POLITECHNIKA LUBELSKA

PODRĘCZNIKI

Energy Effective Materials

edited by
Paweł Żukowski

Lublin 2015

Energy Effective Materials

Podręczniki – Politechnika Lubelska

Politechnika Lubelska
Wydział Elektrotechniki i Informatyki
ul. Nadbystrzycka 38A
20-618 Lublin

Energy Effective Materials

Alexander K. Fedotov

Mirosław Maliński

Aleksy Patryn

Paweł Żukowski

Nikolay Drozdov

Tomasz N. Kołtunowicz

Alexander V. Mazanik

Mikhail Tivanov

Konrad Kierczyński

Lublin University of Technology

Lublin, 2015

This document has been prepared by the financial support of European Union. The authors from Lublin University of Technology, Koszalin University of Technology and Belarusian State University are responsible for the content of this document and it can not be regarded as the European Union's official position.

The book is developed in a frame of the project "Development of Training Network for Improving Education in Energy Efficiency" acronym: ENERGY, grant Nr. 530379-TEMPUS-1-2012-1-LVTEMPUS-JPCR.

Project was approved by the European Commission in frame of the Program Tempus IV – Fifth call for proposals (Programme guide EACEA/25/2011).

This project has been funded with support from the European Commission Project. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.

Chapter *1. Fundamentals of materials science* was developed by Alexander K. Fedotov, Paweł Żukowski, Tomasz N. Kołtunowicz, Konrad Kierczyński, Nikolay Drozdov, Alexander V. Mazanik, Mikhail Tivanov.

Chapter *2. Phase transformations in solid materials* was developed by Alexander K. Fedotov, Paweł Żukowski.

Chapter *3. Properties of materials for power industry and energy saving* was developed by Alexander K. Fedotov, Tomasz N. Kołtunowicz.

Chapter *4. Modern methods of evaluation of parameters for energy effective materials* was developed by Mirosław Maliński, Aleksy Patryn.

Reviewers:

Prof. **Bohdan Andriyevskyy**, Koszalin State University, Poland

Prof. **Victor Bashtavoi**, Belarusian National Technical University, Belarus

Editor: **Paweł Żukowski**

Proofreading: **Tomasz N. Kołtunowicz**

ISBN: 978-83-7947-163-8

Content

Executive summary

The task of optimization of heat, mass and electron transfer is one of the most important in energy technologies (thermal, thermoelectric, nuclear, solar, etc.). Therefore, training of physicists and engineers by the issues of "efficiency", "energy saving" and "renewable and alternative energy sources" leads to the need, among other areas of modern science and engineering, to give students a thorough knowledge in the technology of creation, operation and characterization of energy-efficient materials (EEM) with a set of special properties (thermal, electrical, magnetic, mechanical, etc.), which ensure efficient operation of power plants and systems. For example, without this knowledge it is impossible to design, create, and ensure efficient use of boilers, steam pipes, vapor generators, turbines, transformers, power lines, solar panels and collectors, thermoelectric generators and other units in the systems of generation, transformation, transmission, distribution and utilization of heat and electric power.

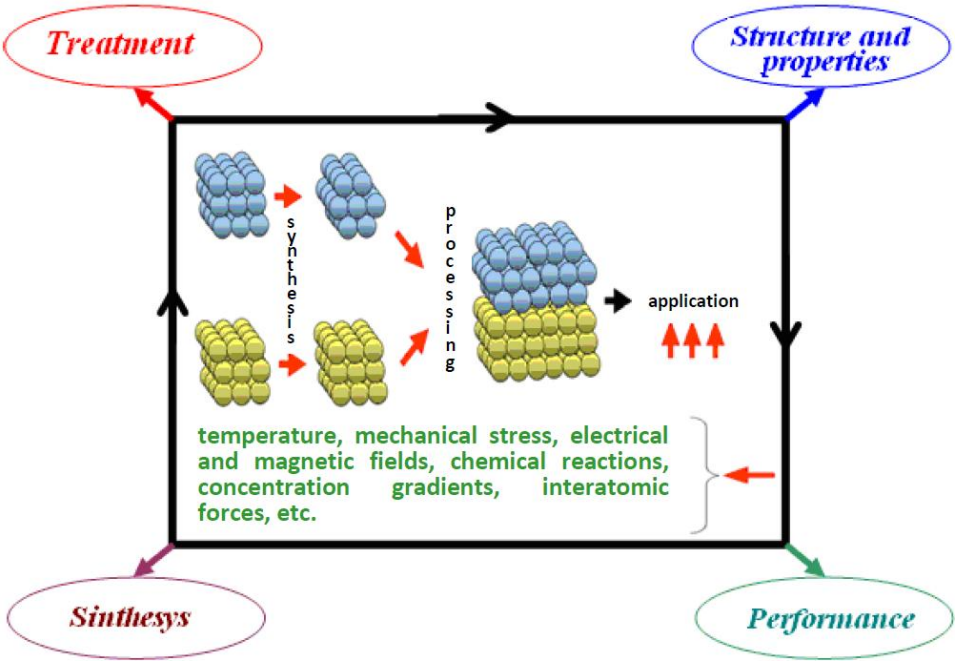
A significant portion of the units in the heat, thermoelectric, solar and other stations operates under intense thermal impacts. Under thermal impacts we mean here every change of temperature due to abstraction/input of heat to the system (media, material, ware) for a certain time, by a certain law, and at a certain speed. Moreover, in some cases, the material making up certain units of power station is subjected not only thermal, but also mechanical and chemical impacts. All this can lead to a significant change in the structure, phase composition and materials properties, including due to the intensification of the processes of heat and mass transfer.

Applied to heat and mass transfer in gaseous, liquid and even solid media, we consider them very often as a continuum abstracting from their discrete (atomic) structure. However, when studying the thermal impacts on the processes of heat transfer and diffusion in crystalline materials, we should take into account atomic and electronic structure of the crystal without fail. This is due to the fact that thermal effects can result, for example, to the intense formation of crystalline defects, their migration and the change of aggregate or phase state of the (for example, the transformation of the crystal in the melt or vapor state, recrystallization of amorphous and polycrystalline materials, etc.). We shall not be able to describe kinetics and dynamics of these processes without taking into account such feature of the crystalline material as its ordered atomic (latticed) structure. Therefore, the change of properties and phase transformations, induced by heat and mass transfer in crystalline materials, can not be described, in principle, without the taking into account the type of crystal lattice and existing defects. In addition, these processes in crystals with a high

concentration of electrons (especially in metals and semiconductors) are essentially dependent on the state of the electronic subsystem. On the other hand, migration of atoms (diffusion) in semiconductors, is determined by the charge state of impurities and defects (vacancies, interstitials, dislocations, grain boundaries, etc.), and, therefore, also depends strongly on the state of the electronic subsystem in the crystal (the Fermi level position). In general, the phenomenon of atomic diffusion in crystalline materials can not, in principle, be described without taking into consideration the atomic and electronic structure of crystals. Mechanisms of diffusion (and, as a result, the activation energy and diffusion coefficients) under the action of heat and other causes (eg, radiation, chemicals and other agents) are completely determined by the material characteristics such as the type of crystal structure (space lattice), the nature and the energy of interatomic bonds, the type, concentration and charge state of the defects in the crystal lattice, the state of the electronic subsystem of the crystal (the concentration and mobility of charge carriers, the position of the Fermi level), etc.

Moreover, there is a deep relationship between the change in the crystalline state of the material as a result of an external (thermal, mechanical, chemical, radiation, etc.) impact and the change in its properties. That is why, to explain the effects of external (including thermal, mechanical, chemical, radiation, and combinations thereof) impacts on the phase state, thermal, mechanical, electrical, magnetic and other properties of crystalline materials we must have a full idea about their atomic and electronic structure. This means that the existence of intense heat and mass transfer in materials used in the units of power systems requires the development of training manuals (handbooks) on EEMs for students of physics and engineering areas identified above.

Block diagram of this book construction derives from interconnection shown in the figure between the "design" (production, growth, synthesis) of the material, its subsequent processing by different methods (thermal, mechanical, thermo-chemical, chemical, etc.) in order to make the hardware with the needed structure and properties which transform the material into a finished product with certain performances required for specific practical applications.



In concordance with the proposed scheme in the figure, in order to create ("construct") material with desired properties, which will provide the optimal performances of the future hardware, we need to know how the structure and properties of the material respond to its composition and external impacts under which it will have to work. This allow to determine the correct technology for synthesis (production) of the material and also its further processing to produce hardware or devices with the required operating parameters.

Under the proposed scheme, the book EEM will consist of three parts.

«Chapter 1. *Fundamentals of material science*» of the book is devoted to presenting the basic physics of condensed matter for students qualified with competence of Physicist-Engineer, who deals with the phenomena and processes in crystalline materials under the impact of different types of thermal, mechanical, magnetic, electrical, radiation, chemical and other influences. This part reviews the crystalline and electronic structure of solids with the different nature of the interatomic bonds, the dynamics of atoms in the crystal lattice, formation of different damages (defects) in ideal crystalline structure, the fundamentals of the band model of the electrons energy spectrum in crystals (including the basic of semiconductor physics), fundamental properties of crystalline materials (magnetic, insulating, semiconducting).

«Chapter 2. *Phase transformations in solid materials*» of the book is devoted to the study how different types of materials response (as a thermodynamic system) on the change of composition, structure and external impacts. In this Chapter we describe some of the technologies for materials synthesis, and also effects of thermal, mechanical, thermal, mechanical, and thermal-chemical, radiation effects (including their joint influences) on the phase state of crystalline materials of different composition, structure and properties. We also describe here equilibrium phase transitions in solid materials, the impact of long-range and short-range order on the processes of phase transitions in condenced matter, as well as the properties of crystalline solids. Just in this Chapter we give a summary of the interconnection between the chemical composition, phase state, structure and properties of materials and the influence of the thermal processes that occur in them both at the stage of materials creation and in the course of their operation in units of power systems with high heat release. Finally, as an example, the phase diagram of the iron-cementite (steels and irons) as the main constructive material is described.

«Chapter 3. *Properties of materials for power industry and energy saving*» devoted to the study of metallic and non-metallic, inorganic and organic EMMs, which are necessary to improve energy efficiency and provide energy saving in the operation of boilers, vapour pipes, vapor generators, turbines, heat transformers, power lines, solar panels and collectors and other units for thermal power generation, transformation, transmission, distribution and use of heat and power, including in conditions of intense thermal, mechanical, chemical, radiation and other impacts. Moreover, some alternative materials and alternative energy (nuclear, geothermal, thermoelectric, hydrogen, etc.), as well as materials for the building envelope and renewable energy are discussed. Finally, review the current methods and techniques for diagnostics of the structure and properties of EEMs.

«Chapter 4. *Modern methods of evaluation of parameters for energy effective materials*» of the book is focused on the study of new methods for characterization of materials which are used in power industry, energy effective technologies and renewable sources. They include a wide spectrum of photoelectric, photoacoustic, photothermal and other methods which allow to estimate recombinational, optical and thermal parameters of metallic, semiconducting and other solids.

The proposed book assumes knowledge by students of modern branches of general and theoretical physics such as molecular and atomic physics, electrodynamics, quantum mechanics, thermodynamics, and quantum statistics. The authors believe that the modern expert in material science should not be limited knowledge regarding only the functional characteristics of materials which are used in certain specific (rather narrow, very often) practical

applications (eg, heat or electricity). A graduate of the classical or modern technical university, working in the field of the use or development of up-to-date materials for energy effective technologies, must have a sound knowledge on the modern physics of solids regardless of whether he is an engineer, technologist or researcher. Without such knowledge, a graduating student will not be able to become expert in the material science and deeply understand the basic principles of materials "construction" that provide their structure (on its atomic, nano, micro and macro levels) and, as a result, the certain functional properties of materials, whether mechanical strength, electrical or magnetic properties. Understanding the nature of the response of the atomic and electronic subsystems of the crystal on various external impacts can not, in principle, be possible without knowing at least the basics of modern quantum theory and statistics of solid.

Of course, to create a relatively brief, but satisfying all the above requirements, the manual – a very difficult task. This book was made possible due to long experience of authors in reading courses in solid state physics, physical material science, and certain branches of semiconductors and a number of related issues in the universities in Belarus (Belarusian State University) and Poland (Lublin University of Technology and Koszalin University of Technology).

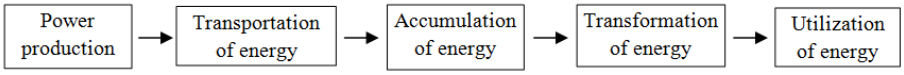
Introduction

This book is overview of 3 directions of modern material science: (*a*) fundamentals of solid state physics, (*b*) phase transformations crystalline materials and (*c*) existing crystalline and non-crystalline, inorganic and organic materials which supply energy effectiveness in heat and electric power production, energy saving technologies, operation of renewable energy resources and performance of devices and hardware which control work of power units, energy saving and materials production. It also gives examples of problem sets, recitation and practices for improving the study of EEM.

1. Fundamentals of materials science

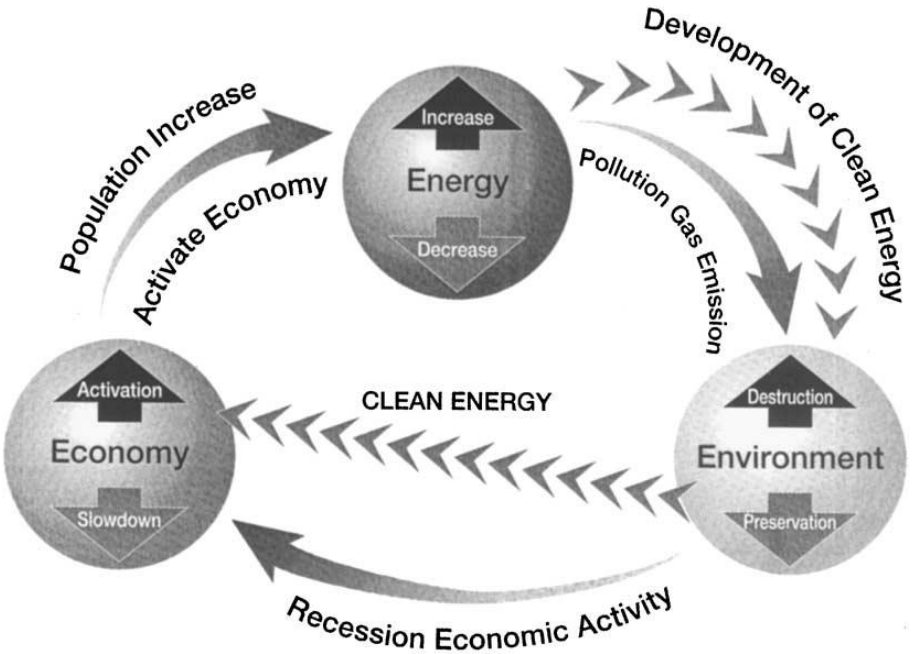
Role of Power in contemporary society

To ensure the economic prosperity of people in contemporary society it is necessary to extract energy, convert it from one form to another, transfer at a distance and consume a variety of ways. This requires an efficient energy system, regardless of the energy source and the means of power production, which should ensure rationalization (energy effectiveness) of the following energy chain:



Energy efficiency of this chain is determined, first of all, by the effectiveness of every of its elements and their interlinks. Improving the efficiency of each element and the chain as a whole can be achieved in different ways. The main ones include lowering of costs for hydrocarbon, increasing of the efficiency of power units and technological equipment with high power consumption and reducing of losses during transport, transformations and use of energy produced. As a result, ensuring the efficiency of the energy chain is provided by a set of technical and organizational methods that are determined by the type of fuel (energy source), the type of energy systems (thermal, nuclear, photovoltaic or solar thermal power, thermoelectric stations, etc.), the nature of energy

consumption (thermal or electric, industry, agricultural sector or municipal engineering, etc.), the economic infrastructure of the state, etc.



Different kinds of Power. Development of energetics and power technologies has led to more and more complex methods of energy production and utilization. Until the 17th century the needs of people in power were provided mainly by the burning of wood, and also the use of wind, water and sun energy. In the 18-19 centuries a coal becomes heavily used for energy production. However, until the 20th century, the use of coal was only in its combustion for heat and electricity production. Currently we use in such a manner only natural gas (after minimal chemical processing), while the oil is used as fuel only after deep petrochemical processing.

Energy development in the 20th century led to strong improving of energy technologies both for coal and other hydrocarbon fuels, as well as the construction of hydroelectric and nuclear power stations. As a result, generation of electric and heat power became the product of a large scale, multi-step and complicated transformation of fuels. This trend to more complex energy processes and technologies, without doubt, will continue in the future.

Power problem in the World. As noted above, the provision of economic well-being of mankind still led to continued expansion of energy consumption. This power production scenario has led to a number of negative consequences –

ecological, first of all. People went to recognition that there is a global energy problem, which is due to exhaustibility of hydrocarbon fuels and environmental sequences of their use. This has forced reconsideration of extensive development of energy consumption as the only way to improve the living standards of people. As a result, in the 21st century global energy strategy, and therefore the whole of civilized life on Earth, is built on the basis of the recognition of the so-called 3E-Trilemma existence (see figure below). 3E trilemma, which is the most important issue for the 21st century's civilization life, establishes the relationship between the development of economy, power production and the state of environment. According to Y. Hamakawa[1], (*Yoshihiro Hamakawa. Solar PV energy conversion and the 21st century's civilization. Solar Energy Materials & Solar Cells 74 (2002) 13–23*) in the regular way along capitalism, for the activation of economic development (E: Economy), we do need an increase of the energy expense (E: Energy). However, it induces environmental problems (E: Environment) by more emissions of pollutant gases. On the contrary, if the political option chooses the suppression of pollutant gas emission, it inactivates the economical development. This is the essence of the 3E-trilemma as the wording of the global availability of the energy problem!

Energy saving and energy effectiveness. A natural way to resolve the above stated 3E-Trilemma, which does not reduce the life quality, is improving the efficiency of energy use. More rational use of energy resources (fuels) can be achieved by the improving of power technologies, including the development of clean technologies and energy saving, on the stages of transmission, transformation, storage and utilization of energy. As a result of this approach, starting from the 70th of the last century, many countries with high energy consumption (U.S., Canada, Germany, France, Holland, etc.) began to aspire to increase the ratio of the gross national product (GNP) in comparable prices to the level of energy consumption (while in the first half of the last century, it remained roughly constant). I was the trend to rationalization of energy consumption, including through the redistribution of the role of various types of production and services in the economy of nations. For example, new types of equipment and technologies have arisen which allowed to lower levels of energy consumption. This is, first and foremost, for the IT technologies, such as computers, telecommunications equipment, mobile phones and many others.

Another obvious and prominent reason is that the resolution of the formulated 3E-Trilemma promotes the creation of new or substantial improvement of old materials that are essential to the production of new energy-

[1] *Yoshihiro Hamakawa. Solar PV energy conversion and the 21st century's civilization. Solar Energy Materials & Solar Cells 74 (2002) 13–23*

efficient technologies and energy saving. *Materials may not only lower the cost of producing energy, but also increase the efficiency (rationality) of its consumption.*

Energy effective materials as basis for effectiveness of power industry and energy saving. As noted above, the effectiveness of any power system is provided by a set of technical and organizational methods, which among others include the use of materials having the appropriate functional characteristics (properties). That is why the life of a modern industrial society is constantly accompanied by the continuous development of the material science and depends on it.

Material – this is a real object with the specified functional purpose. To perform these functions, the material must have a specific composition, atomic structure, micro- or nanostructure and properties. The functions that perform materials – various. This might be ensuring of optimum electric current transport (for example, in conductive and semiconductor materials), maintaining a certain shape under mechanical load (in structural materials), ensuring very low electric current conduction (in dielectric materials), the conversion of electric energy into heat (in materials with high electrical resistance), etc. Typically, the material performs multiple functions. For example, insulating (ceramic or resin) and aluminum conductor materials in an electric cable necessarily undergo mechanical loads, which means they are both structural materials.

The role of materials is essential at all stages of the above energy chain: production and processing of fuel, the conversion of one form of energy into another, the distribution of energy between consumers and its efficient use is essentially dependent on the presence of some special materials. Materials can serve catalysts of chemical reactions (eg, at refining), converters of solar radiation into heat or electricity (in PV panels and solar collectors), electrical conductors in electric power transmission lines, transformers and electric machines. Some of them should be capable to resist corrosion and high temperatures, withstand large radiation impacts, and maintain high strength under mechanical loads and high temperatures. Walling materials must defend our houses against frost or overheating, reflect or absorb heat radiation, etc.

Physical material science – the science that studies the interconnection between the composition, atomic and grain structure, physical properties of materials and their behavior under various external impacts - thermal, electric, magnetic, chemical, radiation, etc., as well as the combination of these effects. Technical progress of mankind is largely based on material science. It also provides new opportunities, methods, tools for the development and diagnostics of materials, allowing us to create their new compositions.

People has learned to use a lot of natural EEMs for their needs. But much more materials we obtain by artificial means. This leads to continuous development and improvement of the various ways to obtain materials that are united by a common name – *technology of materials*.

This book discusses, in general, materials related to power industry and energy saving. However, a large part of them can be used in other fields of modern engineering and technology. In general, the classification of materials will include three main types of materials – metallic, non-metallic (inorganic or organic), and composite materials.

The main materials that are used in the power industry, as well as for energy saving and in systems for control of energy efficiency, can be divided on the basis of their functions:

1. Materials providing stiffness, static and cyclic strength of power units – engines, machinery, gears, turbines, boilers, etc.
2. Wear-resistant materials (resistance to abrasion).
3. Materials with high elastic properties (springs, shock reducers, diaphragms, etc.).
4. Materials with low density and high specific strength (aircraft engines).
5. Materials resistant to impact of high temperatures and aggressive media (heat-resistant, corrosion-resistant).
6. Special-purpose materials (semiconducting, conducting, magnetic, insulating, optical, etc.).

The above illustrates the critical role of power and materials in modern industrial society. Moreover, we can say that our life is constantly accompanied by the continuous development of material science and depends on it. In the history of the power development the increasing role of materials in extraction, transformation, transportation, accumulation and consumption of energy is mainly due to two reasons.

The first reason is that the energy, as far as power industry and technologies developed, is used ever more complicated ways. This trend to energy processes and technologies meshing, without doubt, will continue in the future. Indeed, among all the conceivable and developed energy processes, probably, only the solar power for heating and hot water supply is based on the consumption (extraction) of natural energy in a "pure" form. All other energy processes (nuclear, chemical, mechanical, photovoltaic, etc.) demand their transformation, which requires the use of a variety of materials.

The needs of increasingly complicated power industry in the materials hold both through the development of new types of them, and by modifying or improving the properties of yet existing ones. Properties of a material reflect

those stages and processes that are required for the conversion of raw material into the final product.

An example is a steel as material widely used in power industry. The steel may be forged (as rolled or stamped products produced by mechanical deformation of steel ingots) or cast (as molded slugs). As a consequence of different technologies, propertiesand the cost of forged and cast steel are substantially different: the first possesses greater strength, but the second is cheaper.

Therefore, efforts to develop materials for use in increasingly complicated technologies of power production were directed mainly on the supply of the best their performance at a minimum cost of production (in fact, those with the best energy expenditures), and the fulfillment of certain environmental requirements for the production process and the material itself.

The second reason of the increasing role of materials in power industry is that the development/improvement of materials have ingenuous impact on reducing the cost of power production and energy efficiency. Reducing the cost of energy is due to the use of less power-consuming materials (remember the example of the steel) with significantly improved properties (eg, various composites), as well as materials that can realize a very different, more energy-efficient technologies in power industry.

Solar cells (SCs) can serve an excellent example of how, after a sequence of improving stages, we can reduce the cost of energy produced. First SCs were successfully used in space since about 1974. At the same time, the terrestrial use of SCs in the world, as they do not pollute the environment and do not require the use of fossil fuels, is still low because of their high cost. Nevertheless, thanks to the development of new technologies to produce silicon wafers, the cost of SCs has decreased over 40 years many times (down to about $ 1 per 1 W).

The increase in the efficiency of production and consumption of energy is possible by optimizing the design of power units and improving the functional properties of the materials. The latter can be demonstrated by the example of turbines used to generate electricity on the heat station. The efficiency of any steam turbine can not exceed theoretical limit defined by the second law of thermodynamics. Therefore, to improve efficiency of steam energy consumption we need to raise the temperature T_1 at the entering steam turbine blade, leaving temperature T_2 as low as possible. The temperature of vapor which flows on the blades of a typical electric turbine, coal-fired, usually close to $T_1 \sim 810\,\text{K}$ and the temperature of vapor going out the turbine is $T_2 \sim 370$ K. This means that the maximum efficiency of a turbine will be close to $54\,\%$. Because of the unavoidable losses actual efficiency value is even lower. Therefore necessary to develop materials for turbine blades capable to

operate at higher temperatures, which would approach the thermodynamic limit of efficiency. For example, such ceramic materials as silicon nitride (Si_3N_4), silicon carbide (SiC), and others may be in for a long time of withstand temperatures in excess of 1500 K, without breaking. Influence of materials on the efficiency of power units is also manifested in the increase the reliability and security of production, conversion and use of energy.

