, is the beginning of *fundamental dispersion* caused by the “soft TO -lattice vibrational mode” in BaTiO_3 , which is characterized by very high anharmonicity.

The dispersion, which is clearly observed at frequencies of 60–80 GHz, does not depend on the imperfections in the crystal structure and takes place in crystals, so also in polycrystalline BaTiO_3 . In practice, this means that, if barium titanate would be included to the structure of nonlinear microwave element, one cannot expect from

this paraelectric small dielectric losses, because microwave losses are the consequence of fundamental mechanism of PT in BaTiO₃.

Therefore, microwave dispersion in the paraelectric phase depends on the soft lattice vibration mode. For this reason, the Lorentz oscillator model can be a basic model to describe ϵ^* frequency/temperature dependence:

$$\epsilon^*(\omega) = \epsilon(\infty) + \frac{\epsilon(0) - \epsilon(\infty)}{1 + (\omega/\omega_{TO})^2 + i\Gamma\omega/\omega_{TO}}$$

In this equation, it needs to assume $\epsilon(0) - \epsilon(\infty) = C/(T - \theta)$ and soft mode critical dependence on temperature $\omega_{TO} = A(T - \theta)^{1/2}$. As a result, the relative damping factor is $\Gamma = \gamma/\omega_{TO}$:

$$\epsilon'(\omega, T) - \epsilon(\infty) = CA^2 \frac{A^2(T - \theta) - \omega^2}{[A^2(T - \theta) - \omega^2]^2 + \gamma^2\omega^2}$$

$$\tan \delta \approx \frac{\gamma\omega}{A^2(T - \theta)}$$

From ϵ and $\tan \delta$ temperature dependences at various frequencies, the soft mode-temperature dependence can be calculated. Soft mode-temperature dependence is shown in Table 10.1 and Fig. 10.13.

Table 10.1 shows the main lattice dynamics parameters of paraelectrics and ferroelectrics.

Generalization of displacement-type ferroelectrics spectrum [7]. Frequency-temperature dependence of permittivity in typical displacement-type ferroelectrics is shown in Fig. 10.14. As frequency increases (from radio frequencies to optical frequencies), several mechanisms of permittivity dispersion are clearly seen.

PT has almost no effect on the optical properties (plot 1' and 1 in Fig. 10.14). In the paraelectric phase (above PT temperature T_C), the far-infrared (lattice) polarization dominates (plot 2'). Dielectric contribution of the "soft" lattice vibration mode results in very high permittivity (thousands near T_C). In the ferroelectric phase soft phonon gives rise to relatively small dielectric contribution (plot 2), but there are many other mechanisms to increase permittivity.

Table 10.1 "Soft" Lattice Mode Parameters of Some Ferroelectric Materials [7]

Material	$P_C, \mu\text{Q}/\text{cm}^2$	T_C, K	θ, K	$C \times 10^{-4}, \text{K}$	$A/2\pi, \text{GHz K}^{-1/2}$
CaTiO ₃	–	–	–90	4.5	170
SrTiO ₃	–	–	35	8.4	180
BaTiO ₃	30	400	388	12	75
PbTiO ₃	80	780	730	15	90
KNbO ₃	30	685	625	18	95
LiNbO ₃	70	1500	–	–	–

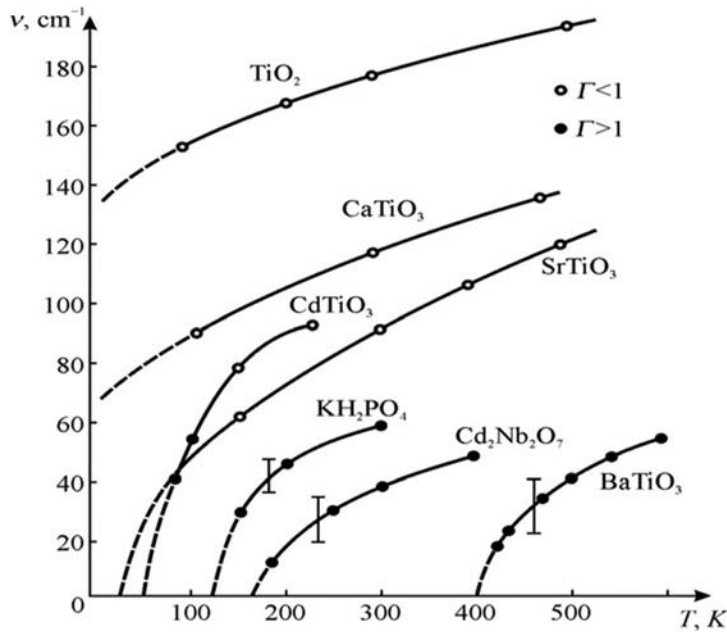


FIG. 10.13

“Soft” lattice mode frequency dependence for various paraelectrics obtained by microwave and far-infrared experiments ($1 \text{ cm}^{-1} = 30 \text{ GHz}$) [7].

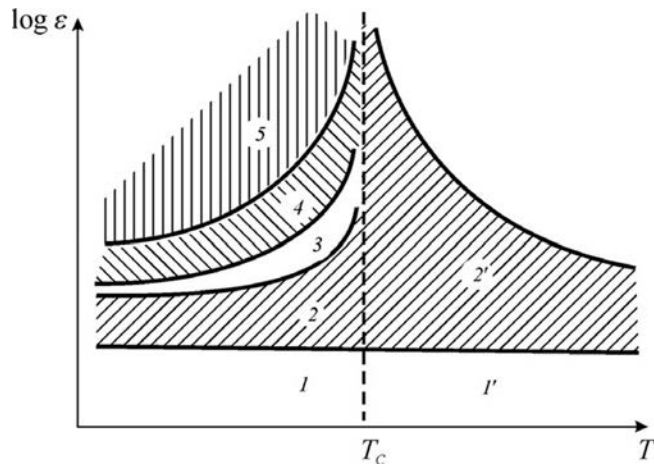


FIG. 10.14

Generalized dielectric spectrum for displacement-type ferroelectrics, that is temperature dependence of main polarization mechanism’s contribution to permittivity: I' , I —electronic polarization contribution at different temperatures; $2'$, 2 —dielectric contribution of “soft phonon” in polar and paraelectric phases; 3 —domain walls dynamical polarization in polar phase; 4 —electromechanical (piezoelectric) polarization mechanism; and 5 —domain switching polarization in strong electrical fields.

The plot 3 characterizes domain wall vibration polarization that leads to microwave dispersion. If the domain structure of crystal (or polycrystal) is not compensated (no neutral), radiofrequency dispersion is observed (plot 4 in Fig. 10.14) in the form of great number of electromechanical resonances. This ϵ -dispersion occurs at frequencies of 10^4 – 10^6 Hz. In the polar phase, low-frequency dielectric contribution is produced by mechanism of domain switching (plot 5 in Fig. 10.14). This dielectric contribution is possible, however, only in the strong electrical fields.

As shown in Fig. 10.14, while cooling from the Curie point, the ferroelectrics, where PT is close to the PT-1, show sharp decrease in their permittivity. This is conditioned by a stepped increase in spontaneous polarization P_C that is accompanied by the coercive field growth. In turn, this results in a stepped increase in “soft” phonon frequency and, consequently, to permittivity graded decrease.

However, *inside* any domain wall, spontaneous polarization and coercive field are very small (close to zero). Therefore, polarizability of domain wall is very high. This phenomenon can be interpreted as a “frozen” paraelectric phase inside domain walls, which, in that way, partially exists below the Curie point. Meaningful frequency of dielectric microwave dispersion (near 10^{10} Hz) enables (using Lyddane-Sachs-Teller relation) finding “effective permittivity” inside domain walls as $\epsilon_{\text{wall}} \sim 3 \cdot 10^5$.

Morphotropic phase transitions. In connection with a search for the most efficient piezoelectric-ferroelectric materials, a new type of PT—morphotropic transitions—was found (and they are widely used). It turned out that highest efficiency in electromechanical properties is seen in the border between antiferroelectric and ferroelectric phases. From the very beginning of this study, the composition lead zirconate-titanate, $\text{Pb}(\text{Zr},\text{Ti})\text{O}_3$ (denoted as PZT), showed piezoelectric module $d_{33} \sim 300$ pC/N and electromechanical coupling factor ($K_{\text{EM}} \sim 0.6$) that was significantly greater than previously known piezoelectric ceramics.

At present, piezoelectric ceramic made on the base of $\text{Pb}(\text{Zr},\text{Ti})\text{O}_3$ with additives are characterized by module $d_{33} = 600$ pC/N and coupling coefficient $K_{\text{EM}} \sim 0.75$. These parameters are achieved using special technology: all compositions are produced near the *morphotropic border*. Correspondingly, the morphotropic PT between rhombohedral and tetragonal ferroelectric phases of $\text{PbZr}_{1-x}\text{Ti}_x\text{O}_3$ is a qualitative leap in the development of piezoceramic materials. Morphotropic PT means that structural transition in solid solutions might be realized just by the changing of *composition*. In the interval of concentrations near such PT *different structures can coexist*. The width of this border depends on technological factors, for example, in the PZT ceramics it can vary from 0.5 to 15 mol%.

It is found that at morphotropic PT the permittivity and also the piezoelectric parameters show pronounced maxims in dependence on their concentration. The maxim of electromechanical activity and permittivity in the vicinity of phase boundaries is associated with the higher degree of domain reorientation in these ceramics. It is also possible that the reason for the piezoelectric activity increase is conditioned by the decrease in unit cell spontaneous deformation near the morphotropic border. It is established that rhombohedral and tetragonal phases

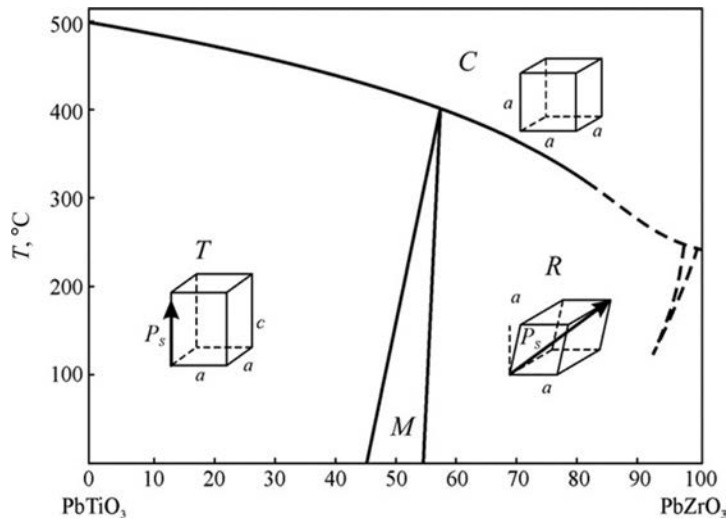


FIG. 10.15

Phase diagram of solid solutions $\text{PbZr}_{1-x}\text{Ti}_x\text{O}_3$: C —cubic, T —tetragonal, R —rhombohedral, and M —monoclinic phase.

on the (x, T) -diagram of $\text{PbZr}_{1-x}\text{Ti}_x\text{O}_3$ is separated by a narrow *intermediate* monoclinic phase (Fig. 10.15).

In this regard, large electromechanical response near the morphotropic border is associated with the fact that rotation of polarization vector in the electrical field during ceramic polarization is facilitated by the presence of one or more intermediate monoclinic phases between rhombohedral and tetragonal phases.

Modern trends in piezoelectric material science relate to solid solution search among such systems, where the morphotropic border exists. In particular, high electromechanical coupling factor and large piezoelectric modules are found in the binary solid solutions $\text{PbB}_{0.5}^{3+}\text{Nb}_{0.5}\text{O}_3$ - PbTiO_3 (where $\text{B}^{3+} = \text{Sc}, \text{Yb}, \text{Lu}$). In the early 2000s based on solid solutions $\text{BiB}^{3+}\text{O}_3$ - PbTiO_3 (where $\text{B}^{3+} = \text{Fe}, \text{Sc}, \text{In}$), high-temperature piezoelectric ceramic materials were developed with increased piezoelectric properties and ferroelectric Curie point of 450–500°C.

Diffuse phase transitions and relaxor ferroelectrics [7]. Crystals and ceramics with diffuse PT have the *nonuniform distribution* of structural ions. In these compositions, owing to structural fluctuations, their PTS are fuzzy. However, these materials (e.g., solid solutions $\text{Ba}(\text{Ti}, \text{Sn})\text{O}_3$ or $\text{Ba}(\text{Ti}, \text{Zr})\text{O}_3$), except decreasing PT temperature and broadening the $\epsilon(T)$ dependence, have no significant features that are peculiar to one other kind of materials—relaxor ferroelectrics.

The *relaxor ferroelectrics* demonstrate an exceptional case of diffuse phase ferroelectric transition. They are ferroelectrics with disordered structure that is characterized by the blurred maximum $\epsilon(T)$ in PT vicinity and by *heightened capability* for electrical controlling by their properties. If ferroelectric materials are applied as

piezoelectric actuators or as electro-optical modulators or other controlling devices, the relaxor ferroelectrics favorably differ from conventional ferroelectrics by the *absence of hysteresis* with electrical control by deformation.

Blurred maximum $\epsilon(T)$ in the relaxor ferroelectrics is not caused only by random distribution of structural components, but also by fundamental properties of these structures. A distinctive feature of relaxors is remarkable “softening” in dielectric, optical, and elastic properties over a *wide temperature range* (other ferroelectrics have such features only in a very narrow temperature range in the vicinity of the Curie point). Relaxor ferroelectric shows large and expanded $\epsilon(T)$ maximum where *giant electrostriction* is observed. The point is that a degree of electrical control by deformation $x(E)$ is dependent on permittivity: $x \sim \epsilon^2$.

All listed properties determine technical applications of relaxor ferroelectrics: they might be used as subminiature capacitors (because of large permittivity), they may be applied as large-strain nonhysteresis actuators (owing to giant electrostriction), they may serve as nonlinear optical devices (thanks to highly disordered “nonergodic” structure), etc. Moreover, relaxor ferroelectrics have potential applications in pyroelectric sensors (because electrically induced pyroelectric coefficient is proportional $d\epsilon/dT$) and can be also used in electrically controlled piezoelectric filters (by electrically induced piezoelectric effect with modulus $d \sim \epsilon^2$).

Extended temperature maximum of permittivity in relaxor ferroelectrics has been investigated in many studies, but usually at comparatively low frequencies (up to 10^6 – 10^7 Hz). It is obvious that ϵ' -maximum shifts with the increase in frequency toward high temperature. Fig. 10.16 shows microwave investigations of $\epsilon'(T)$ for PMN crystal up to $\sim 10^{11}$ Hz [7].

Generally, to distinguish relaxors among other diffused PT ferroelectrics, many microscopic or optical investigations should be used. The experimental test method

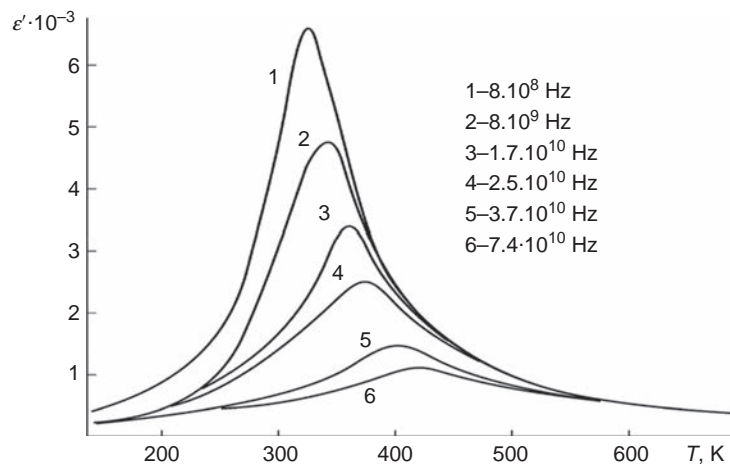


FIG. 10.16

Temperature dependence of the PMN permittivity at high frequencies.

will be described later—how to draw a line between relaxors and usual disordered ferroelectric solid solutions. The study is to measure a family of $\epsilon^*(\nu, T) = \epsilon'(\nu, T) - i\epsilon''(\nu, T)$ dependences in broad frequency-temperature interval.

Generalized dielectric spectra of relaxor ferroelectrics might be compared with spectra of classic displacement-type ferroelectrics (such as BaTiO_3 , Fig. 10.14) and various ferroelectric solid solutions (not relaxors) with the diffused PT, namely, solutions $\text{Ba}(\text{Ti}, \text{Sn})\text{O}_3$, $(\text{Ba}, \text{Sr}, \text{Ca})\text{TiO}_3$, and $\text{Pb}(\text{Zr}, \text{Ti})\text{O}_3$. All these materials are characterized by broad $\epsilon(T)$ maximum, associated with random distribution of *same-valence* cations in correspondent sublattices. The heterogeneity of composition in microregions is accompanied by the *fluctuations of Curie temperature* that leads to broad $\epsilon(T)$ maximum. The nature of dielectric spectra for mentioned compositions is similar to BaTiO_3 .

However, relaxor ferroelectric is characterized by the cations of *different valence* that *occupy randomly* similar structural sites. These materials have two different types of structural disordering. For example, lead magnesium niobate crystal ($\text{PMN} = \text{PbMg}_{1/3}\text{Nb}_{2/3}\text{O}_3$) has $\text{B}^{+2}\text{-B}^{+5}$ -type compositional disordering, while potassium-lithium tantalate crystal ($\text{KLT} = \text{K}_{1-x}\text{Li}_x\text{TaO}_3$) has strongly disordered structure only in lithium ions that are located in the noncentral positions with *various associations* between them.

Generalized temperature/frequency dielectric characteristic of relaxors shown in Fig. 10.17 is constructed by analogy with other type of ferroelectrics (Figs. 10.11

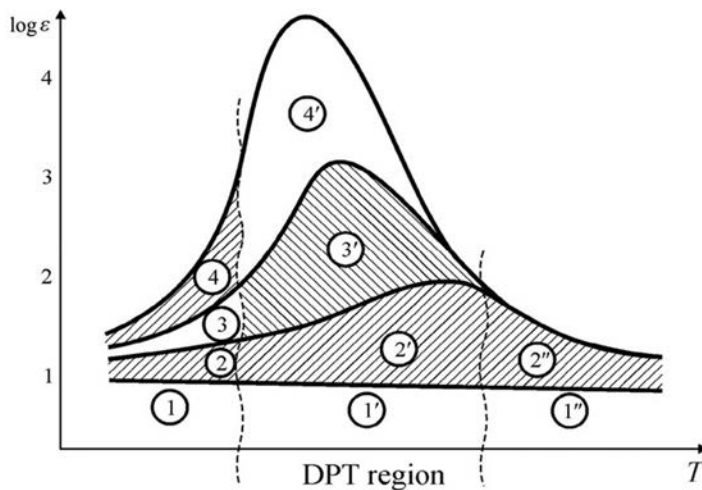


FIG. 10.17

Generalized dielectric spectrum peculiar for relaxor ferroelectrics; temperature dependence of “dielectric contributions” from basic polarization mechanisms: 1—electronic displacement (optical) polarization; 2—lattice (ionic, phonon) polarization mechanism; 3—domain wall polarization mechanism in polar phase; 3′—polarization of interphase boundary displacement; and 4, 4′—polar cluster reorientations (DTP is diffused phase transition region).

and 10.14). Very large ϵ_{\max} is typical for relaxor ferroelectrics in extended temperature interval, while in classical ferroelectrics large ϵ_{\max} is seen only at Curie temperature (note that the narrow peak of ϵ in proper ferroelectrics is a fundamental property of their PT).

With regard to relaxor ferroelectrics, their peculiarities might be explained using the dielectric spectroscopy method. Here it is considered that diffuse $\epsilon_{\max}(T)$ in the relaxor is so large because of heterogeneity in electrical field distribution: while investigating, applied field becomes stronger in the “low- ϵ micro-regions” (clusters). As a result, the polarizability of these clusters increases. Spatial wave of nonlinear polarization extends in the relaxor ferroelectrics with acoustic mode velocity; therefore, a strong ϵ -dispersion is seen in the frequency range of 10^3 – 10^4 Hz.

However, only this mechanism cannot give an exhaustive explanation peculiar to relaxor large dielectric constant ($\epsilon \sim 10^4$ – 10^5) that is observed at a small electrical field over a broad temperature interval. If relaxor ferroelectrics would be a simple composite of low- and high-frequency adjoining polar clusters, the ϵ -value will be formed predominantly by the clusters that have reduced ϵ ; therefore, relaxor permittivity would not be expected to be huge.

It should be noted that at very high frequencies (microwaves) and comparatively low frequencies the rate of $T \epsilon'_{\max}(\log \nu)$ drift is quite different. This circumstance enables to distinguish relaxor ferroelectrics from other types of ferroelectrics. Among other data, frequency dependence of ϵ' -temperature maximum ($T_{\epsilon'_{\max}}$) is analyzed (Fig. 10.18). The ordinary ferroelectric is defined by a sharp ϵ' (T) maximum at Curie temperature $T_C = T_{\max \epsilon'}$, and, as is well known, *no frequency shift* of this maximum is seen (line BT, Fig. 10.18). In contrast, in normal disordered ferroelectrics—solid solutions (but not in the relaxors), the selected parameter $T_{\epsilon'_{\max}}$ *gradually increases* as frequency grows up (line BSnT, Fig. 10.18). In other words,

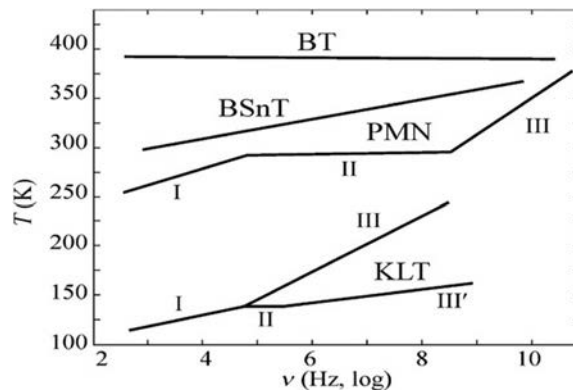


FIG. 10.18

Temperature of ϵ -maximum frequency dependence in relaxor ferroelectric PMN and KLT in comparison with normal displacement-type ferroelectric BT = BaTiO₃ and ordinary diffuse solid solution BSnT = Ba(Ti,Sn)O₃ [7].

the usual ferroelectric solid solution with diffused PT shows a *simple linear rise* in their $T_{\epsilon'_{\max}}(\log \nu)$ dependence.

Relaxor ferroelectrics demonstrate *the broken or the bifurcated* $T_{\epsilon'_{\max}}(\log \nu)$ dependence. That is, the PMN-type relaxor ferroelectric shows gradually increasing but *broken-line* for $T_{\epsilon'_{\max}}(\log \nu)$ dependence (line PMN in Fig. 10.18). However, the relaxor ferroelectric of KLT type demonstrates the *bifurcation* of this line (KLT).

The main reason for the $T_{\epsilon'_{\max}}(\log \nu)$ dependence becoming a broken line or bifurcated line is the *coexistence* in the relaxor ferroelectrics of at least two different types of polar clusters that occupy adjoining nanoscale regions in a crystal.

At low frequencies (below $\sim 10^5$ Hz), dielectric permittivity of PMN decreases with frequency approximately 10 times. Thereafter, as seen on plot I on PMN line in Fig. 10.18, the temperature of observable ϵ'_{\max} increases gradually with a frequency rise. With further frequency increase, temperature maximum of PMN permittivity prolongs its decrease but practically *no shifts* with frequency rise between 10^5 and 10^8 Hz (plot II on PMN line in Fig. 10.18). Finally, starting at about 5×10^8 Hz, the dependence of $T_{\epsilon'_{\max}}(\log \nu)$ again *gradually increases* to higher temperatures (section III on PMN line in Fig. 10.18).

Therefore, Fig. 10.18 demonstrates frequency dependences of $T_{\epsilon'_{\max}}$ for four typical situations:

- normal displacement-type ferroelectric (similar to BaTiO_3) shows no frequency change in its $T_{\epsilon'_{\max}}$ within limits of experimental study (up to 75 GHz);
- ordinary ferroelectric *solid solutions* with diffused PT, such as $\text{Ba}(\text{Ti},\text{Sn})\text{O}_3 = \text{BSnT}$, demonstrates linear shift of $T_{\epsilon'_{\max}}$ to higher temperatures as frequency increases (a simple explanation of this growing $T_{\epsilon'_{\max}}$ is the ϵ -dispersion due to domain wall polarization in polar phase, which is absent in the paraelectric phase);
- relaxor ferroelectric shows a broken-line dependence of $T_{\epsilon'_{\max}}(\log \nu)$. This is seen in frequency interval of (10^5 – 10^8) Hz for PMN, in which the position of $T_{\epsilon'_{\max}}$ practically does not change with frequency (or this change is too small);
- in some relaxor ferroelectrics the $T_{\epsilon'_{\max}}(\log \nu)$ characteristic shows bifurcation; for example, KLT has two different polarization mechanisms that correspond to various ways by which mobile Li^+ ions can move in the multiwell lattice potential.