If we talk about the thermal power-stations, which implement the conversion of thermal energy into electric power and transmit the latter to the consumer, then it requires a great variety of materials with a manifold (and often mutually alternative) properties. For example, the materials for turbine blades should be strength, heat-resistant and corrosion-resistant simultaneously. In systems converting thermal energy into electricity and transporting the latter to the consumer require materials with special electrical properties (light and fast, highly electroconductive, corrosion-resistant, etc.). For example, some units of power generators and power transmission systems either require good electrical conductors (copper, aluminum) or, conversely, insulators having a high dielectric properties (electrical ceramics, plastics, resins, etc.). Transformers, control systems, etc. require special magnetic materials (soft and hard magnetics). In various sensors, transducers and other devices in energy efficiency monitoring and control systems, semiconductor materials and complex metal/semiconductor or dielectric/semicinductor compositions are used.

The contribution that materials science has made and continues to make to the development of technologies for production, transportation, accumulation, conversion and utilization of energy is crucial factor for the economy. We can hope that subsequent achievements of material science will provide development of power technologies to product energy in an amount, needed for sustainable growth of economic prosperity throughout the world, and at the same time reducing the environmental impact on ecology.

1.1. Structure and Bonding

1.1.1. Atomic structure of crystalline and non-crystalline materials

All materials are compositions of one or more chemical elements (atoms) which differ by the electronic structure and hence chemical activity. Forces, which hold atoms and ions in a molecule and/or the crystal are called chemical or interatomic bonds. Interatomic interaction is caused mainly by electrostatic forces of attraction between oppositely charged particles (electrons and cations, cations and anions) and the forces of repulsion between same-charged particles (electrons and electrons, same – charged ions).

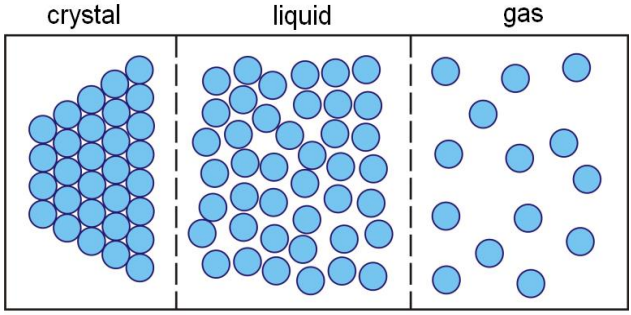


Fig. 1.1.1. Schematic representation of the differences in the three states of aggregation

Depending on the ratio of the potential energy of chemical bonds E_{pot} and the kinetic energy E_{kin} of motion of the atoms or ions at a temperature above absolute zero, all substances in nature are presented in three states (Fig 1.1.1): gaseous, when the kinetic energy of the atoms or ions is much greater than their potential interaction energy – $E_{kin} >> E_{pot}$; liquid – with $E_{kin} > E_{pot}$; solid – if $E_{kin} < E_{pot}$. This ratio of energies determines the physical state of a system of atoms or ions and its properties.

To ensure the transition of a substance from gaseous or liquid to solid state, the particles (molecules or ions) must approach to a specific (optimal) distance when the attractive and repulsive forces between them are in equilibrium. This distance corresponds to the minimum in the $V(r)$ dependence in Fig. 1.1.2. The position of this minimum is exactly corresponds to the equilibrium interatomic distance r_0, which is typically less than 1 nm.

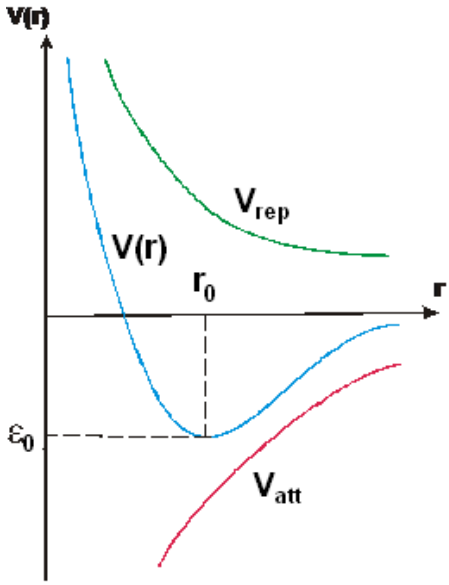


Fig. 1.1.2. The energy of the ions interaction vs interatomic distance. V(r) is total binding energy, Vrep – repulsion energy, Vatt – the energy of attraction, r0 – the equilibrium interatomic distance corresponding to the minimum binding energy

The value of energy in the minimum $\varepsilon_0 = V(r_0)$ is just the energy of chemical bonds or the interatomic interaction energy. The value of ε_0 per atom is defined as the difference between the total energy of the crystal E_{total} and energy and $N \cdot \varepsilon_0$ as in a crystal of N isolated atoms, divided by the number N of atoms:

$$\varepsilon_0 = \frac{\left(E_{total} - N \cdot \varepsilon_0\right)}{N}. \qquad (1.1.1)$$

The value of this bond energy is dependent on the type of chemical bonding (see below) and for the crystalline materials is typically in the range from 0.01 to 5-7 eV.

1.1.2. Chemical bonds in solids

As it was noted above, valence electrons, being on the outer electronic shells in atoms, play the main role in joining of ions in a molecules or solids. Interatomic bonding occurs because the atoms in the material are close enough to each other so that their outer electron shells begin to overlap. As a result of this overlapping, nature of the electrons motion changes dramatically and electrons located at a certain energy level of a single atom, are able to (*a*) pass corresponding to the energy level of the adjacent atoms without energy

expenditure and thus become free to move along the solid, (*b*) move to another atom (forming a cation and/or anion) or (*c*) exchange pairs of electrons between neighboring atoms. Which of these processes is implemented, is primarily determined by the structure of the electron shells of the interacting atoms, i.e. their chemical nature. Depending on the structure of the atomic electron shells, chemical bonds between the atoms (ions)in solids are usually divided into 4 main (ultimate) forms – covalent, ionic, metallic, molecular.

The covalent bond in molecules or crystals is caused by interaction between atoms, when a pair of electrons is shared between two atoms. A major role in the covalent bonding plays a so-called exchange interaction, which has the quantum-mechanical nature. It is due to the Coulomb interaction of electrons with opposite spins, and the influence of the Pauli Exclusion Principle, which takes into account the correlation in the motion of the electrons due to the presence of spin. If diatomic molecule or crystal with a covalent chemical bonds consists of atoms of one element (e.g., H_2, N_2, diamond, Si,Ge), the distribution of electron density between atoms is symmetrical (a pair of electrons belongs to both bounded hydrogen atoms in the same degree). Molecules of this type are called by non-polar or neutral, since the centers of gravity for positive and negative charges (ions) coincide (see Fig. 1.1.3). If diatomic molecule or crystal consists of atoms of different elements, the center of the distribution of electron density in the electron pair can be displaced to one of the atoms. In this case, a covalent bond is called a polar. Molecules with a polar bond, in which the centers of positive charge do not coincide, are called polar or dipole.

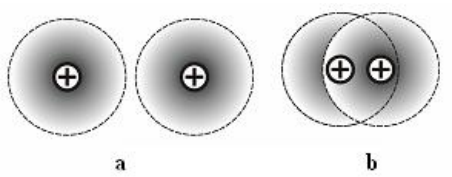


Fig. 1.1.3. Covalent interatomic bond in the hydrogen molecule: *a* – isolated atoms, *b* – a molecule with a non-polar covalent bond

A classic example of covalent crystals are diamond, silicon and germanium. These atoms have two valence electrons in the s- and p-states. When approaching the atoms during formation of the crystal, atoms are rearranged to form 4 joint pairs of electrons, which are common to the nearest neighboring atoms. It can be seen from Fig. 1.1.4 covalent bond in such a structure.

As can be seen from Fig. 1.1.4, electrons in such a diamond-like structure form four covalent bonds linking the neighboring atoms. Since such links are directed along axes of the regular tetrahedron, they are called *tetrahedral*.

The distribution of electron density is found to be highly inhomogeneous, so that covalent bonds are directed along the highest densities of electrons combined. The angles between the bonds in silicon are $109^{o}\ 29'$. Covalent crystals usually have a very high binding energies, which can reach 5-7 eV. As a result, such crystals have the highest strength and refractoriness.

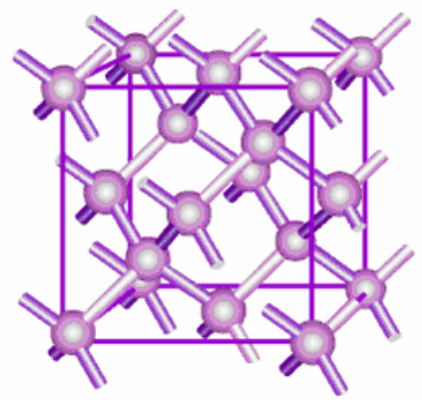


Fig. 1.1.4. Tetrahedral covalent bond in a crystalline diamond (silicon, germanium)

Ionic bonds are due to electrostatic (Coulomb) forces of attraction between positive and negative ions, formed by transfer of electrons from one atom to another (Fig. 1.1.5).

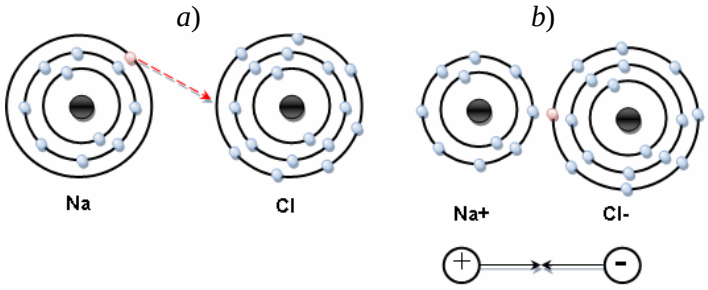


Fig. 1.1.5. Formation of ions (*a*) and ionic bonds (*b*)

The ionic bond is formed in chemical compounds in which one element is a metal and the other is close to the last group of the periodic table (for example, in the crystals of alkali halides such as NaCl, KCl, KBr, LiF, etc., Fig. 1.1.6). In addition, such bonds are typical for many oxides and salts which are composed of ions of opposite signs (e.g., ZnO, CdO, NiO, CuO, etc.).

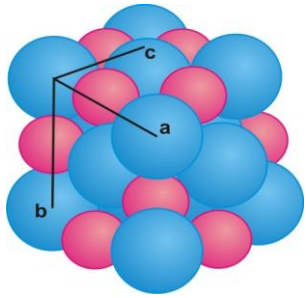


Fig. 1.1.6. Schematic representation of the anions and cations arrangement in an ionic NaCl crystal

The ionic bond is less directed than covalent. Therefore number of nearest neighbors in the materials with ionic bonding is higher than for covalent crystals, and is usually 6 or 8. Ionic bond energy is in the range of 0.5-3 eV. Therefore, ionic crystals are less hard than covalent so that ion yield to covalent crystals in mechanical strength, fusibility and chemical durability.

Metal bonds are formed between atoms of the metal elements, which have the ability to donate valence electrons, becoming positive ions (cations). At the same time valence electrons that leave the atoms become free and called *collective electrons* (Fig 1.1.7). As a result, the metallic crystal can be thought of as a system consisting of positive ions, which are "immersed" in the gas of collective electrons. In this system, there is an electrostatic attraction between cations and the gas of free electrons. This structure causes omnidirectional nature of metallic bonds, leading to the possibility of forming crystals with a maximum number of nearest neighbors (up to 12).

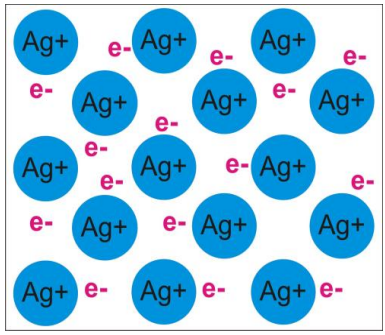


Fig. 1.1.7. A scheme of crystalline lattice with metallic bonding consisting of cations and free electron gas

Energy of metallic bond is in the range of 1-5 eV, which determines a high mechanical strength and high melting points of metals. Due to the large concentration of free electrons metals have high electrical and thermal conductivity and also ductility.

Molecular or van der Waals bonds are formed between the individual molecules or atoms as a result of electrostatic attraction between the charges of opposite signs, which are formed during the formation of dipoles due to shifting of gravity centers of vcalence electrons and atomic nucleus. This electrostatic attraction forces called van der Waals forces. The presence of van der Waals forces is associated with the ability of neutral atoms or molecules to induce in each other instant small electric dipole moments due to fluctuations in the electron density around the nucleus (polarization) due to collisions of atoms or other causes (Fig 1.1.8). On average, the interaction between the induced dipole moments in neighboring atoms will lead to their attraction, which is energetically favorable, as the energy of the system decreases.

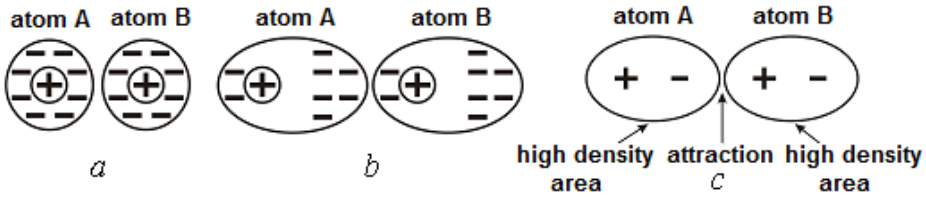


Fig. 1.1.8. The scheme of a van der Waals bonding: the initial atoms A and B (a), the atoms with polarized electron shells (b), attraction of the formed dipoles (c)

Under the influence of van der Waals forces the electrically neutral atoms of inert gases form crystals at low temperatures due to induced dipole-dipole interaction. With such forces the solid molecules of hydrogen H2, nitrogen N2, carbon dioxide CO2 are also formed.

The binding energy values in crystals with the van der Waals interaction is for oneor two orders of magnitude lower (0.01-0.2 eV) than in ionic ones. Therefore, such crystals have low melting and boiling points.

1.1.3. Crystalline and space lattice

Structural elements of solids (atoms, ions or molecules) are located either in an orderly fashion (Fig. 1.1.9a) or randomly (Fig. 1.1.9b). Substances with a regular arrangement of atoms in space, obtained under conditions of thermodynamic equilibrium, are called crystalline. Being synthesized in a non-

equilibrium conditions, these substances become amorphous (highly disordered) and are characterized by a higher energy.

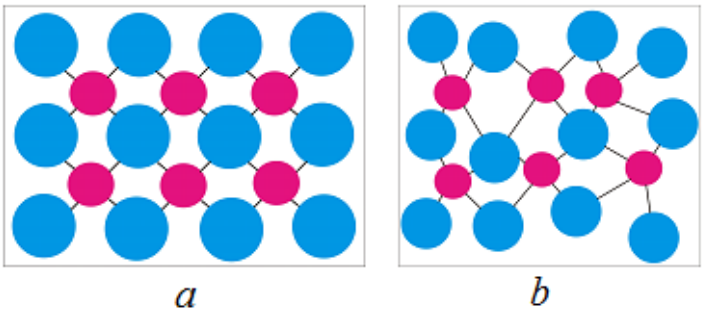


Fig. 1.1.9. Arrangement of atoms in a crystalline (a) and amorphous (b) solid material

Since the internal structure of crystalline substances is characterized by a regular periodic arrangement of their atoms, one say that the latter form a geometrically regular periodic *crystalline lattice* (Fig. 1.1.9a). Due to the spherical symmetry of simple atoms, their crystalline lattice is often portrayed as a contiguous balls (Fig. 1.1.10a). However Bravais (to formalize and simplify the illustration of the atomic structure of crystals) introduced the concept of the *space lattice* in which the system of balls/atoms in space is presented by a scheme in which the centers of balls gravity are replaced by dots, called *lattice sites* (Fig 1.1.10b).

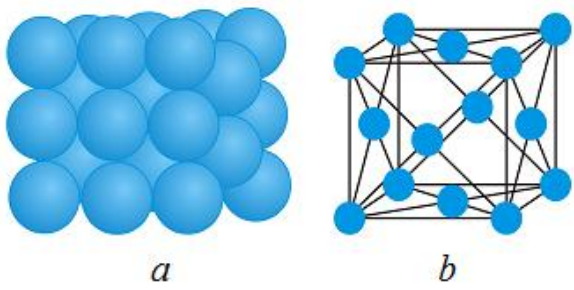


Fig. 1.1.10. Examples of crystalline (a) and space (b) lattices

Thus, the space lattice (SL) is a system of equivalent points, or sites, put in line to crystalline lattice (CR), which reflects its basic symmetry properties. SL also can be defined as set of ordered points (sites) in the space when surrounding of every site is identical to the neighborhood of all the othersites.

Since the lattice structure is the main feature of the crystals so the latter can be defined as a next. Crystal is a solid body in which the structural units (atoms, ions, molecules) are arranged regularly in the sites of SL, being grown under conditions of thermodynamic equilibrium, and have the shape of regular plane surface bodies.

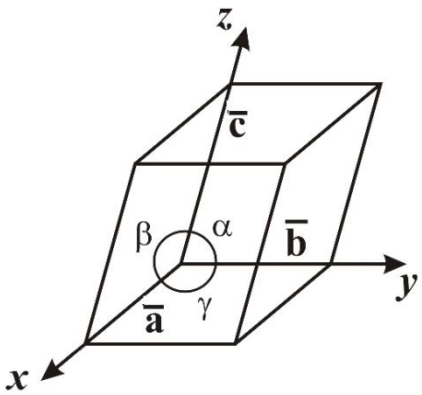


Fig. 1.1.11. An example of an oblique unit cell

Fig. 1.1.11 shows the principle of the SL building with the help of three non-coplanar translation vectors $\vec{a}$, $\vec{b}$, $\vec{c}$, the directions of which respectively coincide with the directions of the axes *x*, *y*, *z*, forming a crystallographic coordinate system. The angles α, β and γ between the axes are called *axial* or *coordinate angles*.The described principle of SL construction allows to describe it mathematically. If one select any SL site as the origin, any other lattice point can be determined by the radius-vector

$$\vec{R} = m\vec{a} + n\vec{b} + t\vec{c}, \qquad (1.1.2)$$

called a translation vector. Here, *m*, *n*, *t* are random integers (from zero to infinity) which are called the site indexes. Lattice constructed using the translation operation (1.1.2) is called *a simple Bravais lattice*, and a parallelepiped, built into the crystallographic coordinates for the three basis-vectors, is called a *unit cell of the Bravais lattice*). Modulo of the translation vectors, which determine the length of the cell edges in Fig. 1.1.11, are called lattice constants or identity periods. Lattice parameters. Lattice parameters can be determined by X-ray analysis (see below) and are measured in nanometers ($1\ \text{nm} = 10^{-9}\ \text{m}$).

Primitive and nonprimitive cell. As can be seen from Figs. 1.1.11 and 1.1.12, structural lattice elements (ions, atoms, molecules) or sites, respectively, should be situated without fail.at the vertices of an unit cell of CD or SL. If the unit cell, based on the shortest translation vectors $\vec{a}$, $\vec{b}$, $\vec{c}$, has a lattice sites

only at the vertices, it is called a ***primitive cell***. Examples of such two-dimensional primitive cells are shown in Fig. 1.1.13a, b, c. Such cell has two important properties: it contains only a single site on one cell and has *the least volume* among all the elementary unit cells.

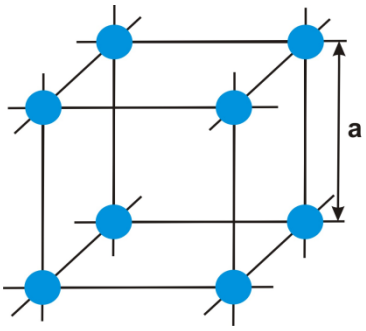


Fig. 1.1.12. Example Brava cell for a simple cubic lattice

However, sometimes the description of the SL (and CR too) through non-rectangular elementary parallele-piped with the smallest unit cell volume (which may be oblique, as in Fig. 1.1.13b, c) is not appropriate. In this case, we can choose non-primitive cell wit the larger the volume, but rectangular, Fig. 1.1.13d,e. Such a cell contains not only the sites at the vertices, but inside itself, as well as on edges, and (or) faces. Therefore, when counting the number of atoms per unit cell,one should keep in mind that each site or atom belogs to many neighboring cells. For example, the cubic lattice shown in Fig. 1.1.12, every atom in the top of the cube, at the same time belongs to 8 cells.

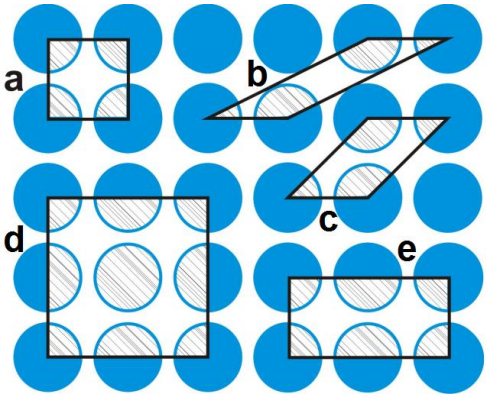


Fig. 1.1.13. Examples of primitive (a, b, c) and non-primitive (d, e) unit cells

Closest ball packaging. As noted above, the spherical symmetry of the electron shells of some atoms (ions) makes it possible to represent them in the CL like a hard, non-compressible spheres of certain radius R, between which there are forces of attraction or repulsion. In this case, the atomic structure of a crystal may be concidered as an ordered spatial packing of hard spheres. The energy of such system will be minimal, if this is the closesest packing.

Fig. 1.1.14 shows the case of close-packing when every ball in a flat bottom (first) layer of the balls is surrounded by six other similar R balls and six, respectively, triangular voids. At the construction of the second (upper) layer of close-packed spheres, we put balls in the voids of the same type. After packing the first two balls layers, two types of voids (pores) are generated – *tetrahedral* and *octahedral*. Tetrahedral pore is surrounded by 4 balls (numbers 1-4 in Fig. 1.1.14 and Fig. 1.1.15a) of the two layers, the centers of which form a tetrahedron (Fig. 1.1.15b). Octahedral are through pores and each of them is surrounded by 6 balls (3-8 in Fig. 1.1.14 and Fig. 1.15b) – by 3 balls in every layers, which are rotated by 60° relative to each other. The centers of the balls form an octahedron (Fig. 1.1.15d). The tetrahedral and octahedral pores have different volumes, which are typically characterized by a maximum size (radius r) of a ball which can be inscribed therein.

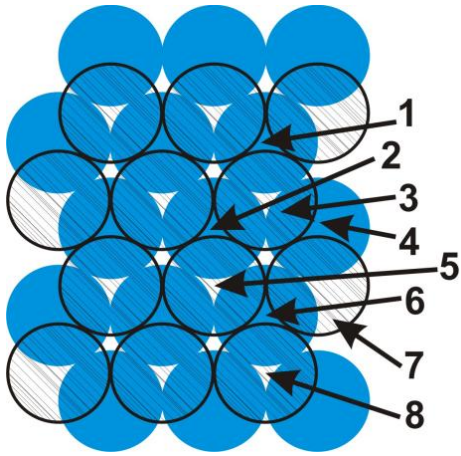


Fig. 1.1.14. Closest packing of two layers of balls. Balls of the second (upper) layer are shaded

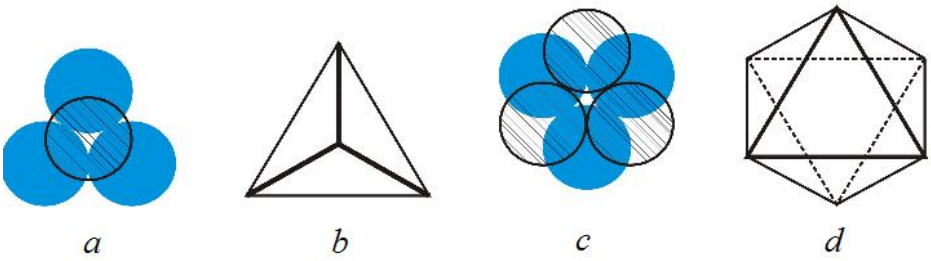


Fig. 1.1.15. Tetrahedral (a) and octahedral (c) pores formed by two layers of balls at their closest packing; polyhedra formed by the centers of the spheres during the formation of tetrahedral (b) and octahedral (d) pores

If the balls of the third layer are laid on the tetrahedral voids of the second layer, their centers will be located just above the centers of the balls of the first layer. Such close-packed structure is *double-layered* and is called *the hexagonal close packed* (hcp), with alternating layers of type ABABAB ... (Fig. 1.1.16a).