The aforementioned is convincing evidence that in relaxor ferroelectrics *two or more types of polar clusters coexist*. Most probable mechanisms of two main dielectric contributions to diffused ϵ -maximum of relaxor are as follows:

- low-frequency (up to 10^5 Hz) dielectric contribution to broad ϵ -maximum that usually is interpreted as the “reorientation of dipole moments of polar nanoregions” that is accounted for *strong electromechanical coupling* of “soft-polarizable” polar regions, in which the electrostriction transfers to the piezoelectric effect;

- high-frequency (10^8 – 10^{10} Hz) contribution to the ϵ -maximum that commonly is explained as the motion of boundaries of polar nanoregions; this is similar to the microwave ϵ -dispersion in normal ferroelectrics, which is due to their multidomain structure, and here it is accounted for the “frozen super-paraelectricity” in the boundaries of polar nanoregions.

Therefore, it is shown that the peculiarity of dielectric spectra of relaxor ferroelectrics is broken or bifurcated $T_{\epsilon'_{\max}}(\log \nu)$ dependence. For this reason, the relaxor ferroelectrics can be recognizable from ordinary ferroelectric solid solution by their frequency investigations. The low-frequency mechanism of the ϵ -dispersion in relaxor ferroelectrics is accounted for strong electromechanical coupling of “soft-polarizable” polar regions, while microwave ϵ -dispersion is due to the “frozen super-paraelectricity” in the boundaries between polar regions (as in ferroelectric domain walls).

10.4 PHASE TRANSITIONS WITH CONDUCTIVITY ANOMALIES

Dielectric conductivity involves a complex function of many parameters: $\sigma(\omega, T, E, p)$. As a rule, conductivity varies with frequency, temperature, electrical field, or pressure *smoothly* and *reversible*. Only electrical breakdown results in a sharp and irreversible jump in the $\sigma(E)$ dependence.

However, there are other parameters like *sudden change* in conductivity, as shown in Fig. 10.19, which are important for electronics engineering and interesting

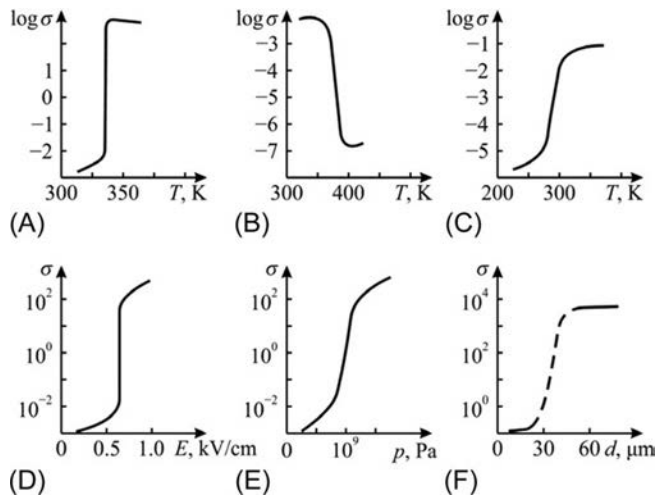


FIG. 10.19

Changing conductivity σ (S/m) depending on temperature (A, B, C), on electrical field (D), on pressure (E), and on film thickness (F): A, D, E—vanadium dioxide; B—doped barium titanate, C—silver iodine; and F—film of semimetal bismuth [5].

for physics. In the event of a change in external conditions and in the case of peculiar size or configuration of studied objects (thin films, ultrathin wires, or small clusters), an abrupt and quite a large change in conductivity can be observed.

Substances that are similar in chemical composition can be in two different steady states: electroconducting or electroinsulating. In the state of *increased conductivity*, the electrical charge can be transferred by electrons (Fig. 10.19A and B) and also by ions (Fig. 10.19C). Usually the conductance increases with increasing temperature, but it may decrease as well (Fig. 10.19B). The jump in conductivity can also be caused by changes in electrical field strength (Fig. 10.19D), or by the increase in pressure (Fig. 10.19E) as well as due to change in geometric dimensions of the studied object (Fig. 10.19F).

The abrupt change in conductivity, in particular its jumps when temperature changes, cannot be explained by the conventional ideas about charge transfer mechanisms in the dielectrics and semiconductors. Electrical conductivity of nonmetal materials (semiconductors and dielectrics) has *activation nature*: charge carriers are generated by thermal, radiation, optical, and other means. Therefore it would be expected that conductivity *gradually* changes with temperature, not *abruptly*. However, large jumps of conductivity are observed experimentally and, apparently, this is caused by PT.

If conductivity has *electronic* nature, it is likely that its jump means the fundamental change in the *energy electronic spectrum* of a crystal. Electronic spectrum peculiar for dielectrics and semiconductors having the *energy bandgap* (i.e., gap in the electronic states distribution) turns into the *continuous spectrum*, which is the characteristic of metallic substances. Conventionally, such PTs are called the *insulator-metal* transition.

If at a PT the *ionic current* abruptly increases, this phenomenon is usually explained as a “partial fusion” of one of sublattices of ionic crystal (usually, it is cationic sublattice), and such a PT is called the *superionic transition*, which suggests that the solid insulator partially goes into the electrolyte state (electrically similar to the liquid melt or solution).

The drastic change in conductivity does not exactly mean PT. For example, a significant jump in electronic conductivity (in millions of times) is observed in *varistors* that are ceramic wide-band semiconductor dielectrics, for example, carborundum (SiC) or zinc oxide (ZnO). Most of varistors are made of zinc oxide doped with bismuth oxide. At both low and high levels of electric field, varistor is characterized by the ohmic behavior ($j \sim E$) (Fig. 10.20). At that, in weak electrical field, varistor has much higher electrical resistance: its resistance is called dielectric, but in strong electrical field, resistance of varistor decreases millions of times, and it becomes a typical semiconductor.

When field strength increases to a critical value, resistance of varistor suddenly decreases, and, therefore, current through it increases sharply [1]. This phenomenon is quite different from conventional electrical breakdown, because the increase in current is *restricted* (so high current does not destroy the varistor), and its nonlinear characteristic $j(E)$ is reversible. It is possible that varistor-type behavior (whose PT is

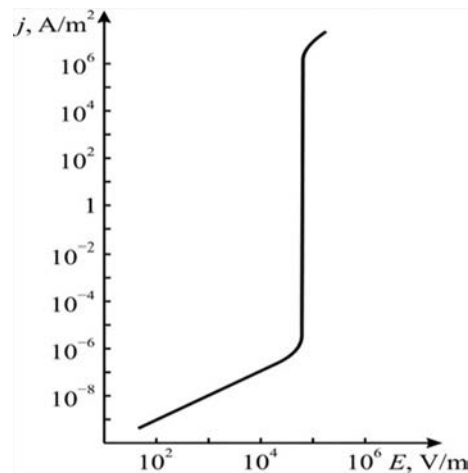


FIG. 10.20

Typical current-voltage characteristic of zinc oxide varistor.

similar to that of insulator-metal type) is caused by the quantum-mechanical *tunneling* through the Schottky barriers that exist on the grain boundaries of ceramics. Varistors are used to protect electrical circuits from the splashes of voltage that often occur in the inductive circuits and can damage circuit's elements: with increase of voltage, the current flows through varistor but not through other elements of a circuit.

Electronic conductivity at insulator-to-metal transition [7]. Transitions of this type are important for solid-state physics as they relate to the fundamental restructuring of the electronic spectrum of a crystal. Studying the nature of the transition from dielectric state to metallic state has not only scientific but also technical interest: sharp variation in electrical properties that are bound with optical properties of crystals is used in electronics and automation devices.

The ability of insulator-to-metal transition was predicted theoretically by H. Mott when analyzing the applicability of energy-band theory, describing electronic spectra of solids. In this theory, usual one-electron approximation is used. It is assumed that each electron moves in the force field of ions and all other electrons (except one under consideration); here, paired interactions are not taken into account even for nearest neighboring electrons (these interactions are included in the average field). In one-electron approximation, the solutions of Schrödinger equation in a crystal are the Bloch functions, while energy eigenvalues form the energy bands that are filled according to Pauli principle.

Before considering any theoretical interpretations of dielectric-to-metal PT, the more complete examination of experimental evidences related to physical characteristics of such crystals near the transition is necessary. A typical example is the transition in three-valence vanadium oxide, whose characteristics are illustrated in

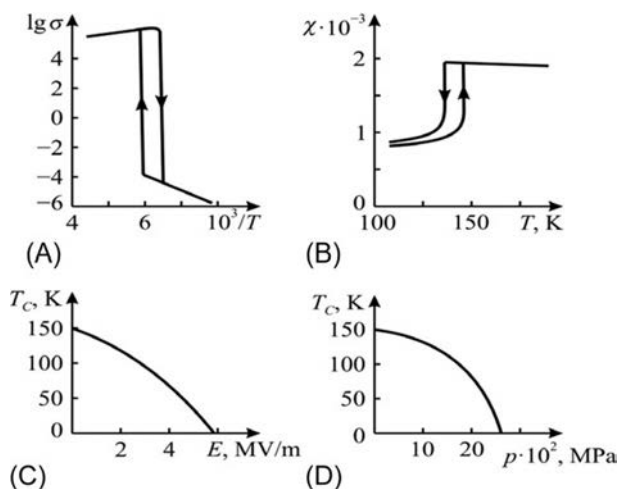


FIG. 10.21

Phase transition in V_2O_3 : (A) temperature jump of conductivity σ (S/M); (B) magnetic susceptibility temperature dependence; (C) temperature-voltage phase diagram; (D) temperature-pressure phase diagram.

Fig. 10.21. At low temperatures, this crystal is dielectric with typical $\sigma(T)$ dependence that has activation character. At critical temperature $T_C = 150$ K, crystal V_2O_3 turns into the metallic phase with a jump in conductivity about ten orders of magnitude (Fig. 10.21A); at that, in the vicinity of PT the temperature hysteresis is observed. Above T_C the $\sigma(T)$ dependence shows metallic character. Critical temperature depends on electrical field intensity (Fig. 10.21C), which promotes metal phase stabilization. Critical temperature depends also on the pressure (Fig. 10.28D), which reduces the distance between atoms that facilitates PT. In the strong electrical fields and at high-pressure dielectric phase V_2O_3 cannot exist. This oxide in its electronic spectrum resembles metal.

It is important that such insulator-metal PT is accompanied by the change in V_2O_3 magnetic properties (Fig. 10.21B). The dielectric phase is simultaneously antiferromagnetic, while the metallic phase is paramagnetic. A similar PT with a jump in conductivity and a change in magnetic ordering is observed in many *transition metal oxides*: in Fe_2O_3 at temperature of 120 K conductivity increases 10^2 times, in WO_3 at temperature of 240 K jump in conductivity is 10^4 times, in VO_2 at temperature of 340 K $\sigma(T)$ increases 10^6 times, and in EuO jump in conductivity at 50 K equals 10^{13} times.

General theory of insulator-metal PT has not been developed: the difficulty lies not only in solving of “many-electron” problem, but also in a wide variety of crystalline structures and chemical compounds, in which these transitions occur. Other than transitions in the metal-oxides, PTs that are similar in nature are found in many chalcogenides.

Three main theoretical models that explain different aspects of this complex physical phenomenon will be considered.

1. **The Mott model** describes the insulator-metal transition in crystals with pronounced *localization* of electrons in their orbitals. Such spatial localization and formation of quasi-bounded states of *d*- and *f*-electrons is peculiar to transition metals and rare earth element compounds. Localization lies in the fact that valence electron is found near its ion rather for a long time τ . Therefore, as it follows from uncertainty relation ($W\tau \leq \hbar$), the width W of correspondent band must be *narrow* in comparison with the band created by *s*- and *p*-electrons, for which the value of τ in orders of magnitude is smaller.

In the dielectric phase *narrow energy band* electronic spectrum complicates any charge transfer: Coulomb repulsion prevents transitions of localized electrons from atom to atom. Energy of this repulsion in the so-called Mott insulator plays the role of a band gap. Localized electrons, without participating in conductivity, significantly increase the polarization of crystal and raise its permittivity.

However, this state of *quasilocalized electrons* is not stable. The impact on crystal by external fields (or small changes in crystal lattice caused by pressure or other factors) may transfer electrons into nonlocalized states. This conversion might be even energy favorable: kinetic energy of free electron movement can match energy of Coulomb repulsion that prevents this movement. This volatility makes possible a transition in which the width of localized states in a band abruptly increases and the energy gap closes.

Localization of electrons in the “Mott insulator” is accompanied by the magnetic ordering (most often, by the *antiferromagnetic* type of ordering). The high-temperature phase, in which localization disappears, corresponds to the nonordered *paramagnetic* phase. At low temperatures (in the ordered phase), this magnetic ordering promotes the formation of specific spectrum of electrons for the “Mott insulator.” Originating below critical temperature T_C , antiferromagnetic superlattice splits the energy band of electronic states on two bands: subfilled and empty, separated by energy gap, which prevents electrical conductivity.

2. **The Wigner model** of “insulator-metal” transition aims to explain the reason for condensation of the gas of free electrons at low temperatures in the nonconducting state. Following this simple model, positively charged ions, localized in the periodic lattice, are replaced by distributed grid of positive charges that compensates the negative charge of electronic gas.

Condensation in this model corresponds to a strong coupling, that is, when potential energy U of interacting charged particles is greater than their kinetic energy E : $e^2/r > \hbar^2/mr^2$, where r is distance, e is charge, and m is mass of particles. From this inequality it is easy to determine that $r > \hbar^2/me^2 = a$, where parameter a corresponds to the Bohr radius of hydrogen-like electronic orbital. Thus conditions of electronic gas condensation in the nonconductive lattice

meet the inequality $r > a$, which means that the average distance between the electrons of the “Wigner insulator” is greater than the radius of their orbit.

This assumption is equivalent to Coulomb energy predominance of over kinetic energy. Kinetic energy of electronic gas increases in the event of temperature rise, and at a certain (critical) temperature this results in “melting” of “electronic sublattice,” and gas of free electrons gives rise to metal conductivity. Wigner model usually can explain metallic type of conductivity in the strongly doped semiconductors, as well as this model can describe some features of the “insulator-metal” transition in the mixed valence oxides (such as V_2O_3).

- 3. *The model of excitons*** can also be used to explain critical increase in the free charge carrier *concentration*, as well as the temperature hysteresis, observed at switching from metal to dielectric phase. As demonstrated by Mott, electronic spectrum with a *slight overlap of two bands* is similar to the semimetals. This promotes localization of electron-hole pairs in the form of excitons (in semimetals any band gap is absent, but the bands do not overlap: valence and conductive bands only come into contact). Formation of excitons in crystal lattice *increases dielectric polarization* (because electronic polarizability is proportional to the cube of distance from electron to positive charge). Accordingly, with the change of polarization in crystal its phonon spectrum also changes because permittivity increases.

Interaction of excitons with each other and with crystal lattice results in the increase of exciton concentration in a crystal, facilitating the process of their formation, as the birth energy of bonded electron-hole pairs decreases. Conditions for growth of *exciton avalanche* can occur even with small changes in crystal lattice due to external influences (such as electrical field or pressure). This is the reason for “*excitonic instability*” of dielectric: above a threshold concentration the excitons disintegrate into free electrons and holes; as a result, the dielectric is transformed into conductive material. The interaction of excitons with phonons and dielectric permittivity increases because excitonic polarization reduces frequency of optical phonons. This is equivalent to the formation of bound exciton-phonon mode (soft oscillation mode) whose frequency is reduced near PT, resulting in lattice instability, which is peculiar to any PT.

Thus explanations of insulator-to-metal transition with a large jump in electronic conductivity might be different. Numerous experimental cases of such conversion can be explained from different positions, including the aforementioned theoretical models. The main factor that unites these models is the instability of the crystal electronic spectrum, for which a jump in conductivity is observed. It is important to note that, sometimes, during PT, instead of conductivity jump, the large polarization and *high permittivity* occur.

DIELECTRICS WITH UNSTABLE ELECTRONIC SPECTRUM [7]

Crystals that have high density of excitonic states are very promising materials for high-temperature superconductor elaboration. In conventional superconductors—metals and metallic alloys—PT to superconducting state is theoretically limited by temperature of 25 K. At higher temperatures, Coulomb repulsion exceeds the attraction between conductive electrons that form Cooper pairs by electron-lattice interaction, thanks to electron-phonon coupling.

However, for a long time theoretical studies pointed out on the possibility of different opportunities of electron attraction: by means of excitonic exchange. With such mechanism, in principle, superconductivity can be obtained even at 300 K (currently high-temperature superconductivity reaches a temperature of about 200 K).

The term “exciton,” while discussing capabilities to get superconductivity, should be interpreted comprehensively: *any polarized excitation* in the electronic subsystem of a crystal might be understood as the exciton, including the variety of vibrational modes of volumetric or surface type. As classic superconductors are seen in the three-dimensional metals, while excitons can be extended only in dielectrics, the excitonic superconductor has to be *both metal and dielectric*, representing a system of “crystal into the crystal.” Such system should be a complex substance, in which metallic subsystem allows free movement of electrons, while dielectric subsystem is the environment for spreading of excitons, binding electronic pairs in metallic subsystem. In this case, the dimensionality of a matter should be decreased.

In order to apply the exciton mechanism to superconductivity, two basic systems are offered:

1. The **one-dimensional system** (the 1D “needle”-like crystal, Little’s model), for example, can be a long well-conductive molecule thread with easily polarizable side radicals that could provide attraction of conduction electrons due to excitonic transfer. The presence of excitons makes appearance of high-temperature superconductivity possible, as they can compensate the Coulomb repulsion of electrons. Unfortunately, obtaining superconductivity in the ideal 1D crystal seems almost impossible owing to thermal fluctuations (this is so-called R. Peierls prohibition¹).

Nevertheless, in the 1D systems PT from the *quasimetal* phase to the *high- ϵ dielectric* phase is possible, and this question will be discussed later.

2. The **two-dimensional system** (the 2D, “sandwich”-type crystal, V.L. Ginzburg model) consists of the “quasi-metal” layers surrounded on both sides by dielectric layers that should provide excitonic exchange, contributing to Cooper-type pairs of electron formation and superconductivity in metallic layers.

¹Note. In one-dimensional organic material (Bechgaard salt, TMTTF₂ClO₄), the superconducting phase was found at ambient pressure with transition temperature $T_C = 1.4$ K. Several other similar salts become superconducting only under increased pressure; all these are the exclusions from Peierls theorem.

The experimental verification of 2D models received full success: the following investigation enables to find many materials with superconductivity, including many *high-temperature* superconductors.

Electrical properties of 1D and 2D systems proved to be extremely interesting, not only for theory of PT but also for new applications in electronics. Results of numerous studies show opening of several classes of quasi-1D, quasi-2D crystals with *sharp anisotropy* of both σ and ε , and various structural PTs, caused by the instability of electronic spectrum with electron-phonon and electron-exciton interactions.

Consider the model of quasi-one-dimensional metal. In some 1D needle-like crystals at rather high temperatures (50–150 K), conductivity shows temperature peak (Fig. 10.22A), in which $\sigma(T)$ is significantly higher than conductivity of any extremely high-conductive metal at the same temperature (such as copper or silver). One might suppose that high conductivity near its temperature peak is a result of superconducting state fluctuations (but this state in the 1D system, as a rule, cannot be realized entirely).

Calculations show that, following one-dimensionality condition (ideally thin wire), the metal-type conductivity is incompatible with 1D system of any valence of metallic atoms in a ground state ($T=0$), but should acquire properties of dielectric (R. Peierls theorem). Indeed, in the experiments the quasi-1D conductive system at

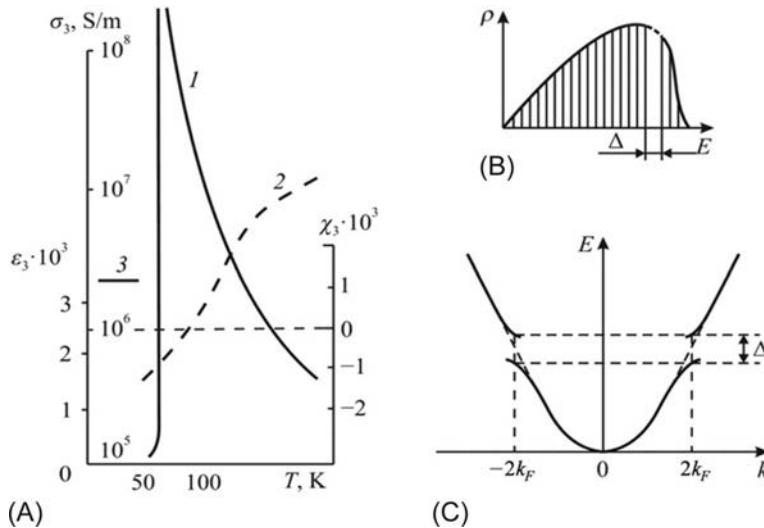


FIG. 10.22

Phase transition in TTF-TCNQ crystals: (A) conductivity σ_3 (curve 1), magnetic susceptibility χ_3 (curve 2) and permittivity ε_3 (curve 3); (B) density ρ of electronic states with energy gap Δ below transition temperature; and (C) energy gap in electronic spectrum of quasi-one-dimensional crystal.

low temperatures mostly turns into dielectric. An example of such behavior is shown in Fig. 10.22 for quasi-1D organic metal crystal: one of most studied 1D conductors—conductive polymer TTF–TCNQ (tetrathiafulvalene-tetracyano-quinodimethane). Anisometric ion (TCNQ)[−] demonstrates strong acceptor properties and it can form salts, in which electronic transfer of charges is possible. Among many such salts that have quasi-1D structure, the combination of TCNQ[−] with TTF⁺ should be noted, which plays the role of well-conductive cation (at 300K). Currently, many crystal polymers of this type are investigated, and they form a class of quasi-1D “metals”; however, most of them transits to the insulating phase at low temperatures.

Needle-like crystals TTF-TCNQ have strong anisotropy of their electrical properties. At high temperatures (on the side of metallic phase) conductivity of this crystal along the “needle” (σ_3) and transverse to it (σ_1) varies about 500 times, while in the point of PT the ratio σ_3/σ_1 reaches 10^5 . This anisotropy characterizes the degree of one-dimensionality of the studied crystal, which looks like a “clot of joint needles.”

At low temperatures (on the side of insulating phase), conductivity of TTF–TCNQ is very small, but a strong anisotropy becomes apparent in *permittivity*. Measured near liquid helium temperature at microwaves (when low-frequency relaxation processes cannot give any dielectric contributions) crystal TTF–TCNQ shows $\epsilon_3 = 3200$ and $\epsilon_1 = \epsilon_2 = 6$.

All characteristics in Fig. 10.22A are shown only for one direction—along peculiar axis 3 of 1D crystal. It is important to note that at above transition temperature the effect of “paraconductivity” is observed: the $\sigma_3(T)$ dependence is close to Curie-Weiss law. However, such a “paraconductivity” cannot turn into superconductivity below PT. At temperature T_C , instead of superconductivity appearance, the $\sigma_3(T)$ dependence shows a discrete-steps break, when conductivity becomes as small as in dielectric. However, instead of large σ_3 below T_C , *permittivity* ϵ_3 becomes abnormally high, and it practically does not vary with temperature (Fig. 10.22A, curve 3).

At low temperatures, the extremely high value of ϵ_3 in TTF–TCNQ reminds the properties of displacement-type paraelectrics (such as SrTiO₃ or KTaO₃). Note that in strontium titanate (and in potassium tantalite as well), a large value of permittivity is observed ($\epsilon \sim 10^3$ – 10^4) without any microwave dispersion. In paraelectrics this is a result of “soft phonon” frequency’s critical decrease, which corresponds to the dynamic instability of *crystal lattice*. However, in case of TTF-TCNQ, the cause of $\epsilon \sim 10^3$ lies in the peculiarities not of a phononic but of the *electronic spectrum*.

Magnetic susceptibility of TTF–TCNQ type-crystals is small, and throughout studied temperatures change from paramagnetic to the *diamagnetic* susceptibility (Fig. 10.22A). This is significantly different from the discussed 3D crystals of the V₂O₅ type that also demonstrate PT from the high-temperature metallic phase to the low-temperature dielectric phase: the V₂O₅-type crystals show the *antiferromagnetic* ordering in their low-temperature phase. Therefore, in TTF–TCNQ the nature of PT is different from Mott transition, where the low-temperature phase corresponds to the dielectric phase with a completely filled valence band.

Phonon spectrum of quasicrystals, based on neutron scattering study, is characterized by a minimum in the $\omega(k)$ dispersion at certain value of phonon quasi-impulse ($p = \hbar k$). This is so-called W. Kohn anomaly that is caused by electron-phonon interaction, and it is observed in such value of phonon quasi-impulse that equals to the double quasi-impulse of Fermi electrons ($k = 2k_F$). Note that in 1D metals the Fermi surface consists only of two planes: $+2k_F$ and $-2k_F$. Electron scattering with energy conservation can occur only between these planes, and it is accompanied by the change in impulse on $2\hbar k_F$. Exactly at this value of impulse the electron-phonon coupling increases.

In TTF – TCNQ crystals, as well as in other type of quasicrystals, as temperature decreases the Kohn anomaly (minimum in $\omega(k)$ dependence) becomes larger when PT approaches. This anomaly results in the *multiplication* of unit cell at transition to dielectric phase with the splitting-off of optical branch (or branches). One of these branches becomes the “soft” mode, and this makes the connection to phonon spectrum with *charge density waves* that leads to high ϵ_3 below transition temperature.

This is illustrated in Fig. 10.23. When approaching PT down from the metallic phase, the Kohn anomaly appears in the acoustic branch of 1D metal and gradually goes down (Fig. 10.23A). To simplify this model, Kohn anomaly is shown here at wave vector $k = \pi/2a$, but, in fact, such coincidence can only be a random and usually is not observed.

Fig. 10.23B shows an *intermediate* (virtual) case of branches splitting: from the acoustic branch “slowing down” the three optical modes and one acoustic mode are

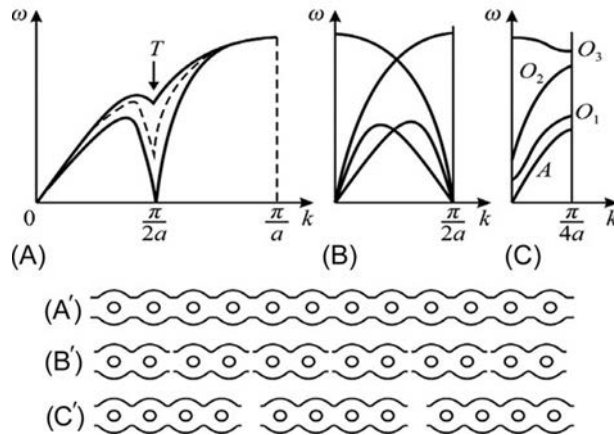


FIG. 10.23

Changes in the phonon's spectrum: (A) temperature-dependent minimums (Kohn anomalies in acoustic mode); (B) mode frequency at $T = T_C$ vanishes at wave vector $k = \pi/2a$ (virtual case); and (C) new phase appearance with four times smaller Brillouin zone and three optical modes, including two “soft” modes (O_1 and O_2); A' , B' , C' —modulated electronic structures in one-dimensional crystal that determines associated electron-phonon modes.

conceived. From Fig. 10.23C it can be seen that after PT unit cell becomes increased four times, two of splintered optical branches (O_1 and O_2) are the soft modes.

In the started cell of 1D metal (with lattice parameter a) all nodes are “neutral,” and electronic density is distributed evenly (Fig. 10.23A') (metal phase). Below PT the displacement of atoms occurs: in the doubling cell (virtual case) atoms are grouped to two (Fig. 10.23B'); in a new formed lattice atoms are grouped by four (Fig. 10.23C' that corresponds to Fig. 10.23C). The redistribution of electronic density gives rise to a situation, when in the region of atoms the “thickening” of electronic density increases, and negative effective charge arises, while in the region of atom “depression” the positive effective charge corresponds.

By this way, the *wave of electronic density* appears, which in this (very idealized) case is commensurate with the lattice parameter (exactly four times). However, in general, what can be implemented in all known experiments, the Kohn anomaly does not arise exactly at $k = \pi/2a$, but in the *arbitrary place* of Brillouin zone. Therefore the structure of electronic density is *modulated*, that is, it is *incommensurate* with the lattice parameter. This is exactly the reason for electronic spectrum instability.

Kohn anomaly in phonon spectrum is a well-known characteristic of many superconducting 3D metals and alloys, wherein electron-phonon interaction is the main cause of superconducting PT. However, in a given case, the lattice instability of Kohn anomaly type does not lead to superconductivity, but to the *dielectric phase* appearance with ultrahigh value of permittivity. This low-temperature “dielectric phase” of quasi-1D metal has some properties similar to superconductors: in such dielectric the energy gap is disclosed in electronic energy spectrum (Fig. 10.22C), but the pair of interacting electrons (or holes) is descended from *different bands*, while Cooper interaction in the superconductors occurs between electrons, belonging to the *same* energy band.

Therefore quasi-1D systems consisting of radicals with quasi-metal phase with large conductivity at lower temperatures turn into the dielectric phase but with the *unstable electronic spectrum*, which at elevated temperatures again transforms into quasimetal. These properties, observed even at microwave frequencies, constitute the evidence of 1D structure proximity to the superconductivity, while the Kohn anomaly is the evidence of electronic nature of this PT.

Theory of 1D conducting systems was developed long before their experimental realization. It was demonstrated for the first time by W. Peierls, who determined that the structure of 1D metal at low temperatures should be changed in such a way that it turns into the insulator. At that, crystal lattice must be reconstructed: lattice distortion, caused by electron-phonon interaction, splits the partially filled energy band of 1D metal to the totally filled energy band and empty sub-band that corresponds to dielectric. In Peierls theory, Coulomb interaction of electrons is not taken into account; however, from the discussed Mott theory it can be seen that this interaction results in low-temperature instability of metal phase. Defects in crystal lattice are also favorable for insulating phase appearance instead of metal phase.

With temperature decrease in the electronic spectrum of Peierls-type insulator, the energy gap Δ forms that tends to have a maximum at temperature

$T=0\text{K}$. This $\Delta(T)$ dependence further extends the analogy between metal-dielectric and metal-superconductor PT. In addition, the phenomenon of “para-conductivity” is observed (curve 1 in Fig. 10.22A) that can be interpreted as the *fluctuations of superconducting phase*, occurring near transition from quasimetal to high-polarized insulator. This transition appears to be disproportionate, because the value $2k_F$ is not a multiple of π/a (a is the lattice parameter in the conductive phase). The modulated charge density waves in the structure of dielectric (originated below PT) causes *large and low-inertia* high contribution to dielectric constant.

In the vicinity of PT, properties of 1D structure are very sensitive to the fluctuations. Theoretically, for the violation of long-ordered structure, the break of ordering, originated *only in one point*, is sufficient. However, in the real quasi-1D structures (thin, needle-like, but still thick *macroscopic* crystals), situation is different, and stability of the system to fluctuations significantly increased due to interaction between the neighboring “threads” of such structure. The degree of “three-dimensionality” is qualitatively assessed by the high value of σ and ϵ anisotropy of these crystals.

This 3D interaction can not only “extinguish” fluctuations, but also suppress the Peierls transition. Due to this suppression, in some quasi-1D structures superconductivity becomes possible: for example, polymer $(\text{SN})_x$ is a quasi-1D superconductor, and in it, dielectric phase does not occur; however, its transition temperature is very low ($T=0.3\text{K}$).

The impact of fluctuations on superconductive PT is minimal in ordinary 3D structures, where, for ordering violation, the break should occur on some *surface* in a crystal. The 2D structures, in terms of resistance to the fluctuations, are in an intermediate position, and as regards destruction of ordering in them, the “fluctuating break” should be on the *line* (but not at a *point*, for 1D structures). Therefore in the quasi-2D structures the probability to obtain superconducting state is much greater than in 1D structures (while the probability of high-polarizable dielectric occurrence is lower than in the case of quasi-1D systems).

Currently, 2D structures of metal-insulator type are widely implemented, both in the macroscopic level (in the form of alternating metal and dielectric films) and in the microscopic level (atomic layers). The latter are the most interesting systems with electron-phonon instability, leading to PT. The high-temperature superconductivity discovered in 2D structures offers great opportunities for technical applications in cryogenic electronics.

Electron-phonon instability in 2D systems, as in 1D systems, might result in PTs, analogous to the Peierls transition. However, if in the 1D structures the metal-insulator transition is most typical, then in case of 2D structures the transition of metal-semimetal looks more usual. In both cases, transitions are conditioned by the electron-phonon interaction.

The *dielectrics with superconducting phase transition*. The peculiarity of electronic spectrum of semimetals leads to electron-electron pairing with Cooper pair formation, in addition to their electron-hole “pairing” (energy gap formation as the “forbidden band”, i.e., “dielectrization”), resulting in superconductivity. In latter

case, the electron attraction (arising due to exchange by phonons) exceeds Coulomb repulsion of electrons.

Not only in metals, but also in crystals with energy gap in their electronic spectrum (semiconductors and dielectrics that have *increased concentration* of charged carrier), the conductivity at low temperatures, instead of a smooth decrease to zero, can in spurts increase up to the infinite value. Coupling energy of Cooper pairs is caused by the electron-phonon interaction; as a result, the pair of electronic cannot be scattered by lattice vibrations, which leads to superconductivity. The *very high permittivity* facilitates the formation of Cooper pairs: increased ϵ dramatically reduces Coulomb repulsion of electrons.

It should be noted that that large ϵ at low temperatures is peculiar to paraelectrics and virtual ferroelectrics. Indeed, the superconducting PTs in dielectrics was first discovered in the doped strontium titanate (i.e., virtual ferroelectric with $\epsilon \approx 40,000$ at $T = 4$ K) and in the narrow-gap ferroelectric semiconductor (SnTe possessing $\epsilon \approx 2000$). Although in these materials the temperature of superconducting transition does not exceed 0.3 K, the possibility of superconductivity in them is crucial. This is used in many investigations and, finally, resulted in the invention of high-temperature superconductivity in the mixed oxides that are typical for ferroelectric perovskite structure (for details see Section 5.7).

Phase transitions with ionic conductivity jump. In the ionic dielectrics, PT may occur with abrupt change in ionic component of electrical current (Fig. 10.24). As in the case of electronic conductivity jump, ionic conductivity may be increased with temperature by several orders of magnitude. Usually ionic conductors are *liquid electrolytes*, that is, the melts or solutions of salts, acids, or alkalis, where charge transfer is caused by mobile ions (positive or negative; accordingly, cations or anions).

Therefore conductivity of *electrolytes* ($\sigma = 10^{-3} - 10^3$ S/m) is significantly lower than electronic conductivity of metals, but large enough for technical use in electrolytic conductors (as well as in the sources of electrochemical power). While direct current passes, the chemical composition of electrolyte varies due to the *electrolysis*.

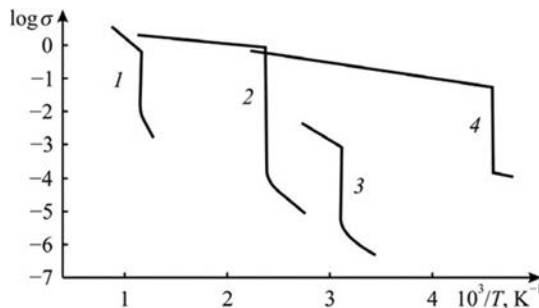


FIG. 10.24

Conductivity σ (S/cm) temperature jumping at phase transitions in different superionic crystals: 1— Li_2SO_4 ; 2— AgI ; 3— $\text{Ag}_2\text{Hg}_4\text{I}_6$; and 4— $\text{Rb}_4\text{Ag}_5\text{I}_9$.

Ionic conductivity of usual solid compounds (crystals, glass, or ceramics), as a rule, is small (at 300 K $\sigma = 10^{-1} - 10^{-8}$ S/m) due to durable connection of ions in crystal lattice. Exceptions are *solid electrolytes*, known also as *superionic conductors*.

Some ions in these electrolytes, owing to peculiarities of crystalline structure (disordering of lattice structure, between-framework channels, etc.), may take several equivalent positions, and, therefore, they can easily migrate when an electrical field is applied to the material. For example, in the β -alumina ($\text{NaO} \cdot n\text{Al}_2\text{O}_3$ at $n = 5 \dots 11$), the *cations* Na^+ can easily move between Al_2O_3 blocks; by the same way, in combined crystal Ca_2YF_3 the *anions* of fluoride F^- can easily migrate. Solid electrolytes are widely used in modern power sources, which have increased reliability and power consumption.

A special case is *phase transition* in the ionic dielectric having low conductivity ($\sim 10^{-10}$ S/m) to superionic state with high conductivity: $1 - 10^2$ S/m (Fig. 10.24). Increased ionic conductivity in superionic conductors is observed at temperatures much lower than their melting temperature. Moreover, after these crystals are melt, their ionic conductivity sometimes even slightly decreases, as compared with the superionic solid phase.

From a microscopic point of view, PT to the superionic state is considered as a “fusion” of one of sublattices of crystal (“partial melting”). It looks like the superionic conductor consists of two sublattices. One of them is the hard regular structure of ions (as usual, anions) while the second represents a highly disordered structure, including mobile ions (usually—cations), and has a large number of vacancies. It is assumed that the creation of internode vacancies (Frenkel defects) plays a major role in superionic conductivity. Increased mobility of cations in the superionic phase is explained by the *cooperative interaction* of Frenkel defects: the more the cations, the easier they are transferred through internode states. When temperature increases to a certain value, the *avalanche* of structural defects occurs, and conductivity rises sharply (corresponding jumps are shown in Fig. 10.22).

The occupancy of internode states by cations may also critically rise, thanks to restructuring of crystal near the PT. Thus basic lattice is changed by such a way that cations are “squeezed out” in the internodes, thereby increasing electrical conductivity. The sharp increase in cation mobility at temperatures higher than PT is conditioned by the decrease in energy barriers between equivalent unoccupied internodes.

Depending on structural features of the superionic conductor, its conductivity may be not only isotropic, but also anisotropic. Strong anisotropy of ionic conductivity is possible in the peculiar directions of a crystal, along which potential barriers are substantially reduced (down to ~ 0.1 eV). One-dimensional superionic conductivity occurs at PT in crystals such as $\text{Li}_2\text{Ti}_3\text{O}_7$, LiAlSiO_4 , and others. However, superionic conductivity in silver iodide-type crystals (AgI , Rb_4AgI_5 , and similar others) is almost independent of the direction (as it is expected in the crystals of cubic symmetry).

At present, many crystals are studied with superionic conductivity, in which the ions Li^+ , Na^+ , Ag^+ , and H^+ take part. A separate but important case of such conduction is the *protonic conductivity*. Stripped of electronic shell, the ion H^+ , as compared

with other cations, has not only a smaller mass, but also a very small dissipation section during collisions. This causes unusually high mobility of protons that determine electrical conductivity of many crystals that hold hydrogen (ice, some of order-disorder-type ferroelectrics, polymers, etc.).

Technical devices with high ionic conductivity (solid electrolytes with jump in conductivity near superionic PT) are used in automation and instrumentation. Solid electrolytes are applied as elements of subminiature high-energy accumulation. Using solid electrolytes, it is possible to create miniature electrical capacitors. Many properties of superionic conductivity, such as transport of electrode material (usually, silver electrode can be transferred through a crystal), allow the use of superionic crystals in the electrical elements (i.e., as electricity limiters or electric quantity measurement). The jump in ionic conductivity near the superionic PT can be used to control temperature.

Posistor effect in ferroelectric semiconductors. The anomaly in temperature dependence of electrical conductivity in ordinary ferroelectrics near ferroelectric PT usually is small; this transition does not change significantly the general nature of activation-type increase in conductivity $\sigma(T)$ with increasing temperature. The point is that energy gap (forbidden band) in most ferroelectrics is large, and, therefore, they usually are considered as dielectrics but not as semiconductors.

However, by intense doping, it is possible to prepare ferroelectric semiconductors with rather high conductivity. In some special cases near ferroelectric Curie point a sharp decrease in conductivity can be seen (i.e., 10^2 – 10^4 times) with *increasing temperature*. Experimentally observed multiple increases in *resistivity* in the doped BaTiO₃-type ferroelectric ceramics are shown in Fig. 10.25.

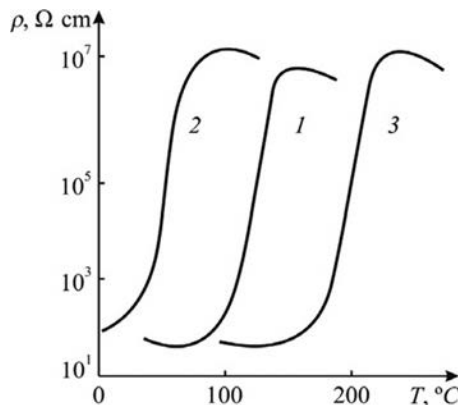


FIG. 10.25

Critical changes in specific volume resistance of ceramic posistors, which are doped with neodymium and manganese ferroelectric solid solutions: 1—BaTiO₃; 2—Ba(Ti,Sn)O₃; and 3—(Ba,Pb)TiO₃.

Unusual temperature change in conductivity in some ferroelectric semiconductors near transition from the polar (ferroelectric) phase to the nonpolar (paraelectric) phase is observed in *posistors* (thermistors with positive temperature coefficient). This effect can be seen in the doped oxide-type ferroelectric of barium titanate type. Increased conductivity (rather low resistance) is peculiar only to the *polar phase* (when temperature is below the Curie point). At transition to the nonpolar (paraelectric) phase, the electrical resistance *increases* significantly. Typical to dielectrics and semiconductors, it is possible to observe activation-type conductivity increase only at further heating to temperatures much higher than the Curie point.

As shown in Fig. 10.25, resistance of posistor increases only within a narrow temperature range—close to PT. It should be noted, that ferroelectric Curie temperature can be changed using solid solutions; for instance, in ferroelectric solid solutions (Ba, Pb)TiO₃ transition temperature can be increased from 400 up to 600 K, while in solid solution (Ba,Sr)TiO₃ it falls down from 400 up to 200 K. Thus ceramic ferroelectric semiconductors with effect of *posistor* can be elaborated for a very wide temperature range—depending on technical requirements for such elements in electronics.

To obtain posistor, the ferroelectric *ceramics* based on barium titanate are commonly used. In the BaTiO₃ doped by donor ions (La³⁺, Dy³⁺, or Ce³⁺) that partially replace Ba²⁺ (or using ions Nb⁵⁺ to replace ions Ti⁴⁺), the conductivity in the polar phase, as compared with pure BaTiO₃, increases billions of times ($\sigma \approx 10^{-2}$ S/cm). Doped by lanthanum, BaTiO₃, after synthesis and sintering at high temperature, turns into a semiconductor with *n*-type conductivity, providing the *electronic transfer* between titanium ions (Ti⁴⁺ and Ti³⁺).

The *crystallites* in this doped barium titanate ceramics are conductive as in the ferroelectric, so also in the paraelectric phase. However, *boundaries* between crystallites are able to change their electrical conductivity—depending on the presence or absence of *spontaneously polarized state*. This is because during synthesis of ceramics the oxygen is adsorbed by crystallite surface; next, oxygen diffuses into the pores in crystallite boundaries. Additional oxygen atoms attract electrons from nearby Ti³⁺ ions; these electrons are captured in traps, and thus the *insulating barrier* arises between crystallites that can decrease conductivity of crystallite *surface*. At the same time, the regular structure *inside* crystallites prevents diffusion of oxygen and always remains highly conductive.

Below PT temperature (in BaTiO₃ it is 130°C) its perovskite structure is spontaneously polarized, and this polarization creates a *neutralizing* effect for electrons on the borders of crystallites, capturing in the traps in the boundaries of crystallites. For this reason, the resistance of doped BaTiO₃ in its *ferroelectric phase* is low, because boundaries of crystallites have the same conductivity as the crystallite's body. From this model it is clear that in the ferroelectric *single crystal* posistor effect is impossible.

Above PT barium titanate becomes paraelectric, and spontaneous polarization disappears. The boundaries of crystallites become insulating, and the resistivity increases, demonstrating posistor effect. Thus the increased conductivity in the polar

phase of ferroelectric semiconductor of barium titanate type is due to the *activating influence* of spontaneous electrical field.

Practical applications of posistors—ceramic elements with low “cold” resistance and high “hot” resistance—are very wide ranging. Posistors are used in thermal-controlling systems, in measurement technology, in devices that prevent thermal and current overload, in engine-starting systems as well as in auto-control systems. They are also used to protect overvoltage and short circuit currents, being connected in series with the load: the current in posistors is limited by a safe level.

Thus posistors can be considered as one different type of *thermistors*—ceramics elements, characterized by a large-temperature coefficient of resistance.

10.5 PHASE TRANSITIONS IN LIQUID CRYSTALS

It is essential for condensed matter physics that, in addition to crystals and liquids, the “transitional” state exists—between the crystalline (anisotropic) and the liquid (isotropic) substances. The *liquid crystal (LC)* can be defined as a thermodynamically stable physical state, in which material maintains the *anisotropy* of physical properties as solid crystal and the *fluidity* that is characteristic of liquid. This unusual combination of properties enables widespread use of LC in modern electronics and computer sciences.

It is assumed that LC state arises in the process of PT, which is degraded in a certain temperature range. As is known, near PT the structure of a matter (and most of its physical properties, conditioned by this structure) is ready to be changed even under a weak influence of external fields—electrical, thermal, magnetic, optical, or mechanical. This compliance of the LC opens a possibility for easy control of the LC physical properties (first of all, control its optical properties) even in case of very small controlling impact (mainly, using electrical field).

The first observations of liquid crystalline *mesomorphic* properties in some organic compounds were carried out by F. Reinitzer about 100 years ago. Currently, thousands of organic compounds are known that are capable of forming LCs. The main requirement of mesomorphism manifestation is, particularly, the *geometric anisotropy of molecules* that should be anisometric, for example, being a long and relatively narrow “spindle.” Depending on the geometry of molecules, LC can pass through one or more *mesophases* (intermediate phases) before the final transition into isotropic liquid occurs. Transitions to intermediate states can be caused by temperature change (thermotropic mesomorphism) or due to influence of solvents (lyotropic mesomorphism).

Classification of liquid crystals. Intermediate LC phase that exists between crystal and normal liquid is often called the mesophase. The main types of molecular ordering are the *nematic* (“filamentary”), the *smectic* (“polygonal”), and the *cholesteric* (“screw”). The term “nematosis” in Greek means “thread” and describes the 1D ordering; “smegma” (also Greek word) means “soap” and corresponds to 2D ordering; it is typical for flat layers of strongly bounded molecules. A spiral structure in

orientation of LC molecules is observed in the derivatives of well-known organic matter—cholesterol.

The *nematic liquid crystal* (NLC) can be described by 1D arrangement of anisometric molecules that form in a volume of NLC, the ordered “swarms”; they consist of a large number (10^5 – 10^7) in the same way as targeted molecules (Fig. 10.26A).

Definite order in the direction of these molecules that tends to be installed parallel to a common axis is characterized by the unit vector (or director) \mathbf{n} . This is true for all macroscopic anisotropic properties of NLC that are described by tensors. Orientation of director in a space might be arbitrary; however, in practice, it is determined by weak forces (such as orienting effect of vessel walls). Optical axes \mathbf{n} and $-\mathbf{n}$ are indistinguishable. For example, if any separate molecule has a permanent dipole moment, the number of dipoles directed “up” is exactly equal to the number of dipoles directed “down,” so that the NLC system is usually not pyroelectric (or ferroelectric).