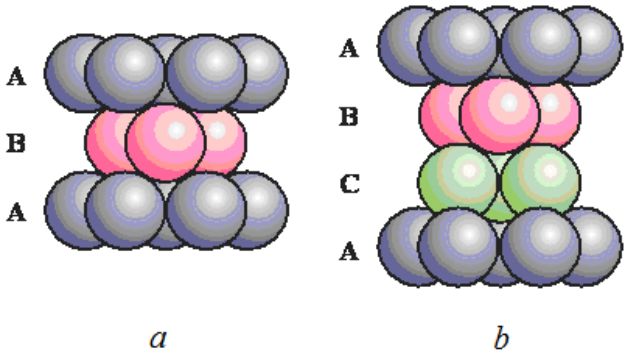


Fig. 1.1.16. The hexagonal (a) and cubic (b) close packings

If the balls of the third layer (layer C) are laid on the octahedral voids, and the balls of the 4th layer directly over the balls of the first layer, then this package is called a trilaminar or a *cubic close packing* with alternating layers of type ABCBCA... (Fig. 1.1.16b).

The main types of crystal structures. The nature of packaging (type of lattice) for each solid material depends on the electronic structure of its constituent atoms and type of chemical bond between them. Examples of metals have often found these types of crystal structures (types of unit cells) as a body-centered cubic (*bcc*) – Fig. 1.1.17a, face-centered cubic (*fcc*) – Fig. 1.1.17b and hexagonal close-packed (*hcp*) – Fig. 1.1.17c. As shown, the elementary bcc cell is not primitive because contains 2 atoms for one cell, not only in the vertex 8

carbon atoms, but one atom in the center of the cube. In the fcc lattice every unit cell (also non-primitive) contains 8 carbon atoms at the vertices and 6 atoms in the centers of the faces of the cube (i.e., 4 atoms per unit cell). In the hcp lattice unit cell has the form of a hexagonal prism and contains 12 atoms at the vertices, 2 atoms at 2 faces and 3 atoms inside the prism (ie, 6 atoms per unit cell).

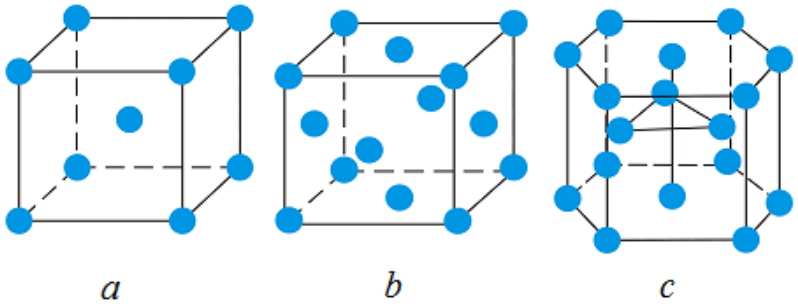


Fig. 1.1.17. The main types of crystal (spatial) lattices of metals: *a* – body-centered cubic (*bcc*), *b* – face-centered cubic (*fcc*), *c* – hexagonal close-packed (*hcp*)

In the case of cubic (bcc and fcc) structures all the edges of the cell (the length of translation vectors) are the same and form right angles to each other (Fig. 1.1.17a, b). Therefore, the parameters of cubic lattices are characterized by long edges of the cube and are denoted by the letter *a*. To characterize the hcp lattice we take two parameters – a hexagon side (along the *x*-axis) and height of the prism (along the axis *z*). When the ratio $c/a = 1.633$, the atoms are packed most tightly, forming the hcp structure. Some metals have a hexagonal lattice with a less dense packing of atoms: for example, for zinc $c/a = 1.86$, for cadmium $c/a = 1.88$.

The simplest structures of semiconductors are diamond-like (Fig. 1.1.18a), sphalerite (Fig. 1.1.18b) and wurtzite (Fig. 1.1.18c). The diamond-type structure is typical for diamond (consisting of carbon atoms), Ge and Si. The unit cell of a diamond refers to the fcc structure.

However, in addition to atoms at the 8 vertices and 6 centers of the cube faces, diamond crystal cell contains 4 extra atoms (gray balls are shown in Fig. 1.1.18a). These atoms are in 4 of 8 octants in which fcc unit cell can be divided in Fig. 1.1.18a. As a result, such cell contains 8 carbon atoms, and each of this atoms adjoins with 4 nearest neighbors (tetrahedral neighboring).

The sphalerite-type structure (ZnS, CdS, AlP, AlAs, GaAs, etc.) looks similar to diamond-like (Fig. 1.1.18b). However, 8 atoms belonging to the cell

are divided into two classes: 4 atoms of one type belong the vertices and centers and 4 atom of another sort are in the centers of 4 (of 8) above mentioned octants.

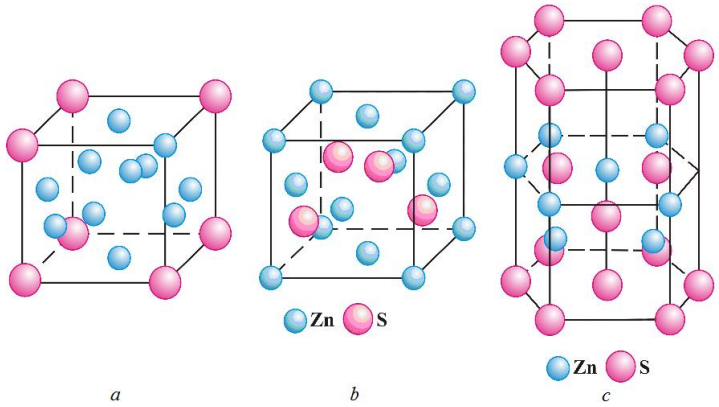


Fig. 1.1.18. Examples of unit cells for the diamond (*a*), sphalerite (*b*) and wurtzite (*c*) structures

Wurtzite structure (CdS, ZnS, ZnO, InSb, etc.) corresponds to the hcp lattice (Fig. 1.1.18c). However, in the unit cell of this structure there is an additional number of atoms of a different sort on the edges and inside the hexagonal prism. For this structure is also typical a tetrahedral environment. Knowing the type of the lattice, we can determine a coordination number and atomic (or ionic) radius for the structure.

Coordination number of the lattice structure is the number of the nearest neighboring atoms surrounding every atom. As can be seen, in a bcc lattice coordination number is 8. For fcc and hcp lattices coordination number is 12. The *atomic* (or *ionic*) *radius* in the crystal lattice is defined as half the distance between the centers of the balls-atoms (ions) belonging to the edge on unit cell (Fig. 1.1.12).

The close packing model allows us to calculate the *fill factor* or the *compactness Q* of the crystal structure. It is defined as the fraction of the volume of the unit cell occupied by *n* atoms or ions of radius R:

$$Q = \frac{4\pi R^3 n}{3V} \cdot 100\%. \qquad (1.1.3)$$

where: *V* – volume of the unit cell, *n* – number of atoms per cell. According to (1.1.3), for a simple cubic lattice *Q* = 52 %, for the bcc structure – 68 %, and for the fcc and hcp lattices – 74 %. Thus, among these four structures the fcc and

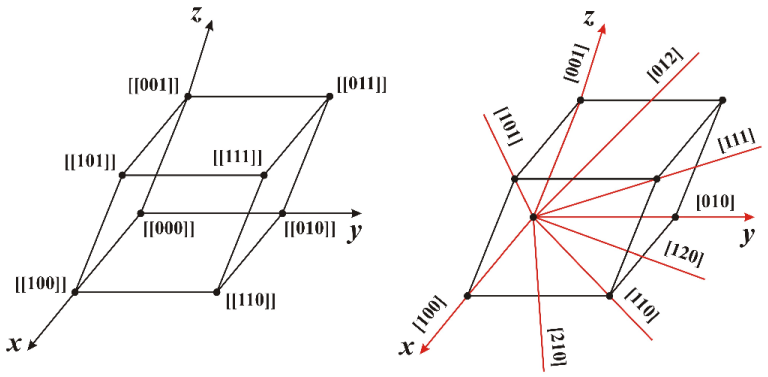
hcp lattices are the most closely packed. The changes in fill factor of lattices result in changes of interatomic pores dimensions, that is important for incorporation of foreign atoms into the crystal lattice (for example, the formation of complex compounds or alloys). As is seen, in the fcc and hcp lattice atoms occupy 74 % of the total volume of the crystal lattice, and interatomic voids – 26 %. In bcc lattice atoms occupy 68 % of the total volume, and nly 32 % – pores.

From geometrical considerations, we can easily find that the radius *r* of the octahedral pores in the close-packed fcc and hcp lattices, having the same coefficient of compactness, is 0.41 of atom (ion) radius R, and the radius of the tetrahedral one is only 0,22R. Since the bcc lattice has a lower *Q* values, so octahedral pore radius therein is only 0.154R. In the same tetrahedral pores of bcc structure atoms can be held with a radius of 0.29R. In accordance with the theory of the close packings, spherical monoatomic crystal structure should be formed by hexagonal (*hcp*) or cubic (*fcc*) laws, as it is observed in many crystals of inert gases and the most metals. Deviations from the close packings (for example, the appearance of more loose bcc or diamond-like structures) may be associated with features of the interatomic interaction (the nature of the chemical bonds) in crystals. In non-monoatomic crystals, most large ions (usually anions) are arranged in the space under the laws of the close packing, whereas smaller ions (cations) are situated in the octahedral pores, such as in the structure of NaCl.

1.1.4. Methods for atomic structure description

As can be seen from Section 1.1.3, the unit cells of solid crystalline materials can fit the form of a rectangular and oblique parallelepipeds. Therefore, to descript the crystal structures with different kinds of unit cells, special *crystallographic indexing method* has been developed. It allows to describe the SL (CL) in the symbolic form. This method allows to describe uniformly positions of sites (atoms), and also specific directions and planes in lattices. This metod is irrespective to the crystallographic coordinate system (rectangular or oblique), and specific values of modulo for basic translation vectors (the lattice constants) in the coordinate axes.

In the crystallographic indexing method for describing the atomic (or site) coordinates in the lattice one can use the formula (1.1.2), where the radius vector is a linear combination of the basic translation vectors $\vec{a}$, $\vec{b}$, $\vec{c}$. In this case, the set of three indices in (1.1.2), written in double square brackets, gives the coordinates of every site. The [[mnt]] is called the *symbol of the site*. Examples of such symbols for the vertex atoms of the primitive unit cell are shown in Fig. 1.1.19a.



a) b)

Fig. 1.1.19. Crystallographic simbols of sites (*a*) and directions (*b*) in the primitive unit cell

In crystallography all directions are passed (in the form of straight lines) through the sites (atoms) of the lattice. Therefore, every crystallografic direction in the crystal is always possible to characterize by the coordinates of two sites (atoms), through which it passes. If one of the sites is chosen as the origin, the coordinates of the second site will determine the preferred direction. Therefore, the second site coordinates [mnt], recorded in single brackets, can be called a symbol of direction. The same symbol, enclosed in curly braces, describes the whole family of directions (lines) parallel to this. Some examples of the major crystallographic directions are shown in Fig. 1.1.19b.

Every crystallographyc plane also should pass through a system of the space lattice sites (see Fig. 1.1.20a). In this case, any plane that passes through the three lattice sites, that do not lie on a straight line, contains an entire grid of sites. In this case, to describe the position of every crystallographic plane (and system of parallel planes too) in the space, one should define a set of the intercepts, which are cutted off by the crystallographic plane on the coordinate axes x, y, z. According to Fig. 1.1.20a, these intercepts are equal ma, nb, tc, where the numbers m, n, t determine the lengths of the intercepts (in fractions of the modulo of unit translation vectors $\vec{a}$, $\vec{b}$, $\vec{c}$). Fig. 1.1.20a a system of parallel planes is given, where the first (the nearest to the origin) is characterized by $m = n = 1/2$ and $t = 0$. Fig. 1.1.20b shows two planes, labeled as (111) and (112), which cut the axes with intercepts having indices $m = n = t = 1$ (the non-shaded plane) and $m = n = 1$ and $t = 1/2$ (shaded plane).

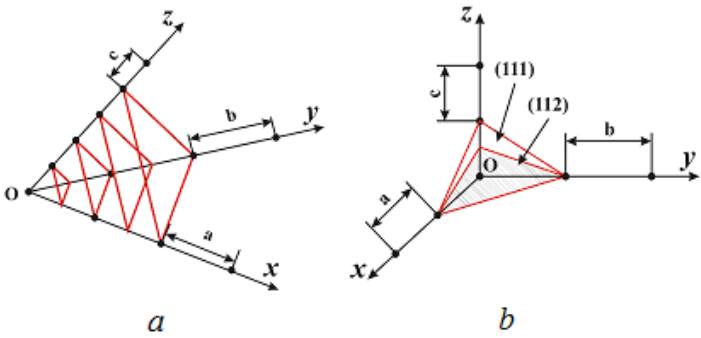


Fig. 1.1.20. Examples of the crystallographic planes construction in the space lattice: a – (221)–type plane intersects the axes x and y in the points with coordinates [[1/2,0,0]] and [[0,1/2,0]], respectively; b – planes (111) and (112) intersect the z-axis at the points [[0,0,1]] and [[0,0,1/2]], respectively. The coordinates are given in fractions of the modulo of unit translation vectors $\vec{a}$, $\vec{b}$, $\vec{c}$

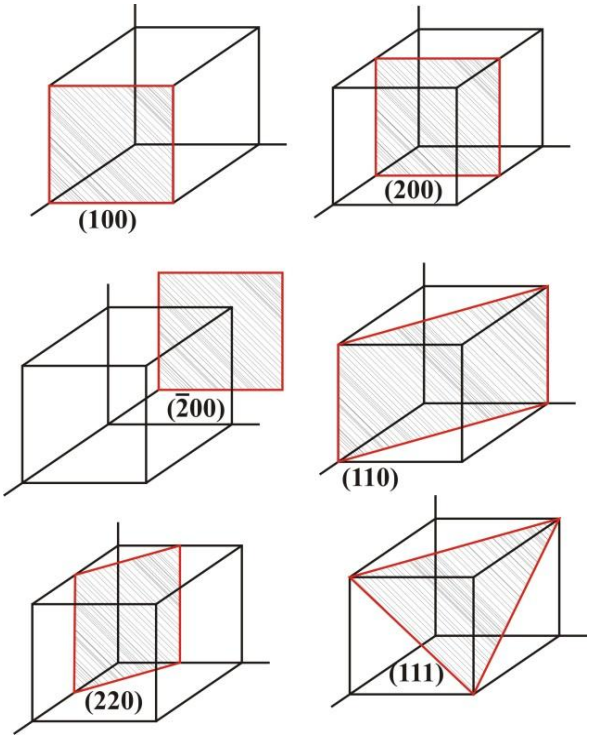


Fig. 1.1.21. (100), (200), ($\bar{2}$00), (110), (220) and (111) planes in the unit cell

However, in crystallography for convenience, the spatial arrangement of the planes is not characterized by *m*, *n*, *t* numbers or intercepts by the axes. They are usually characterized by the so-called *crystallographic Miller indices*. Miller indices (*hkl*) are enclosed in parentheses and indicate the reciprocal lengths of the intercepts, measured in units of the lattice parameter, namely – $h = 1/m$, $k = 1/n$, $l = 1/t$. Therefore for the planes in Fig. 1.1.20a, Miller indices are (221). In Fig. 1.1.20b the non-shaded plane is (111) and the shaded plane – (112).

If the plane cuts off negative intercepts by any axis, a minus sign is put above the corresponding index. It follows that the geometric meaning of Miller indices is that the system of parallel planes wit the indices (hkl) dissects identity periods (or unit cell edges) *a*, *b*, *c* on *h*, *k*, *l* portions, respectively. This means that the symbols of the planes, shown in Fig. 1.1.20b, will (111) and (112). Fig. 1.1.21 shows images of major planes and their corresponding Miller indices for cubic unit cells.

The interplanar spacing *d* is the main characteristic of crystallographic planes family. In particular, for the crystals with cubic symmetry *d* is defined by the formula

$$d^2 = \frac{a^2}{h^2 + k^2 + l^2}, \tag{1.1.4}$$

where: *a* is *the lattice period* (*lattice constant*).

1.1.5. The Bragg method to determining the structure and parameters of the crystal lattice

To determine the atomic structure of solids the diffraction of X-rays, electrons and neutrons is used. All these methods are based on the general principles of the diffraction for waves on crystalline lattice if their wave lengths are close to the average interatomic distance in the material.

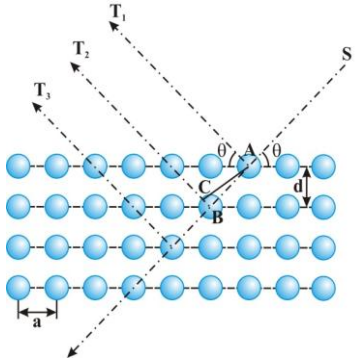


Fig. 1.1.22. The experimental scheme for Wulf-Bragg scattering of X-ray beams by a crystal

In X-ray diffraction (XRD) method wave lengths about 0.7-3 nm are used to determine the crystalline structure of solids,. The diffraction patterns of the crystal characterize the intensity of X-rays scattered from or passed through the crystal that are fixed with special counters of X-ray quanta.

Let's derive a condition that describes the position of the interference maxima of the X-ray beams scattered by a crystal assuming its mirror reflection by the set of parallel atomic planes. Schematic view of the experiment, shown in Fig. 1.1.22, allows to determine the path difference Δ for two parallel monochromatic X-ray beams T_1 and T_2 (with a wavelength λ), which are angled θ to the family of parallel atomic planes. After reflection from the two top planes (at the same angle θ), parallel beams T_1 and T_2 interfere, i.e. amplify or attenuate each other depending on a path difference Δ between them. If the Δ = (AB + BC) for the given angle θ is an integer n of wavelength λ number, then one observe the Bragg interference maximum. The condition of maximum occuring

$$\Delta = n\lambda = 2d\sin\theta \qquad (1.1.5)$$

has been called Bragg formula.

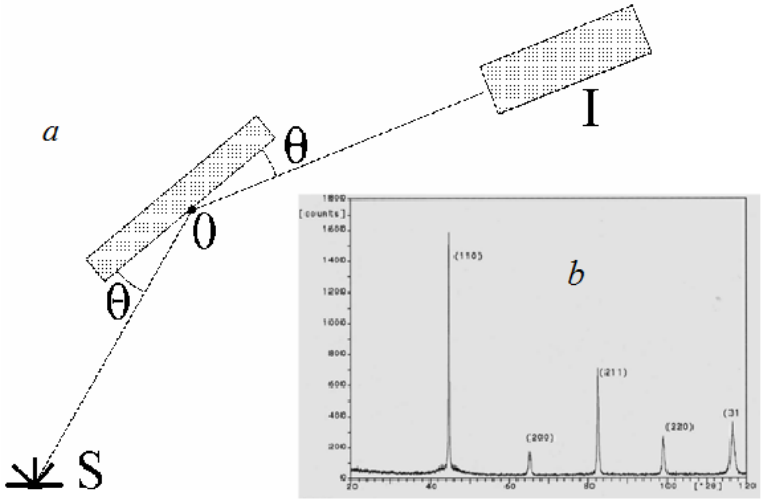


Fig. 1.1.23. The scheme of X-ray experiment using polycrystalline rotation method (*a*) and XRD pattern for polycrystalline iron (*b*). *S* – X-ray source, *I* – detector of X-ray photons reflected from the rotated crystal

The scheme for obtaining of X-ray patterns by polycrystalline rotation method is shown in Fig. 1.1.23a. Bragg reflection angles θ can be determined from the diffraction pattern in Fig. 1.1.23b. Knowledge of Bragg angles θ

allows one to estimate the distance *d* between the crystallographic planes from Eq. (1.1.5). If you know the structure type and the Miller indices of the reflecting planes, the *d* values allow to extract the lattice constant of the crystal using, for example, relation (1.1.4).

1.1.6. Reciprocal lattice

Up to now we have considered a crystalline solid as a system of atoms rigidly fixed in the sites of Bravais space lattice. However, actually the atoms of the lattice are always oscillating around the equilibrium positions. These thermal atomic vibrations lead to the formation of a system of elastic waves propagating through a crystal. If a crystal consists of N identical atoms, 3N such waves arises. They are called *normal* (or *intrincic*) *oscillations*, or *modes*.

In order to describe a system of waves (elastic, electron, electro-magnetic) in a crystal, the concept of the reciprocal lattice is used. This concept is very useful in the analysis of many phenomena caused by the particle-wave nature of atoms, electrons, photons, etc. propagating in a periodic lattice.

The reciprocal lattice is closely related to the direct lattice. Relationship between the basic translation vectors $\vec{a}$, $\vec{b}$, $\vec{c}$ of normal (or direct) lattice and the basic translation vectors of the reciprocal lattice can be expressed as

$$\vec{a}^* = \frac{[\vec{b}\vec{c}]}{V}, \quad \vec{b}^* = \frac{[\vec{c}\vec{a}]}{V}, \quad \vec{c}^* = \frac{[\vec{a}\vec{b}]}{V}, \tag{1.1.6}$$

where: $V = \vec{a}[\vec{b}\vec{c}]$ is the volume of the cell unit in direct lattice.

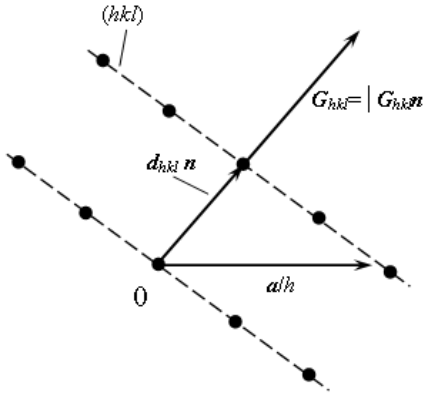


Fig. 1.1.24. To the reciprocal lattice definition

There is a relationship between the crystallographic planes of the direct lattice, determined by Miller indices (*hkl*), and the vector $\vec{G}^*$ of the reciprocal

lattice, which is a linear combination of the basic translation vectors $\vec{a}^*$, $\vec{b}^*$, $\vec{c}^*$. This relationship is defined by equation

$$\vec{G}^*_{hkl} = h\vec{a}^* + k\vec{b}^* + l\vec{c}^*. \qquad (1.1.7)$$

As follows from (1.1.6) and (1.1.7), the $\vec{G}^*$ vector is always perpendicular to the (*hkl*) plane and its modulo is proportional to the inverse interplanar spacing $1/d_{hkl}$ between the equivalent {*hkl*} planes of the direct lattice (see Fig. 1.1.24).

The definition (1.1.7) of the basic translation vectors of the reciprocal lattice gives the next important rules:

The heteronymous main translation vectors of both lattices are always perpendicular to each other so that

$$\left(\vec{a}^*\vec{b}\right) = \left(\vec{a}^*\vec{c}\right) = \left(\vec{b}^*\vec{c}\right) = \left(\vec{c}^*\vec{a}\right) = \left(\vec{c}^*\vec{b}\right) = \left(\vec{b}^*\vec{a}\right) = 0. \qquad (1.1.8)$$

Having the same names basic translation vectors of both lattices are always parallel because

$$\left(\vec{a}^*\vec{a}\right) = \left(\vec{b}^*\vec{b}\right) = \left(\vec{c}^*\vec{c}\right) = 1 \qquad (1.1.9)$$

The product of the primitive cells volumes of both lattices is equal to unity:

$$VV^* = 1, \qquad (1.1.10)$$

where $V^* = \vec{a}^*[\vec{b}^*\vec{c}^*]$.