The anisometric form is in the geometrical shape of molecule that is very different from the usual quasispherical form; usually this shape looks like the elongated ellipsoid. In the arrangement of such molecules, any long-range ordering is absent, and, therefore, the X-ray photograph does not show Bragg-type diffraction peaks. Correlation in the centers of gravity of neighboring molecules is similar to

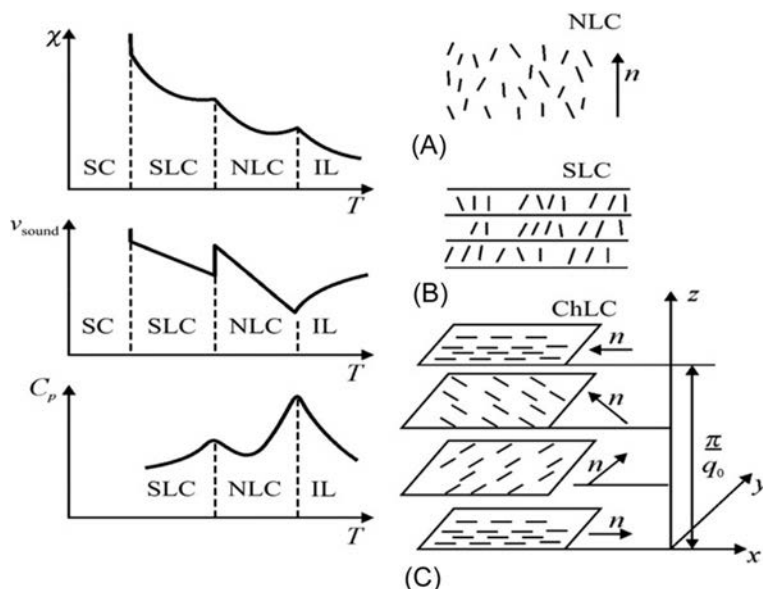


FIG. 10.26

Temperature dependence of thermal conductivity χ , speed of sound v_{sound} , and heat capacity at constant pressure C_p for LCs in sequential process of change mesophase state; location of LC molecules is also shown: (A) nematic phase, NLC; (B) smectic phase, SLC; and (C) cholesteric phase, ChLC (SC—solid crystal, IL— isotropic liquid).

correlation that exists in common liquid. At that, NLC can flow, in spite of very high viscosity.

The dual nature of nematic phase (which is similar to liquid, but optically is uniaxial) is most evident in the nuclear magnetic resonance (NMR) spectrum. Uniaxial symmetry is defined by the splitting of lines (that is absent in conventional isotropic liquid phase).

The *smectic liquid crystals* (SLCs) are characterized by a 2D molecular arrangement. In smectic crystals, a stable flat of molecular layers exists with clearly defined distance between the layers that can be measured by X-ray diffraction. In the volume of SLCs these layers may be deformed (bent or curled), while maintaining strict ordering between the nearest neighboring molecules (Fig. 10.26B).

There are many types of smectics, representing a variety of macroscopic textures that vary during optical observations. In the SLC classification, *three basic cases* are defined by letters A, B, and C. For example, many cholesterol compounds during temperature decrease are converted into the smectics of type A, the main characteristics of which have the following features:

- thickness of layers is close to full length of molecules;
- the “centers of gravity” of molecules have no long-range ordering inside of each layer;
- each layer is a 2D liquid;
- system is optically uniaxial with optical axis z , perpendicular to the plane of layers;
- directions z and $-z$ are equivalent.

Requirement of sustainability between layers imposes a condition: $\text{rot } \mathbf{n} = 0$. Thus spiral arrangement of molecules is prohibited.

In smectics of A and C types each layer behaves as a 2D liquid, and smectics C is optically biaxial. In smectics of B type the layers reveal periodicity and strength of 2D solids, that is, similarity to crystalline film. Of the three major phases of smectics, type B is most ordered.

The *cholesteric liquid crystals* (ChLCs) in their molecular ordering resemble nematics, but they have *spiral orientation* of spindle molecules that spontaneously form the helical structure (Fig. 10.26C). In this case, long-range ordering in the arrangement of “centers of gravity” of molecules is absent; they are primarily oriented along axis, oriented by the director \mathbf{n} .

However, \mathbf{n} has no fixed direction in a space. If the helix axis is denoted through z , then n will be described by the following components:

$$\begin{aligned}n_x &= \cos(q_0z + \phi), \\n_y &= \sin(q_0z + \phi), \\n_z &= 0.\end{aligned}$$

Direction of helix axis and angle ϕ are arbitrary; therefore, in the ChLCs another type of symmetry violation is observed different from that in NLC. Their structure is

periodic on z and, since \mathbf{n} and $-\mathbf{n}$ again are equivalent, the spatial period L is a half-step spiral:

$$L = \pi/|q_0|.$$

Typical value of L is about 3000 \AA that is much larger than molecules. As L is comparable with the wavelength of light in the visible optic range, this periodicity leads to *light scattering*. For helical arrangement of molecules, a condition $\text{rot } \mathbf{n} = -q_0 \mathbf{n} \neq 0$ can be implemented.

For different substances, the temperature range of the LCD phase existence can be from 0.01 to 100 degrees. In this temperature range the PTs occur, determining many characteristics of LCs. The first transition is *partial fusion* of solid crystal with appearance of one of LC phases; in Fig. 10.26, this process is shown symbolically: PTs can occur between different mesomorphic phases. Finally, with increasing temperature, any partial ordering of anisometric molecules disappears and LC is transformed into the *isotropic liquid* (IL).

Typical consecutions of PTs in LCs with temperature increase are:

SC \rightarrow SLC \rightarrow NLC \rightarrow IL

SC \rightarrow ChLC \rightarrow NLC \rightarrow IL

SC \rightarrow SLC-A \rightarrow SLC-C $\rightarrow \dots \rightarrow$ ChLC \rightarrow IL

SC \rightarrow SLC \rightarrow ChLC \rightarrow IL

(SC—solid crystal, SLC—smectic phase; NLC—nematic phase; ChLC—cholesteric phase; IL—*isotropic liquid*).

In some cases, LCs might have several smectic and cholesteric phases. All transitions between phases are usually *reversible*. In the neighborhood of mesomorphic conversions almost all physical properties of substance change, including electrical and optical properties; therefore the features of LC are *very pliable*. They vary greatly in case of applied external fields, as well as under changing of light conditions, temperature, and pressure. This is caused by the imbalance between molecular interactions in LCs that is facilitated by the proximity of PT, when the structure of LCs is unstable.

Therefore LCs are organic compounds, whose molecules are anisometric, having sufficient “hardness” (that ensures stability of their orientation) and attracted to each other by the forces of electrical nature (dipole, quadrupole, and more complex interactions). Temperature interval ΔT of LC mesophase existence in pure compounds is usually small. For example, in LC MBBA (*p*-methoxybenzaldehyde-*p'*-*n*-butylamine), temperature interval ΔT might exist from 20 to 46°C; in other LC EBBA (*p*-ethoxybenzaldehyde-*p'*-*n*-butylamine), interval ΔT is from 35 to 80°C. However, in the *mixtures* of LCs this interval might be extended; for example, in the eutectic MBBA-EBBA LC phase exists within the range, acceptable for technical applications from -16 to $+55^\circ\text{C}$.

There are several theories to explain the existence of LC state, and the most common is the “statistical” theory. According to this theory, the most important factor that determines the likelihood of LC state is forces of intermolecular interaction,

quite essential in some substances. At the transition from the anisotropic crystal state to the LC state, molecules save their orientation (e.g., parallel to each other).

Typically, molecules of a substance, which have a tendency to form LCs, have anisometric form. In such cases, the contribution of dipole-dipole component of intermolecular interaction increases significantly, resulting in the additional gain in energy that stabilizes the LC state. The magnitude of this energy is largely characterized by the anisotropy of molecule polarizability that is characterized by $\Delta\alpha = \alpha_{\parallel} - \alpha_{\perp}$, where α_{\parallel} and α_{\perp} are polarizabilities in the directions parallel and perpendicular to the long axis of molecule.

Among a large variety of types of LCs, *NLCs* occupy a special place in terms of their practical significance. They differ from other types of LCs by the uniaxial ordering, in which molecules are arranged in such way that their long axes are parallel to each other.

In some substances, their molecules themselves might have no anisometric properties, but these molecules have the ability to *aggregate* with other molecules by the creation of partially *ordered structures*. This ability, called the *mesagenic*, is the main feature of some LC compounds. The study of relationship between molecules and their building is the topic of *mesagenic chemistry* that is the basis of LC science. One of most important parameters that characterizes mesagenic ability of any compound is the width of the temperature range of mesophase existence. It is determined that the more different the energies of intermolecular interaction are, directed along a long axis of molecule and perpendicular to it, the wider is the temperature range of mesophase.

For theoretical consideration of PT in the LCs Landau's theory of PTs is usually applied. As already mentioned, this theory is based on the assumption of analytical capabilities to describe the state by using a conception of *ordering parameter*, as well as the analytical description of free energy in the neighborhood of phase transition. Full theoretical description of PT requires experimental determination of ordering parameter and main thermodynamic properties of LC material.

The ordering parameter is not the same for all types of mesophase states. For example, to describe transitions from the nematic phase to the isotropic phase, the anisotropic part of diamagnetic susceptibility tensor as the ordering parameter can be used. However, to describe transition from the nematic phase to the smectic phase (A-phase), two quantities are usually used: the density of molecules in smectic layers and relative displacement of layers.

Experimental methods for determining order parameters are very complex and cumbersome. The measurement of thermodynamic parameters of LCs near their PTs in most cases indicates that these are first-order transitions (transitions, accompanied by a jump in heat capacity with latent heat calorification). Therefore experimental determination of transition temperature needs to use the differential thermal analysis (DTA) method, when heat capacity change is determined by experimental investigation of object while temperature changes.

Detection of PTs in the LC requires using the highly sensitive thermal detectors, as the released transition heat PT is very small (in order of magnitude less than

melting heat of other organic substances). For example, the transition “nematic phase \implies isotropic liquid” is accompanied by melting heat of no $>4\text{J/m}^3$, while the transition between “smectic A \implies smectic C” phases is characterized by the energy change of only 0.4J/m^3 . It is also worth noting that most studies should be provided only with very small amounts of LC substance.

Requirements of high sensitivity of detection and strict linearity in the change of temperature measuring unit should also take into account that a large group of LC compounds is characterized by *multiple transitions* in a narrow temperature range. It should be noted that near the PT all parameters of material are very sensitive to the presence of impurities in LC material.

10.6 LARGE PARAMETERS AND GIANT EFFECTS IN ELECTRONIC MATERIALS

During the last decades, several physical phenomena in electronic materials have been discovered; some of them relate to the large values of main parameters in materials that many times *exceed usual values*. In part, but not always, these anomalies are due to the proximity of the substance to PTs.

In some materials such electrophysical effects were found to be tens and hundreds of times stronger than previously known effects: researchers named these effects “*giant effects*.” A number of these unusual effects already found important applications in science and technology, allowing design technical devices with very high performance.

Large electrical and magnetic parameters [8]. As is well known, main electrodynamic parameters of material are electrical *conductivity* σ (characterizing charge transfer in electrical field), dielectric *permittivity* ε (describing polarization, or charge separation in electrical field), and magnetic permeability μ (determining magnetic induction in matter in magnetic field). Recently it was discovered that in some materials these parameters can take very large values that can find important applications in electronics.

Hyperconductivity can be seen in some conductors at low temperatures. It is different from the well-known superconductivity. In fact, conductivity σ is proportionality factor between electrical current and applied field: $j = \sigma E$. However, in the superconductors, $E = 0$; therefore σ might be *considered* as infinite quantity. In *hyperconductors* σ is a definite quantity but very large. Unlike superconductivity, which can be destroyed by a strong magnetic field (Meissner effect), hyperconductivity cannot be suppressed by a magnetic field.

At certain temperatures used in cryogenic technologies, namely, 77 K (liquid nitrogen boiling point), hyperconductor Be shows higher conductivity than any other metal (even Cu), and this large σ can be applied in cryogenics. More convenient for presentation is not σ but resistivity $\rho = 1/\sigma$ (Fig. 10.27): it can be seen that $\rho_{\text{Be}} < \rho_{\text{Cu}}$ at 77 K. In the same way, at cryogenic temperatures below 40 K the best hyperconductor is Al: $\rho_{\text{Al}} < \rho_{\text{Cu}}$.

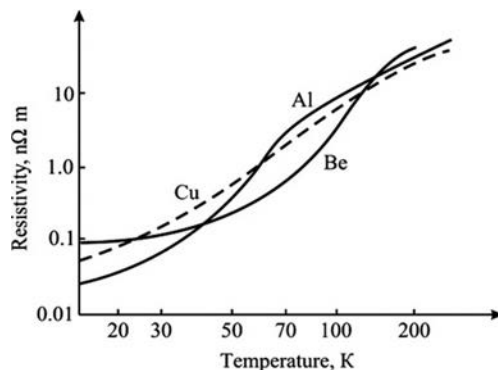


FIG. 10.27

Temperature dependences of resistivity of Cu, Be, and Al.

The reason for such behavior of beryllium is its largest *concentration* of free electrons among metals: $n_e \sim 2.5 \times 10^{23} \text{ cm}^{-3}$. However, at room temperature, resistivity of beryllium significantly surpasses resistivity of copper, which is explained by the peculiarities in electron scattering in crystal lattice. However, at low temperatures, this scattering significantly decreases, and the density of electrons becomes the prevailing factor. However, below 40 K the mechanism of electron scattering on lattice defects comes out on top, and again ρ_{Be} becomes larger than ρ_{Al} and ρ_{Cu} . At much lower temperatures, the aluminum becomes the best hyperconductor that ultimately is followed by its PT into the superconducting phase at lower temperatures.

Large dielectric permittivity is of great interest in electronics, as it allows decreasing the size of device components. In usual solid dielectrics $\epsilon = 2\text{--}10$; however, in the ferroelectrics a record value of *high-frequency* permittivity ϵ_{max} in the Curie point was previously set: in BaTiO_3 $\epsilon_{\text{max}} \approx 10,000$ and in SrTiO_3 $\epsilon_{\text{max}} \approx 40,000$ (Fig. 10.28). Yet, the industry needs such components that have *thermally stable*

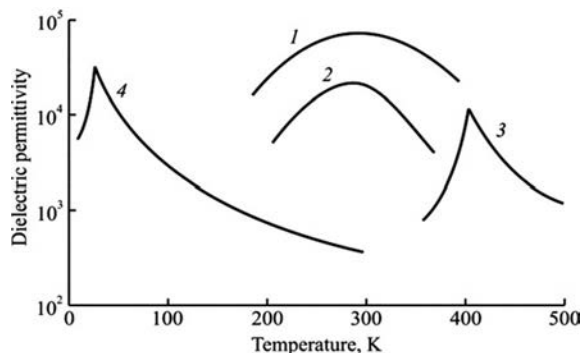


FIG. 10.28

Permittivity temperature dependences in PMN (1), $\text{Ba}(\text{Ti},\text{Sn})\text{O}_3$ (2), BaTiO_3 (3), and SrTiO_3 (4).

parameters in technical interval of about 200–400 K, as very high peak values ϵ_{\max} in ferroelectrics usually cannot be used in practice.

Therefore in high-capacity miniature capacitors the ferroelectrics with *diffused phase transition* are applied: they have $\epsilon \sim 10,000$ in the relevant temperature interval ($-60\dots+80^\circ\text{C}$). These relatively thermostable ferroelectric ceramics consist of many ferroelectric districts with a different PT temperature that widens $\epsilon(T)$ dependence, curve 2 in Fig. 10.28.

Recently the *relaxor ferroelectrics* are elaborated in which $\epsilon > 50,000$ (Fig. 10.29, curve 1). The physical reason for such large permittivity in relaxors is the creation of such a structure in which highly polarizable *nanoclusters* stably exist and contribute greatly to $\epsilon(T)$ over a wide temperature range (e.g., in compound $\text{PMN} = \text{Pb}(\text{Mg}, \text{Nb})\text{O}_3$). As was shown in Chapter 9, the *theoretical limit* for high-frequency permittivity in solids can be estimated as $\epsilon_{\max} \sim 3 \times 10^5$. It corresponds, firstly, to ϵ_{\max} in ferroelectric Curie point and, secondly, to instantaneous value of permittivity that can be achieved while switching of ferroelectric polarization (from hysteresis loop $\epsilon_{\max} \sim \partial P / \partial E$).

Large magnetic permeability is a very important parameter for application of soft magnetic materials. It should be noted that most materials in nature have very weak magnetics with $\mu \approx 1$: in diamagnetics $\mu \leq 1$ and in paramagnetics $\mu \geq 1$; correspondent magnetic inductance B dependences on magnetic field H are shown in Fig. 10.29A. Ferromagnetics demonstrate hysteresis in the $B(H)$ dependence where $\mu_{\max} \sim \partial B / \partial H$ is seen at coercive field.

High magnetic permeability is observed in the ferrites: $\mu_{\max} = 100\text{--}1000$ and in the ferromagnetic metals: in nickel $\mu_{\max} \sim 500$ and in iron $\mu_{\max} \sim 10,000$. The record value of $\mu_{\max} \sim 80,000$ was obtained previously in the FeNi alloy, and for a long time this material has been used in electrical engineering. Recently, using nanotechnologies, this value has been surpassed by >10 times. Physical mechanisms of such super high permeability are, firstly, *superparamagnetism* (Fig. 10.29A) and, secondly, technologically made metallic *spin glasses* (nanoscale glassy alloys exhibit excellent soft magnetic properties).

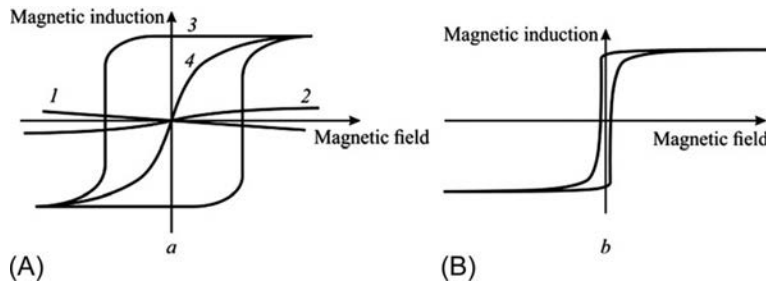


FIG. 10.29

Magnetic induction on field dependence in: (A) diamagnetic (1), paramagnetic (2), ferromagnetic (3); superparamagnetic (4); (B) Metglas.

At present, the maximal permeability of $\sim 1,000,000$ is obtained in Metglas that is an amorphous alloy produced by rapid solidification process. As shown in Fig. 10.29B the very narrow rectangular hysteresis loop provides giant $\mu_{\max} \sim \partial B / \partial H$. In the disordered glassy-like structures, electronic spins can be easily reoriented even in a small magnetic field.

Giant effects in electronic materials [8, 9] are not only interesting for the physics of materials but can be successfully used in technical devices.

Giant electrostriction. In the dielectric electrical polarization $P = \epsilon_0 \chi E$ always is accompanied by a reversible shift of connected electrical charges—electrostriction: $x = R E^2$, where x is strain, E is electrical field, and R is coefficient of electrostriction. This square dependence of strain on the electrical field in majority of dielectrics is very small ($x \sim 10^{-5}$ in a strong field). Therefore in most cases the electrostriction is negligible as compared with the linear piezoelectric effect: $x = d E$, where d is the piezoelectric module.

Nevertheless, electrostriction increases in the dielectrics that have large permittivity (such as paraelectrics and ferroelectrics), but it grows to an outstanding value in the relaxor ferroelectrics, for instance, in the PMN type: $(\text{Pb}, \text{Mg})\text{NbO}_3$ and its solid solutions (Fig. 10.30).

Electrostriction proportional to square of permittivity in the PMN-type crystal ϵ is large, so their deformation in electrical field can reach even 2%, which is much more than that in best piezoelectrics. In contrast to the piezoelectric effect, in strong electrical fields giant electrostriction shows *no hysteresis*, which is one important advantage of active dielectric application in microactuators.

Besides ceramics and crystals, exceptionally large electrostrictive response (up to 4%) is observed in the electron-irradiated polyvinylidene fluoride-trifluoroethylene copolymer.

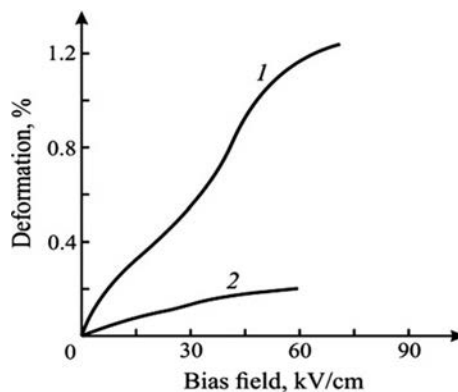


FIG. 10.30

Electrically controlled deformation: 1—electrostriction in crystal PZN-4.5% PT; 2—piezoelectric effect in ceramics PZT-8.

Giant magnetostriction is used for application in sound generators, for vibration suppression, for direct and indirect control of motions, etc. This effect produces a change in the shape and size of magnetic sample when it is put in the magnetic field. Previously, magnetostriction was considered as a very weak effect (induced strain is only 0.003% in usual magnetic).

However, in the rare earth metals (terbium, dysprosium, and some of their alloys) the *giant* magnetostriction effect is discovered at a point where strain is greater in two orders of magnitude (Fig. 10.31). The reason for this is that the electronic clouds of atoms that form lattice of such crystals are highly elongated, and they behave as non-deformable. Under the influence of external magnetic field, electronic clouds of such atoms forcefully rotate and push adjacent atoms, thus strongly deforming entire lattice. The largest magnetostriction effect is seen near the morphotropic boundaries between different phases. The alloy TbFe_2 (with 2.6% strain) is now considered as the best magnetostrictive material in modern engineering.

Giant magnetocaloric effect is manifested as essential cooling (or heating) of magnetic material in the external magnetic field. Magnetic cooling has an extensive area of possible applications at normal conditions with good prospects in production of refrigerators. The point is that magnetic refrigerators can work at room temperature; these devices use magnetic materials with low Curie temperature (like gadolinium) or various alloys of rare earth elements. The operating temperature range is convenient to apply the magnetocaloric effect in devices such as home refrigerator, air conditioner, as well as in various devices for cooling. Compared to traditional gas-compression refrigeration, magnetic refrigeration is safer, more compact, has a higher cooling efficiency, and it is more environmentally friendly, because it does not use harmful ozone-depleting coolant gases.

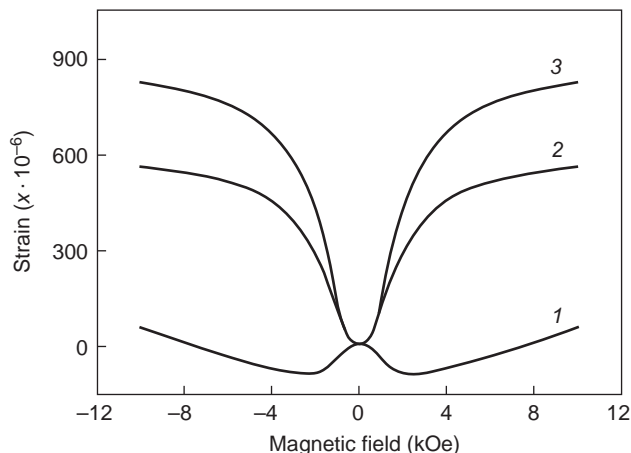


FIG. 10.31

Magnetostriction in $(\text{Tb,Dy})\text{Co}_2$: 1—80% Dy, 2—60% Dy, and 3—70% Dy.

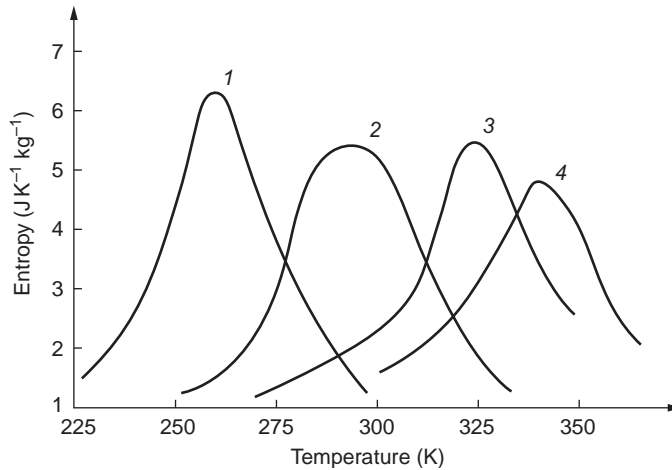


FIG. 10.32

Temperature dependence of entropy in $\text{La}_{0.8}\text{Na}_{0.2-x}\text{Me}_x\text{MnO}_3$ for x : 1—0.15; 2—0.10; 3—0.05; and 4—0.00.

Giant magnetocaloric effect is typical for intermetallic compounds based on the rare earth elements, such as silicide-germanide $\text{Gd}_5(\text{Ge},\text{Si})_4$. These materials provide promising applications in magnetocaloric cooling. Yet there are other prospective materials: Fig. 10.32 shows giant magnetocaloric effect in the compound $\text{La}_{0.8}\text{Na}_{0.2-x}\text{Me}_x\text{MnO}_3$ ($0 \leq x \leq 0.15$). In this composition, a second-order magnetic PT from ferromagnetic to nonmagnetic state occurs at Curie temperature (T_c), wherein temperature T_c can be controlled in the range from 335 to 260K when the sodium deficiency rate is increased. The ability to tune temperature transition close to room temperature is possible by changing sodium-deficiency content as well.

Giant thermistor effects in critistors and posistors. Fig. 10.33 demonstrates different types of thermistors that use $R(T)$ dependence. In metals, this effect is positive and small, curve 1; in semiconductors the thermistor effect is much larger and negative, curve 2.

The *critistors* use very temperature-sensitive material near dielectric-metal phase transition (curve 3 obtained for vanadium dioxide). Composition based on VO_2 below 60°C is close to dielectric, but above 70°C it exhibits metallic behavior. In a rather narrow temperature range, the resistance of critistor falls thousands of times. At that, the critical temperature of critistor can be changed by selecting its chemical composition.

The opposite but also critical $R(T)$ dependence is seen as the *posistor effect* in ferroelectric semiconductors (curve 4 in Fig. 10.33) [1]. Ferroelectric ceramics of BaTiO_3 type, owing to intense doping, acquires low resistance. With temperature increase near the Curie point T_C , this material shows a sharp increase in resistance

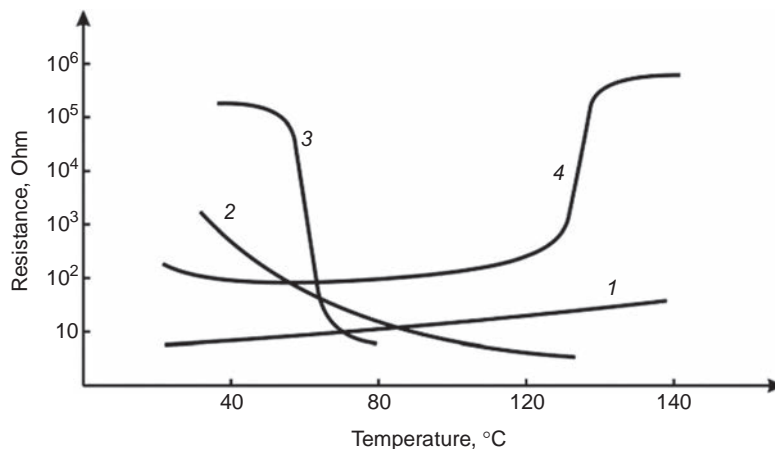


FIG. 10.33

Resistance dependence: 1—platinum wire; 2—transition-metal oxides thermistor; 3—VO₂-based critistor; and 4—BaTiO₃-based posistor.

(10²–10⁴ times). This effect is described in detail in Section 10.4. The technical application of posistor ceramic elements with low “cold” resistance and high “hot” resistance is very wide.

Field-controllable switching elements [9]. Electrical and magnetic fields may act as the controlling factors on resistivity, which changes into tens and hundreds of times.

From the beginning, a possibility of resistivity control by means of *electrical* field is considered. A decrease in resistance in thousand times corresponds to the PT of “insulator-metal” type, for example, in the VO₂, as shown in Fig. 10.33 curve 3. Below temperature ~60°C, vanadium dioxide is wide-gap semiconductor (almost dielectric), but above 65°C it exhibits metallic behavior: energy bandgap in VO₂ almost disappears. As a result, in a narrow temperature interval, resistance of VO₂ decreases rapidly, as shown in Fig. 10.35A, curve “*E* ≈ 0.”

However, if external *electrical field* is applied to VO₂ (usually thin film is used with interdigital electrodes), then the transition from “dielectric” phase into “metallic” phase will start at a much lower temperature, for example at ~45°C, and under controlled field, resistance decreases by thousand times, as seen by vertical line 1 → 2 in Fig. 10.34A. The displacement of the PT in electric field is the physical basis for vanadium oxide application as the *switching elements*: externally applied electrical field promotes charge carrier liberation from their bounded state; thus fast switching takes place at the conducting state. Electrically controlled metallic-type *optical reflection* from VO₂ film is used in electrically controlled optical devices.

Further, the possibility of controlling resistivity by means of a *magnetic field* is considered. One example of effective *magnetic control* of resistance can be seen in the perovskite structure manganites (Fig. 10.34B). These manganites are an

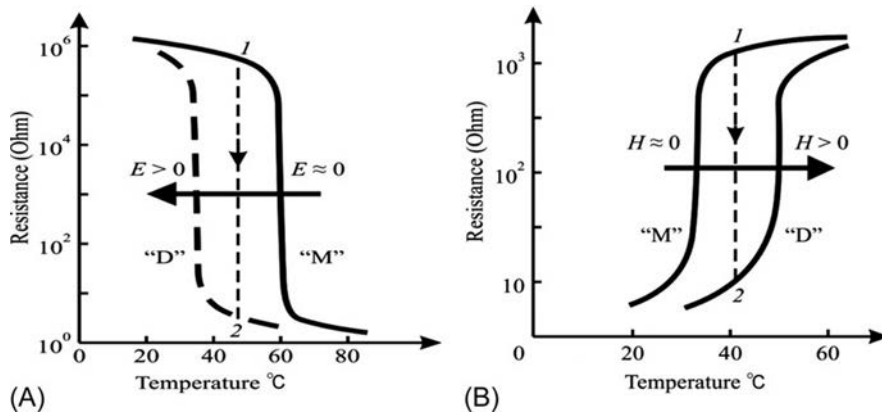


FIG. 10.34

Resistance control by external fields: (A) electrical field manages R in VO_2 ; (B) magnetic field manages R in $(\text{La}_{1-x}\text{Ca}_x)\text{MnO}_3$ (dotted line $1 \rightarrow 2$ shows large fall in resistance).

important family of oxides having *colossal magnetoresistance*. Of particular interest are the compounds of $(\text{La}_{1-x}\text{Ca}_x)\text{MnO}_3$ type, where concentration of Ca can vary over a wide range: $0 \leq x \leq 1$, wherein physical properties of manganite change dramatically. This system undergoes a series of PT with different types of ordering: magnetic, structural, electronic.

In contrast to the previous case of VO_2 electrical resistance control, in case of $(\text{La}_{1-x}\text{Ca}_x)\text{MnO}_3$ the *low-temperature* phase is *highly conductive* (“quasimetallic”) with the ferromagnetic ordering but the *high-temperature* phase becomes a wide-band semiconductor (*nearly dielectric*).

The mechanism of magnetic ordering is *double exchange*: $\text{Mn} \Leftrightarrow \text{O} \Leftrightarrow \text{Mn}$ (double electronic transfer via intermediate oxygen ion). Parallel ordering of spins in the “M”-phase corresponds to low resistance, Fig. 10.34B, *left part*, $H=0$ (low resistance appears also at external magnetization). When temperature rises, growing thermal movement in manganite lattice destroys the double-exchange mechanism, and the ferromagnetic-type (conducting) phase turns into the nonmagnetic (nearly dielectric) phase; correspondingly, the sharp increase in resistance occurs.

However, at increased temperature magnetic ordering in manganite can be forcibly *returned* by the external *magnetic field* applying (see Fig. 10.34B, *curve* $H > 0$). At that, the resistance falls down 100 times, as shown by the dotted line $1 \rightarrow 2$. Induced by magnetic field “insulator-conductor” transition is otherwise called *colossal magnetoresistance*. Grate change in resistivity under controlled magnetic field (Fig. 10.35A) can be used in many electronic devices.

Thus both electrical and magnetic fields may shift transition temperature of the “insulator-conductor” PTs, leading to enormous changing of resistivity. Therefore, high-conductivity phase may be seen at higher temperature and low-conductivity phase at lower temperature.

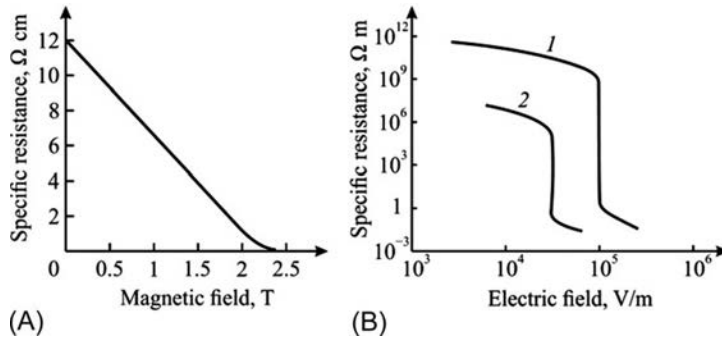


FIG. 10.35

External field control resistivity: (A) magnetic field influence at $\text{La}_{0.67}\text{Ca}_{0.33}\text{MnO}_x$ near the Curie point, (B) electrical field influence at ZnO varistor (1) and SiC varistor (2).

However, a large change in resistivity does not obviously correspond to the PT, as it was in the case of VO_2 critistors or in BaTiO_3 posistors or in $(\text{La}_{1-x}, \text{Ca}_x)\text{MnO}_3$ magnetoresistor. An interesting example of resistance change by 1000 times by means of the *electrical field* application is shown in Fig. 10.36B. A Significant jump in resistance (millions of times) is observed in the ceramic dielectric-semiconductor elements—the *varistors*, for example, in the polycrystalline carborundum (SiC) or in zinc oxide (ZnO).

Similar to low levels, at high levels of electrical field, varistors are characterized by the *linear* (ohmic) behavior: $j \sim E$. In comparatively small electrical field, varistor has high resistance (close to dielectrics), but in a strong electrical field the resistance of varistor sharply *decreases*, and it becomes a typical semiconductor. Also, no sign of electrical breakdown in this material can be seen: a large increase in current is naturally restricted and cannot destroy varistor. Moreover, the characteristic $j(E)$

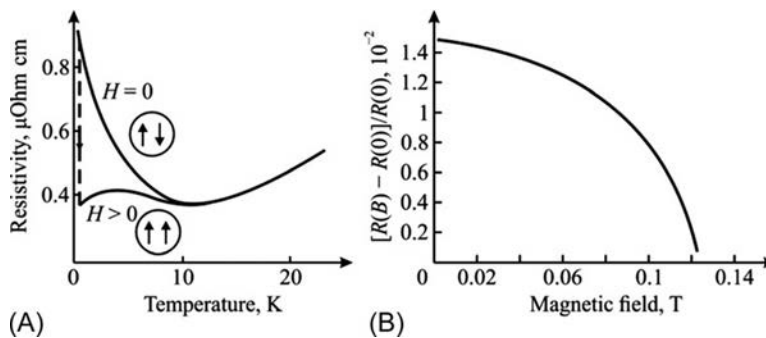


FIG. 10.36

Magnetic field manages resistivity: (A) suppressing of the Kondo effect (at $H=0$) by external magnetic field ($H>0$); (B) magnetoresistance in Fe-Cr multilayered structure.

in varistors is reversible; increased current passing through varistor remains stable. Therefore the phenomenon of varistors is quite different from conventional electrical breakdown.

The features of carborundum varistors and zinc oxide varistors are similar, but for zinc oxide the change in resistance is much more. Zinc oxide shows some other advantages, associated with its broadband gap, a consequence of which is the high breakdown voltage and the utility in high-power operations. For these reasons, ZnO can be used as components in many electronic devices: in the piezoelectric nanogenerators, luminescent materials, light-emitting diodes, lasers, as well as in various sensors, solar elements, etc.

Giant magnetoresistance is the effect observed in thin-film structures, composed of alternating layers of high conductive **ferromagnetics** and low conductive antiferromagnetics. External magnetic field strongly decreases the resistance of such layers.

It is necessary to note that *positive* magnetoresistance (Gauss effect) is always observed in metals and semiconductors, because it is conditioned by the twisting of electron trajectory in the magnetic field. But the *negative* magnetoresistance is a feature of ferromagnetics only: ordering of domains reduces the possibility of scattering of conduction electrons on domain walls.

However, the *giant effect* (being also *negative*) far surpasses in size of all mentioned effects. To explain the giant magnetoresistance effect, it is better to start by considering the Kondo effect. This effect is seen in nonferrous metals, containing impurities of magnetic ions. The scattering of conduction electrons by magnetic impurities becomes noticeable only at very low temperatures, resulting in the uncommon *increase* in **electrical resistivity** when temperature *decreases* *curve* $H=0$; at low temperatures electron scattering on phonons is no longer significant). When temperature tends to zero, magnetic moments of impurities can capture some of free electrons, creating the *opposite-spin* clusters (whose structure is similar to that of antiferromagnetics). Effective mass of moving through these clusters of electrons largely increases, so the resistivity increases as well *curve* $H=0$).

Applied magnetic field changes the *opposite* orientation of coupled spins into their *parallel* orientation that partially suppresses the Kondo effect (Fig. 10.36A, *curve* $H>0$). At that, the *negative* magnetoresistance appears (shown in Fig. 10.36A) by the vertical dashed line. Therefore, at low temperatures, when background noise from phonons does not hinder observing this effect in a pure form, the Kondo effect testifies that *opposite* orientation of adjacent spins promotes the increase in resistivity, while external magnetic field gives rise to negative magnetoresistance by means of magnetically induced parallel orientation of spins. To realize the giant magnetoresistance effect in normal circumstances, several methods are developed (see Chapter 6); see Fig. 10.36B for an example. In the magnetic field, nanosized mixture of paramagnetic material (Cr) ferromagnetic (Fe) changes its resistivity, because magnetic field orients spins in parallel and in this way enhances conductivity.

10.7 SUMMARY

1. Degrees of freedom of atomic particles in solids can be divided into two groups. For some degrees of freedom the interaction energy of particles U_{int} is small in comparison with thermal motion energy $k_B T$. When $U_{\text{int}} \ll k_B T$ then appropriate degrees of freedom behave as the collection of particles, that is, as “almost ideal gas,” and applicability to use model of *quasiparticles* is justified. Conversely, when $U_{\text{int}} \gg k_B T$, then appropriate degrees of freedom are usually ordered, but this movement, too, can be described by introduction of quasiparticles.
2. If interaction energy $U_{\text{int}} \sim k_B T$, then theoretical description of solids becomes complicated: this case usually corresponds to the phenomenon of *phase transitions (PT)*. At specified temperature, in almost all substances the change of physical properties is not gradual but abrupt: this spasmodic change is the phase transition. The “liquid-vapor” (vaporization) phase transition is a typical example, and another example is the “liquid-crystal” (crystallization) transition. Both transitions refer to first order (PT-I), in which phases before and after the transition point *differ significantly* from each other. One phase replaces another phase just because it is more favorable energetically. To make this change happen, the potential barrier, separating these phases, should be overpass. Therefore, in the neighborhood of first-order phase transition overcooling (or overheating) is expected.
3. In condensed state physics, mainly PTs *within same physical state* are investigated (in solid crystals or liquid crystals (LCs)). Of particular interest are PTs at which a *new property* appears in crystal, for example spontaneous electrical moment in case of transition from the paraelectric to ferroelectric phase. This type of transition is related to PT-II: at Curie temperature ($T = T_C$) one phase ceases to exist, but it is replaced by another phase. At a point of such transition both phases cannot be distinguished, but when system moves away from this point, the difference between properties of phases gradually increases.
4. Near the point of *second-order phase transition* crystal behaves in such a way that conventional concept, based on quasiparticles, cannot adequately describe experimental situation. Normally, exactly closest neighboring particles in a crystal are considered as the *strongly interacting* particles, while the interaction of distant particles might be neglected. However, near the phase transition, in contrast, the interaction of neighboring particles *compensates one another*, and, on this background, the interaction of those particles, which are *at a distance* from one another, appears dominant. This interaction has a special character: the probability of *collective movements* is larger than the probability of *individual movements*. Abnormal increased role of collective movements is confirmed by experiments: at temperature $T = T_C$ crystal shows maximum of specific heat, permittivity in ferroelectric tends to infinity as well as

permeability in ferromagnetic (in superconductors conductivity actually becomes infinite), and so on.

5. PTs in solids do not obviously correspond precisely to second-order type of transitions (PT-II), but sometimes in some respects they can approach phase transition of first order (PT-I). Energy in the vicinity of PT-II varies continuously, so any temperature hysteresis should be absent; at that, *derivatives* of energy functions are changed abruptly. However, in case of PT-I, just the basic energy functions of crystal show a stepped changing, and in the vicinity of a transition temperature hysteresis is observed. These transitions in solids include the PTs of insulator-metal type, when small conductivity of dielectric becomes large (almost of metallic type).
6. *Microscopic mechanisms* of PTs are conditioned by various electrical, mechanical, thermal, and other interactions in crystals. PTs are very often caused by the interaction of electronic and phonon subsystems. Therefore such transition occurs with a significant change as in phonon spectrum, so also in electron energy spectrum.
7. *Macroscopic theory* of PTs is based on thermodynamic potential expansion on the powers of ordering parameter. As the ordering parameter, the most important characteristic of crystal is selected—which changes at transition. For example, in most of ferroelectrics the ordering parameter is spontaneous polarization, in ferroelastics it is mechanical deformation, in ferromagnetics it is spontaneous magnetization, etc.
8. In ferroelectrics of *order-disorder-type* behavior of a crystal shows large microwave absorbance as in the ferroelectric so in the paraelectric phase. With regard to practical applications, some interest might have strong anisotropy of microwave absorbance that is possible to use in millimeter and submillimeter wave polarizers.
9. Ferroelectrics of *displacement type* above their phase transition practically have no dielectric dispersion (in paraelectric phase); correspondingly, their losses above the Curie point are small that might have application in the microwave tunable devices. Below the Curie point microwave absorbance of ferroelectrics increases owing to domain wall relaxation (losses in single domain crystals in displacement-type ferroelectrics remain small).
10. Ferroelectrics with *partially disordered structure* exhibit the degraded (diffused) phase transition that is characterized by extended $\epsilon(T)$ maximum. With few exceptions, ferroelectrics with diffused phase transition are common solid solutions, characterized by nonuniform allocation of structural ions. An exceptional case of diffused phase transition crystals is the *relaxor ferroelectrics* whose distinctive feature is extraordinary “softening” in their dielectric, elastic, and optical properties over a wide range of temperatures.

11. In *relaxor ferroelectrics* electrically induced piezoelectric effect is challenging for many electromechanically controlled devices. Based on relaxors, electrostrictive actuators are used in precision positioners, miniature ultrasonic motors, adaptive mechanical dampers, etc. Promising fields of application of nonhysteresis electrostrictive actuators are mechanically tunable microwave phase shifters and dielectric filters. An important attribute of most mentioned electronic devices is their high-low inertia of operation.
12. Relaxor ferroelectrics have *large ϵ* in a wide temperature interval; this property is very important for applications in many electronic devices (micromechanics and microdrivers). Displacement, produced by actuator, is determined by value of relaxor ferroelectric permittivity. To estimate the limitations in the operation rate of relaxor-based electronic devices, frequency dispersion of relaxor ϵ should be studied in a wide frequency range. It is assumed that dominating contribution to relaxor ferroelectric permittivity is electromechanical coupling of polar clusters. Therefore the quickness of response in relaxor-based devices is determined by sound velocity of relaxor ferroelectric.
13. In addition to solid active dielectrics, the *liquid-crystal dielectrics* are widely used in electronics. In this case the “transitional” state between crystalline (anisotropic) and liquid (isotropic) substance is used. Liquid crystal (LC) state is a thermodynamically stable physical state, in which a matter maintains anisotropy of physical properties (typical for solid crystals) and can flow (that is characteristic of liquids). This unusual combination of properties makes a widespread application of LCs in modern electronics and information technology. The fact is that LC is in a permanent state of phase transition that is “blurred” in certain temperature range, making the LC is very sensitive to controlling fields.
14. Intermediate LC phase (between normal crystal and ordinary liquid) is often called as the *mesophase*, and cardinal types of molecule ordering are the *nematic* ordering (“filamentary”), the *smectic* (“polygonal”), and the *cholesteric* (“screw”) ordering. In the neighborhood of phase transition, low-intensity external influences can greatly change electrical, mechanical, and optical properties of LCs.
15. During the last decades, many physical phenomena in electronics materials have been discovered; some of them relate to the large values of main parameters in materials that many times *exceed usual values*. In part, but not always, these anomalies are due to the proximity of the substance to the PT. Conductivity, permittivity, and permeability in some materials can hundreds of times exceed the parameters of conventional electronic materials.
16. In some materials such electrophysical effects are found that are tens and hundred times more strongly manifested than previously known effects:

researchers named these effects “*giant effects*.” They are giant electrostriction, magnetostriction, magnetoresistance, magnetocaloric effect, critistor and posistor effects, as well as controllable by electrical and magnetic field switching properties of electronic elements.

REFERENCES

- [1] R.E. Newnham, *Properties of Materials: Anisotropy, Symmetry, Structure*, Oxford University Press, Oxford, 2004.
- [2] G.A. Smolensky, V.A. Bokov, V.A. Isupov, N.N. Krainik, R.E. Pasinkov, M.S. Shur, *Ferroelectrics and Antiferroelectrics*, Nauka, Leningrad, 1971.
- [3] B.A. Strukov, A.P. Levanyuk, *Physical Fundamentals of Ferroelectric Phenomena*, Nauka, Moscow, 1983.
- [4] M.E. Lines, A.M. Glass, *Principles and Application of Ferroelectrics and Related Materials*, Clarendon, Oxford, 1977.
- [5] Y.M. Poplavko, *Physics of Dielectrics*, Vischa Shkola, Kiev, 1980.
- [6] I.S. Rez, Y.M. Poplavko, *Dielectrics, Main Properties and Electronic Applications*, Radio i svyaz, Moscow, 1989.
- [7] Y.M. Poplavko, *Dielectric Spectroscopy of Solids, Basic Theory and Method Application*, LAMBERT Academic Publishing, Saarbruken, 2013.
- [8] Y.M. Poplavko, Y.I. Yakimenko, Giant Effects in Electronic Materials, Intern. Conference Electronics and nanotechnology, ELNANO, 2017, pp. 16–20.
- [9] Y.M. Poplavko, A.V. Borisov, In: Polarization influence on conductivity, Intern. Conference Electronics and nanotechnology, ELNANO, IEEE, 2017, pp. 22–26.

Index

Note: Page numbers followed by *f* indicate figures and *t* indicate tables.