Design of the reciprocal lattices. We give examples of reciprocal lattices construction for oblique-angled, simple cubic, fcc and bcc Bravais space lattices. For two-dimensional (flat) oblique-angle lattice (Fig. 1.1.25) basic translation vectors $\vec{a}$ and $\vec{b}$ may be expressed in *xy* rectangular coordinate system as follows:

$$\begin{cases} \vec{a} = 2\vec{x} \\ \vec{b} = \vec{x} + 2\vec{y}, \end{cases}$$

where $\vec{x}$ and $\vec{y}$ are the basis vectors by the *x* and *y* axes. The third axis *z* and vector $\vec{c}$ we assume perpendicular to the drawing, asuming $\vec{c} = \vec{z}$. In this case, the reciprocal lattice vectors $\vec{a}^*$ and $\vec{b}^*$ lie in a plane *xy*. Then from Eq. (1.1.8) follows:

$$\begin{aligned} &[\vec{b}\vec{c}] = [(\vec{x} + 2\vec{y})\vec{z}] = [\vec{x}\vec{z}] + 2[\vec{y}\vec{z}] - \vec{y} + 2\vec{x} = 2\vec{x} - \vec{y} \\ &[\vec{c}\vec{a}] = [\vec{z}2\vec{x}] = 2[\vec{z}\vec{x}] = 2\vec{y} \\ &\left(\vec{a}[\vec{b}\vec{c}]\right) = \left(2\vec{x}\left(-\vec{y} + 2\vec{x}\right)\right) = -2\vec{x}\vec{y} + 4\vec{x}\vec{x} = 4 \end{aligned} \qquad (1.1.11)$$

and then:

$$\vec{a}^* = \pi\vec{x} - \frac{1}{2}\pi\vec{y};$$
$$\vec{b}^* = \pi\vec{y}. \qquad (1.1.12)$$

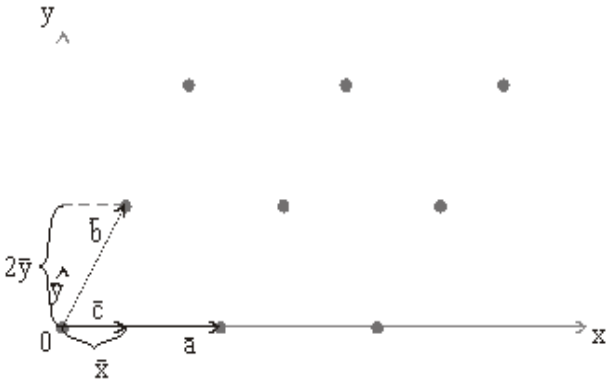


Fig. 1.1.25. The construction of the reciprocal lattice for the oblique direct lattice

Now let's construct the main vectors $\vec{a}^*$ and $\vec{b}^*$ and also reciprocal lattice itself. It is clear from Eq. (1.1.8), that $\vec{a}^* \perp \vec{b}$ and $\vec{b}^* \perp \vec{a}$, because $(\vec{a}\vec{b}) = 0$ and $(\vec{b}\vec{a}) = 0$.

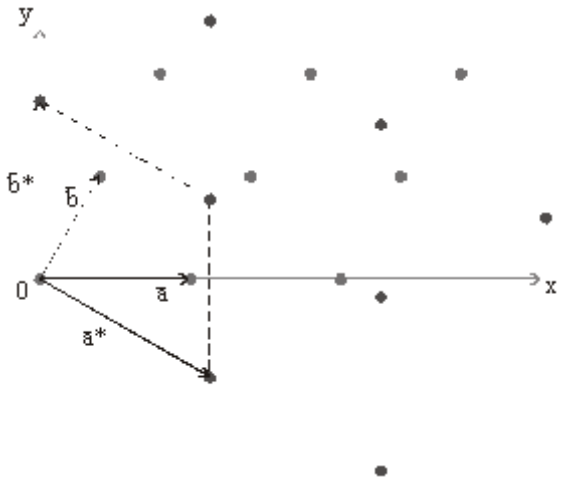


Fig. 1.1.26. The construction of the reciprocal lattice for a simple cubic lattice line

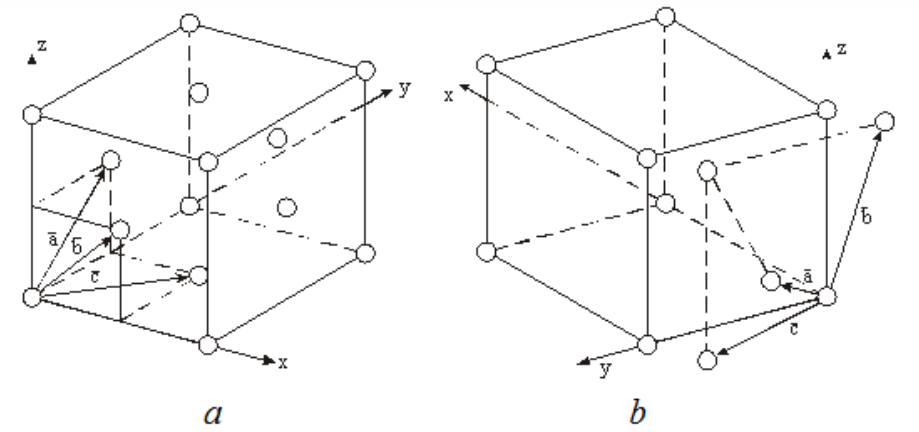
Simple cubic lattice ($a = b = c$). According to Fig. 1.1.26, we can define the basic vectors of the direct simple lattice: $\vec{a} = a\vec{x}$; $\vec{b} = a\vec{y}$; $\vec{c} = c\vec{z}$, where $\vec{x}$, $\vec{y}$, $\vec{z}$ are the unit vectors along the axes *x*, *y*, *z*. Then, by definition (1.2.1) we have

$$\vec{a}^* = \frac{2\pi[\vec{b}-\vec{c}]}{\vec{a}[\vec{b}\vec{c}]} = \frac{2\pi a^2[\vec{y}-\vec{z}]}{a^3} = \frac{2\pi}{a}\vec{x}. \qquad (1.1.13)$$

This means that simple cubic lattice has the simple reciprocal lattice, but with the edges $2\pi/a$ of the unit cell.

The basic vectors of direct bcc lattice is defined as shown in Fig. 1.1.27a by the following ratio:

$$\vec{a} = \frac{a}{z}(\vec{y}+\vec{z}-\vec{x}); \vec{b} = \frac{a}{z}(\vec{z}+\vec{x}-\vec{y}); \vec{c} = \frac{a}{z}(\vec{x}+\vec{y}-\vec{z}). \qquad (1.1.14)$$



a *b*

Fig. 1.1.27. The construction of the reciprocal lattice for the bcc (*a*) and fcc (*b*) direct lattices

As shown in Fig. 1.1.27b, the basic vectors of direct fcc lattice are defined by the following relations:

$$a = \frac{a}{z}(\vec{y}+\vec{z}); \vec{b} = \frac{a}{z}(\vec{z}+\vec{x}); \vec{c} = \frac{a}{z}(\vec{x}+\vec{y}). \qquad (1.1.15)$$

Hence, we can get the trio of translation basic vectors for the corresponding reciprocal lattice in form

$$\vec{a}^* = \frac{4\pi}{a}\frac{1}{2}(\vec{y}+\vec{z}); \qquad (1.1.16)$$

for *bcc* lattice, and

$$\vec{a}^* = \frac{4\pi}{a}\frac{1}{2}\left(\vec{y}+\vec{z}-\vec{x}\right). \qquad (1.1.17)$$

for *fcc* lattice. As follows from a comparison of (1.1.14) - (1.2.17), the bcc lattice has a reciprocal lattice of fcc type (and vice versa) with the edge $2\pi/a$ of the cubic unit cell.

1.1.7. Defects in crystalline lattice

At thermodynamic equilibrium, the location of particles (atoms, ions, molecules) in a perfect crystal is characterized by a strict three-dimensional periodicity. However, real crystals, due to violation of the thermodynamic equilibrium conditions in the process of their production, are always contain distorted parts of the periodic lattice.

Any local deviation from the periodic structure of the crystall is called *the defect of crystal lattice*. These distortions are characterized by changes in the coordination of the atoms, a violation of the lengths and angles of atomic bonds, the introduction of foreign atoms, the formation of foreign phases, etc. The presence of such defects always distorts the whole region of the crystal lattice.

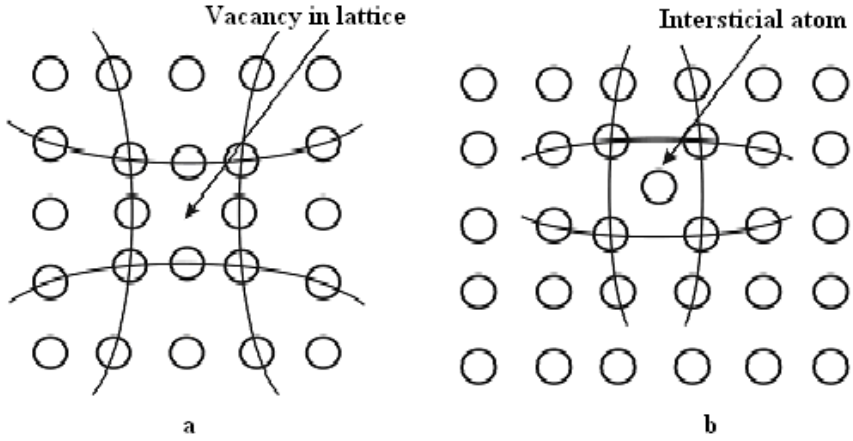


Fig. 1.1.28. Schematic representation of the vacancies (*a*) and interstitial atoms (*b*)

Low mobility and a large (almost infinite) lifetime of structural defects under normal conditions allow one to describe them using the visual geometric images. Classification of defects is usually carried out by the number of dimensions in which the distortions of the crystal structure are extending over

distances exceeding the characteristic parameters of the lattice. Such approach allows to devide all defects on four classes: point (zero-dimensional), linear (one-dimensional) surface (two-dimensional) and volumetric (three-dimensional).

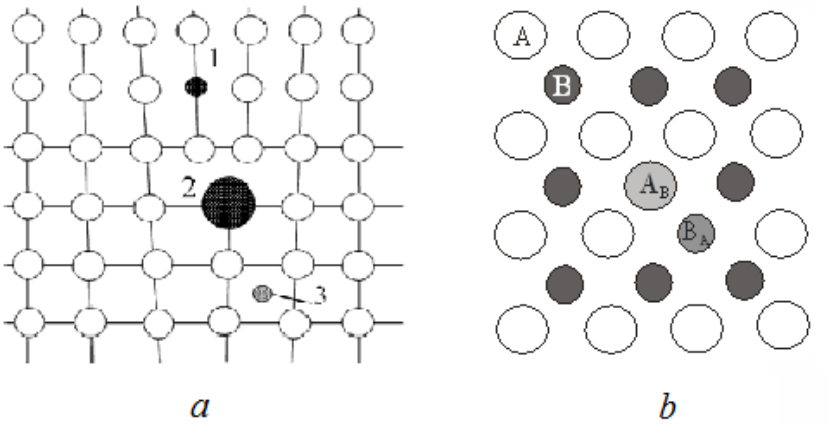


Fig. 1.1.29. Schematic representation of the simple point defects in a single-atom (a) and diatomic (b) crystal lattices. In a: 1 – impurity atom substitution of short-range, 2 – impurity atom substitution of long-range, 3 – introduction of an impurity atom. In *b*: A, B – atoms of different sorts; AB, BA – antisite defects

Point defects. Under point defects we shall understand the isolated from each other damages of periodicity which are extended on a few interatomic distances in all three crystallographic directions. The so-called *intrinsic point defects* of the crystal lattice include *vacancies* (Fig. 1.1.28a) and own *interstitial atoms* (Fig 1.1.28b).

Impurity atoms in the crystal lattice are important types of point defects. They can substitute the lattice sites, defects 1 and 2 in Fig. 1.1.29a, or be incorporated into intersticials of the lattice, Fig. 1.1.29a-3. Simple point defects can interact creating their combinations, such as vacancy-interstitial atom (*Frenkel defect*), the vacancy-atom on the surface (*Schottky defect*), *double and triple vacancies*, *vacancy-impurity*, etc.

There are the so-called antisite defects which occur in chemical compounds AB, whose lattice consists of two sublattices of type A and B (see Fig. 1.1.29b). *Antisite defect* A_B (B_A) is a point defect, which is obtained when atom A (B) occupies the site of sub-lattice of B (A) atoms. Once again, we emphasize that since the introduction of a point defect is always a distortion of the crystal lattice in its vicinity (see Fig. 1.1.28), the concept of a point defect is the whole area of the distorted crystal lattice.

Point defects usually arise due to impact of heat or irradiation. The energies of the point defects formation are typically a few electron volts: ~ 2 eV for vacancies in Germanium, ~ 2.3 eV for the vacancy in silicon, ~ 1 eV for vacancies in copper, ~ 2.4 eV for the interstitial atoms in copper, ~ 2 eV for Schottky defects for NaCl and ~ 1.5 eV for Frenkel defects for NaCl.

Linear defects. Linear defects include damages of periodicity in the crystal lattice, which are extended on much interatomic distances along one crystallographic direction (axis), and along other directions – only on some of the interatomic distances (as for point defects). There are two main types of linear defects – edge and screw dislocations. Besides the chains of nano-and micro-cracks, and point defects are also related to linear defects. Linear defects can be both thermal and mechanical (generated as a result of plastic deformation) nature.

Edge dislocation presents linear range of lattice distortions, which occur formally along half-plane incorporated into the crystal lattice (Fig. 1.1.30b and c). Really edge dislocation is formed due to the incomplete shift of top part of the crystal elative the low one upon application of shear stress τ, perpendicular to the edge of this *extra plane* (*dislocation line*). The edge of this half-plane will coincide with the line of the edge dislocation. If the shift was not pass to the end of the crystal, dislocation can be represented as a certain distorted region of the crystal lattice around the edge of extra plane. As follows from Fig. 1.1.30, the dislocation line (edge of extra plane) is perpendicular to the shear direction.

If the extra plane is at the top part of the crystal, the dislocation is conventionally called *positive* if at the low part – then *negative*. Sign of the dislocation allows to evaluate the character of their interaction. Experiments show that the dislocations of the same sign repel each other, but the opposite – attract. The attraction of two dislocations of opposite signs accompanied by their annihilation (disappearance of defect region), as the two half-planes form a complete plane.

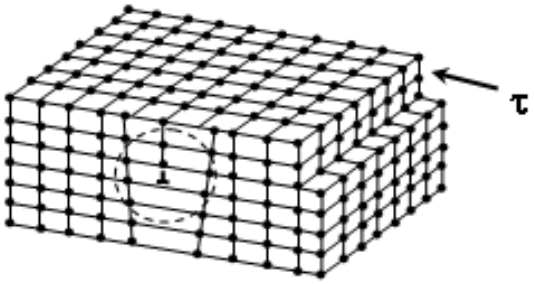
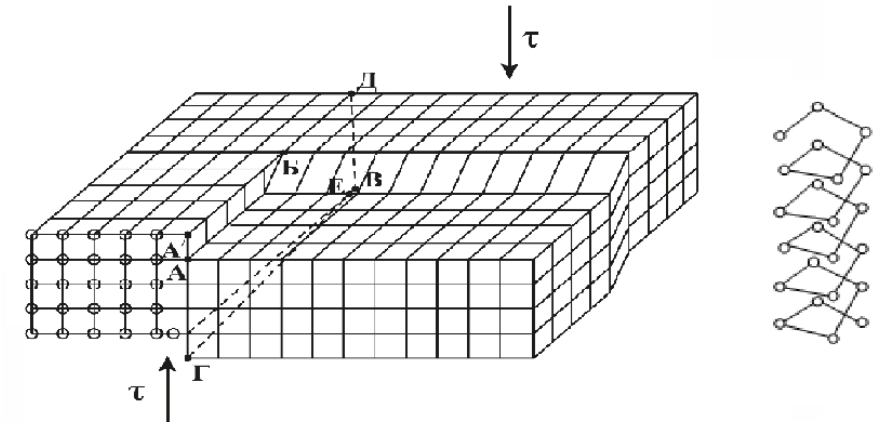


Fig. 1.1.30. Three-dimensional diagram of the crystal lattice with an edge dislocation arising from the action of shear stress τ, perpendicular to the dislocation axis (line)

Screw dislocation is a distortion region in the crystal lattice, which occurs in the crystal due to subjection to shear stress τ, parallel to the line *B-B`* (see Fig. 1.1.31a). It is accompanied by a partial (on one interatomoc distance) shift of the atomic layers in the plane AA`BB` (Fig. 1.1.31a). The atoms of the crystal lattice, which are shifted from their equilibrium positions along the axis B-B` (called the *screw dislocation line*), are arranged by twisted (spiral-like) curve along the dislocation line (Fig. 1.1.31b).



a) b)

Fig. 1.1.31. Scheme of screw dislocations (*a*) and the arrangement of atoms along the dislocation line (*b*) under shear of two parts of the crystal relative to one another

The dislocation line separates shifted part of the crystal from the part of the crystal, where the shift has not yet occurred. Screw dislocation, formed by rotating clockwise, called the *right* and counterclockwise – *left*.

As follows from Fig. 1.1.30, crystal lattice is distorted around the dislocation line, causing the formation of the stress fields of compressing and stretching. Screw dislocations are formed in the crystallization process, and also due to plastic deformation and phase transformations. By the *density of dislocations* we usually understand the total length of dislocations per unit volume.

The lattice distortions (stress fields) around dislocations results in attraction of imputity atoms. Accumulation of impurity atoms around the dislocations form the so-called *Cottrell atmospheres* or clouds that hinder the movement of dislocations during plastic deformation (see below).

Two-dimensional defects. Two-dimensional or surface defects in the crystal are extended along two directions on distances which are many times greater

than the typical values of the lattice parameter, while in the third direction – only a few interatomic distances. *Surfaces* of the crystals, *stacking faults*, and internal surfaces (*interfaces*), like *grain boundaries* and *interphase boundaries* are related to two-dimensional defects. *Stacking faults* occur when hexagonal and cubic crystalline layers alternate in the closed-packing structure (e.g., ABCABCABABC ...), Fig. 1.1.32a.

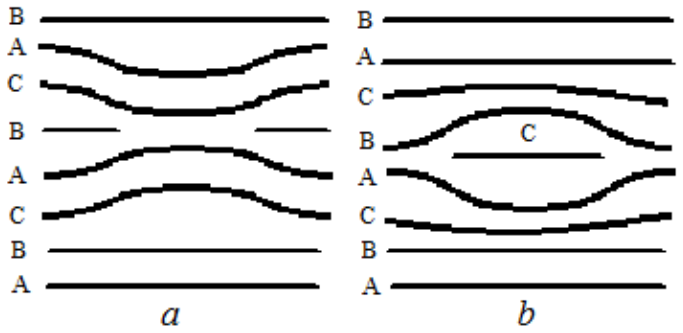


Fig. 1.1.32. Stacking faults

Interfaces are typical places of crystal lattices matching neighboring grains in the polycrystalline material (Fig. 1.1.33b) or *interphase boundaries* in multiphase alloys (Fig. 1.1.33a).

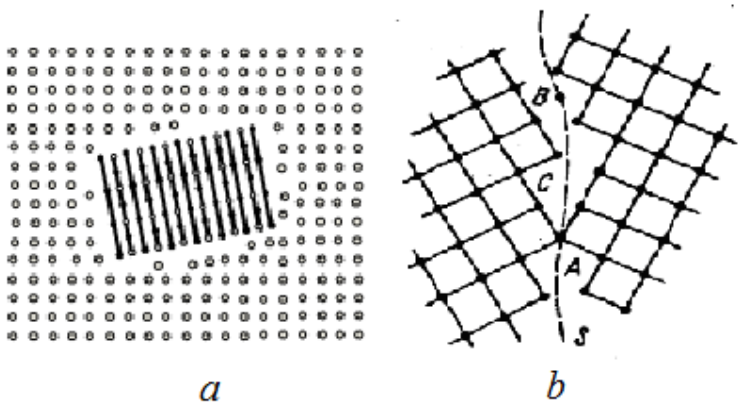


Fig. 1.1.33. Interphase (*a*) and intergranular (*b*) interfaces

The structure of the grain boundary depends on the angle of misorientation of the crystal lattices in the neighboring crystallites. Boundaries with the large misorientation angles represent crystallographicly strongly mismatched area S

of two neighboring lattices (Fig 1.1.33b). In some cases, part of atoms at the grain boundary may belong to both lattices (for example, atom A), while the other atoms can be unmatched (B atom).

The structure of the grain boundary exerts a great influence on the properties of the crystals. Grain boundaries interact with dislocations being within the grains. Depending on the location of dislocation relative to the boundary and their sign, forces of attraction and repulsion can occur. Grain boundaries also attract point defects located at a few interatomic distances. Atmospheres of attracted impurity atoms (Cottrell atmospheres) inhibit the migration of boundaries. Two-dimensional defects arise typically during the crystal growth due to violation of local thermodynamic equilibrium conditions.

Bulk defects. Bulk defects include micro- and nanopores, and also the inclusions of the other phases (Fig 1.1.33a), if their dimensions are much higher than the typical values of the lattice parameters in all three crystallographic directions. Bulk defects arise typically during the crystal growth due to violation of local thermodynamic equilibrium conditions or at intensive plastic deformation, heat treatment and other impacts.

1.1.8. Diffusion in crystalline solids

One of the important consequences of heating the crystalline system is an initiation of the migration of atoms by the crystalline lattice, which is called *diffusion*. Diffusion in condensed systems is very spread phenomena. It plays an important role in various technological processes such as crystallization, doping, oxidation, chemical and thermal processing, gettering of impurities, manufacturing of *pn*-junctions, etc. In addition, the diffusion determines the phase and structural transformations, the formation of defect complexes and redistribution of point defects, formation of impurity atmospheres around dislocations and grain boundaries (Cottrell atmospheres), etc.

Thus, during the crystallization of solids with complex crystalline structure (chemical compounds) homogeneous stoichiometric crystals can not be obtained without the diffusion. Diffusion allows to supress such an undesirable phenomenon in the crystallization of alloys as segregation, liquations, etc.

Diffusion is very important in the synthesis of complex composite materials, for example, at sintering of refractory metal powders. Just a diffusion enables to obtain the compositions of metals having melting points higher than 2500-3000 K (like tungsten, rhenium, etc.), which in principle can not be obtained by crystallization from the melt. Particularly diffusion processes are important in thin-film compositions, which are widely used in solid state electronics. In bulk materials near surface changes in composition occuring due to the mutual diffusion do not cause practically changes in the integral properties of these

materials. At the same time, in thin film compositions, especially in 1-100 nanometers thick layers, diffusion processes may lead to changes in the composition and structure of these layers as a whole. As a consequence, it results in a degradation of the parameters of the devices and integrated circuits, up to their failure. Diffusion of impurities in single crystalline silicon or germanium is one of the main processing step to create semiconductor devices. Diffusion is used for the formation of *pn*-junctions, solar cells, as well as diode and transistor structures, resistors and other elements of microelectronics.

We can say that many of the technological processes, which are used to obtain crystals and finished products from them, and give them the necessary properties, are not possible in principle without the use of diffusion.

General information about the diffusion. The thermal vibrations of atoms in solids are generally reduced to the low-amplitude oscillating near their average equilibrium positions. However, interaction between a vibrating atoms leads to the fact that, even at low average kinetic energy of atoms in the crystal, a certain number of atoms will always have sufficient kinetic energy to overcome the interatomic potential barrier and to come to a new equilibrium state. Such an exited atom can leave its equilibrium position (lattice site) and, after overcoming the potential barrier, move into a new, unoccupied equilibrium position (interstitial or vacant position). During this transition atom looses extra energy, donating it to surrounding atoms of the crystal lattice. After a while, the atom can again get extra energy to go back to the next free vacancy or interstitial. Just such thermal movement of the atoms makes a basis of diffusion processes in solids.

Therefore, we can consider atomic diffusion in solids as directed thermal migration of atoms. Elementary act of diffusion always consists in jumps of single atoms or their small groups on the range of the order of interatomic distances. A general reason which causes this directed movement of atoms is difference of chemical potentials at different points in the system. The occurrence of this difference is usually connected with inhomogeneous distribution of the atoms in space, i.e. the presence of a concentration gradient, but it can also be caused by the existence of gradient of temperatures or fields (electrical, mechanical, etc.). In an ideal (nondamaged) crystal lattice, in which the atoms only oscillate around their equilibrium positions, in general, the diffusion processes are unlikely. Migration of impurity or native atoms by the lattice is mostly associated with the presence of defects – vacancies, interstitials, divacancies and other more complex point defects and also dislocations, grain boundaries, and vacancy-impurity agglomerations (clusters), etc.

The first and second Fick` laws. The diffusion of atoms in solids represents one of three main processes of directed transfer of mass and energy. These three processes also include the charge transfer processes (electrons, ions) and the

propagation of acoustic and optical lattice waves (phonons), which define electrical and thermal conductivity. However, if the last two processes are mainly associated with the movement of electrons and thermal energy, atomic diffusion in solids is always associated with directed movement of substances (mass transfer). All three processes are described by equations with similar form: the flow (of electrons, holes, ions, heat due to diffusing atoms) is directly proportional to the gradient of the parameter that defines the basic paramters of the process:

$$I = -\sigma \frac{dU}{dx}, \quad q = -\kappa \frac{dT}{dx}, \quad j = -D \frac{dN}{dx}, \tag{1.1.18}$$

where: $\frac{dU}{dx}$, $\frac{dT}{dx}$, $\frac{dN}{dx}$ are respectively gradients of the electrostatic potential, temperature and concentration; I – the flow of electrons; q – heat flux; j – flow of diffusing atoms, σ – electrical conductivity; κ – specific thermal coefficient; D – diffusivity. The negative signs in the right- side of (1.1.18) indicate the direction of the process in the direction of gradient decreasing.

Thus, the diffusion coefficient can be represented as some "specific amount" of diffusing atoms. Laws of (1.1.18) have both a close sense and the same type of restrictions. For very large values of gradients of potential, temperature and concentration these relations are not applicable. In addition, the parameters σ, κ and D, determining the transfer of charge, mass and heat consequently, are not always constant and may depend on many factors, the main of which is the temperature. This dependence is given by

$$\sigma = \sigma_0 \exp\left(-\frac{E}{k_B T}\right), \qquad D = D_0 \exp\left(-\frac{E}{k_B T}\right), \tag{1.1.19}$$

where: E – the activation energy of the process. The absence of a similar correlation for the thermal conductivity in (1.1.19) is due to more complex mechanisms of heat transfer electrons and phonons.

As noted above, in general terms the driving force of diffusion is the presence of the chemical potential gradient $\overrightarrow{\text{grad}}\,\mu$ (or $d\mu/dx$ one-dimensional case), which can be caused by different reasons – the inhomogeneous distribution of temperature, composition, mechanical stresses, and external or internal (intrinsic) electric fields in the crystal. It is established that the amount of matter, diffused in unit time through unit area of the surface in a direction perpendicular to the direction of diffusion – the so-called diffusion current density of atoms J – is proportional to the gradient. This dependence is described by the first law of diffusion, which for three-dimensional diffusion is described by the relation

$$\vec{J} = -L\overrightarrow{\text{grad}}\ \mu, \tag{1.1.20}$$

where: L – proportionality coefficient that determines the rate of equalization of the chemical potential μ. Minus sign before the right-hand side of equation (1.1.20) means that the flow is going in the direction opposite to the chemical potential gradient. The mostly known is the so-called free diffusion when the system (material) consists of at least two kinds of atoms distributed inhomogeneously. In this case, the chemical potential gradient is due to the presence of the gradient of concentration C of atoms ($\overrightarrow{\text{grad}}\, C \neq 0$). For non-zero gradient of concentration directed diffusion flow occurs in the material. It tends to equalize the concentration in the system (material) of atoms, i.e. make their distribution uniform over the crystal. The diffusion flux J, due to the presence of the concentration gradient of the atoms, is described by the following differential equation:

$$\vec{J} = -D\overrightarrow{\text{grad}}\ C, \tag{1.1.21}$$

or in case of one-dimensional diffusion by equation:

$$J = -D\ \frac{\partial C}{\partial x}. \tag{1.1.22}$$

Here C is concentration of the diffusing particles and a proportionality factor D is the diffusivity which has the dimension cm^2/s.

Equations (1.1.21) and (1.1.22) present the first Fick's law. The conclusion of these expressions is given in several well-known handbooks. Note that the diffusivity in Fick's first law (1.1.22) in the simplest case is given by $D = av^2$, where a is the length of the elementary jump of the atom (the order of the interatomic distance), and ν – the speed of atomic diffusion.

The second Fick's law is as follows:

$$\frac{\partial C}{\partial \tau} = D\frac{\partial^2 C}{\partial^2 x}. \tag{1.1.23}$$

For a given diffusivity, it describes the distribution of the concentration of diffusing particles in different points of the medium as a function of time. Equations (1.1.22) and (1.1.23) are derived under the assumption of equal probability of diffusion displacement of atoms in all crystallographic directions of the crystal, which is determined by symmetry and form of the potential barriers. For this limitation the diffusivity D is scalar isotropic value independent on x, y, z.

The general solution of the 2nd Fick's equation (1.1.23) for an unlimited body with a given initial distribution of dopant concentration $N(x,0) = f(x)$ at the moment $\tau = 0$ has a complicated form. However, with the introduction of

additional limitations, some solutions can be simplified allowing direct quantitative calculation of the diffusant profiles as function of time. Examples of such profiles are discussed below.

Interconnection between diffusivity and physical-chemical properties of crystals. The diffusivity is a material constant which characterizes the rate of diffusion. As was mentioned above, when temperature increase, the diffusivity in solids usually increases drastically by an exponential law

$$D(T) = D_0 \exp(-Q/k_B T). \qquad (1.1.24)$$

Here k_B is the Boltzmann constant, T – absolute temperature, a D_0 and Q – the main parameters that are dependent on the mechanism of diffusion. The Q value is called the *activation energy of diffusion*, a pre-exponential factor D_0 is often called *the frequency coefficient (factor).* They are associated with physical-chemical properties both the material, in which the diffusion occurs (which is sometimes called *matrix*), and also diffusing atoms (the diffusant). In some cases, the diffusivity also depends on the concentration of the diffusant.

The exponential behavior of diffusivity with temperature was established empirically, and then was proved theoretically. Exponential (activation) depndence $D(T)$ means that to make the elementary act of diffusion atom has to overcome a certain energy barrier, which value is close to Q (but not equal, as we shall see below).

Depending on the specific mechanism of diffusion in the crystal (see below) Q values are between 0.5 and several electronvolts. At the same time, the average thermal energy ($k_B T$) of the vibrating atoms in solids are between 0.02-0.03 eV and 0.2 eV, that is significantly less than Q values. Therefore, for a jump atom should get a very large local fluctuation of the thermal energy of the vibrating atoms in the crystal lattice is needed. So the probability f of atom jump from one to another position, as an elementary act of diffusion, is proportional to probability of such fluctuations. The latter is proportional to $\exp(-Q/k_B T)$, because the distribution of thermal energy quanta in the crystal at high temperatures obey Maxwell-Boltzmann statistics. In addition, f is proportional to the frequency of atomic vibrations ν (i.e. the number of attempts to jump) and the number of vacant seats (vacancies or interstitials) nearby to which atom can jump in case of overcoming the activation barrier. This part of the jump probability determines pre-exponential factor in (1.1.24).

Jump frequency dependence on the, which As follows from (1.1.24), drastic (exponential) temperature dependence of f can be clearly illustrated on the example of steel (an alloy of iron and carbon) as one of the most important materials. As for carbon in iron $Q \approx 0.9$ eV, it means that at 393 K (20°C) and 1818 K (1545°C) the carbon atom performs $2.5 \cdot 10^2$ and $2 \cdot 10^{11}$ jumps/s respectively.

In case of free diffusion in the crystal parameters D_0 and Q depend on the atomic diffusion mechanism, and hence on the lattice type. In addition, in semiconductors they are determined by charge states of diffusing atoms (impurities) and, therefore, by the concentration of free electrons (holes).

Types of diffusion. Speaking of free diffusion, one accepted to distinguish its two types – heterodiffusion and self-diffusion. *Heterodiffusion* is the diffusion of foreign atoms (doping impurity) when gradient of their concentration. *Self-diffusion* is a special case of diffusion in pure substance or ordered solid solution of constant composition of the atoms of that substance or solution components.

Directed self-diffusion can be performed under the influence of accidentally occurred local gradients of the isotopic composition of atoms in the direction of equalization of the gradient. Artificial creation of such a gradient is often used for the experimental determination of self-diffusion coefficients

Besides free diffusion, the so-called forced diffusion is distinguished, when the chemical potential gradient appears under the influence of directed external or internal forces. Forced diffusion leads no to equalization, and on the contrary, to increase of the concentration gradient. There are the following main types of forced diffusion.

Thermal diffusion (or *Soret effect*) occurs under the influence of a temperature gradient *dT/dx*. In this case, the flow of migrating particles is directed toward lower temperatures. An example of the thermal diffusion is the diffusion of charge carriers from the hot to the cold junction in thermoelectric energy converters.

Electrodiffusion (or *electromigration*) takes place under the influence of an electric field (potential gradient *dU/dx*) on charged particles (ions) of the crystal. The ions diffuse towards the electrode of opposite sign – to anode cations and anions – the cathode. Electrodiffusion examples are migration of impurity ions in semiconducting crystals under exposion of the internal electric fields (like in *pn*-junctions), aluminum ions in the thin film conductor in metallization of integrated circuits, and electrolytic processes at electrochemical corrosion.

Barodiffusion occurs under exposion of gravity or pressure *dP/dx*. An example is the liquation (segregation) of impurities due to the difference in the density of compomemts in the alloy ingot crystallizing from the melt. It, in particular, leads to the enrichment of the bottom portion of the ingot by more heavy elements.

The *uphill diffusion* occurs under the action of elastic stresses. It is usually characterized by the value *da/dx*, where *a* is the lattice parameter. Thus the flow of larger-size atoms is directed to a regions being under action of tensile stress,

whereas the flow of ions with smaller radii – to the compressed domain. Just this type of diffusion leads to the formation of impurity atmospheres around dislocations, grain boundaries and dispersed phases under formation of crystals.

When acting the external driving forces, the first Fick's law is transformed into the general diffusion equation of the form

$$J = - D \text{ grad } C + Cv = - D \Delta C + Cv, \quad (1.1.25)$$

where: J, grad C (ΔC) and v are vector of glow, concentration gradient and mean drift velocity of migrating particles and D – diffusivity. Using the continuity equation, one can conclude from (1.1.25) for one-dimensional case the following diffusion equetion

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - v\frac{\partial C}{\partial x}. \quad (1.1.26)$$

Kinetics of self diffusion and diffusion of impurities in crystals. To perform an elementary act of diffusion (jump), the atom always has to overcome a certain energy barrier. Atom hopping probability P_m is proportional to the product of the probability of the local fluctuations of atom's energy $\exp(-Q/k_BT)$, the atom vibration frequency ν and also the number of empty sites (vacancies) or interstitials z in the nearest coordination sphere (on which the atom can jump). In addition, this probability is also proportional to the increase of entropy ΔS after the jump, associated with the breaking of the ordered arrangement of the atoms in the lattice (so $\Delta S > 0$). Hence it follows that

$$P_m = Cz\exp(\Delta S/k_B)\cdot\exp(-Q/k_BT), \quad (1.1.27)$$

where C is proportionality coefficient close to unit.

In a more rigorous approach, it is necessary to take into account such additional factors as dependency of the diffusion coefficient on the concentration, which can change in the process of diffusion, on the presence of internal electric fields, on possible complex formation between diffusant and other impurities or defects, etc.

For all cases of diffusion in crystals common feature consists of parabolic interconnection between the diffusion time and the distance which diffusant passes for this time (it is called *diffusion length*). It is evident that the length of a single jump δ is the same for all atom because it is close to the interatomic distance in the crystal ($\delta \sim a$). We can also assume that the directions of the jumps have the character of random walks, i.e. jumps in all directions are equally probable. In this case, for a large number n of jumps, average distance $\bar{x}$ that every atom migrates from the start of the jumps will be zero. However, mean square distance $\bar{x}^2$, which is a measure of the total distance passed by n atoms from the origin, will be not equal to zero. It is easy to show that

$$\bar{x}^2 = n\delta^2. \quad (1.1.28)$$

The total number of jumps n equals to the product of the frequency (probability) of jumps P_m and the time t spent on n jumps. Hence

$$\sqrt{\bar{x}^2} = \sqrt{ft\delta^2}. \quad (1.1.29)$$

From the theory of gas diffusion we have the relation for diffusivity

$$D = \frac{f\delta^2}{6}, \quad (1.1.30a)$$

for the three-dimensional diffusion, whereas for the one-dimensional diffusion

$$D = \frac{f\delta^2}{2}. \quad (1.1.30b)$$

Comparing (1.1.29) and (1.1.30b), we can show that for the one-dimensional diffusion the following relation holds

$$\sqrt{\bar{x}^2} = x = \sqrt{2Dt}. \quad (1.1.31)$$

The relation (1.1.31) means that $t \sim x^2$ (a parabolic law), i.e. the diffusion length x is proportional to the square root of the diffusion time.

Diffusive migration of the atoms in the crystal lattice (selfdiffusion or impurity diffusion) should be always associated with the presence of defects (see below in detalis). However, the most known and studied mechanisms of atomic diffusion are connected with the diffusion by *interstitials* (free lattice sites) or *vacancies*. In the latter case the movement of atoms (ions) can be regarded as *the diffusion of vacancies*. Consider the elementary model for calculating the parameters of diffusion in crystals with the formation of vacancies or interstitial atoms developed by Frenkel.

The Frenkel theory of self-diffusion. The diffusion process in Frenkel model, developed for the vacancy diffusion mechanism on the basis of the classical kinetic theory of crystals, includes the following atomic processes:

Oscillations of atoms in a lattice site around the equilibrium positions with frequency ω_0 (or ν_0);

The transition of an atom from a neighboring lattice site to an interstitial position with the vacancy V formation;

Oscillation of atom in an interstitial position in Fig. 1.1.34a (or, that is the same, on top of the barrier in Fig. 1.1.34b) for a long time before it jumps to a free neighboring position (the vacancy);

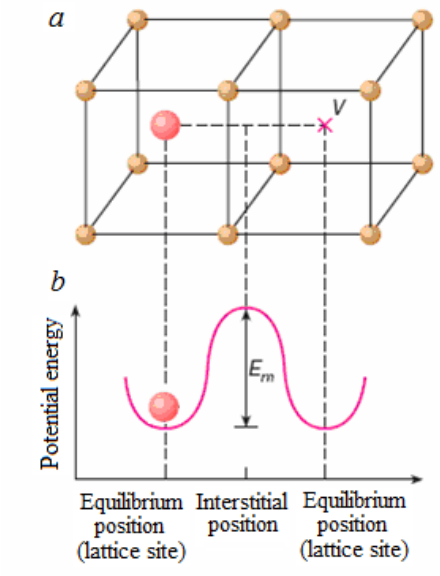


Fig. 1.1.34. The jump of the diffusing atom into the neighboring vacant site for BCC lattice (*a*) and the potential energy dependence of the coordinate of diffusing atom (*b*). Em is the activation energy (potential barrier) of diffusion

The transition of an atom to the vacancy through the interstitial position (Fig. 1.1.34a) by overcoming the potential barrier E_m (Fig. 1.1.34b).

The last means movement of vacancies to the place of diffused atom. In that sense, Frenkel model can be considered as diffusion of vacancies in the crystal. Because of the above-described thermal diffusion processes, the running mixing of the atoms in the crystal occur. The mixing rate is determined by the probability of atomic transitions from one equilibrium position to another (e.g., from sites to vacancies). Due to the statistical nature of this process, the probability of a jump f (or P_m) changes with temperature according to the known exponential law

$$P_m = f = \omega_0 \exp\left(-\frac{E_m}{k_B T}\right). \qquad (1.1.32)$$

Here E_m is the height of the potential barrier that must be overcome by an atom to jump from one equilibrium position (site) in the lattice to another (vacancy); ω_0 – the natural frequency of atomic vibrations (of the order of the Debye characteristic frequency 10^{12}-10^{13} Hz).

According to the Frenkel theory, the barrier height E_m can be considered as the *migration energy of an atom* or a *vacancy* that depends on the strength of

the interatomic bonds in the lattice. Without going into details, which will be discussed later, we note only that the value of ω_0 is also determined by atomic mass and the degree of anharmonicity of the atomic oscillations in the crystal lattice (i.e. the deviation from Hooke's law).

Starting from Eq. (1.1.32), Frenkel instead of the transition probability f or P_m, introduced inverse value, which have a sense of the mean time τ of the atom life in one of the equilibrium positions. So τ is called *the mean settled lifetime*, for which the following relation is true

$$\tau = 1/P_m \approx \tau_0 \exp(E_m/(k_B T). \qquad (1.1.33)$$

Here τ_0 is constant, which is close to the period of the characteristic atomic vibrations in the crystal lattice sites ($\sim 10^{-12}$-10^{-13} s) by the order of magnitude. The mean lifetime of the atoms (vacancies) in the lattice sites depends on the temperature. For example, for germanium, which has the migration energy $E_m \sim 1$ eV for its own atoms, this time is $\sim 2\cdot10^4$ s for 300 K and $\sim 3\cdot10^{-8}$ s for 900 K. Thus, it follows from (1.1.33) at 900 K atom (vacancy) in the site perform $\sim 10^8$ jumps per 1 s, passing the length (as a broken line) $L_0 = \delta\cdot10^8$ cm $= 1\cdot10^{-8}\cdot10^8$ cm $= 1$ cm with an average velocity

$$<v_B> = L_0/1 \text{ sek} = \delta/\tau = \delta f = (\delta/\tau_0)\exp[-E_m/(k_B T)], \qquad (1.1.34)$$

which equals to 1 cm/s for germanium at 900 K.

Under thermodynamic equilibrium conditions, relative concentration of vacancies has probabilistic meaning

$$P_B = f/\omega_0 = \tau_0/\tau = V = n/N = \exp[-E_B/(k_B T)], \qquad (1.1.35)$$

where n is the number of vacancies and N – the number of atoms per unit volume. If the energu of vacancy formation for metals with close-packed structure is $E_B \sim 1$ eV, the relative concentration of vacancies (1.1.35) is very small value of the order of 10^{-3}- 10^{-4}. Such low concentration of equilibrium vacancies leads to the relatively large average distance between them

$$l = 1/n^3 = 1/(VN)^3 = 1/[V(1/a^3)] = a/V^{1/3} \approx 3\cdot10^{-7} \text{ cm}, \qquad (1.1.36)$$

and, consequently, to the absence of interaction between them. This allows one to consider the set of equilibrium vacancies in a crystal as an ideal gas. Therefore we can estimate diffusivity from the well-known formula for the kinetic theory of gases

$$D = (1/3)<v_B>\cdot <\lambda> = (1/3)<\lambda^2>/\tau, \qquad (1.1.37)$$

where $<\lambda>$ and $<v_B>$ are the mean free path and the average velocity of the diffusing atoms (vacancies).

Unlike gas here is that an elementary act (jump) of vacancy (or atom) diffusion is always present the same, discrete value $\delta \sim a$. Therefore, to study diffusion process in crystalline solids it is more convenient to transit from the mean free paths to mean squared free paths.

Let x be the actual (variable) mean free path of the particles in the gas. Then the probability that the particle will pass this path x without collisions is equal $\exp\left(-\frac{x}{\lambda}\right)$, and the mean squared value (by the mean value theorem) will be equal

$$\left\langle x^2 \right\rangle = \frac{\int_0^\infty x^2 \exp\left(-\frac{x}{\lambda}\right) dx}{\int_0^\infty \exp\left(-\frac{x}{\lambda}\right) dx} = 2\left\langle \lambda^2 \right\rangle. \qquad (1.1.38)$$

Considering that in crystals $\langle x^2 \rangle = \delta^2$ and using Eq. (1.1.37), we get diffusivity of vacancies

$$D_B = (\delta^2/6\tau) = (\delta^2/6\tau_0)\exp[-E_m/(k_B T)]. \qquad (1.1.39)$$

If one considers vacancy mechanism of atomic diffusion in the crystal, then in all relationships one should take into account the fact that before jump of atom from the lattice site to the vacancy, it should "wait" until the vacancy would come close to this site. In other words, there are two processes in Frenkel diffusion: generating of vacancies and then jumps of atoms on it. Therefore, when calculating the diffusion coefficient (the probability of the jump), one must take into account probability of vacancy approaching to the jump distance.

Since the probability of vacancy formation P_V is equal to its relative concentration V, then the total probability P_V, that vacancy will appear around the atom and this atom perform jump in it, is equal to the product of the probabilities

$$P = P_m P_V = (1/\tau_0)\exp[-E_m/(k_B T)]\exp[-E_V/(k_B T)] = (1/\tau_0)\exp[-Q/(k_B T)], \qquad (1.1.40)$$

where E_B is the vacancy formation energy, and

$$Q = E_m + E_B \qquad (1.1.41)$$

is the *activation energy for self-diffusion*. As a result, we get a new expression for the diffusion of atoms in the crystal through the vacancies

$$D = P_B D_B = [\delta^2/(6\tau_0)]\exp[-Q/(k_B T)] = D_0 \exp[-Q/(k_B T)], \qquad (1.1.42)$$

where the pre-exponential factor is just equal to

$$D_0 = \delta^2/(6\tau_0) \qquad (1.1.43)$$

(it depends on the diffusion mechanism). This formula is in full agreement with the previously established for solids empirical temperature dependence of the diffusivity. The same considerations can be applied to the mechanism of interstitial diffusion of atoms or ions in the crystal. In this case, the activation

energy Q is simply equal to the migration energy of the atoms E_m by interstitials.

In real crystals, the diffusion processes occur in much more complicated ways, so that there may be several mechanisms of diffusion. Therefore, derived equations are suitable only for rough estimates when performing diffusion experiments.

Direct theoretical calculation of the activation energy is extremely difficult. Therefore, attempts have been made by various authors to associate it to some other properties permitting calculation or experimental determination. Thus, according to Frenkel, for self-diffusion process activation energy should be close in magnitude to the evaporation heat of the crystal. In the experiment, however, typically lower values are measured. Activation energies of self-diffusion for different crystals lie in the range (10^3-10^5 J/mole), and the pre-exponential multiplier D_0 in the most cases varies from 10^{-6} to 10^{-4} cm^2/s.

Structure defects and atomistic diffusion. Point defects and their interaction with other defects play a decisive role in the mechanisms of diffusion. Not less important influence on diffusion, as will be shown below, is rendered by extenfed defects (like linear, surface and volume), because they affect the behavior and the concentration of point defects.

Speaking about the influence of defects on the diffusion, it should also be mentioned a change of the state of the crystal as a whole under their influence. In particular, the formation of vacancies is accompanied by elastic relaxation of the atoms around the vacancy, which is manifested in the displacement of the atoms surrounding the vacancy in the direction to its center. Thus, displacement of atoms is higher in the close-packed structures than in more loose crystals (see below). Displacement of the atoms couses the local static lattice distortions, local stresses that results in the decrease of the vacancy effective radius. Estimations show that the effective radius of vacancy in the diamond lattice is about 0.8 of covalent radius R of the atom. In lattices with metallic bonds, it is even less (about 0.5–0.6 R). Therefore vacancies in most cases lead to a decrease in the average lattice period.

Atoms in interstitials also cause local elastic distortions due to the symmetrical displacement of the surrounding atoms in the direction from the interstitial atom center, i.e. the effect is opposite in sign to the mentioned above for vacancies. As a result, the presence of interstitial atoms increases the average lattice period.

In s emiconducting materials with a diamond-like lattice (less densely packed crystals), the self-interstitials effect of the lattice period increasing is greater than the effect of reducing the lattice period due to vacancies. Therefore, the formation of Frenkel pairs is accompanied by an increase of lattice

parameter of the matrix. This is important in the study of the nature of point defects in semiconductors.

The impurity atoms distort the matrix lattice because their sizes are always different from the size of the atoms of the matrix (when they are dissolved by substitution mechanism) or the size of the interstitials into which they are introduced (when they are icorporated into interstitials). Therefore, the average lattice constant of the matrix with interstitial impurities always increases with impurity concentration growth. In the presence of substitutional impurities average lattice period of the matrix increases if $R_{imp} > R_{matrix}$ and decreases if $R_{imp} < R_{matrix}$.

Mechanisms of diffusion in crystalline solids. In the crystal lattice, where many types of defects are present, there are a large number of diffusion mechanisms (types of atomic jumps), such as paired exchange, ring exchange, vacancy mechanism; interstitial mechanism; crowdions mechanism; dissociative mechanism; migration by dislocations, migration by stacking-fault defects; migration along the grain boundaries, etc.

The simplest mechanism of the elementary diffusion jump, requiring no the presence of structural defects, is a pair exchange with direct exchange of two adjacent atoms (Fig. 1.1.35a). Such exchange shall be hindered in close-packed crystals due to strong compressions of the surrounding atoms during an elementary jump. That is why such a mechanism requires 3-4 times higher activation energy than the vacancy mechanism.

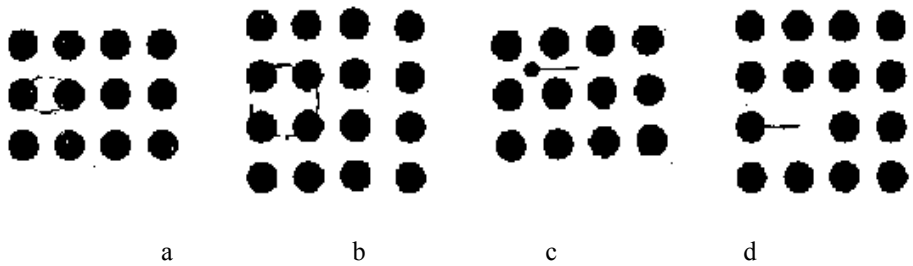


Fig. 1.1.35. The mechanisms of atomic migration: *a)* – pair exchange, *b)* – ring exchange, *c*) – interstitial mechanism, *d*) – vacancy mechanism (arrows indicate the direction of the elementary jumps of atoms)

If the direct exchange, involving simultaneously several atoms, is arranged by a circle, such mechanism is called *ring exchange* (Fig. 1.1.35b). For such a mechanism, every atom in a ring is shifted on one interatomic distance. Although exchange in ring of atoms results in less lattice compressions, as for pair exchange, nevertheless migration of atoms for ring exchange mechanism in

the crystal is also sufficiently weak. Both exchange mechanisms are realized only in perfect (defect-free) crystals with loose lattice (e.g., in diamond and BCC structures).