A

- Absolute temperature, 97
- Absorption edge, 459
- A^{III}B^V compounds, 411–412
- A^{III}B^{VI} compounds, 412
- A^{IV}B^{VI} compounds, 412
- A^{III}B^V semiconductors, 426–427, 426*f*
- Acoustoelectronics, 88, 510
- Acousto-optics, 510
- Actinoelectricity, 587
- Active dielectric applications, 567
- Adiabatic approximation, 121–122, 415
- Adiabatic demagnetization, 258
- Allotropic elements, 36
- Allowed energy levels, 419–420
- α - γ phase transition, 193
- Amorphous materials, 52
- Amorphous semiconductor, 412–413
- Anharmonicity coefficient, 102
- Anions, 121–122
- Anisotropic crystals, 91
- Anisotropic magnetoresistance, 275
- Anisotropic magnetostriction, 256
- Anisotropy, 102
 - of magnetic properties, 254
- Anomalous Hall's effect, 470
- Anoporesis, 362
- Antiferroelectrics, 575, 593, 615
- Antiferromagnetic interaction, 259
- Antiferromagnetic Neel point, 279
- Antiferromagnetics, 144, 144*f*, 636
- Antimony sulfide (SbSI), 567
- Antisymmetric wave function, 245
- Artificial pyroelectric coefficient, 541
- Atomic bonding in metals, semiconductors, and dielectrics
 - attraction, 1
 - chemical bonds, 1
 - common wave function, 2
 - coordination number, 4
 - covalent bond, 7
 - diamond, 7
 - dielectrics and semiconductors, 4–5, 5*f*
 - donor-acceptor bond, 8
 - electrical charge distribution, 3, 4*f*
 - electronic orbitals, 2, 3*f*
 - energy, 2
 - heteropolar bond, 8
 - homeopolar bond, 8
 - hybrid ionic-covalent bond, 9
 - hydrogen bond, 12
 - impenetrability, 5
 - ionic bond, 5–6
 - ionic residues, 2
 - length, 2
 - low ionization energy, 2
 - metal bond, 4
 - molecular bonds, 11
 - nature of attraction forces, 5
 - nearest neighboring atoms, 7
 - nonpolar (simple) covalent bond, 7
 - parameter of lattice, 4–5
 - π -bond, 8
 - polar covalent bond, 8
 - polar crystals, 10
 - polarizability, 7
 - repulsion, 1
 - σ -bond, 8
 - spherical symmetry, 6
 - valence electrons, 2
 - valence shells, 2
 - various models of atomic bonds, 2, 3*f*
- Atomic crystals, 40–41
- Atomic scale, 122–123
- Autoelectronic emission, 167
- Axial vector, 224–225

B

- Backward-wave media, 60
- Ballistic, 494
- BaLn₂Ti₄O₁₂ (BLTs), 564–565
- Band engineering, 498
- Band theory, 163
 - and heavy fermions, 196
- Barium titanate (BaTiO₃), 311
- Binary compounds, 411
- Bipolar injections, 376, 380–381
- Bipolar thermal conductivity, 446–447
- “Black-body” radiation, 129
- Bloch wall, 250, 251*f*

Bloch wave, 416–417, 419–420
 Bohr's magneton, 141
 Boltzmann constant, 107
 Boltzmann distribution, 321
 Bose-Einstein condensation, 144
 Bose-Einstein distribution, 125
 Bose-Einstein statistics, 128
 Bosons, 123, 125
 Bouguer-Lambert law, 452
 Bravais cells, 18, 19*t*
 Bravais lattices, 17
 Brillouin zone, 24–25, 423
 Bulk modulus of elasticity, 80–81
 Butterfly wings-type hysteresis, 612*f*, 613

C

Cadmium selenide (CdSe), 494
 Carbine, 54*f*, 55
 Carbon nanotubes, 55, 56*f*, 232
 Carriers, 168
 Cataphoresis, 362
 Cations, 121–122
 Cauchy relations, 81
 Ceramics, 529, 563–564
 Cesium chloride, 30
 Charge carrier diffusion, 42
 Charge carrier mobility, 171
 Charge carrier scattering, 442
 Charge transfer description, 170
 Chemical compound semiconductors, 410
 Cholesteric liquid crystals (ChLCs), 645*f*, 646
 Classes of symmetry, 17
 Classic electronic theory of metals, 180
 Clausius-Mosotti-Lorentz equation
 dielectric permittivity of gas
 nonpolar gases, 333
 polar gases, 333
 local (acting) electrical field in dielectric
 ionic crystal, 332–333
 nonpolar dielectrics, 331–332
 Lorentz model for local field calculation
 continuous medium, 333–334
 macroscopic field, 334
 permittivity calculation, 335
 physically infinitesimal volume, 334
 polar molecule, 335
 solid isotropic dielectrics, 335
 temperature parameter, 336
 Cochran law, 561–562
 Coefficient of transparency, 496
 Coerced zone of wave vectors, 418
 Collective motions, 122–123
 Colossal magnetoresistance, 277–278, 656
 Common wave function, 2
 Composites, 56–62
 Composite superlattices, 498
 Compressibility, 80
 Compression effect, 472–474
 Conduction band, 419–420
 Conduction quantum, 495–496
 Coordination number (CN), 4
 Coordination polyhedron, 26
 Corundum (Al₂O₃), 31
 Cotton-Mouton effect, 267
 Coulomb law, 317–318
 Coupling coefficient, 533, 535
 Covalent bond, 7
 Covalent crystals, 64
 Cryoconductivity, 176
 Crystal heat capacity
 anharmonicity, 106
 arbitrary waves, 106
 Born's theory of lattice dynamics, 106
 Debye's model of specific heat, 110
 Debye's quantum theory of heat capacity, 106
 Einstein's quantum theory of heat capacity, 106, 109
 law of heat capacity constancy, 106
 law of specific heat constancy, 106
 mass specific heat, 104–105
 specific heat, 104–105
 volumetric specific heat, 104–105
 Crystallography, 13
 Crystals
 lattice defects in
 defect formation, 37
 1D defects-dislocations, 43
 2D and 3D defects, 44
 zero-dimensional (point) defects, 38–39
 structural peculiarities of
 basic structures of dielectrics, 30
 basic structures of semiconductors, 29
 coordination polyhedron, 26
 covalent crystals, 27
 ionic crystals, 27
 solid solutions, 35
 typical structures of metals, 28
 symmetry of
 Bravais cells, 18, 19*t*
 Bravais lattices, 17
 classes of symmetry, 17
 crystallography, 13
 crystal syngonies, 20, 20*t*
 Curie principle, 22

- elementary parallelogram, 17
 - enantiomorphism, 16
 - highest category of symmetry, 20
 - highly symmetric crystals, 14
 - inversion axes, 16
 - lattice translation vectors, 13
 - limiting symmetry group, 14, 15*f*
 - linear periodic structure, 18–19
 - lowest category of symmetry, 20
 - middle category of symmetry, 20
 - mirror-rotary axis, 14
 - physical and crystallographic installations, 22
 - vs. physical properties, 21
 - point symmetry operations, 13
 - polar axis, 14–16
 - reciprocal lattice, 24
 - rotation symmetry axis, 14
 - symmetry element, 16–17
 - symmetry elements and operations, 13
 - symmetry plane, 14
 - Curie constant, 233
 - Curie-Weiss law, 561–562
 - Current density, 168
 - Cyclotron effective mass, 479–480
 - Cyclotron resonance, 479
- D**
- Damped waves, 89*f*, 90
 - De Broglie wavelength, 488
 - Debye model, 320–322
 - Debye's model of specific heat, 110
 - Defect formation
 - Frenkel mechanism, 38
 - linear defects, 37
 - planar defects, 37
 - point defects, 37
 - Schottky defects, 38
 - screw dislocation, 37–38, 37*f*
 - structurally insensitive properties, 38
 - structurally sensitive properties, 38
 - thermal fluctuations, 38
 - volume defects, 37
 - Deformation type-induced polarization, 290
 - Dember effect, 455–456
 - Density of elastic energy, 80
 - Density of thermal energy, 168
 - Diamagnetics, 279
 - Diamagnetic susceptibility, 635*f*, 636
 - Diamagnetism, 226, 229
 - Diamond, 53–54
 - Diamond-type semiconductors, 410–411
 - Dielectric containing defects, 378
 - Dielectric crystals, 198
 - Dielectric losses
 - classification of, 349, 350*f*
 - conductivity, 353–354, 353*f*
 - loss tangent, 351–352, 352*f*
 - magnitude of, 349
 - quasielastic polarization, 356–358, 357*f*
 - relaxation polarization
 - Debye equations, 351*f*, 354
 - frequency dependence, 354
 - power density, 354–355
 - reactive conductivity, 354–355, 355*f*
 - temperature dependences of, 355–356, 356*f*
 - Dielectric-metal electrode, 366
 - Dielectrics, 114
 - Clausius-Mosotti-Lorentz equation, 331–336
 - dielectric losses, 349–360
 - electrical polarization (*see* Electrical polarization)
 - electrical properties, 288
 - electroconductivity in, 360–383
 - mechanical properties, 287–288
 - spectroscopy, 358
 - thermally activated polarizations, 319–330
 - thermal properties, 287–288
 - transport phenomena, 288–289
 - Dielectrics and semiconductors, bonding in, 4–5, 5*f*
 - Dielectric spectroscopy
 - dielectric spectrum, 359–360
 - dispersion frequency, 360
 - Kramers-Kronig relations, 358–359
 - Differential nonlinearity parameter, 296
 - Differential permittivity, 612*f*, 613
 - Differential thermal analysis (DTA)
 - method, 648
 - Diffused phase transition, 651
 - Diffusion, 41
 - Diffusion coefficient, 42–43
 - Diffusion transfer, 121
 - Dipole thermal polarization mechanism
 - Boltzmann distribution, 321
 - Debye model, 320–322
 - electrical field, 319–320
 - Langevin function, 322
 - nonlinearity of, 322–323
 - polar crystals, 320
 - relaxation time of, 320
 - Dipole-type structural polar motive, 588, 588*f*
 - Direct intrinsic absorption, 457–460
 - Direct lattice, 24
 - Discreteness of energy spectrum, 486

- Disordered magnetics
 - diamagnetism, 229
 - Lanzheven-Curie paramagnetism, 234
 - nuclear magnetism, 240
 - paramagnetism, 232
 - Pauli paramagnetism in metals, 239
 - spin of electron, 239
 - temperature dependence of magnetic susceptibility, 238
 - Dispersion forces, 11
 - Displacement-type paraelectrics, 561
 - Displacement type PT, 607*f*, 609–610
 - Donor-acceptor bond, 8, 411–412
 - Doped semiconductors, 471. *See also* Extrinsic semiconductors
 - Doped superlattice, 498–499, 499*f*
 - Double refraction, 483
 - Drift velocity, 168
 - Drude-Lorentz equation, 293
 - Drude's model, 181
 - Dulong-Petit law, 109, 117–118
 - Durability, 71
 - Dynamic computer memory (DRAM) devices, 349
 - Dynamics of electrical polarization
 - electrical conductivity, 337, 337*f*
 - far-IR (ionic lattice) polarization
 - displacement-type paraelectrics, 347–348, 347*f*
 - electromagnetic waves, 346
 - electronic and ionic contributions, 345, 345*f*
 - integrated dielectric films, 349, 349*t*
 - ionic crystal permittivity, 343, 344*f*
 - longitudinal oscillations, 343
 - mechanisms of losses, 350, 351*f*
 - one-dimensional crystal, 343
 - oscillator model, 342
 - permittivity of, 344–345
 - quasielastic displacements of ions, 342
 - quasielastic lattice polarization, 343
 - semiconductors of $A^{III}B^V$ and $A^{II}B^{VI}$, 348
 - sensitive indicators, 350
 - temperature dependence of permittivity, 346–347, 347*f*
 - fast and slow polarization processes, 336–337
 - geometrical capacity, 336–337
 - optical polarization, 340, 341*f*
 - relaxation polarization
 - complex value, 339–340
 - Debye dispersion formula, 338
 - equilibrium state, 338–339
 - reactive current, 339
 - relaxation character, 338
 - surface charge density, 338
 - saturation current, 337
 - Dynamic theory of Born, 112
- ## E
- Edge dislocation, 43
 - Effect of magnetoresistance, 173
 - Eight electrothermal effects, 514–515
 - Einstein's quantum theory of heat capacity, 106, 109
 - Elastic energy, 534
 - Elasticity, 71
 - Elastic stiffness and elastic compliance
 - bulk modulus of elasticity, 80–81
 - Cauchy relations, 81
 - compressibility, 80
 - density of elastic energy, 80
 - Hooke's law, 78
 - Lame parameters, 82
 - Poisson's ratio, 81
 - shear modulus, 81
 - Young's modulus, 81
 - Elastic waves in crystals
 - acoustoelectronics, 88
 - amplification, 91
 - anisotropic crystals, 91
 - bulk piezoelectric elements, 85, 85*f*
 - damped waves, 89*f*, 90
 - discrete atomic chain, 84
 - dispersion relation, 84
 - elastic continuum, 83
 - elastic force, 83
 - elastic piezoelements, 87, 88*f*
 - longitudinal waves, 64
 - Love waves, 89*f*, 90
 - MEMS, 87
 - nondamping waves, 90
 - piezoconsole, 87–88, 88*f*
 - piezoelectric motor, 86*f*, 87
 - piezoelectric transformer, 85–86, 85*f*
 - piezoelectronics, 86
 - Rayleigh waves, 89, 89*f*
 - SAW filter, 91, 91*f*
 - space-time periodic process, 82
 - Stoneley waves, 89*f*, 90
 - ultrasonic and hypersonic surface waves, 91–92
 - wave propagation, 82–83
 - Electrets
 - residual (quasipermanent) polarization of
 - additional activation impact, 552
 - electroelectrets, 555
 - electrostatic field, 552

- heterocharge formation, 553
- homocharge formation, 553
- photoelectrets, 554–555
- polarized state exists, 551
- radioelectrets, 555–556
- thermoelectrets, 553–554, 554*f*
- triboelectrets, 556
- thermally stimulated depolarization, 556
- Electrical aging, 294
- Electrical boundary conditions, 521
- Electrical breakdown
 - in crystals
 - AHCs, 389, 389*f*
 - criterion of, 391
 - defect-free crystals, 390
 - dielectric samples, 387–388, 387*f*
 - elastic shock wave, 390
 - electrical field, 389–390
 - experimental data, 387*f*, 388
 - exponential increase, 388
 - features of, 389
 - longitudinal optical mode, 390–391, 390*f*
 - one-avalanche mechanism, 391
 - optical (laser) breakdown, 392–393
 - pinching, 391
 - temperature, 391
 - thin films, 386*f*, 391–392
 - in dielectrics
 - dependence of, 393–394, 393*f*
 - electrodegradation (aging)
 - (*see* Electrodegradation (aging) of dielectrics)
 - fragile dielectrics, 396
 - frequency, 396, 396*f*
 - physics of, 394
 - second kind, 394
 - solid dielectrics, 393
 - temperature of, 394–395, 395*f*
 - thermal equilibrium, 394
 - general regularities of
 - dependence of, 386–387, 386*f*
 - development time of, 387
 - electrochemical processes, 385
 - electrothermal breakdown, 384–385
 - first stage, 384
 - identity of principal mechanism, 386
 - maximal voltage for, 386
 - second stage, 384
 - volt-second characteristic, 385, 385*f*
 - local destruction, 383–384
- Electrical charge transfer, 168
- Electrical conductivity of heavy-fermion systems, 195–196
- Electrical conductivity of metals
 - carriers, 168
 - charge transfer description, 170
 - current density, 168
 - density of thermal energy, 168
 - drift velocity, 168
 - effect of magnetoresistance, 173
 - electrical charge transfer, 168
 - electrical current, 168
 - frequency dependence, 170
 - Hall's effect, 172
 - hyperconductivity, 176
 - mobility, 168
 - Ohm's law, 168
 - response time and free path of electrons, 174
 - skin effect, 175
 - temperature dependence, 169
 - Wiedemann-Franz law, 169–170
- Electrical current, 168
- Electrical polarization, 59
 - capacitance measurements, 309
 - conductivity, 291–292
 - dielectric contributions, 307, 308*f*, 310, 310*f*, 312*f*
 - dielectric losses, 293
 - dielectric spectra, 311
 - dielectric spectroscopy, 307
 - dipoles thermally activated polarization, 306
 - dynamics of, 336–349
 - electrical breakdown, 293
 - electrochemical breakdown, 294
 - electroconductivity, 290–291
 - electronic thermally activated polarization, 305
 - electrothermal breakdown, 293–294
 - field-induced electrical moment, 301
 - impact ionization, 292
 - ionic crystals, 311
 - ionic thermally activated polarization, 305
 - Lorentz dielectric permittivity, 309–310
 - macroscopic description of, 289
 - dielectric anisotropy, 295*f*, 297–298
 - dielectric constant, 296
 - dielectric nonlinearity, 296–297
 - dielectric susceptibility, 294–295
 - electrical induction, 295
 - tensor of permittivity, 298–301, 300*f*
 - migratory polarization, 306–307
 - molecular crystals, 311
 - polar dielectrics, 308–309
 - quasielastic (*see* Electronic quasielastic polarization)
 - quasielastic polarization, 301–304, 302–303*f*
 - relative permittivity, 290

- Electrical polarization (*Continued*)
 - relaxation dispersion, 293
 - space-charge, 290
 - surrounding cations, 305
 - thermal chaotic motion, 305
 - thermally activated jumps, 304
 - thermally supported, 290
- Electrochemical breakdown, 385
- Electrochemical potential, 125–126
- Electroconducting, 628*f*, 629
- Electroconductivity in dielectrics
 - charge carrier classification, 362, 363*f*
 - classification of, 362, 363*f*
 - electronic conduction, 362
 - ionic conduction, 363
 - SCLC, 364
 - charge carrier generation, 364–365
 - charge carrier injection and nonlinear conductivity
 - bipolar injections, 376, 380–381
 - electronic emission into vacuum, 377
 - metal-insulator, 377
 - monopolar injections, 376–380
 - Ohmic conductivity, 376
 - SCLC, 376
 - electronic conduction, 366–367, 366*f*
 - frequency dependence of conductivity
 - delay of slow polarization, 381
 - plasma resonance, 383
 - ionic conductivity
 - in crystal, 367
 - electrolytes, 367
 - experiment, 367
 - movement of, 367
 - potential barriers, 368–369, 368*f*
 - supported by electrical field, 368
 - surface conductivity, 370–372
 - temperature dependence of, 369–370, 370*f*
 - mechanisms of charges transport, 365–366, 365*f*
 - mobility of electrons, 373
 - polaron (hopping) electrical conductivity, 374
- Electrodegradation (aging) of dielectrics, 385
 - current density, 398–399, 398*f*
 - dielectric durability, 397
 - direct voltage, 397
 - durability of, 396*f*, 397
 - F-centers, 399
 - inorganic dielectrics, 397–399
 - irreversible, 397
 - reversible, 396
 - solid dielectric, 398
 - thermally activated processes, 397
 - titanium-containing ceramics, 399
- Electroelectrets, 555
- Electroinsulating, 628*f*, 629
- Electromechanical coupling
 - coefficient, 533
 - in piezoelectric, 533–537
- Electronic paramagnetic resonance (EPR), 238
- Electronic properties of metals
 - band theory of metals, 185
 - classic electronic theory of metals, 180
 - Fermi energy level, 182
 - Fermi surface, 183–184
 - quantum distribution of electronic gas, 181
- Electronic quasielastic polarization
 - Bohr's model, 313–314, 313*f*
 - definition, 303, 311
 - energy band theory, 315
 - Gauss system, 314
 - least inertial polarization mechanism, 312–313
 - negative ions, 314
 - permittivity, 315
 - semiconductors of diamond structure, 315
 - universal mechanism, 312–313
- Electronic semiconductors, 471
- Electronic thermal conductivity, 446
- Electronic thermal polarization
 - chaotic movement, 326
 - F-centers, 328
 - photoactivation, 328
 - relaxation time of, 326–327
 - rutile, 327, 327*f*
 - weakly bounded electrons, 326
- Electronic-type semiconductor, 444
- Electron-phonon interaction, 209
- Electrons, 121–122
 - in atoms
 - Bohr radius, 147
 - Coulomb force of attraction, 146
 - electronic cloud, 145
 - fundamental constants, 146
 - helium atom, 149
 - lithium atom, 149
 - quantum oscillator, 147–148
 - Schrödinger's equation, 146
 - in crystals
 - copper, 152, 153*f*
 - crystal lattice constant, 150–151, 151*f*
 - electronic energy spectra, 151, 152*f*
 - emission spectra, 151
 - Fermi level, 151
 - overlapping areas, 149–150
 - sodium atom, 149–150, 150*f*
 - valence band, 153

- in metals, dielectrics, and semiconductors
 - acoustical phonons, 162
 - adiabatic approximation, 163
 - bands of electronic spectrum overlap, 164
 - bands of electronic spectrum without overlapping, 158
 - band theory, 163
 - big orbital space, 156
 - bosons, 161
 - comparison of energy bands, 158, 159*f*
 - covalent crystals of semiconductors, 160
 - electrical properties, 154
 - electron-hole plasma, 161
 - energy bands, 157
 - energy-band structure, 155
 - fermions, 161
 - long acoustic waves, 162
 - macroscopic crystal, 163
 - magnon/spin wave, 162
 - mechanical properties, 155
 - one-electron approximation, 155–156
 - optical phonons, 162
 - optical properties, 154–155
 - phonon, 161
 - photon, 161
 - quantum fluid, 161
 - quasicontinuous energy bands, 156
 - quasiparticle, 161
 - strongly interacting particles, 160
 - thermal properties, 155
 - valence electrons, 159
 - weakly interacting waves, 160
 - Electron scattering of metals
 - on crystal defects, 187
 - electrons collision with each other, 191
 - on phonons, 189
 - resistance caused by scattering, 188
 - Electrosonic waves, 91
 - Electrostriction
 - coefficient, 538, 543
 - definition, 516
 - electrical controlling, 539–540, 540*f*
 - electrically induced piezoelectric effect, 540–544
 - electromechanical coupling, 540
 - electromechanical effects, 521
 - feature of, 519
 - giant, 520, 538–539
 - limits by boundary conditions, 538
 - linearized electrostriction, 519–520, 541–542, 543*f*
 - mechanically free dielectrics, 538
 - piezoelectric module, 541
 - quadratic (“even”) effect, 537–538
 - tensors of, 538
 - Elementary crystal cell, 263
 - Elementary movements in solids
 - adiabatic approximation, 121–122
 - anions, 121–122
 - atomic scale, 122–123
 - bosons, 123
 - cations, 121–122
 - collective motions, 122–123
 - diffusion transfer, 121
 - electrons, 121–122
 - exciton, 123–124
 - fermions, 123
 - hole, 124
 - local vibrations, 122–123
 - magnons, 123–124
 - movement of structural defects, 121
 - phonon, 124
 - photon, 123–124
 - plasma, 122
 - plasmon, 125
 - polariton, 125
 - polaron, 124
 - wave clots, 122, 122*f*
 - Enantiomorphism, 16
 - Energy spectrum of electrons, 416, 419–421, 420*r*
 - Enthalpy, 96
 - Entropy, 96
 - Equatorial effect, 268
 - Ettingshausen effect, 467, 471–472
 - Exciton, 123–124
 - Excitonic absorption, 457, 458*f*, 463–464, 464*f*
 - Excitonic instability, 633
 - Excitons, 41
 - thermal conductivity, 446, 448
 - Explosive electron emission, 167
 - External photoelectric effect, 453
 - Extrinsic semiconductors
 - acceptors, 433–434, 434*f*
 - donors, 434–438, 434*f*, 436–437*f*
 - high-purity crystals, 432–433
 - problems, 433
- F**
- Faraday effect, 482–483, 483*f*
 - Far-infrared polarizations, 315–319
 - Far-infrared region, 465
 - Fermi-Dirac distribution, 125, 373
 - Fermi-Dirac energy distribution function, 443

- Fermi-Dirac statistics, 127
 - Fermi energy, 171, 182
 - Fermions, 123, 125
 - heavy, 195
 - Fermi surface, 183–184
 - Fermi velocity, 171
 - Ferrimagnetism, 260–261
 - Ferrites, 228, 263
 - based on rare-earth elements, 263
 - Ferroelastics
 - acoustic lattice vibrations, 579
 - definition, 576
 - gadolinium molybdate, 578
 - improper ferroelectric, 578–579
 - lead orthophosphate, 579
 - mechanical stiffness of, 577
 - mechanical stress, 577
 - optical diffraction grating, 579
 - phase transition, 578–579
 - spontaneous deformation changes, 577*f*, 579
 - temperature dependence, 577*f*, 579
 - Ferroelectrics, 509–510, 566–575, 567–568*f*, 570–571*f*, 573*f*, 576*f*, 580–585
 - Ferroelectric semiconductors
 - posistor effect, 642, 642*f*
 - Ferrogarnets, 265
 - Ferromagnetic Curie point, 248, 279
 - Ferromagnetic phase transition, 258
 - Ferromagnetism
 - anisotropy of magnetic properties, 254
 - antiferromagnetic structures, 241
 - antisymmetric wave function, 245
 - collinear magnetic structures, 242–243, 242*f*
 - domain structure, 250
 - exchange integral, 245
 - heat capacity, 249
 - intrinsic magnetic moment, 241
 - magnetic hysteresis, 253
 - magnetic sublattices, 241
 - magnetization curve, 252
 - magnetocaloric effect, 258
 - magnetomechanical effect, 243
 - magnetostriction, 255–256
 - manganese, 246
 - Mendeleev's periodic table, 241
 - noncollinear magnetic structures, 242–243, 242*f*
 - noncompensated antiferromagnetism, 242
 - nonequivalent positions, 242
 - overlapping orbitals, 245
 - permanent magnets, 254
 - physical nature of, 243
 - spontaneous magnetization, 245
 - strong magnetism, 240
 - temperature characteristics, 246–247
 - thermal expansion and invar effect, 257
 - Fick's law, 42–43
 - First-order phase transition (PT-I)
 - Ehrenfest's classification, 602
 - entropy, 603–604, 604*f*
 - heat $C_p(T)$ temperature dependence, 604
 - Landau's theory, 606
 - overheating and overcooling, 603
 - polynomial form, 606
 - thermal expansion coefficient, 604
 - thermal hysteresis, 603
 - thermodynamic potential, 602, 603*f*
 - Flexoelectricity, 587
 - Fluorite (CaF₂) structure, 31
 - Fourier's heat conduction law, 113
 - Free electronic gas, 28
 - Free energy, 125–126
 - Frenkel defects, 641
 - Frenkel exciton, 41, 464
 - Frenkel mechanism, 38
 - Frequency dependence of conductivity, 170
 - Fullerenes, 55, 56*f*, 232
 - Fullerites, 55
- ## G
- Gallium arsenide, 426–427, 426*f*
 - Galvanomagnetic effects
 - compression effect, 472
 - definition, 466
 - even and odd effect, 467
 - Hall's effect, 467–468
 - longitudinal, 472
 - magnetoresistance, 470
 - particle motion, 467
 - transversal, 471
 - Galvanothermomagnetic
 - longitudinal effect, 472
 - transversal effect, 471–472
 - Gapless semiconductors, 158
 - Germanium
 - bandgap energy profiles, 425*f*, 426
 - effective mass tensor for, 424, 424*f*
 - intrinsic concentration in, 429, 429*f*, 432
 - Giant and colossal magnetoresistance, 275
 - Giant effects in electronic materials
 - electrostriction, 652
 - field-controllable switching elements, 655, 656–657*f*
 - magnetocaloric effect, 654, 654*f*
 - magnetoresistance, 657*f*, 658