The vacancy migration mechanism (Fig. 1.1.35d) was described above for Frenkel model. As was shown, when the transition of atom in the neighboring position occur, it must overcome a certain energy barrier (see, Table 1.1.1). This is due to the necessity of temporary breaking of bonds between the neighboring atoms and the presence of elastic displacements of the atoms surrounding the vacancy.

Table 1.1.1. The energy of vacancy formation (E_V), the migration energy of vacancy (E_{mV}) and the activation energy of diffusion (Q_V) for some crystals, eV

Activation energy, eV	Si	Ge	Cu	Au	α-Fe	Zn
E_V	3.5-4.0	2.0-3.0	-	-	-	-
E_{mV}	1.0-1.5	0.5-1.0	-	-	-	-
Q_V	4.0-5.0	2.5-3.5	2.2	1.5	2.7	1.2

As shown above, for the vacancy diffusion mechanism the energy has to be expended on generation of vacancy (E_V) and its migration to the diffusing atom (E_m). The total energy gives full energy of diffusion activation Q_V, see Table 1.1.1). The vacancy mechanism is fundamental for self-diffusion and heterodiffusion of substitutional impurities. The activation energies of heterodiffusion by vacancies are usually smaller, and the rate of diffusion is greater than for self-diffusion.

The most important consequence of the above mentioned consists of high sensitivity of the vacancies diffusion rate to the type and concentration of structural defects. The presence of vacancies and large-scale defects (dislocations, grain boundaries, pores, etc.) in crystals, makes easier the vacancies generation reducing the activation energy of diffusion and enhancing the last. Therefore, large-scale defects are often referred to as sources of vacancies.

Interstitial mechanism is observed in the case when the atom-diffusant is migrated by the crystal from one interstitial to another (see Fig. 1.1.35c). In this case, the rate of diffusion is limited by activation energy E_{mi} of atoms migrating by the interstitials. The E_{mi} value usually increases with the increase of ion radius of diffusing impurity. This is due to the value of the lattice distortions that are caused by incorporation of atoms in the interstitials. Last are the greater the close-packed crystal lattices. Therefore E_{mi} values for metals are greater than

in crystalline semiconductors with diamond-like structure. As a result, the self-diffusion in metals and heterodiffusion by interstitials for large atoms, forming the substitutional solutions, are so small that it can be ignored. Small size impurity atoms diffuse in the metallic crystals mostly by interstitials, forming the interstitial solid solutions.

In semiconductors, the packing density of lattice is typically about half as much as that of metals. Therefore, the probability of heterodiffusion and even self-diffusion over interstitial sites in semiconductors is large enough, and in practice it is often observed, especially at elevated temperatures.

Practically important case of interstitials diffusion in semiconductors is realized for impurity atoms and even own components at the decomposition of supersaturated solid solutions, as well as the formation of precipitates of the second phase. The values of the activation energy E_{mi} for migration of interstitials are much more than activation energy E_{mV} for vacancy migration, but significantly less than the activation energy Q for vacancy mechanism of diffusion. As a first approximation, $E_{mi} \approx (0.5\text{-}0.7)E_{m\mathrm{V}}$ so that they are about 1.0-1.5 eV for Si, Ge and Cu and 1.0 eV for Fe.

One of the kind of interstitial mechanism is the so-called *interstitial push-out mechanism*: the interstitial atom makes hopping to the site atom pushing it to the next interstitial position (Fig. 1.1.36-1). This mechanism is performed, for example, at the self-diffusion of silver in the silver halides.

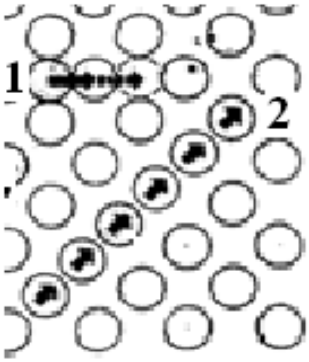


Fig. 1.1.36. Possible diffusion mechanisms in crystals: *1* – interstitial pushing-out; *2* – crowdion

Under *crowdion* a group of atoms is meant which is arranged along closed-packed direction and contains one extra atom in comparison with an ideal lattice (Fig. 1.1.36-2). When *crowdion diffusion mechanism* implementing, small displacements of atoms occure along the whole set of these atoms.

The activation energy for this process is relatively small. Crowdions mechanism occurs during the annealing of radiation defects as well as at local deformations of the lattice under stresses. In metals and elemental semiconductors (such as Ge, Si, etc.) the migration of impurities are mostly carried out by the vacancy or interstitial mechanisms.

The situation becomes more complicated when considering the mechanism of impurities migration in binary semiconductor compounds such A^3B^5, A^2B^6, etc. In such compounds, atoms of sort A are surrounded by atoms of sort B, and vice versa, creating two sub-lattices – A and B. Therefore, the vacancy in sublattice A, which is the closest to the atom in the site of the sublattice A, is in the second coordination sphere (Fig. 1.1.36). Naturally, in this case the probability of exchange between the nearest site atom and vacancy located in the sublattice A is smaller in comparing with the same in elementary semiconductors. Therefore, in the binary semiconductor compounds, diffusion can occur by *dissociative mechanism* that is a combination of vacancy and interstitial mechanisms (Fig. 1.1.36). This diffusion mechanism as if consists of several successive processes: the formation of vacancies and interstitial atoms, their independent (separate) migration, and then their recombination. This means that part of the impurity atoms is diffused by interstitial mechanism, and part by the vacancy mechanism.

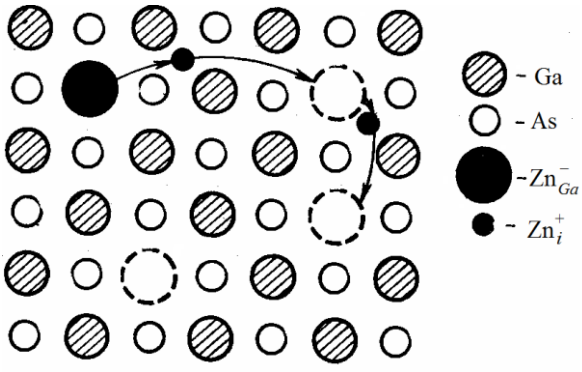


Fig. 1.1.37. A schematic presentation of dissociative mechanism on the example of Zn diffusion in GaAs

Typical examples of this mechanism of diffusion is the diffusion of copper in silicon, germanium and GaAs, as well as the diffusion of zinc in GaAs (see Fig. 1.1.37). If zinc ions are placed in the gallium sub-lattice, they behave as acceptors (accepting electrons) and diffuse through the gallium vacancies. If

zinc ions are located in the intersticials, they behave as donors (donating electrons) and diffuse through the intersticials.

However, the actually observed in the experiments rate of diffusion of zinc in gallium arsenide is much more than would be expected on the basis of purely vacancy mechanism, but much less than it should be for the interstitial mechanism. Such behavior allowed to suggest a dissociative mechanism of zinc migration. In this case, to the elementary event of diffusion of acceptor-like zinc Zn_{Ga}^{-} by gallium vacancies precedes its dissociation (zinc atom gives electron to the crystal) to form an interstitial donor-like zinc Zn_i^{+}, which has a smaller ionic radius than Zn_{Ga}^{-}. Then Zn_i^{+} and vacancy V_{Ga} diffuse through the crystal separately and at different rates. This Zn_i^{+}, of course, diffuses faster than that Zn_{Ga}^{-}, but only to meet with the other vacancy V_{Ga}. After the recombination with a vacancy zinc ion Zn_i^{+} again is placed in a lattice site, capturing an electron and becoming an acceptor-like Zn_{Ga}^{-}. As a result, such mechanism creates the appearance of an abnormally fast diffusion of the acceptor-like impurity zinc by sites.

It should be noted that the diffusion mechanism for impurities in many elementary and complex semiconductors to date is sufficiently studied, which cannot be said about the mechanism of self-diffusion, especially in the binary semiconductor compounds.

Let us briefly consider possible mechanisms of self-diffusion in binary semiconductors AB:

antisite self-diffusion model, in which the atom of A-sublattice goes to the nearest vacancy of B-sublattice, is performed due to generation of *antisite defect* (atom A in the site of the sublattice B, and vise versa (Fig. 1.1.38a);

divacancy model, when vacancies are generated in two neighboring atomic sites, so that atom migrates to the nearest neighboring vacancy of its own sublattice (see Fig. 1.1.38b);

model of pushing-out when the atom is moved from the site in the direction to the atom in neighboring site of own sublattice pushing it to an interstitial position (see Fig. 1.1.38c). This mechanism was considered above in the discussion of possible variations of the interstitial mechanisms;

model of displacement of the atom from the site over the intersticial to the closest vacancy in its own sub-lattice (see Fig. 1.1.38g), similar to the dissociative mechanism.

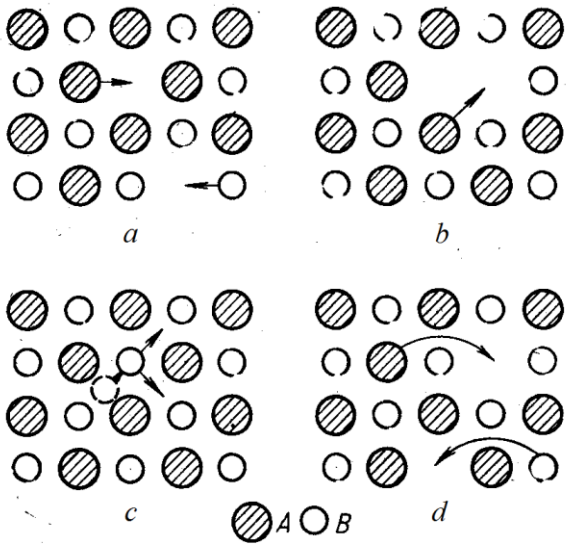


Fig. 1.1.38. Self-diffusion mechanisms in binary semiconductot compounds A3B5: *a*) antisite; *b*) divacancy; *c*) pushing-out; *d*) site-interstitial-site

Influence of structural defects on the diffusion parameters. One manifestation of the high sensitivity of diffusion to the presence of structural defects is the effect of the vacancy concentration and their possible sources on the rate of diffusion.

One of vacancies sources in the crystal are the so-called large-scale defects (LSD), such as dislocations, grain boundaries, stacking faults, precipitates, pores, etc. In particular they stimulate the generation of point defects within the crystal, which eliminates the diffusion to the external surface of the crystal. Even in the case where the generation and recombination of defects occur within the crystal (e.g., the formation and annihilation of Frenkel pairs), LSD often contribute to this process. Therefore, diffusion along LSD goes generally easier than in defect-free crystals due to the presence of damages introduced by LSD to the crystal lattice. The examples of LSD (grain boundaries and surface) influence on impurities diffusion is shown in Fig. 1.1.39. Discuss this question in more details.

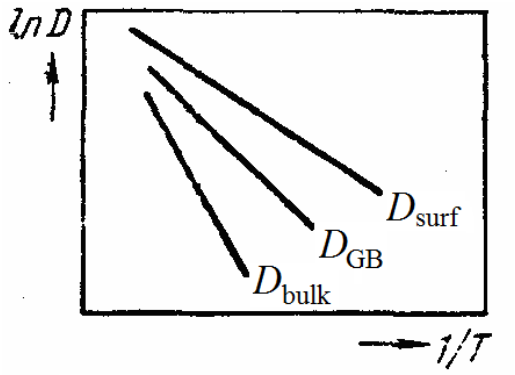


Fig. 1.1.39. The temperature dependences of the bulk (Dbulk), interface (DGB) and surface (Dsurf) diffusion of the same substance

If to coat the surface of a polycrystalline specimen by the diffusant layer (Fig. 1.1.40) and to subject it by high temperature annealing, the diffusion will occure both in grain bulk and grain boundaries (GBs). The ratio of diffusivities Dbulk/DGB is determined by the type of GBs (their atomic structure). From this viewpoint, all of GBs are usually divided into two types: special and non-special (or random, or general).

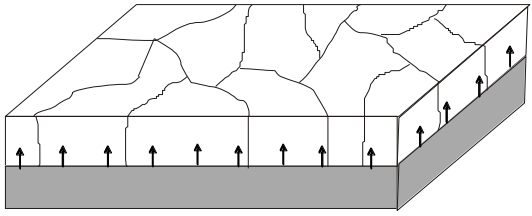


Fig. 1.1.40. Scheme of diffusion from the surface of polycrystalline films with a columnar grain structure

Special boundaries are more perfect (due to the ordered atomic structure) and therefore diffusion occurs along them at a lower rate than along random GBs. Higher D_{GB} values for random GBs are caused by imperfections of their structure, in particular, due to the increased content of vacancies. As a result,

the diffusion along random GBs often occurs at a rate several orders of magnitude greater than in the bulk due to the diffusion activation energy of $Q_{GB} \approx (0.3\text{-}0.7)Q_{bulk}$. This, in particular, leads to the fact that the temperature dependence of the grain-boundary diffusion is significantly lower than that in the bulk. As a result, the lower the temperature of diffusion the greater excess grain boundary diffusion over bulk.

Fig. 1.1.41 shows a diagram of GB diffusion from the source of diffusant with the concentration of C_o at the surface of polycrystalline film. In this scheme grains (crystallites) are distinguished by GBs with thichkness δ, shown as vertical lines. From this scheme follows that, simultaneously with the diffusion along grain boundaries may go both outflow and inflow of impurity atoms from or to GB (depending on expansion or contraction of crystal lattice around GB). As a result, as can be seen, the diffusion front can be non-smooth.

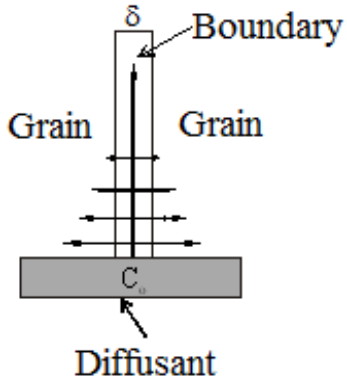


Fig. 1.1.41. Scheme of diffusion along a single GB from the crystal surface

The roughness of the diffusion front is determined by D_{GB}/D_{bulk} ratio, which depends on the GBs structure, diffusion temperature, and other factors. When $D_{GB} >> D_{bulk}$, so that the diffusion length $l = \sqrt{D_{bulk}t}$ in the grain bulk for the time t is small, the side outflow from the GB is virtually equal to zero at a relatively short distance from the GB core (plane). Just this reason, the diffusion front in a polycrystalline sample will be non-smooth. It was also established that the enhanced grain-boundary diffusion is more effective for the samples with low-size grains. The influence of the temperature on the D_{GB}/D_{bulk} ratio can be seen from the following example. In the case of silver self-diffusion at 480°C, $D_{GB}/D_{bulk} \approx 10^6$, whereas at 950°C it decreases dramatically (down to 10^3).

Note that in the literature experimental data on diffusivities in polycrystalline solids are given very often as the averaged values taking into account both GBs and bulk diffusivities. Dislocation bunches affect the rate of diffusion similarly to the grain boundaries. Often *p*-*n*-junctions have non-smooth front of diffusion (non-even interface between *n*- and *p*-regions) due to the enhanced diffusion of dopants along dislocations until their high concentrations is achieved due to accumulation. This dramatically reduces the local electric resistance along GBs and creates the preconditions for the *p*-*n*-junction breakdown.

Surface diffusion runs along the surface as the boundary between solid state and vacuum or gas or liquid. The fact that atoms arranged near the surface of the crystal are surrounded by neighboring atoms not from every side, changes their behavior in the under surface layers. In particular, in semiconductor crystals the interatomic distances and also phonon and electron spectra are changed which alters the position of the Fermi level and some properties of crystals. Especially dramatically, these differences are manifested in the properties of low-dimension samples (nanoparticles, ultrathin films, nanopowders, etc.), which are free from foreign (adsorbed) atoms. In the near-surface layers of such materials are often formed phase states, which differ from those in the bulk. For example, weakening the chemical bonds in the surface layers leads to a substantial reduction of the melting temperature and the thermal conductivity of ultrathin films. Just for all of these reasons, the activation energy of surface diffusion is much lower than by the grain boundaries, the interface (lateral) and, especially, bulk (see Fig. 1.1.39).

One of the characteristics of real surfaces of crystals is that they are almost always have different roughness, as well as adsorbed atoms. If there are roughnesses, the latter play the role of trapping centers (sinks) for migrating impurities, retarding their diffusion.

The diffusion of atoms adsorbed on the crystal surface depends on many factors. It is usually assumed that the activation energy Q_{surf} for the jump of adsorbed atoms along the surface is equal to the sum of two energies – its capturing by surface Q_a and energy of jump (displacement) activation Q_{dis}.

Features of diffusion in semiconductor crystals. In semiconductors heterodiffusion has several distinctive features compared to heterodiffusion in metals. These features are explained by several factors, the main of which is the fact that the impurities and point defects in semiconductors are electrically active (charged).

Since the diffusion of impurity atoms typically occurs in the presence of another (or several) types of impurities, as well as electrically active vacncies, there are different types of interaction (the change in the concentration of electrically active point defects, the formation of complexes, the formation of

internal electric fields and fields of elastic stresses, etc.) between them. This should affect the activation energy of the process.

The presence of point defects and several types of impurities in different charge states in semiconductor crystal often has also a decisive influence on their solubility and diffusivities. Significantly lower packaging density (greater looseness) of crystal lattices in semiconductor crystals as compared with metallic materials makes interstitial diffusion in semiconductors more probable. Let us consider these features of diffusion in semiconductors.The influence of the charge state on impurities solubility and diffusion in semiconductor crystals is easy to understand on the basis of concept which allows to describe interactions between defects and also charge carriers in semiconductors on the language of quasi-chemical reactions.

Let us consider the p-type semiconductor, which is in equilibrium with the acceptor impurities both in gas and solid phases (when crystal growing from the gas phase). This equilibricity can be described by the following quasi-chemical reaction

$$A_{\text{gas}} \Leftrightarrow A_{\text{solid}} \Leftrightarrow A^{-} + e^{+}, \tag{1.1.44}$$

where A_{gas} and A_{solid} are non-ionized atoms in the gas phase and solid semiconductor respectively, and the A^{-} and e^{+} – ionized acceptors and holes in semiconductor.

For donor impurity quasichemical reaction can be written as

$$D_{\text{gas}} \Leftrightarrow D_{\text{solid}} \Leftrightarrow D^{-} + e^{+}, \tag{1.1.45}$$

where D^{+} and e^{-} are ionized donors and electrons in semiconductor respectively.

In Eqs. (1.1.44) and (1.1.45) it is assumed that the impurity is introduced either in site or in interstitial positions in the same charge state.

In addition to the reactions of the type (1.1.44) and (1.1.45, the reaction describing the interaction between electrons and holes (their recombination) can also take place in a semiconductors

$$e^{-} + e^{+} \Leftrightarrow e^{-}e^{+}. \tag{1.1.46}$$

From Eq. (1.1.46) follows that increasing of electron density (due to ionization of donors) in the presence of ionized acceptors will reduce the number of the latter due to mutual electron-hole recombination in accordance with reaction (1.1.46). This, consequently, will result in the increase the solubility of the donor impurities in the semiconductor (as size of ionized donor is always smaller than in neutral state).

The study of the effect of the Fermi level position on the solubility of impurities, being in both interstitial and substitutional positions in different charge states, have shown that the equilibrium ratio of impurities solubilities in

these charge states depends both on the parameters of the crystal band structure and also on the concentration of electrons and holes in semiconductor.

Heterodiffusion features in elementary semiconductors. The rate of impurities diffusion in semiconductor crystals may vary by 4-6 orders of magnitude. Therefore, from this point of view, all the impurities in semiconductors can be divided into *slow* and *fast diffusing*.

The *slowly diffusing impurities* in silicon and germanium include, above all, the group III elements of the periodic table (boron, aluminum, gallium, indium), and group V (phosphorus, arsenic, antimony), which diffuse mainly by vacancy mechanism. Their diffusivities are within 10^{-10}-10^{-14} cm^2/s, that is approximately two orders of magnitude more than the diffusivities of selfdifusion in silicon and germanium at working temperature range. Activation energies of these impurities diffusion comprise 3.5-4.0 eV (Table 1.1.2).

To the *rapidly diffusing impurities* in semiconductors, significantly more items is related: primarily the elements of Group I (lithium, copper, gold) and Group VIII (iron, cobalt, nickel) of periodic table, and such transition metals as titanium, chromium, etc. This type of impurities diffuses mainly along interstitials. Their diffusivities are in the range 10^{-4}-10^{-8} cm^2/s (see Table 1.1.2), and the activation energy – 0.35-1.0 eV.

Table. 1.1.2. Diffusivities and activation energies of some impurities in silicon and germanium

Chemical element	*T*, K	*D*, cm^2/s					*Q*, eV	
		Ag	Zn	Ga	As	Fe	Ga	As
Si	1600	$1.5 \cdot 10^{-6}$	$1.0 \cdot 10^{-6}$	$2.0 \cdot 10^{-11}$	$2.0 \cdot 10^{-12}$	$1.0 \cdot 10^{-5}$	3.9	4.2
Ge	1100	$8.8 \cdot 10^{-7}$	$1.0 \cdot 10^{-12}$	$1.0 \cdot 10^{-13}$	$1.0 \cdot 10^{-11}$	$1.6 \cdot 10^{-6}$	3.1	2.5

Note that fast and slow diffusing impurities in silicon and germanium also differ sharply by their solubility limits. Slowly diffusing impurities are relatively soluble in silicon and germanium, so that their solubility limits normally reaches concentrations of 10^{19}-10^{21} cm^{-3}. At the same time, rapidly diffusing impurities are typically dissolved very few, in the range of 10^{14}-10^{16} cm^{-3}.

Electrical activity of impurities in semiconductors. Electrical activity of diffusing impurities causes a change in the concentration of excess vacancies. This leads to the fact that if the vacancies in silicon behave as acceptors, then any impurity diffusing by the sites in the material will have higher diffusivities in n-type than in p-type material. This is due to the electrons in electronic semiconductor with high electron concentration *n*, when passing to vacancies

from donor impurities, cause a decrease in the ionic radius (and therefore, the growth of the diffusivity) of the impurity.

Another important factor is that the relative rate of diffusion of shallow donors (Group V element) and acceptors (Group III elements) in germanium and silicon are different. As can be seen from Table 1.1.3, in silicon $D_{III} > D_V$ and $Q_{III} < Q_V$, while vice versa in Germany.

Since the atomic radii difference ΔR between the elements of groups III and IV is more than for V and IV groups elements, rapid diffusion of the acceptors (with smaller ionic radii) in silicon is well explained by their radii difference $\Delta R/R$ (as in the case of metals). However, this explanation is not acceptable to germanium, where the enhanced diffusion of donors (compared with acceptors) requires a different explanation.

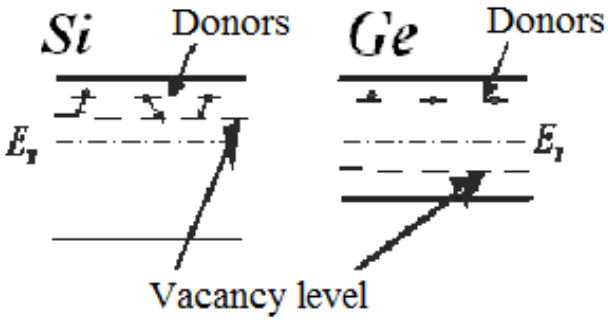


Fig. 1.1.42. The scheme of donor and vacancy levels in silicon and germanium

Apparently, the reason here consists in the fact that the acceptor level of vacancies in silicon lies much higher than in germanium (Fig. 1.1.42). Because it lies in Si by 0.16 eV below the bottom of the C-band, whereas in Ge – by 0.26 eV above the V-band, so the charge states of vacancies in Si and Ge are different. It can be assumed that the portion of the ionized vacancies in germanium at 800 °C is higher than in silicon at 1300 °C. Just due to this reason, the impact of Coulomb forces results in the fact that the regions around the donor impurities in germanium are enriched by acceptor-like vacancies to a greater extent than the same in silicon.

Features of diffusion in semiconductor and dielectric chemical compounds. The presence of various kinds of charged impurities and structural defects in the binary semiconductor compounds can significantly influence the processes of diffusion, self-diffusion and solubility. Thus, as in elementary semiconductors, the interaction between the impurities and defects, on the one hand, can lead to the formation of complexes of impurities and defects. On the other hand, the

interaction between the charged impurities or structural defects associated with changes in the conditions of semiconductor electric neutrality, i.e., the position of the Fermi level, can have a significant impact on the processes of diffusion and solubility of impurities in the compound.