- magnetostriction, 653
 - thermistor effects in critistors and posistors, 654, 655*f*
 - Giant electrostriction, 520, 538–539, 652
 - Giant magnetocaloric effect, 258–259, 653, 654*f*
 - Giant magnetoresistance, 657*f*, 658
 - Giant magnetostriction, 256, 265, 653
 - Glassy semiconductors, 412
 - Graphene, 55
 - Graphite, 54, 232
 - Gyromagnetic resonance, 264
- H**
- Hall's effect, 172, 437, 467–470, 470*f*
 - Hall sensors, 469
 - Heat capacity, 97–98
 - Heat capacity of ferromagnetic, 249
 - Heat transfer in semiconductors, 446
 - Helium atom, 149
 - Heteropolar bond, 8
 - High-permittivity dielectrics
 - electronic polarizability, 558
 - electronic polarization, 559
 - electronic shell displacements, 558
 - thermal stable microwave dielectric, 563*t*
 - ceramics, 563–564
 - permittivity and quality factor for, 549*t*, 563
 - special quality factor, 563
 - High-temperature superconductivity (HTS), 211
 - High-temperature superconductors (HTS), 205
 - Hole, 124
 - Hole-type semiconductor, 444
 - Homeopolar bond, 8
 - Hooke's law, 78
 - Hot probe method, 437
 - Hund's rule, 235
 - Hybrid ionic-covalent bond, 9
 - Hydrogen bond, 12, 64
 - Hyperconductivity, 176
- I**
- Ideal gas of magnons, 144–145
 - Impact ionization, 292
 - Improper ferroelectrics, 578–579, 615
 - Indeterminacy principle, 375
 - Indirect bandgap semiconductors, 426
 - Indirect intrinsic transitions, 457, 460–462, 460–462*f*
 - Inelastic scattering of neutrons, 140
 - Inert gas crystals, 33–34
 - Inorganic crystalline semiconductors
 - A^{III}B^V compounds, 411–412
 - A^{III}B^{VI} compounds, 412
 - A^{IV}B^{VI} compounds, 412
 - amorphous, 412–413
 - binary compounds, 411
 - chemical compound semiconductors, 410–411
 - conductors and dielectrics, 409
 - monoelement semiconductors, 410
 - oxide glassy semiconductors, 412
 - silicon carbide (SiC), 412
 - ternary compounds, 411
 - Inorganic dielectrics, 294, 397–398
 - Insulator-to-metal transition
 - definition, 629
 - energy-band theory, 630
 - model of excitons, 633
 - Mott model, 632
 - one-electron approximation, 630
 - vanadium oxide, 630–631, 631*f*
 - Wigner model, 632
 - Integral of movement, 419
 - Integrals of motion, 418–419
 - Interacting nearly impenetrable charged spheres, 6
 - Interband absorption, 457, 462–463, 462*f*
 - Interband magneto-absorption, 481
 - Internal force field, 419
 - Internal photoelectric effect, 454
 - Intrinsic absorption
 - direct, 457–460
 - indirect, 457, 460–462, 460–462*f*
 - valence bond photoionization, 458
 - Intrinsic polarization, 588
 - Intrinsic semiconductors
 - band diagram and flat model, 427, 428*f*
 - concentration of electrons and holes in, 432
 - definition, 431
 - electronic and hole conductivity components, 430–431
 - Fermi level in, 444
 - generation and recombination, 431
 - thermal ionization, 428–429
 - valence bond violation, 429
 - Invar alloys, 283
 - Invar effect, 257
 - Invar-type metallic alloys, 257
 - Inverse piezoelectric effect, 517, 527–533
 - Inversion in time, 225
 - Ion-electron emission, 167
 - Ionic bond, 5–6
 - Ionic (far IR) polarization, 315–319
 - Ionic quasielastic polarization, 311, 316–317
 - Ionic residues, 2
 - Ionic thermally activated polarization, 305

- Ionic thermal polarization mechanism
 - electrostimulated local diffusion, 325
 - equiprobable positions of, 324, 324*f*
 - nonlinearity of, 325–326
 - relaxation time, 326
 - saturation of, 326
 - structural defects, 323
 - Irreversible dielectrics, 288
 - Isomorphism, 36
 - Isothermal permittivity, 551
- J**
- Josephson effects, 209
- K**
- Kikoin-Noskov effect, 478–481, 478*f*, 481*f*
 - Kinetic energy, 96
 - Kinetic processes in semiconductors
 - electrical conductivity, 438*f*
 - charge carrier scattering, 442
 - drift velocity, 438
 - in electrical field, 445
 - electron and hole mobility, 442–443, 443*f*
 - Fermi level, 443–445
 - generation and recombination of charge carriers, 439–441
 - thermal conductivity, 438*f*
 - bipolar, 446–447
 - electron, 446
 - excitons, 446, 448
 - heat transfer, 446
 - phonon, 446
 - photon, 446–447
 - thermoelectric effects, 438*f*
 - mechanisms of, 448–449
 - Peltier effect, 450–451
 - Seebeck effect, 448
 - Thomson effect, 451
 - Kohn anomaly, 637
 - Kondo effect, 169, 658
- L**
- Lame parameters, 82
 - Landau diamagnetism, 231
 - Landau's theory, 605
 - Langevin function, 322
 - Lanthanoid compression, 236
 - Lanzheven-Curie paramagnetism, 234
 - Lattice defects in crystals
 - defect formation, 37
 - 1D defects-dislocations, 43
 - 2D and 3D defects, 44
 - zero-dimensional (point) defects, 38–39
 - Law of equipartition, 106–107
 - Law of specific heat constancy, 106
 - Least inertial polarization mechanism, 312–313
 - Left-handed materials (LHM), 58
 - Liddeyn-Sachse-Teller relation, 558, 561–562
 - Light absorption mechanisms
 - excitonic, 457, 458*f*; 463–464, 464*f*
 - impurity absorption, 457, 458*f*, 463
 - interimpurity, 457
 - intragand, 462–463, 462*f*
 - intrinsic absorption
 - direct, 457–460
 - indirect, 457, 460–462, 460–462*f*
 - valence bond photoionization, 458
 - phonon absorption, 465
 - plasma absorption, 457, 458*f*, 465
 - Linear double refraction, 267
 - Linear electromechanical phenomena, 514
 - Linear expansion coefficient, 98
 - Linearized electrostriction, 519–520, 520*f*, 541–542, 543*f*
 - Linear pyroelectric, 566
 - Linear thermal expansion coefficient, 179
 - Liquid conductors, 165
 - Liquid crystals (LCs)
 - in phase transitions, 601–602
 - ChLCs, 645*f*, 646
 - definition, 644
 - mesagenic, 648
 - mesomorphic properties, 644
 - mesophases, 644
 - multiple transitions, 649
 - NLC, 645, 645*f*
 - ordering parameter, 648
 - phases, 647, 650*f*
 - SLCs, 646
 - statistical theory, 647–648
 - Lithium atom, 149
 - Lithium niobate crystal, 539, 540*f*
 - Local vibrations, 122–123
 - London's penetration depth, 207
 - Longitudinal galvanothermomagnetic effect, 472
 - Lorentz dielectric permittivity, 309–310
 - Lorentz model for local field calculation
 - continuous medium, 333–334
 - macroscopic field, 334
 - permittivity calculation, 335
 - physically infinitesimal volume, 334
 - polar molecule, 335
 - solid isotropic dielectrics, 335
 - temperature parameter, 336

- Lorentz oscillator model, 620
 Lorenz number, 169–170
 Love waves, 89*f*, 90
 Low ionization energy, 2
 Lithium-ion batteries, 62
- M**
- Magnesium, 28, 28*f*
 Magnetically hard nanomaterials, 272
 Magnetic cooling, 258
 Magnetic electronics, 222
 Magnetic field, 222–223
 Magnetic field sensors, 173–174, 471
 Magnetic field strength, 223
 Magnetic flux, quantization of, 207
 Magnetic hysteresis, 253
 Magnetic induction, 223
 Magnetic moment, 223
 Magnetism
 antiferromagnetic, 228
 antiferromagnetic interaction, 259
 axial vector, 224–225
 diamagnetism, 226
 disordered
 diamagnetism, 229
 Lanzheven-Curie paramagnetism, 234
 nuclear magnetism, 240
 paramagnetism, 232
 Pauli paramagnetism in metals, 239
 spin of electron, 239
 temperature dependence of magnetic susceptibility, 238
 ferrimagnetism, 260–261
 ferrites based on rare-earth elements, 263
 ferromagnetic, 227
 ferromagnetism
 anisotropy of magnetic properties, 254
 antiferromagnetic structures, 241
 antisymmetric wave function, 245
 collinear magnetic structures, 242–243, 242*f*
 domain structure, 250
 exchange integral, 245
 heat capacity, 249
 intrinsic magnetic moment, 241
 magnetic hysteresis, 253
 magnetic sublattices, 241
 magnetization curve, 252
 magnetocaloric effect, 258
 magnetomechanical effect, 243
 magnetostriction, 255–256
 manganese, 246
 Mendeleev's periodic table, 241
 noncollinear magnetic structures, 242–243, 242*f*
 noncompensated antiferromagnetism, 242
 nonequivalent positions, 242
 overlapping orbitals, 245
 permanent magnets, 254
 physical nature of, 243
 spontaneous magnetization, 245
 strong magnetism, 240
 temperature characteristics, 246–247
 thermal expansion and invar effect, 257
 field strength, 223
 giant magnetostriction, 265
 inversion in time, 225
 low-inertia reorientation, 221–222
 macroscopic examination, 225
 magnetic dipole, 221
 magnetic electronics, 222
 magnetic field, 222–223
 magnetic moment, 221
 magnetic semiconductors and dielectrics, 265–266
 magnetization, 223
 magneto-optical phenomena, 266
 microscopic processes, 225
 monocrystalline ferrites, 265
 moving charges, 221
 nanomagnetic materials
 computer memory devices, 272–273
 giant and colossal magnetoresistance, 275
 interstices, 274
 magnetically hard nanomaterials, 272
 magnetoelectronics, 274–275
 ratio of surface to volume, 269
 soft magnetic nanomaterials, 269–270
 tunneling magnetoresistance, 278
 nuclear demagnetization, 225
 nuclear magnetic resonance, 225
 paramagnetics, 226–227
 permanent magnets based on rare-earth alloys, 264
 polar vector, 224–225
 rare-earth microwave ferrites, 264
 Magnetic semiconductors and dielectrics, 265–266
 Magnetic sublattices, 241
 Magnetite, 263
 Magnetization, 59, 223
 Magnetocaloric effect, 258
 Magneto-dielectric composites, 262
 Magneto-dipole mechanism, 256
 Magnetoelastic effect, 256–257, 283
 Magnetolectrets, 556

- Magnetoelectrical ceramic composite material, 58
- Magnetoelectronics, 274–275
- Magneto-optical effects
 - birefringence, 483
 - electrical properties at oscillations, 484–485
 - Faraday effect, 482–483, 483*f*
 - impurity magnetoabsorption, 481–482, 482*f*
 - interband magneto-absorption, 481
 - magnetoplasma phenomena, 483–484
 - optical phenomena, 484
 - photoelectromagnetic effect, 478–481, 478*f*, 481*f*
- Magneto-optical Kerr effect, 267–268
- Magneto-optical phenomena, 266
- Magnetoplasma resonance method, 483–484
- Magnetoresistance effect, 173
- Magnetoresistive Gauss effect, 470–471
- Magnetostriction, 255–256, 283
- Magnons, 123–124
 - antiferromagnetics, 144, 144*f*
 - Bohr’s magneton, 141
 - Bose-Einstein condensation, 144
 - conductive electrons, 145
 - electronic spin precession, 142–143
 - ideal gas of magnons, 144–145
 - integer spin, 143–144
 - intrinsic property, 141
 - long-wave acoustic phonons, 142–143
 - one-dimensional crystal, 142, 142*f*
 - ordered magnetic crystals, 141
 - source of a magnetic field, 141
 - source of the electrical field, 141
 - spin waves, 142
 - temperature dependence, 145, 145*f*
- Mass specific heat, 104–105
- Maxwell-Boltzmann distribution, 125
- Maxwell-Boltzmann statistic, 125
- Maxwell distribution, 126
- Maxwell law, 466
- Mechanical boundary conditions, 521–522
- Mechanical properties, 71
 - continuous homogeneous medium, 72
 - durability, 71
 - elasticity, 71
 - elastic stiffness and elastic compliance, 78–82
 - elastic waves in crystals, 82–92
 - failure, 71
 - Hooke’s law, 72
 - mechanical strain tensor, 75–78
 - mechanical stress tensor, 72–75
 - MEMS, 71
 - of metals, 180
 - plasticity, 71
 - wave propagation, 71–72
- Mechanical strain tensor
 - elastic deformation, 78
 - inelastic deformation, 78
 - intrinsic property, 77
 - limit of proportionality, 78
 - linear strain, 76
 - one-dimensional model, 75–76, 76*f*
 - perfectly elastic body, 78
 - “pure shear” strain, 77
 - residual deformation, 78
 - second-rank tensor, 76–77
 - thermal deformation, 77
 - thermal expansion, 77
 - three-dimensional model, 75–76, 76*f*
 - two-dimensional model, 75–76, 76*f*
- Mechanical stress tensor
 - field tensor, 74
 - imaginary hyperboloid, 74
 - material tensors, 74
 - one-dimensional model, 72–73, 73*f*
 - shear stresses, 74
 - static equilibrium, 72
 - surface acoustic waves, 73
 - three-dimensional model, 72, 73*f*
 - two-dimensional model, 72, 73*f*
- Mechanisms of lattice thermal conductivity, 115
- Mechanoelectrets, 556
- Meissner effect, 206, 206*f*
- Meridional Kerr effect, 268
- Mesagenic, 648
- Mesophases, 644–645
- Metacoverings, 61
- Metafilms, 61
- Metal bond, 4
- Metallic luster, 167
- Metals, 113
 - defining features of, 166–168
 - electrical conductivity
 - carriers, 168
 - charge transfer description, 170
 - current density, 168
 - density of thermal energy, 168
 - drift velocity, 168
 - effect of magnetoresistance, 173
 - electrical charge transfer, 168
 - electrical current, 168
 - frequency dependence, 170
 - Hall’s effect, 172
 - hyperconductivity, 176
 - mobility, 168

- Ohm's law, 168
 - response time and free path of electrons, 174
 - skin effect, 175
 - temperature dependence, 169
 - electronic conductivity, 165
 - electronic properties of metals and Fermi surface
 - band theory of metals, 185
 - classic electronic theory of metals, 180
 - Fermi energy level, 182
 - Fermi surface, 183–184
 - quantum distribution of electronic gas, 181
 - electron scattering
 - on crystal defects, 187
 - electrons collision with each other, 191
 - on phonons, 189
 - resistance caused by scattering, 188
 - liquid conductors, 165
 - mechanical properties, 180
 - special electronic states
 - band theory and heavy fermions, 196
 - electrical conductivity of heavy-fermion systems, 195–196
 - heavy-fermion system features, 202–203
 - magnetic properties of heavy-fermion metals, 194–195
 - metals with intermediate valence, 192
 - specific heat in metals with heavy fermions, 193
 - specificity of rare-earth metals, 198
 - valence instability of rare-earth elements, 199
 - superconductivity
 - anomaly of heat capacity at phase transition, 207
 - electron-phonon interaction, 209
 - high-temperature superconductivity, 211
 - Josephson effects, 209
 - magnetic field influence, 205–206
 - Meissner effect, 206
 - quantization of magnetic flux, 207
 - zero resistance, 205
 - thermal properties
 - linear thermal expansion coefficient, 179
 - temperosensitivity, 179
 - thermal capacity, 177
 - thermal conductivity, 177
 - thermoelectromotive properties, 178
 - very low resistive, 165
 - Metals with intermediate valence, 192
 - Metamaterials, 56–62
 - Microelectromechanical systems (MEMS), 71
 - Microscopic Lorentz model, 308–310
 - Microwave magnetic resonance, 264
 - Microwave photon, 130
 - Microwave technique, 354
 - Middle-field theory, 415–416
 - Migratory polarization
 - circuit of, 329–330
 - dielectric contribution of, 330
 - effective permittivity, 329
 - for example, 329
 - frequency dispersion, 307
 - inhomogeneous dielectric, 329*f*, 330
 - macroscopic mechanisms of, 328–329
 - piezoelectric and pyroelectric textures, 307
 - space-charged polarization, 306–307, 329, 329*f*
 - Miller indices of edge, 23
 - Minibands, 500
 - Mobile point defects, 41
 - Mobility, 168
 - Molar specific heat, 104–105
 - Molecular bonds, 11
 - Monochromatic acoustic waves, 133–134
 - Monocrystalline ferrites, 265
 - Monoelement semiconductors, 410
 - Monomers, 35
 - Monopolar injections, 376–380
 - Mott-Gurney law, 378
 - Mott model, 632
 - Movement of structural defects, 121
 - Mudgee-Righi-Leduc effect, 477–478
- N**
- Nanocomposite, 62
 - Nanomagnetic materials
 - computer memory devices, 272–273
 - giant and colossal magnetoresistance, 275
 - interstices, 274
 - magnetically hard nanomaterials, 272
 - magnetoelectronics, 274–275
 - ratio of surface to volume, 269
 - soft magnetic nanomaterials, 269–270
 - tunneling magnetoresistance, 278
 - Nanomaterials
 - amorphous materials, 52
 - Bernal polyhedra, 52, 53*f*
 - carbide, 54*f*, 55
 - carbon nanotubes, 55, 56*f*
 - classification of carbon structures, 53–54, 54*f*
 - diamond, 53–54
 - fullerenes, 55, 56*f*
 - fullerites, 55
 - graphene, 55
 - graphite, 54
 - MgO structure, 53, 53*f*

- Nanomaterials (*Continued*)
 weak van der Waals forces, 55
- Neel temperature, 259
- Negative index materials (NIM), 58
- Negative refractive index, 58, 59*f*, 60
- Negative temperature coefficient, 166
- Nematic liquid crystal (NLC), 645, 645*f*
- Nernst-Ettingshausen effect, 474, 475*f*
 longitudinal electrical field, 476
 transverse electrical field, 474–476
- Nernst's effect, 467, 472
- Nine linear effects, 513, 513*f*
- Noncentrosymmetric dielectrics, 297
- Noncompensated antiferromagnetism, 242
- Nondamping waves, 90
- Nondegenerated doped semiconductors, 469
- Nondegenerate semiconductors, 466, 466*f*
- Nonlinearity of ionic polarization, 318
- Nonlinear pyroelectric, 566
- Nonpolar (simple) covalent bond, 7
- Nonpolarized (isotropic) ferroelectric ceramics, 529
- Nonselective (indiscriminate) absorption, 462–463
- Nuclear demagnetization, 225
- Nuclear magnetic resonance, 225
- Nuclear magnetism, 240, 278
- Nuclear paramagnetism, 233
- O**
- Ohm's law, 168
- 1D defects-dislocations, 43
 edge dislocation, 43
 screw dislocation, 44
- One-electron approximation, 465
- One-electron band theory, 185
- One-electron task, 416
- One-ion mechanism, 256
- Optical nonlinearity, 296
- Optical phenomena in semiconductors
 absorption coefficient, 452–453
 external photoelectric effect, 453
 internal photoelectric effect, 454
 light absorption (*see* Light absorption mechanisms)
 photoluminescence, 453–454
 photoresistivity, 454–455, 455*f*
 photovoltaic (Dember) effect, 455–456
 reflection coefficient, 452
- Optical polarizations, 311–319
- Orbital magnetism, 278
- Orbital paramagnetism, 233
- “Order-disorder” phase transitions, 36
- Order-disorder type ferroelectrics, 571, 572*t*, 573*f*, 615–617, 616*f*
- Ordering parameter
 critical degree of freedom, 607
 definition, 608
 displacement type, 607*f*, 609
 in ferroelectrics, 607
 Landau's theory, 605
 in liquid crystals, 648
 macroscopic value, 606
 microscopic characteristics, 602
 order-disorder type, 607*f*, 609
 temperature dependence of, 608, 608*f*
 thermal motion, 607
 zero-rank tensor, 607
- Organic metals, 166
- Organic semiconductors, 413–414, 413*f*
- Oscillator model and elastic waves, 108
- Oscillatory movements, 457
- Oxide glassy semiconductors, 412
- P**
- Paraelectrics, 509–510
 of displacement type
 Curie-Weiss law, 560
 displacement-type paraelectrics, 561–562
 “order-disorder” type, 560–561
 perovskite-type structures, 561
 second-order phase transition, 560–561
 shell model, 562
 soft vibration mode, 561–562
 nonlinearity of, 580–585, 583*f*
- Paramagnetics, 226–227
- Paramagnetism, 232
- Parameter of lattice, 4–5
- Pauli paramagnetism in metals, 239
- Pauli principle, 630
- Peierls prohibition, 211–212
- Peltier effect, 450–451
- Penrose mosaic, 50, 51*f*
- Perfect dielectrics, 160
- Permanent magnets, 254
 based on rare-earth alloys, 264
- Permittivity frequency dispersion, phase transition
 displacement-type ferroelectrics, 420*t*, 618, 619*f*, 620, 621*f*
 order-disorder-type ferroelectrics, 615–617, 616*f*
- Perovskite (CaTiO₃), 33
- Phase transitions (PTs)
 with conductivity anomalies
 electroconducting or electroinsulating, 628*f*, 629

- electronic spectrum, 629
- insulator-to-metal transition (*see* Insulator-to-metal transition)
- sudden change in, 628–629, 628*f*
- crystal symmetry, 601–602
- in dielectric properties
 - displacement type, 609–610
 - ferroelectrics with PT-II, 610, 612–613
 - morphotropic phase transitions, 622–623, 623*f*
 - order-disorder type, 609–610
 - permittivity frequency dispersion
 - (*see* Permittivity frequency dispersion, phase transition)
 - relaxor ferroelectrics, 623–624, 624–626*f*, 628
- dielectrics with unstable electronic spectrum
 - ionic conductivity jump, 640
 - one-dimensional system, 634
 - posistor effect in ferroelectric semiconductors, 642, 642*f*
 - quasi-one-dimensional metal, 635–639, 635*f*, 637*f*
 - superconducting phase transition, 639–640
 - two-dimensional system, 634
- external influence, 601–602
- first-order PTs (PT-I)
 - Ehrenfest's classification, 602
 - entropy, 603–604, 604*f*
 - heat $C_p(T)$ temperature dependence, 604
 - Landau's theory, 606
 - overheating and overcooling, 603
 - polynomial form, 606
 - thermal expansion coefficient, 604
 - thermal hysteresis, 603
 - thermodynamic potential, 602, 603*f*
- giant effects in electronic materials
 - electrostriction, 652
 - field-controllable switching elements, 655, 656–657*f*
 - magnetocaloric effect, 654, 654*f*
 - magnetoresistance, 657*f*, 658
 - magnetostriction, 653
 - thermistor effects in critistors and posistors, 654, 655*f*
- large electrical and magnetic parameters
 - hyperconductivity, 649, 650*f*
 - large dielectric permittivity, 650–651, 650–651*f*
 - large magnetic permeability, 651
- in liquid crystals, 601–602
 - ChLCs, 645*f*, 646
 - definition, 644
 - mesagenic, 648
 - mesomorphic properties, 644
 - mesophases, 644
 - multiple transitions, 649
 - NLC, 645, 645*f*
 - ordering parameter, 648
 - phases, 647, 650*f*
 - SLCs, 646
 - statistical theory, 647–648
- lowest nonzero partial derivative, 602
- ordering parameter, 602, 605–609
- second-order PTs (PT-II)
 - heat $C_p(T)$ temperature dependence, 604
 - Landau's theory, 606
 - polynomial form, 606
 - symmetry changes abruptly, 604
 - temperature dependence, 605–606
 - thermal expansion coefficient, 604
 - thermodynamic potential, 602, 603*f*
- solid materials, 601
- π -bond, 8
- Phonons, 124
 - absorption, 457, 465
 - acoustic modes, 133
 - center of the Brillouin zone, 137–138
 - chaotic acoustic waves, 133–134
 - Debye frequency, 133
 - Debye's theory, 131
 - diatomic crystal diamond, 138, 138*f*
 - dispersion, 134
 - electrical force, 136
 - first Brillouin zone, 133
 - fluctuations of density, 139
 - gas of quasiparticles, 131
 - group velocity, 132–133
 - inelastic scattering of neutrons, 140
 - ionic crystal, 135, 135*f*
 - longitudinal optical waves, 135, 135*f*
 - macroscopic field, 136
 - monochromatic acoustic waves, 133–134
 - one-dimensional monoatomic crystal, 131
 - optical phonons, 134
 - oscillatory spectrum of aluminum, 140, 140*f*
 - permittivity, 136
 - phase velocity, 132–133
 - quantization of elastic waves, 139
 - spatial dispersion, 135–136
 - thermal conductivity, 446
 - transverse optical waves, 135, 135*f*
 - wave of displacements, 131–132
- Phosphorescence, 453–454
- Photoconductivity, 414
- Photoelectrets, 554–555

- Photoelectromagnetic effect, 478–481, 478*f*, 481*f*
 Photoelectronic emission, 167
 Photogeneration, 431
 Photoluminescence, 453–454
 Photonic crystals, 62
 Photonic/optical metamaterials, 60
 Photons, 123–124
 “black-body” radiation, 129
 corpuscular theory, 129
 de Broglie ratio, 130
 dispersion law, 130
 duality in the nature, 129
 microwave photon, 130
 neutral particle, 129
 spin of a photon, 130
 thermal conductivity, 446–447
 wave theory, 129
 wave vector, 130
 zero mass, 130
 Photoresistivity, 454–455, 455*f*
 Photovoltaic (Demmer) effect, 455–456
 Piezoelectric effect
 application of, 510–511
 barium titanate, 525
 electrical boundary conditions, 521
 electrically induced, 519–520
 electromechanical coupling in, 533–537
 electromechanical effects, 516
 electromechanical properties, 521
 electrostriction, 516, 520, 537–544
 intermediate conditions, 522
 inverse piezoelectric effect, 517, 527–533
 linear effects, 518, 518*f*
 linearity, 519
 linearized electrostriction, 519–520, 520*f*
 matrix representation, 522–523
 mechanical boundary conditions, 521–522
 mechanical stress, 513*f*, 516–517, 517*f*
 microsystems and microelectronic components, 515–516
 modulus components, 522–523, 523*t*, 524*f*, 525
 practical application of, 515
 in quartz, 517–519, 518*f*, 523, 524*f*
 Rochelle salt crystal, 527
 scientific and technical fields of, 510–511
 spatial distribution of, 519
 technical application of, 515
 transverse, 518
 Piezoelectricity, 288, 516
 Piezoelectric modulus, 517
 Piezoelectric square, 530
 Piezoelectric transformer, 85–86, 85*f*
 Piezoelectronics, 86, 510
 Pinch effect, 472–474
 Plasma, 122
 Plasma absorption, 457, 458*f*, 465
 Plasma resonance, 465
 Plasmon, 125
 Plastic deformation, 288–289
 Plasticity, 71
 Poisson’s ratio, 81
 Polar covalent bond, 8
 Polar crystals, 10
 Polar dielectrics, 308–309
 electromechanical and electrothermal properties, 510
 ferrielectrics and ferroelastics, 575–579
 ferroelectrics
 and antiferroelectrics, 566–575
 and paraelectrics (*see* Paraelectrics)
 high-permittivity dielectrics (*see* High-permittivity dielectrics)
 interdependence in
 cause of noncentrosymmetry in crystals, 586
 experimental illustrations, 589, 590–592*f*
 intrinsic polarity modeling, 587, 588*f*
 micromachining, 510
 piezoelectric effect, 510–511, 515–527
 primary effects in
 elastic, electrical, and thermal effects, bonding diagram for, 513, 513*f*
 impact-response method, 511, 511*t*
 linear and nonlinear, 512, 513*f*
 magnetolectrical effects, 512
 trivial/conventional responses, 511
 vector field, 512
 pyroelectrics and electrets, 544–557
 Polar effect, 268
 Polariton, 125
 Polarization switching, 580
 Polaron, 41, 124
 Polaron (hopping) electrical conductivity
 Coulomb interaction, 374
 movement of, 375
 self-trapping electron, 374
 size of, 374–375, 375*f*
 Polaron-type charge carriers, 362
 Polar vector, 224–225
 Polymorphism, 21, 36
 Polytypicism, 35–36
 Posistor effect, 642, 642*f*
 Potassium dihydrogen phosphate (KDP), 311
 Potential energy, 96
 Power actuators, 265

- Practically free electrons, 170
 Precession, 264
 Primary pyroelectric, 547
 Primary pyroelectric coefficient, 547
 Principle of Le Chatelier, 230–231
 Protonic conductivity, 641–642
 Pseudolinear “piezoelectric effect,” 541–542
 Pure/undoped semiconductors. *See* Intrinsic semiconductors
 Pyroelectric effect, 541
 Pyroelectric group, 22
 Pyroelectricity, 288
 Pyroelectrics, 509
 ancient sources, 545
 capacitance of, 551
 classes of, 549, 549*f*
 crystal symmetry, 549
 current flows, 545
 efficiency of, 550
 electrical boundary conditions for, 549–550
 electrically free, 549–550
 electrical phenomenon, 545
 electrocaloric effect, 551
 electrodes, 545
 hidden (or latent) internal polarity, 546
 isothermal permittivity, 551
 linear pyroelectric crystals, 551
 nonlinear, 550
 one-dimensional model of, 546–547, 547*f*
 piezoelectric conversion of thermal strain, 547
 polar polymers, 551
 secondary effect, 548
 single-domain structure, 545–546
 solid-state energy converter, 544
 spatial distribution of, 548, 548*f*
 symmetry and physical mechanism of, 545
 temperature change, 546
 thermal energy in, 544
 tourmaline and lithium sulfate crystals, 545, 547
 volumetric piezoelectric effect, 546
- Q**
- Quantization of magnetic flux, 207
 Quantum dimensional effects, 488
 ballistic conductivity of nanoscale conductors, 494–496, 495*f*
 de Broglie wavelength, 488
 density of states, 488–489
 energy spectrum of electrons, 489
 infinite crystal quantum properties, 489–490, 489–490*f*
 quantum dot, 493–494, 493*f*
 quantum well, 490–492, 490–491*f*
 quantum wire, 491*f*, 492–493, 493*f*
 resonance tunneling, 496–497, 497*f*
 superlattices (*see* Superlattices)
 Quantum-dimensional levels, 490
 Quantum distribution of electronic gas, 181
 Quantum-mechanical tunneling, 273
 Quantum statistics
 Bose-Einstein statistics, 128
 collectivization, 128
 distribution function, 127
 electrochemical potential, 127–128
 Fermi-Dirac statistics, 127
 laws of quantum mechanics, 126–127
 many-boson system, 128
 Maxwell-Boltzmann statistic, 125
 Pauli principle, 127
 quantum gas, 128
 temperature of degeneracy, 127
 Quasicrystals
 “amorphous” inclusions, 51
 icosahedral clusters, 49
 irrational numbers, 50
 linear chain of atoms, 50
 long-range order, 49–50
 metal glasses, 49
 microsymmetry, 52
 molten aluminum-manganese, 47–48
 order of rotary symmetry axes, 47
 Penrose mosaic, 50, 51*f*
 regular convex polyhedra, 48, 48*f*
 shechtmanite, 48–49, 51
 structure of, 49
 Quasielastic electronic polarization. *See* Electronic quasielastic polarization
 Quasielastic polarization, 301–304, 302–303*f*
 Quasi-impulse, 418–419
 Quasiparticles in solids
 different elementary movements in solids
 adiabatic approximation, 121–122
 anions, 121–122
 atomic scale, 122–123
 bosons, 123
 cations, 121–122
 collective motions, 122–123
 diffusion transfer, 121
 electrons, 121–122
 exciton, 123–124
 fermions, 123
 hole, 124
 local vibrations, 122–123
 magnons, 123–124

- Quasiparticles in solids (*Continued*)
- movement of structural defects, 121
 - phonon, 124
 - photon, 123–124
 - plasma, 122
 - plasmon, 125
 - polariton, 125
 - polaron, 124
 - wave clots, 122, 122*f*
 - electrons in atoms and in crystals, 145–153
 - magnons, 141–145
 - phonons, 131–140
 - photons, 128–131
 - quasiparticle statistics, 125–128
- Quasipermanent (residual) polarization of electrets
- additional activation impact, 552
 - electroelectrets, 555
 - electrostatic field, 552
 - heterocharge formation, 553
 - homocharge formation, 553
 - photoelectrets, 554–555
 - polarized state exists, 551
 - radioelectrets, 555–556
 - thermoelectrets, 553–554, 554*f*
 - triboelectrets, 556
- R**
- Radioelectrets, 555–556
- Radioelements, 58
- Rapid solidification process, 651*f*, 652
- Rare-earth metals, 198
- Rare-earth microwave ferrites, 264
- Rayleigh waves, 89, 89*f*
- Reciprocal lattice, 24–26
- Reflection coefficient, 452
- Reflection spectrum, 465
- Regular convex polyhedra, 48, 48*f*
- Relative permittivity, 290
- Relaxation polarization
- dielectric losses
 - Debye equations, 351*f*, 354
 - frequency dependence, 354
 - power density, 354–355
 - reactive conductivity, 354–355, 355*f*
 - temperature dependences of, 355–356, 356*f*
 - temperature–frequency dependence, 355, 356*f*
 - dynamics of electrical polarization
 - complex value, 339–340
 - Debye dispersion formula, 338
 - equilibrium state, 338–339
 - reactive current, 339
 - relaxation character, 338
 - surface charge density, 338
- Relaxor-ferroelectric ceramics, 515–516
- Relaxor ferroelectrics, 623–626, 624–626*f*, 628, 651, 651*f*
- Residual (quasipermanent) polarization of electrets. *See* Quasipermanent (residual) polarization of electrets
- Resonance tunneling, 496–497, 497*f*
- Resonant dispersion, 311
- Response time and free path of electrons, 174
- Righi-Leduc effect, 477
- Rochelle salt crystal, 527
- Rochelle salt dielectric spectrum, 359–360, 360*f*
- Rutile (TiO₂) structure, 31
- S**
- s- and p-electronic orbitals, 185
- Schottky defects, 38
- Schrodinger equations, 414–415
- Screw dislocation, 37–38, 37*f*, 44
- Secondary effects, 515
- Secondary electron emission, 167
- secondary pyroelectric coefficient, 548
- Secondary pyroelectric effect, 547
- Secondary radiation, 453–454
- Second-order phase transition (PT-II)
- ferroelectrics with
 - nonpolar (paraelectric) phase, 610–612, 611*f*
 - polar phase, below the Curie point, 612–613, 612*f*
 - PT-I close to PT-II, 613–615, 613–614*f*
 - heat $C_p(T)$ temperature dependence, 604
 - Landau's theory, 606
 - polynomial form, 606
 - symmetry changes abruptly, 604
 - temperature dependence, 605–606
 - thermal expansion coefficient, 604
 - thermodynamic potential, 602, 603*f*
- Seebeck effect, 448
- Seeming permittivity, 309
- Semiconductors, 113–114
- characteristic feature of, 409
 - conductors and dielectrics, 409
 - electronic bonds in, 410
 - extrinsic semiconductors (*see* Extrinsic semiconductors)
 - fundamentals of band theory of
 - adiabatic approximation, 415
 - band structure of A^{III}B^V semiconductors, 426–427, 426*f*
 - conduction band of silicon and germanium, 425*f*, 426

- effective mass, 422–426, 422*f*, 424*r*, 425*f*
- electronic energy bands in, 414–415
- energy spectrum of electrons, 419–421, 420*r*
- filling of permitted bands, 421–422
- one-electron approximation, 415–416
- quasi-impulse, 418–419
- impurities or defects, 410
- inorganic crystalline
 - A^{III}B^V compounds, 411–412
 - A^{III}B^{VI} compounds, 412
 - A^{IV}B^{VI} compounds, 412
 - amorphous, 412–413
 - binary compounds, 411
 - chemical compound semiconductors, 410–411
 - conductors and dielectrics, 409
 - monoelement semiconductors, 410
 - oxide glassy semiconductors, 412
 - silicon carbide (SiC), 412
 - ternary compounds, 411
- insulator property, 409
- intrinsic (*see* Intrinsic semiconductors)
- kinetic processes in, 438–451
- in magnetic field
 - compression effect, 472–474
 - free charge carriers, 465
 - galvanomagnetic effects (*see* Galvanomagnetic effects)
 - Hall's effect, 468–470, 470*f*
 - magneto-optical effects (*see* Magneto-optical effects)
 - magnetoresistance, 470–471
 - nonequilibrium conditions, 466
 - thermomagnetic effects (*see* Thermomagnetic effects)
- optical phenomena in, 452–465
- organic, 413–414, 413*f*
- potential wells, barriers, and tunneling
 - Broglie wave, 487
 - discreteness of energy spectrum, 486
 - isoenergetic tunneling, 486
 - microparticle in, 486–488
- quantum dimensional effects
 - ballistic conductivity of nanoscale conductors, 494–496, 495*f*
 - de Broglie wavelength, 488
 - density of states, 488–489
 - energy spectrum of electrons, 489
 - infinite crystal quantum properties, 489–490, 489–490*f*
 - quantum dot, 493–494, 493*f*
 - quantum well, 490–492, 490–491*f*
 - quantum wire, 491*f*, 492–493, 493*f*
 - resonance tunneling, 496–497, 497*f*
 - superlattices (*see* Superlattices)
- Semimetals, 232
- Shear modulus, 81
- Shechtmanite, 48–49, 51
- Shockley-Read-Hall recombination, 441
- Short elastic waves, 115
- σ -bond, 8
- Silicon
 - bandgap energy profiles, 424, 425*f*
 - drift mobility of electrons in, 442–443, 443*f*
 - effective mass tensor for, 424, 424*r*
 - intrinsic concentration in, 429, 429*f*, 432
- Silicon carbide (SiC), 412
- Skin effect, 175
- Smectic liquid crystals (SLCs), 646
- Soft magnetic nanomaterials, 269–270
- Solid solutions, 35
- Sommerfeld constant, 240
- Space-charged polarization, 306–307
- Space-charge (migration)-induced polarization, 290
- Space charge limited current (SCLC), 364
- Special electronic states of metals
 - band theory and heavy fermions, 196
 - electrical conductivity of heavy-fermion systems, 195–196
 - heavy-fermion system features, 202–203
 - magnetic properties of heavy-fermion metals, 194–195
 - metals with intermediate valence, 192
 - specific heat in metals with heavy fermions, 193
 - specificity of rare-earth metals, 198
 - valence instability of rare-earth elements, 199
- Specific heat, 96, 104–105
- Sphalerite and wurtzite structures, 29–30
- Spin magnetism, 278
- Spin of a photon, 130
- Spin of electron, 239
- Spin paramagnetism, 233
- Spin-polarized electrons, 266
- Spintronics, 274–275
- Spontaneous magnetization, 245
- Spontaneous polarization, 566.
 - See also* Pyroelectrics
- Spontaneous thermostriction, 257
- Stable thermodynamic state of polar dielectrics, 566
- Standing waves, 535
- Stoneley waves, 89*f*, 90
- Superconductivity of metals
 - anomaly of heat capacity at phase transition, 207
 - electron-phonon interaction, 209

- Superconductivity of metals (*Continued*)
 - high-temperature superconductivity, 211
 - Josephson effects, 209
 - magnetic field influence, 205–206
 - Meissner effect, 206
 - quantization of magnetic flux, 207
 - zero resistance, 205
 - Superionic conductors, 640–641
 - Superionic electrical conduction, 366
 - Superionic transition, 629
 - Superlattices
 - acoustic properties of, 500–508
 - band engineering, 498
 - composite, 498
 - definition, 498
 - doped, 498–499, 499*f*
 - energy diagram of, 498, 499*f*
 - heterojunctions, 497–498
 - minibands, 500
 - solid-state structures, 500
 - Superparamagnetics, 270–271
 - Synthetic metals, 166
- T**
- Temperature dependence of conductivity, 169
 - Tensosensitivity of metals, 179
 - Ternary compounds, 411
 - Tetrathiafulvalene-tetracyano-quinodimethane (TTFTCNQ), 635–636
 - Thermal capacity of metals, 177
 - Thermal conductivity, 98
 - bipolar, 446–447
 - dielectrics, 114
 - electron, 446
 - excitons, 446, 448
 - Fourier's heat conduction law, 113
 - heat transfer, 446
 - mechanisms of lattice thermal conductivity, 115
 - metals, 113
 - of metals, 177
 - phonon, 446
 - photon, 446–447
 - semiconductors, 113–114
 - Thermal expansion, 98, 257
 - anisotropy, 102
 - Debye temperature, 99
 - linear expansion coefficient, 98
 - simple model, 99
 - thermal expansion coefficient, 99, 99*f*
 - volumetric coefficient, 98
 - Thermal ionization, 428–429, 439
 - Thermally activated polarizations
 - dipole
 - Boltzmann distribution, 321
 - Debye model, 320–322
 - electrical field, 319–320
 - Langevin function, 322
 - nonlinearity of, 322–323
 - polar crystals, 320
 - relaxation time of, 320
 - electronic
 - chaotic movement, 326
 - F*-centers, 328
 - photoactivation, 328
 - relaxation time of, 326–327
 - rutile, 327, 327*f*
 - weakly bounded electrons, 326
 - ionic
 - electrostimulated local diffusion, 325
 - equiprobable positions of, 324, 324*f*
 - nonlinearity of, 325–326
 - relaxation time, 326
 - saturation of, 326
 - structural defects, 323
 - migratory polarization
 - circuit of, 329–330
 - dielectric contribution of, 330
 - effective permittivity, 329
 - for example, 329
 - inhomogeneous dielectric, 329*f*, 330
 - macroscopic mechanisms of, 328–329
 - space-charge polarization, 329, 329*f*
 - Thermally stimulated current (TSC), 556–557, 557*f*
 - Thermally supported (relaxation)-induced polarizations, 290
 - Thermal properties
 - absolute temperature, 97
 - constant temperature, 96–97
 - constant volume, 96–97
 - crystal heat capacity
 - anharmonicity, 106
 - arbitrary waves, 106
 - Born's theory of lattice dynamics, 106
 - Debye's model of specific heat, 110
 - Debye's quantum theory of heat capacity, 106
 - Einstein's quantum theory of heat capacity, 106, 109
 - law of heat capacity constancy, 106
 - law of specific heat constancy, 106
 - mass specific heat, 104–105
 - specific heat, 104–105
 - volumetric specific heat, 104–105
 - enthalpy, 96
 - entropy, 96
 - free energy, 96
 - heat, 97
 - heat capacity, 97–98

internal energy, 96
 internal energy of movement, 95
 kinetic energy, 96
 potential energy, 96
 specific heat, 96
 thermal conductivity, 98
 thermal expansion, 98
 anisotropy, 102
 Debye temperature, 99
 linear expansion coefficient, 98
 simple model, 99
 thermal expansion coefficient, 99, 99*f*
 volumetric coefficient, 98
 Thermal properties of metals
 linear thermal expansion coefficient, 179
 tensosensitivity, 179
 thermal capacity, 177
 thermal conductivity, 177
 thermoelectromotive properties, 178
 Thermoconductivity, 113
 Thermocouple, 448
 Thermoelastic effects, 514
 Thermoelectrets, 553–554, 554*f*
 Thermoelectric effects, 448–450
 Thermoelectricity, 438*f*
 mechanisms of, 448–449
 Peltier effect, 450–451
 Seebeck effect, 448
 Thomson effect, 451
 Thermoelectromotive (thermal EMF) properties,
 178
 Thermogeneration, 431
 Thermoionic emission, 167
 Thermomagnetic effects
 longitudinal temperature gradient, 477–478
 Nernst-Ettingshausen effect, 474, 475*f*
 longitudinal electrical field, 476
 transverse electrical field, 474–476
 transversal temperature gradient, 477
 Thermostriction, 283
 Thomson effect, 451
 Traditional theory of ferroelectrics, 566
 Transversal galvanothermodynamic effect,
 471–472
 Transverse electrical field, 474–476
 Traveling waves, 535
 Triboelectrets, 556
 Tunneling magnetoresistance, 278
 2D and 3D defects
 physics of modulated structures, 46
 surface of a crystal, 45
 3D (volumetric) defects, 46

U

Ultrasound emitter, 530
 Ultrasound receiver, 530
 Uniform heating, 102–103
 Unit cell of inverse lattice, 418

V

Valence angles, 7
 Valence electrons, 2
 Valence instability of rare-earth elements, 199
 Valence shells, 2
 van der Waals bonds, 11, 64
 Van Vlack paramagnetism, 230*f*, 238
 Voigt effects, 483
 Volumetric coefficient, 98
 Volumetric piezoelectric effect, 546, 588

W

Wannier-Mott excitons, 464
 Wave clots, 122, 122*f*
 Wave function, 415–418
 Wave propagation, 71–72
 Weighting factor, 202
 Wiedemann-Franz law, 169–170
 Wigner model, 632

Y

Young's modulus, 81

Z

Zeeman effect, 267
 Zero-dimensional (point) defects
 atomic crystals, 40–41
 diffusion, 41
 excitons, 41
 impurity atom, 39–40, 40*f*
 interstitial atom, 39–40, 40*f*
 ionic crystal, 40
 1D defects-dislocations, 43
 edge dislocation, 43
 screw dislocation, 44
 polaron, 41
 Schottky defects, 39, 39*f*
 structure violation, 38–39
 2D and 3D defects
 physics of modulated structures, 46
 surface of a crystal, 45
 3D (volumetric) defects, 46
 vacancy, 39–40, 40*f*
 Zero-dimensional (0D) nano-object, 493, 493*f*
 Zero energy, 487–488
 Zero resistance of superconductors, 205

ELECTRONIC MATERIALS

Principles and Applied Science

Yuriy M. Poplavko

Electronic Materials: Principles and Applied Science addresses the materials properties that are essential for researchers and engineers entering the field of electronics. The book uses simplified mathematical treatment of theories, with emphasis placed on the basic concepts of physical phenomena in electronic materials. Most chapters are devoted to the advanced scientific and technological problems of electronic materials; in addition, some new insights into theoretical facts relevant to technical devices are presented.

There is discussion of the basics of symmetry and internal structure of crystals, and the main properties of metals, dielectrics, semiconductors, and magnetic materials, while mechanical and thermal properties are reviewed with an emphasis on electrical and magnetic properties. Theory and modern experimental data are presented, as well as the specifications of materials that are necessary for practical application in electronics. The modern state of research in nanophysics of metals, magnetic materials, dielectrics, and semiconductors is considered, with particular attention to the influence of structure on the physical properties of nanomaterials.

- Provides important overview of the fundamentals of electronic materials properties significant for device applications along with advanced and applied concepts essential to those working in the field of electronics
- Takes a simplified and mathematical approach to theories essential to the understanding of electronic materials and summarizes important takeaways at the end of each chapter
- Interweaves modern experimental data and research in topics such as nanophysics, nanomaterials, and dielectrics

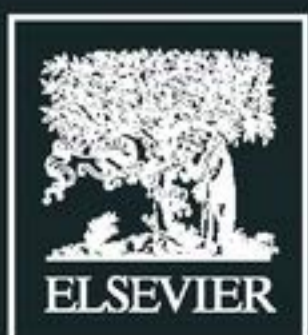
Electronic Materials is an essential reference for newcomers to the field of electronics, providing a fundamental understanding of important basic and advanced concepts in electronic materials science.

Related Titles

Ohring and Kasprzak / Reliability and Failure of Electronic Materials and Devices, 2nd Edition / 9780120885749

Stehr, Buyanova, and Chen / Defects in Advanced Electronic Materials and Novel Low Dimensional Structures / 9780081020531

Fornari / Single Crystals of Electronic Materials / 9780081020968



elsevier.com/books-and-journals