Let us consider physical principles lying inherently in electron-hole equilibrium between the introduced charged impurity ions and shallow donors and acceptors uniformly distributed throughout the compound semiconductor crystal. Diffusion in compounds proceeds with much lower speed than in elementary substances and has a number of important features. Let us consider reasons for such behavior.

The first important feature is that the self-diffusion and heterodiffusion of substitutional impurities can occur in binary compounds through only their sublattices. This is usually observed more strictly, the greater the percentage of ionic bonds in the compound is.

This means that the difusing ion jumps not on the closest site, but on the site of the second coordination sphere, in which atoms of the same sublattice are arranged. To make such jump, the diffusing ion must overcome the Coulomb attraction of the ions of opposite sign, located at the sites of the nearest (first) coordination sphere. That is why, the activation energy for diffusion in the compounds is significantly higher then activation energy in elementary semiconductors. The activation energy for self-diffusion and heterodiffusion for several $A^{III}B^{V}$ compounds are given in Table 1.1.3.

As is evident from the data shown in Table 1.1.3, the second important feature is that the the activation energy for diffusion tends to rise with the increasing covalent and ionic bonds (reduction of metallic one) and decreases sharply for interstitial mechanism of diffusion (see the data for the diffusion of Cu and Au).

Table 1.1.3. The activation energies of self-diffusion and heterodiffusion for some semiconductor compounds

Compound	AlSb				GaSb			InP			GaAs		
Diffusing element	Al	Sb	Zn	Cu	Ga	Sb	Au	In	P	Cu	Ga	As	Au
Q, eV	1.8	1.5	1.93	0.36	3.15	3.45	1.13	3.85	5.65	0.69	5.6	3.45	2.1

The third important feature is that the diffusion usually runs faster by the sub-lattice with weaker bonds and accordingly larger amplitudes of the thermal vibrations. In $A^{III}B^{V}$ compounds, diffusion typically proceeds faster in sublattice A (metal).

The forth important feature is that, in the case of deviations from the stoichiometry, the diffusion is enhanced in that sublattice, which contains extra stoichiometric vacancies. This regularity is extremely important to explain such processes as oxidation and its differences in metals and semiconductors. Let's consider an example of oxidation of iron to form FeO (see Fig. 1.1.43a) and silicon to form SiO_2 (see Fig. 1.1.43b).

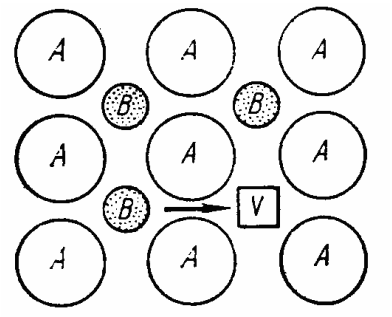


Fig. 1.1.43. Diffusion in semiconductor compounds through vacancies in one of sublattices

Both of FeO and SiO_2 compounds are phases with the broken stoichiometry due to excess of vacancies. Iron oxide has, as a rule, an excess of oxygen, so that vacancies are in the iron sub-lattice (and therefore more accurate formula is $F_{1-x}O$). In the silicon oxide, an excess of silicon is present, so that the excess vacancies are located in the oxygen sublattice (more accurate formula SiO_{2-x}).

As a consequence of the presence of vacancies in different sublattices, the growth of the iron oxide film is due to preferential diffusion of iron through the sub-lattice with vacancies in the oxide layer. As a result, actually oxidation (thickening of the oxide layer) is proceeding at the interface iron-oxygen. In silicon, on the contrary, the oxide film grows due to preferential diffusion of oxygen in its sublattice, so that the oxidation occurs at the interface silicon oxide-silicon because oxygen diffuses into the silicon oxide by the oxygen sublattice. The silicon content in the oxide layer is variable and increases the closer to the silicon.

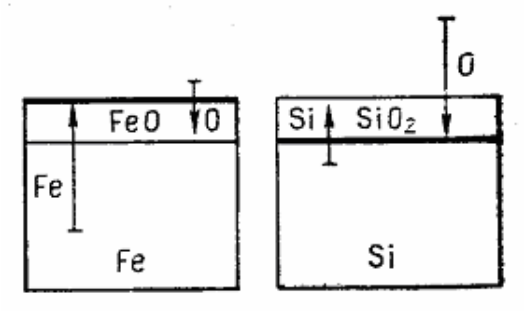


Fig. 1.1.44. Scheme of the iron (a) and silicon (b) oxidation. Arrows indicate the directions of the respective components diffusion through the oxide layer

The effect of external electric fields on diffusion in semiconductors. In many applications (e.g., solid-state microelectronic devices) processes of forced diffusion (electromigration), caused by an external electric fields applied to the sample, take place. When electric field is applyed to semiconductor, except drift of impurity ions in the external field, their interaction with free charge carriers (electrons and holes) arises. This leads to the *drag effect of ions* by electrons (holes) "wind". Under certain conditions, the latter has a significant effect on the migration of impurity ions. The study of electromigration allows (along with diffusion parameters, data about the signs, magnitudes of the charge and mobilities of the diffusing ions) to obtain information about the interaction of free charge carriers with diffusing ions.

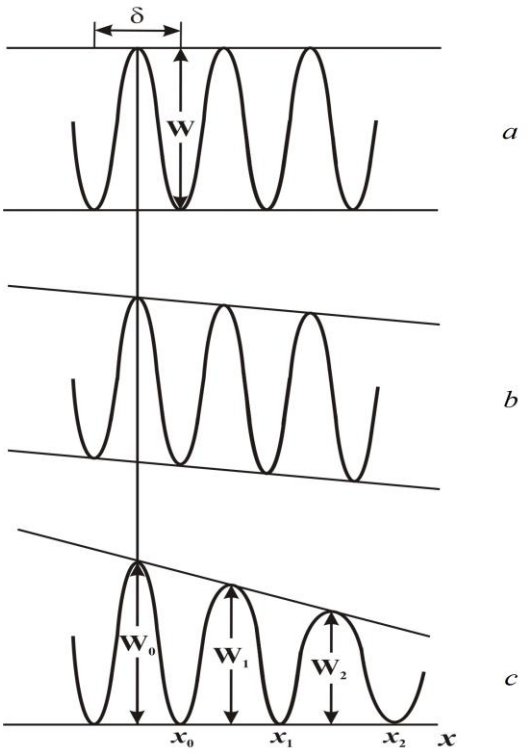


Fig. 1.1.45. The changes in periodic (with periodicity δ) of the potential relief in homogeneous (*a*, *b*) and inhomogeneous (*c*) semiconductor: *a*, *c* – in zero electric field; *b* – the electric field with intencity E is applied to homogeneous crystal

When applying an external electric field of strength E to a semiconductor crystal along the x-axis (in the positive direction), the potential energy W of a periodic lattice is increased by the amount of $\sim qEx$. This leads to the fact that the height of the potential barrier W between atoms changes by $qE\delta/2$ (see Fig. 1.1.456). This results in impact of two forces on the impurity ion in semiconductor – the force of the external electric field $F_{dr} = qE$ (*drift contribution*) and force from the electrons (electron wind) moving toward the anode. Just the second force creates a drag effect of ions with electrons, due to the scattering of electrons by ions.

Let us define the expressions for both (*drift* and *drag*) *forces* in the free-electron approximation. Without electric field, distribution of the electrons by energies and velocities is isotropic and the resultant force acting on an ion from the side of the electrons is equal to zero. When an external electric field E applying, the electrons are accelerated between collisions for a time $t = l_n/v$

acquiring additional momentum $\Delta p = eEl_n/v$, where l_n – the mean free path of electrons. It is assumed that an electron loses this extra momentum during a collision with an ion, transmiting it to the ion. The drag force, acting on ion by the electrons, in this case is

$$F_{ei} = eEnl_n\sigma_i, \tag{1.1.47}$$

where n is the electrons concentration, and σ_i – scattering cross section of electrons by ion.

Taking into account the impact of drift force from the external electric field eE, acting in the opposite direction, the resultant force F, acting on the ion is

$$F = eE(1 - nl_n\sigma_i). \tag{1.1.48}$$

The experimentally observed mobilities of ions, that are effective mobilities μ_{eff}, are associated with true (drift) mobility μ_o by the following relationship:

$$\mu_{eff} = \mu_0 (1 - nl_n\sigma_i) \tag{1.1.49}$$

As foolows from Eq. (1.1.48), if $nl_n\sigma_i \neq 1$, the non-zero total force acts on impurity ion. Moreover, if $nl_n\sigma_i > 1$, this force is directed toward the anode, and at $nl_n\sigma_i < 1$ – to the cathode.

Thus, the necessary condition for movement of impurity ions through the crystal lattice at $E = 0$ consists of the transmission to them the sufficient kinetic energy (from the thermal vibrations) to overcome the potential barrier of the lattice and to jump to the next available equilibrium position. If $E \neq 0$, such activated ions can also perform jumps under additional influence of an electric field and being dragged by electrons. The drag of the ions with electrons in metals would be apparent at a relatively low temperature, as the concentration of electrons in metals is weakly dependent on temperature. In the lightly doped semiconductors dragging effect should be observed at a higher temperatures, because only for sufficiently high temperatures electron concentration reaches a significant level, and the product $nl_n\sigma_i$ could be close to unity. Neutral atoms, when the external electric field applying, will be dragged by the electrons to the anode, so that their effective mobility will be determined by the relation

$$\mu_{eff} = nl_n\sigma_i/\mu_0. \tag{1.1.50}$$

For negative ions dragging effect is added to the motion of electrons in an electric field

$$\mu_{eff} = \mu_0(1 + nl\sigma_i) \tag{1.1.51}$$

i.e. the effective mobility of negative ions increases. In the case of compensated semiconductor (with two types of impurities ions – donors and acceptors), the experimentally measured mobility of donors will be

$$\mu_{eff} = \mu_0[1 - (nl\sigma_i)_n + (nl\sigma_i)_p], \tag{1.1.52}$$

where $(n,\ l,\ \sigma_i)_n$ and $(n,\ l,\ \sigma_i)_p$ are concentrations, the mean free paths and the scattering cross sections of carriers by ions (the indices *n* and *p* refer to electrons and holes, respectively).

The drag force F_{ei} (see, Eq. (1.1.47) was derived in the free-electron approximation, when the momentum transmitted to scattering centers is equal to the difference of the electron momentum in the initial and final states. However, in a real crystal, the electron is subjected to the force from the periodic lattice field V_r. Therefore, in accordance with the conservation law for quasi-momentum, the momentum that the electron transfers to impurity ion will not be equal to the difference between average electron momentum before and after scattering, because part of this momentum can be transferred to the lattice as a whole (see Chapter 1).

For the arbitrary electron dispersion law, the drag force F_{ei} has the same form as for free electron approximation, if the total average momentum of an electron is replaced by average quasi-momentum.

1.1.9. Mechanic properties of crystals

Mechanical properties of crystals reflect their reaction to external mechanical stresses such as compression, extension, bending, blow, torsion, incorporation of a foreign bodies, etc. The mechanical properties of crystals primarily belong to the most important material properties such as *strength, plasticity, hardness* and *brittleness*. The mechanical properties of any solid material are determined by the type of external forces and also interatomic bonding, as well as the type and concentration of structural defects.

Mechanical stress and strain. When acting of external forces on crystalline material, the interatomic distances are changed, so that the crystal lattice is distorted. This leads to a violation of the balance of interatomic interaction forces (attraction and repulsion) that is typical for the equilibrium state of the atoms in the lattice. The arising, in so doing, internal forces tend to return the atoms in the initial equilibrium positions. In this case, we say that the body is in a *stressed state*.

When the body is in the stressed state, two types of forces act on every its volume element:

1. *Body forces* (such as gravity) acting on all elements of the body and proportional to the volume of the body.

2. The forces acting on the surface of the allocated element from the surrounding body parts and proportional to the element surface area. Such force, per unit surface, are called *stresses*.

If the stresses are distributed uniformly over the cross section of the body, the relationship between force and stress is given (for example, the axial

extension of isotropic homogeneous cylindrical rod in Fig. 1.1.46a) by the relation:

$$F = \sigma S\,. \tag{1.1.53}$$

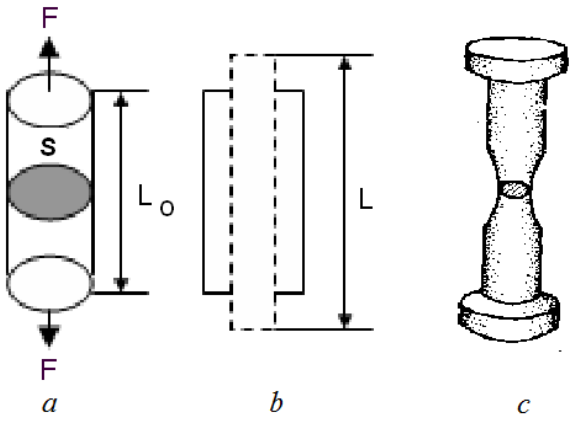


Fig. 1.1.46. Tensile scheme (*a*), the result of deformation (*b*) of a homogeneous isotropic cylindrical rod-like sample and real stretch of inhomogeneous cylindrical sample with a "neck" formation (*c*)

However, we should differentiate the true and engineering (conventional) stresses. *Engineering (conventional) stress* is defined as the ratio of the tensile forces F acting on the rod-like sample with the initial cross-section $S = S_0$ in the entire range of deformations, including the destruction of the sample (Fig. 1.1.46a). *The true stress* is defined as the ratio of the force F, applied to a solid sample, to the actual (lowest) value of its cross-section S, which may change at sufficiently high stresses (Fig. 1.1.45b).

As shown in Fig. 1.1.46c, for an inhomogeneous distribution of stresses during sample elongation due to its large deformation, a local contraction can be gradually formed in the form of a "neck". In this case, the true and engineering stress differ essentially (see below). In the case of such an inhomogeneous distribution of stresses in the isotropic body relation between the force F and stress σ in static equilibrium is given by relation

$$F = \int \sigma \mathrm{d}S. \tag{1.1.54}$$

The stressed state that occurs in a solid under load test, substantially affects the processes of its deformation and destruction.

Deformation is a change of volume and/or shape of solid body without change its mass under the action of an external force. During the deformation

the distances between any points of the body is always changing. Elementary deformation under *uniaxial extension (compression)* of the cylindrical sample consists of its lengthening (shortening), Fig. 1.1.46b.

Typically, the deformation of the sample with an initial length L_0 and the length after deformation L is characterized by *strain*

$$\varepsilon_d = \frac{L - L_0}{L_0}, \tag{1.1.55}$$

which shows the relative deformation of the sample at tensile stresses.

Stress-strain diagram. Fig. 1.1.47 shows dependence of strain ε_d from conventional stress $\sigma = \frac{F}{S_0}$, where F is a load applied to the sample and S_0 – its initial cross section. Such curve is called *the stress-strain diagram* for the tensile load.

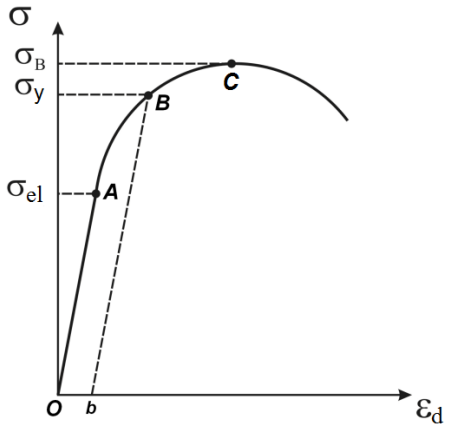


Fig. 1.1.47. The stress-strain dependence for a defect free crystal

Under infinitely slow (reversible) and minor stretch, internal forces acting on any area, ideally selected in the crystal, balance the external load applied to this area. For small displacements u of atoms in crystal lattice, forces of reaction F tend to bring them back into the equilibrium position. In the first approximation, this force can be taken as proportional to these displacements ($F \approx ku$). As a result, at small deformations the initial part of stress-strain curve (section OA in Fig. 1.1.47) in the crystal is approximately linear

$$\sigma = E \cdot \varepsilon_d, \tag{1.1.56}$$

where E is elastic modulus. The elastic modulus E is the material constant, which depends only on the nature of chemical bonds and structure of the crystal. The relation (1.1.56) is called Hooke's law.

Thus, as follows from the stress-strain curve in Fig. 1.1.46, in its initial part the deformation will be elastic, i.e. it should disappear after removal the load. This elastic deformation exists until the applied stress does not exceed the certain conditional stress corresponding to the point A in Fig. 1.1.47. So, the stress σ_{el} is called *the limit of elasticity*.

When the applied stress is increased above the elastic limit σ_{el}, we observe either destruction of the crystal or dependence $\sigma(\varepsilon_d)$ deviates from linearity (Hooke`s law is violated). For the first case the material is considered as *brittle (fragile)*, whereas in the second case as *ductile (plastic)*.

If the material under tension is sufficiently plastic, the increase of the load over the σ_{el} (beyond the point A in Fig. 1.1.47) results in the nonlinear stretching of the sample, so that the $\sigma(\varepsilon_d)$ curve goes by line OAB. If we stop the loading at the point B and start to reduce the stress to zero, the curve $\sigma(\varepsilon_d)$ will follow the line of B_b. This means that after unloading the sample becomes longer than was prior to loading. This means the presence of *residual* (*plastic*) deformation $\varepsilon_{res} = \varepsilon_b$ in the sample after removal the external load.

Nonlinear progress of stress-strain diagram in Fig. 1.1.47 will continue until approaching the maximum point *C*, which is called *the tensile strength* $\sigma_в$. With further load increase the $\sigma(\varepsilon_d)$ curve goes down until the sample is breaked. For ductile metallic materials, the magnitude of $\sigma_в$ can reach the values up to 10^8 Pa.

Let us consider the causes which arouse the stress drop beyond the point C in the $\sigma(\varepsilon_d)$ diagram in Fig. 1.1.47 with the strain increase. As was mentioned above, when applying the tensile load σ_C to real ductile metallic material, a strong local stretching of the sample can arise due to the inhomogeneity of the mechanical properties by the sample length. This results in constriction of the loaded sample in the form of a "neck" (Fig. 1.1. 46) and, correspondingly, the increase the actual stress in the "neck", since the cross-section of the sample is reduced. But if we use, as before, formula $\sigma = \frac{F}{S_0}$ to calculate σ values, this leads to decrease in stress-strain curve beginning from the point C in Fig. 1.1.47. If we shall use the real cross section of the "neck" in formula $\sigma = \frac{F}{S_0}$, the curve $\sigma(\varepsilon_d)$ go up beyond the point *C* until the sample will broken.

Deformation mechanisms in defect free crystals. Elastic and plastic deformations are of a different nature. Elastic deformation changes the distances

between atoms in a crystal lattice. Since the interatomic forces in the lattice are very large, elastic deformation can only be very small and generally does not exceed values $\varepsilon_d \sim 0.01$ %. Plastic deformation in the perfect (defect-free) crystal can occur as a result of *shear stresses* that can cause irreversible shift in some parts of the crystal relative to the other without breaking the bonds between them.

As can be seen from Fig. 1.1.48, in the crystal under impact of external shear stress σ higher than elastic limit σ_{el}, the top part of the crystal is shifted relative to its bottom part on one or several atomic distances along a specific plane *S*, called the *slip plane* (Fig. 1.1.48c). After removal the shear load $\sigma > \sigma_{el}$, residual deformation of the crystal is conserved (Fig. 1.1.48d). The residual deformation of a defect-free crystal as a whole is composed from such small irreversible shifts, occurring in many slip planes.

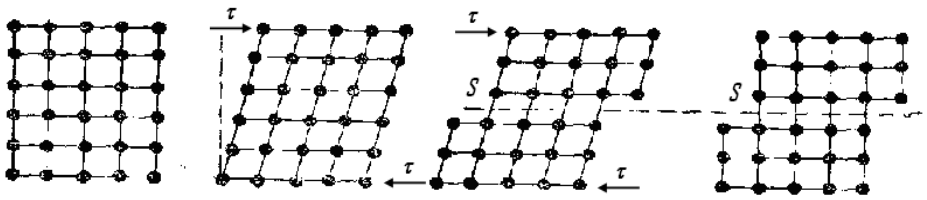


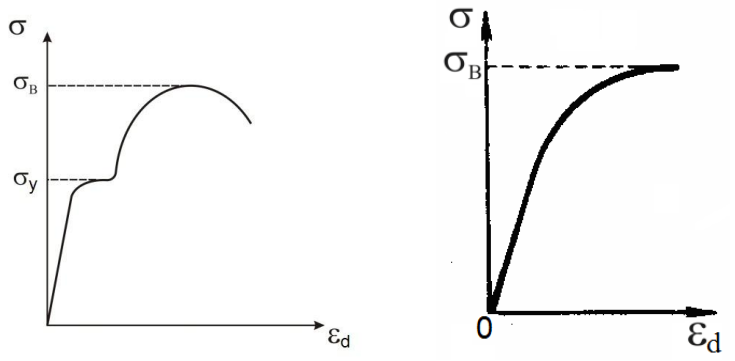
Fig. 1.1.48. Deformation of defect-free (ideal) crystal under impact of external shear stress σ: *a*) initial, unstressed, crystal; *b*) elastic shear deformation at the applying of $\sigma < \sigma_{el}$; *c*) plastic shear deformation due to sliding of the top part of the crystal relative to its bottom part along the plane S under $\sigma > \sigma_{el}$; *d*) an ideal crystal after removal the load $\sigma > \sigma_{el}$ conserves residual deformation due to shift of the top part of the crystal relative to its bottom part along the plane S

Plastic deformation of crystals with defects. In real crystals defects are always present. From the point of view of the mechanical properties, edge dislocations are the most important defects because just they determine the character of stress-strain diagrams (the reasons for this are inducated below.)

The presence of edge dislocations leads to the fact that at some point, even a small change of stress causes a sufficiently strong deformation of the crystal (see rapid growth ε_d on the segment BC of the stress-strain diagram in Fig. 1.1.47). In some cases, particular ductile metallic materials have plateau in the $\sigma(\varepsilon_d)$ curve (Fig 1.1.49a) at the stresses beyond the segment BC. The stress σ_y, at which the crystal begins to "flow" under impact of tensile load, is called *the yield stress*. For the steels, in the technical measurements of their

mechanical properties, it is accepted to estimate σ_y as $\sigma_{0.02}$, which corresponds to the stress, giving the residual strain of 0.2 %.

Note, just metallic materials, which display the "neck" appearance under loading, are characterized by yield effect. In the case of non-yield (very often, brittle) materials, sample may be broken without formation of the "neck" (Fig 1.1.49b). In this case, $\sigma_в$ is the maximum stress reached on the stress-strain curve.



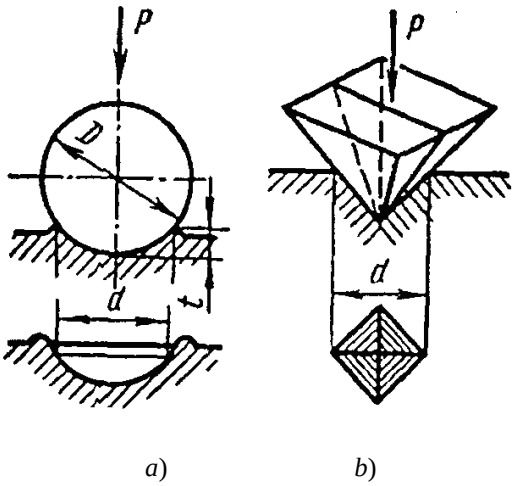
a) *b)*

Fig. 1.1.49. Stress-strain diagrams for metallic materials with high (*a*) and low (*b*) ductility

In amorphous inorganic materials deformation beyond the elastic limit has the character of a homogeneous viscous flow, and its speed is determined by the strain and temperature. Glass is destroyed in the cold state without a preliminary plastic deformation, which corresponds to an ideal brittle fracture. It usually occurs where there is any defect in glass, which is then expanded and breaks the sample. In the fragile metals when the load is greater than elastic limit, there is always plastic deformation (even if small). Before the destruction cracks begin to appear, which then are developed resulting in breaking the sample. Metal is ductile, if you need to do work to create crack, and is fragile, if the crack is expanded after its formation without further increasing the load.

Hardness of crystals. In technology, the hardness of the material is often used as an indicator of its strength. Hardness testing can be defined both on a special samples of materials and finished products (parts). Under *hardness* we understand the resistance of the material to incorporation (penetrating) to it of foreign, more hard, solids.

There are several methods to test the materials or products for hardness. All of them are based on the indentation of a foreign body (indenter) at a certain load into the sample. The indenters are in the form of a balls or a diamond pyramids of various sizes. The hardness testing consists in measurement of the size of indenter image (imprint) obtained at the surface of the tested sample after indentation (Fig. 1.1.50). The softer the material, the deeper indenter penetrates into the sample and the greater the size of the resulting imprint. Measuring the size of the imprint on the surface by one or another way and relating it to the load value, we can estimate the magnitude of hardness.



a) *b*)

Fig. 1.1.50. Testing of hardness of the crystalline sample usind ball-like (*a*) and piramide-like (*b*) indenters

The ratio of the load force *P* to the imprint area *S* after indentation is used as a measure of the hardness *H*. The *S* value is determined either by the imprint diameter *d*, which is formed at the indentation of the ball (see Fig. 1.1.50a), or the diagonal of the imprint for the prism-like indenter (see Fig. 1.1.50b). Typically, the numerical hardness values are determined by the corresponding *H* numbers using conversion factors between *P* and *d* or special tables connecting their quantities with *H*.

To evaluate the local hardnesses of single grains or phases in polycrystalline and/or multiphase samples, the micro- or nanohardness methods are used. In microhardness method, a diamond tetrahedral pyramid (with an apex angle of 136°) is used as indenter. The load on the indenter is small in this case (of about 0.05-5 N), so that the imprint sizes are about 5-30 μm. Microhardness H_μ measurements are performed on a special H_μ–meters equipped with an optical

microscope and a special mechanism of loading. Microhardness H_μ is evaluated via the averaged diagonal of several imprints. Since the numbers of hardness (H or H_μ) obtained by various methods differ, to enable their comparisons there exist conversion tables to transform corresponding hardness numbers from one method to another.

Mechanisms of strength and ductility in crystals with structural defects. The presence of defects in the crystal lattice has a strong (sometimes decisive) influence on the mechanical properties of crystals (especially, on their strength and ductility). For example, the hardness of damaged crystals is typically increased. As to the ability of crystals to plastic deformation, it may both increase and decrease due to introducing of defects into the crystal.

Using dislocation theory provided an explanation for a large difference between the theoretical and the real strength of ductile metallic materials. It can be assumed that the theoretical specific strength of the crystal should be proportional to the product of the interatomic bonding forces on number of atoms in the unit cross section of the crystal (about 10^{14} atoms/cm^2). However, based on this assumption, the estimated σ_b values needed to shift one part of the crystal relative to the other, is 2-3 orders of magnitude higher than those that actually cause plastic deformation of the crystal in experiments.

For example, the theoretical strength of iron is about $1.3 \cdot 10^5$ MPa, while the real – only 250 MPa. Such a discrepancy of theoretical and real strengths indicates that the deformation can not be explain by simultaneous breaking of atomic bonds at the whole atomic plane S in Fig. 1.1.48c. This effect was explained by the gradual sliding of dislocation along the slip plane (see Fig. 1.1.51a-c). For such a model, shift of dislocation on one interatomic distance leads to a temporary breaking of the only 10^7 bonds per 1 cm^2 which are situated along the dislocation extra plane. Net result of dislocations motion in the crystal (i.e. deformation process) changes the form of the crystal, as is seen from Fig. 1.1.51d.

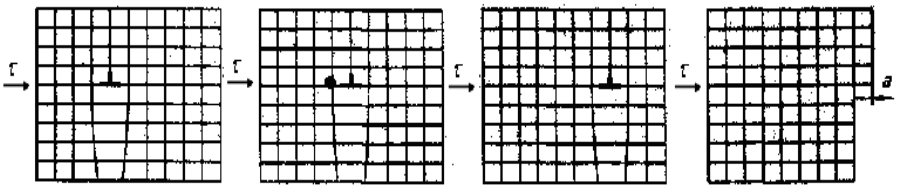


Fig. 1.1.51. The sequential stages of the edge dislocation sliding during plastic deformation under impact of shear stress τ

The model in Fig. 1.1.51 shows that dislocations can easily move normally to the extra plane under an applied shear stress. The easier dislocations move, the lower the strength of the crystal (the easier plastic deformation). Plastic deformation of crystalline materials is closely related to the density and mobility of dislocations, their width (the size of lattice distortions along the dislocation line) and the degree of interaction with the lattice defects, etc. The nature of the interatomic bonds also affects the crystal ductility. Thus, the non-metallic crystals with the directed bonds (e.g., covalent semiconductors) dislocations are very narrow, and require 10^3 times larger, than for metals, starting stresses to begin the dislocations movement. As a result, in non-metallic crystals brittle fracture occurs earlier than dislocations start their motion (see Fig. 1.1.52).

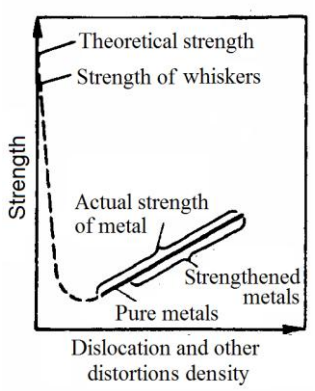


Fig. 1.1.52. The influence of lattice distortions on the strength of crystals

Growth of dislocation-free crystals leads to a sharp increase of the materials strength. It is seen from in Fig. 1.1.52, where dependence of strength on dislocations density N_D of the crystalline materials is shown. Really, the left branch of the curve tends to the theoretical strength with decrease that corresponds to the perfect dislocation-free crystals (the so-called "whiskers"). With the limited (but not too high) N_D values, the process of deformation occurs stronger due to the easier dislocation motion and strength goes down.

When increasing the applied load, the density of dislocations increases. In doing so, in addition to the multiplication of dislocations in parallel planes of their best sliding, they are also generated in other (intersecting) planes and directions. As a result, dislocations begin to interact with each other. On one hand, some of dislocations annihilate (cancel). On the other hand, with the N_D increasing, their motion becomes more difficult (because the intersecting

dislocations prevent the movement of each other), that requires increasing the applied load to continue deformation. This retardation of dislocations movement corresponds to the hardening of the crystal, which is shown on the right branch of the curve in Fig. 1.1.52.

The other damages in the crystalline structure also contribute to the strengthening of the crystal, as well impeding the motion of dislocations. These imperfections include the dissolved impurities and alloying elements, the particles of the second phase, grain boundaries, etc. The reduced temperature also prevents the free movement of dislocations. Therefore, at low temperatures, the strength increases and the ductility fall down, so that the metallic material becomes more strong (hard), but brittle. Thus, as follows from Fig. 1.1.52, increase in strength of ductile crystalline materials can be achieved in two ways: either by obtaining materials with nearly-ideal structure of the crystal lattice (defect-free crystals) or, on the contrary, by increase of structural imperfections blocking the dislocations movement.

Brittle fracture of the crystals. Apart from ductile fracture, accompanied by yielding, crystalline materials may experience brittle fracture. Last occurs either after preliminary plastic deformation (Fig. 1.1.49a) or immediately beyond the tensile strength on the stress-strain diagram (in particular, in the case of metals or non-metals at low temperatures), as was seen in Fig. 1.1.49b. For the case of an ideal crystal is considered that brittle fracture occurs due to instantaneous break of interatomic bonds in the plane normal to the shear stresses (see Fig. 1.1.48). Theoretical estimates under this assumption show that the fracture stress should depend on the modulus of elasticity according to the law

$$\sigma_{el} \propto \frac{E}{10}. \tag{1.1.57}$$

However, comparison with the experiments indicates that the actual stresses for brittle failure of the crystal is 2-3 orders of magnitude lower than the theoretical values. The explanation was given by Griffiths through the mechanism of nucleation of cracks in brittle solids, which can play the role of stress concentrators. The generation of crack leads to its avalanche-type growth at a certain load and the destruction of the crystal.

1.2. Atomic Dynamics

As noted in section 1.1.6, at finite temperature, the atoms of the crystal lattice are always oscillate around their equilibrium positions – SL sites. As a result, in a crystal, being in thermal equilibrium with the environment, a steady state in the form of standing or progressing waves is established.

The character of these oscillations depends on the crystal symmetry (structure), the number of atoms in the unit cell, the type of chemical bonds, as well as the type and concentration of lattice defects. Displacements of atoms from the equilibrium positions during these oscillations is greater the higher the temperature, but they are much smaller than the lattice parameter up to the melting temperature when transforming of solids into liquid state.

The forces, that tend to hold the atoms in the equilibrium positions, as a first approximation can be considered proportional to their relative displacements as if they were connected to each other by elastic springs (Fig. 1.2.1). Representation of the crystal as an ensemble of particles bounded by purely elastic forces is called a *harmonic approximation*.

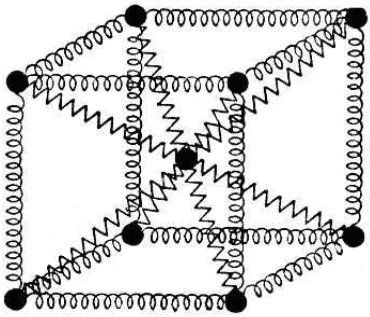


Fig. 1.2.1. Presentation of the bcc crystal in the form of a set of atoms bound to each other by spring bonds

The harmonic approximation leads to the fact that a system of elastic waves, connected with displacements of atoms from their equilibrium positions, propagates in the lattice. In a crystal consisting of N identical atoms, there are 3N such waves that are called *normal* (or *intrincic*) *oscillations*, or *modes*.

1.2.1. Introduction to quantum mechanics of atoms and electrons in solids. Wave mechanics

Wave properties of microparticles. One of the most important manifestations of the quantum-mechanical properties of the crystal forming microparticles (atoms, ions, electrons, etc.) is formulated by the French physicist Louis de Broglie's principle of wave-particle duality. He expressed the wavelength λ as a basic wave characteristics of quantum-mechanical particles through their momentum $p = hv/c$ which is their basic corpuscular characteristics:

$$\lambda = h / p, \qquad (1.2.1)$$

where h is Planck constant, and v – frequency. The (1.2.1) means that the substance particles have not only corpuscular but also wave properties (and vice

versa). As a result, the movement of any quantum micro particle, having a momentum *p*, can be described by a wave with the wavelength λ given by (1.2.1). Note that the de Broglie relation (1.2.1) is often more convenient to represent in the form of two interrelated equations:

$$E = \hbar\omega \tag{1.2.2a}$$

and

$$p = \hbar k, \tag{1.2.2b}$$

where $\hbar = h/2\pi$ and k is wave vector. In these equalities corpuscular and wave properties relationship of micro particles becomes quite obvious – at the left are corpuscular characteristics of wave-particle (the energy E and the momentum vector p), and at the right – wave characteristics (frequency v and the wave vector k). Note that the motion of such wave-particle is described by the Schrödinger equation.

The de Broglie ratio is applicable to any particles, but the diffraction phenomenon can be observed only for micro particles (electrons, neutrons, protons, etc.). The first confirmation of the de Broglie hypothesis was the discovery of the diffraction of electrons by the crystal lattice in the experiment C. Davisson and L. Germer.

In accordance with quantum mechanics, wave properties of micro particles lead to a principal impossibility to know exactly both their coordinate x and momentum p simultaneously when their route through the small aperture. This uncertainty for one dimension case is expressed by the famous *Heisenberg uncertainty relation*:

$$\Delta p_x = \frac{2h}{\Delta x} \tag{1.2.3}$$

or

$$\Delta x \cdot \Delta p_x \geq h, \tag{1.2.4}$$

where Δx and Δp_x are uncertainties of our simultaneous knowledge of the particle position and momentum. The substitution of the exact value of the momentum (this means that $\Delta p_x = 0$) leads, according to (1.2.4), to full uncertainty of a particle position in space $\Delta x \geq h/\Delta p_x \to \infty$.

Let the electron velocity $v = 10^8$ cm/s is determined with an accuracy of 1 % (i.e. $\Delta v_x = 10^6$ cm/s). Then the uncertainty of its position Δx is not less than 10^{-4} cm, which is much larger than the size of an atom (~10^{-8} cm). But this means that the location of an electron within an atom completely undefined. On the other hand, this uncertainty is quite insignificant in vacuum devices of normal size (from millimeters to tens of centimeters). In this case, each time an electron is always possible to assign a certain speed for a certain point in space.

This means that, upon transition from micro to macro bodies, uncertainty of their position in the space becomes small, so that the motion of a particle can be described in terms of classical mechanics (we have possibility to attribute the trajectory by particle, that is, at the same time with sufficient accuracy to determine their speed and position). The uncertainties of energy ΔE for quantum particle with energy E and Δt (the time which particle stays in the state with this energy E) will be as follows

$$\Delta E \cdot \Delta t \geq h. \tag{1.2.5}$$

From (1.2.5) follows that if we want to absolutely know the energy of any state of the system (this means that $\Delta E = 0$), this state must exist infinitely long ($\Delta t = \infty$). If Δt is finite, for such a state, there is a whole set of energies in an interval ΔE. Just this is observed for the excited states of electrons in atoms, when $\Delta t \sim 10^{-8}$ s.

The Schrödinger equation. The uncertainty principle leads to the need of the probabilistic approach to describing the relationship between the position of the micro particles and time on the basis of the Schrodinger equation. The latter describes the movement of the micro particles using a wave function Ψ (x, y, z, t). The physical meaning of the latter is that probability dW of finding a particle at a given time in a volume dV near the point with coordinates (x, y, z) is determined by

$$\mathrm{d}W = \left|\Psi(x,y,z,t)\right|^2 \mathrm{d}V. \tag{1.2.6}$$

According to this relation the square of the wave function gives the probability density dW/dV to find the wave-particle in a certain space range around some point with the coordinates (x, y, z) in a time moment t. This means that wave function Ψ only gives information about the presence of particles in this volume dV. This also means that in quantum mechanics particle location in space is determined by the probability density (that is not precisely defined) and therefore the notion of definite position (a trajectory) of the particle in the space does not exist.

When movement of a particle in the field of changing external forces, Schrödinger equation is:

$$-\frac{\mathrm{h}^2}{2m}\Delta\Psi + U\Psi = i\mathrm{h}\frac{d\Psi}{dt}, \tag{1.2.7}$$

where $\hbar = h/2\pi$, m – particle mass; $U(x,y,z,t)$ – the potential energy of the particle in the field of external forces; $i = \sqrt{-1}$ – the imaginary unit, and

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial \mathrm{z}^2}$$

is called the Laplace operator. For the special case of a particle along the *x*-axis Schrödinger equation is:

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2}+U\Psi=i\hbar\frac{d\Psi}{dt}. \qquad (1.2.8)$$

In case of the stationary field of external forces (when $U(t) = 0$), the Schrödinger equation is independent on time and has the form

$$\frac{\partial^2\Psi(x)}{\partial x^2}+\frac{2m}{\hbar^2}(E-U)\Psi(x)=0, \qquad (1.2.9)$$

where E – the kynetic energy of the particle. The solution of (1.2.9) for a free particle ($U = 0$) will have the form of two monochromatic plane de Broglie waves

$$\Psi(x)=Ae^{-i(\omega t-kx)}+Be^{-i(\omega t+kx)}, \qquad (1.2.10)$$

extending along the x axis in two opposite directions with the amplitudes A and B, respectively. Here, $k = 2\pi/\lambda = p/\hbar^2 = \sqrt{2mE}/\hbar$ is wave number, and $\omega = E/\hbar$ – circular frequency.

The tunnel effect. The solution of the Schrödinger equation for a free particle ($U = 0$) gives the same probability density to find the particle at the each point of space. Situation will be different in the case of a particle in the field of external forces, when $U(x) \neq 0$.

Consider the behavior of the particle with the kinetic energy E, which meets on its way to a one-dimensional rectangular potential barrier with the width a and the height U_0 (Fig 1.2.2). According to classical ideas, if $E < U_0$, particle should be reflected from the barrier, but if $E > U_0$, it overcomes the barrier.

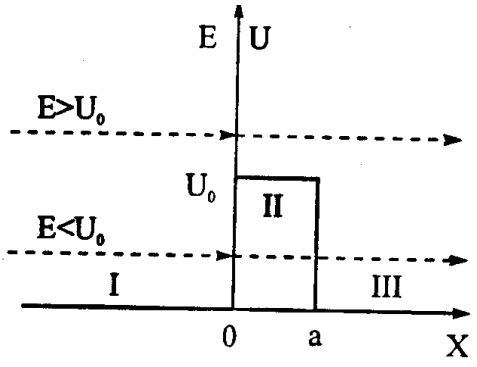


Fig. 1.2.2. Schematic representation of the movement of particles through a rectangular potential barrier with different E/U_0 ratio

From the standpoint of quantum mechanics, for regions I and III of the barrier, the solution of the Schrödinger equation (1.2.9) for free particle ($U = 0$) has the form

$$\Psi_{|}(x) = A_1 e^{ik_1 x} + B_1 e^{-ik_1 x}, \tag{1.2.11}$$

$$\Psi_{|||}(x) = A_3 e^{ik_1 x}, \tag{1.2.12}$$

where

$$k_1 = \frac{\sqrt{2mE}}{\hbar}, \tag{1.2.13}$$

and A_1, B_1 and A_3 are the amplitudes of the waves that fall to the barrier, reflected from it and pass through it, respectively. For the region II of the potential barrier, solution of the Schrödinger equation takes the next form

$$\Psi_{||}(x) = A_2 e^{ik_2 x} + B_2 e^{-ik_2 x}, \tag{1.2.14}$$

where

$$k_2 = \frac{\sqrt{2m(E - U_0)}}{\hbar}, \tag{1.2.15}$$

and A_2 and B_2 are the amplitudes of the waves to be moved in opposite directions. For the case of $E > U_0$, solution of Eq. (1.2.9) gives $B_1 \neq 0$. This means that, unlike the classical case, when the quantum particle motion, the wave reflected from the barrier appears.

A particle in a rectangular potential well. Let us now consider a particle localized in a potential well (of a width a) with impermeable borders, when the height of the barrier tends to infinity. Therefore in the barrier region I and III the potential energy $U \to \infty$ in Fig. 1.2.2, whereas in II $U \to 0$. As the probability to find the particle in regions I and III is zero, so the wave function of a particle is also zero ($\Psi \to 0$) in these regions. The solution of the Schrödinger equation for the region II can be written either in the form (1.2.10), or after the transformation from exponentials to trigonometric functions, as

$$\Psi(x) = C\sin(kx + \alpha), \tag{1.2.16}$$

where $k = \frac{\sqrt{2mE}}{\mathrm{h}}$, and C and α are constants. It follows from the continuity condition $\Psi(0) = 0$ for wave function at the boundary between I and II regions that $\alpha = 0$, so that $\Psi(x) = C\sin(kx)$. It follows from the condition of continuity $\Psi(a) = 0$ at the boundary between regions II and III that the

$$k = \frac{\pi n}{a} \tag{1.2.17}$$

where n = 1, 2, ... From (1.2.17) follows that the wave vector (momentum) of the micro particles, which are in an infinitely deep potential well, is quantized, i.e. takes only discrete but not arbitrary values, unlike in classical mechanics. This leads also to quantization of the kinetic energy values of the micro particles in accordance with the relation

$$E_n = \frac{k^2\hbar^2}{2m} = \frac{\pi^2\hbar^2}{2ma^2}n^2 = \frac{h^2}{8ma^2}n^2. \qquad (1.2.18)$$

These quantized energy values E_n in each (n-th) state particles, shown in Fig. 1.2.3a, are called by *energy levels*.

As follows from (1.2.18), the energy E_n of the particle in any n-th state, and the energy difference between the two neighboring states (distance between adjacent energy levels E_n and E_{n+1} increases with the level number n increasing (see, Fig. 1.2.3). The minimal energy of the particle, localized in an infinitely deep potential well, corresponds to n = 1 and is equal to

$$E_{\min} = E_1 = \frac{h^2}{8ma^2}. \qquad (1.2.19)$$

This means that a quantum particle in the lowest energy state has non-zero energy, in contrast to classical particles. This also means that the motion of a particle in a potential never stops. So this kind of energy is called a "zero energy".

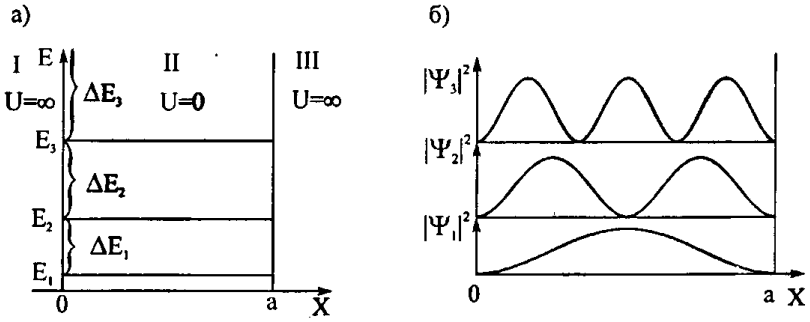


Fig. 1.2.3. The energy levels (a) and the probability density distributions (b) for the infinitely deep potential well with the width a for the first three states of particle

In accordance with (1.2.16), the wave function eigenwaves for particle with energy E_n is equal to

$$\Psi_n(x) = \sqrt{\frac{2}{a}}\sin\frac{\pi n}{a}x. \qquad (1.2.20)$$

This implies that the probability density distribution $|\Psi_n(x)|^2$ by the width of the potential well for different energy states (corresponding to different values of n), are significantly different. The form of these distributions for the first three states of micro particle within the potential well is shown in Fig. 1.2.3b.

Note that the quantization condition (1.2.17) for wave vector corresponds to the formation of a standing wave in region II ($0 < x < a$) in Fig. 1.2.3a with the wavelength $\lambda = \frac{2\pi}{k}$, if an integer number of half wavelengths $n = \frac{a}{\lambda/2}$ is settled within this region.

Quantum oscillator. The quantization of energy has find wide application for the study of the atomic vibrations in the crystal lattice. In the classical approximation, if the particle, displaced from its equilibrium position by a distance x, oscilates under explosion of the elastic force $F = -kx$, its motion can be described by the oscilating process, characterized by the angular frequency $\omega = \sqrt{\frac{k}{m}}$. The potential energy of such classical (harmonic) oscillator is described by a parabolic expression

$$U = \frac{kx^2}{2} = \frac{m\omega^2 x^2}{2}, \qquad (1.2.21)$$

where k is elastic constant (coefficient of elasticity), and m – mass of the particle.

In quantum mechanics, if the particle oscillates under the action of elastic force, it is considered that it is in a parabolic potential well (1.2.21), see Fig. 1.2.4. The solution of the Schrödinger equation for such a quantum harmonic oscillator gives the quantized energies

$$E_n = (n + \frac{1}{2})\hbar\omega \; ; n = 0,1,2,3... \; , \qquad (1.2.22)$$

for the quantum harmonic oscillator, where the minimal energy of the oscillator is in the state with $n = 0$, by analogy with the above case of a quantum particle in a rectangular potential well:

$$E_{\min} = E_0 = \frac{1}{2}\hbar\omega \; . \qquad (1.2.23)$$

Just this energy is called a "zero energy".

The electron energies in atoms. Energy of an electron in an atom can be determined by the Schrödinger equation, if we assume that the atomic electron is in a potential well. The simplest example of this calculation is an estimation of the energy of electron interaction with the positively charged proton in

a hydrogen atom. In accordance with the Coulomb law potential energy of this interaction is $U(r) = -\frac{q^2}{4\pi\varepsilon_0 r}$, where r is radius of the electron orbit, q – electron charge, and ε_0 – dielectric constant (Fig. 1.2.5). The solution of the Schrödinger equation for the potential well of the hyperbolic type gives the quantized values of the kinetic energy of the electron in the hydrogen atom

$$E_n = -\frac{q^4 m}{8h^2 \varepsilon_0{}^2} \frac{1}{n^2} = \frac{E_1}{n^2}, \tag{1.2.24}$$

where numbers n = 1, 2, 3, ... are called *the principal quantum number*. From (1.2.24) it follows that the minimal energy of the electron in the hydrogen atom is

$$E_1 = -\frac{q^4 m}{8h^2 \varepsilon_0^2} = -13{,}6 \text{ eV}. \tag{1.2.25}$$

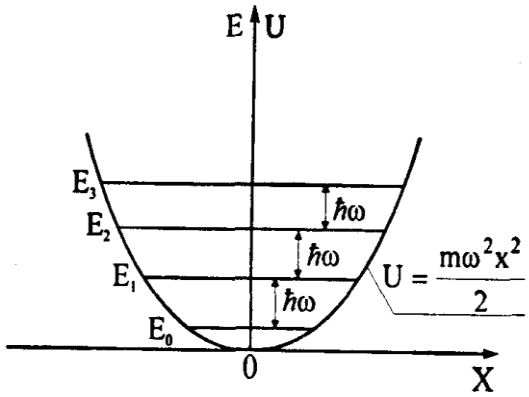


Fig. 1.2.4. The potential energy U and the energy levels En for harmonic oscillator

This energy of the electron in the hydrogen atom is measured from the value of $E_\infty = 0$ in the potential well in Fig. 1.2.5 and is called *the energy of the ground state*. Since just this energy electron should be expended for transition from the ground state with n = 1 in the state with $n = \infty$ (when electron leaves the potential well), the E_1 is also called *the ionization energy of the atom*.

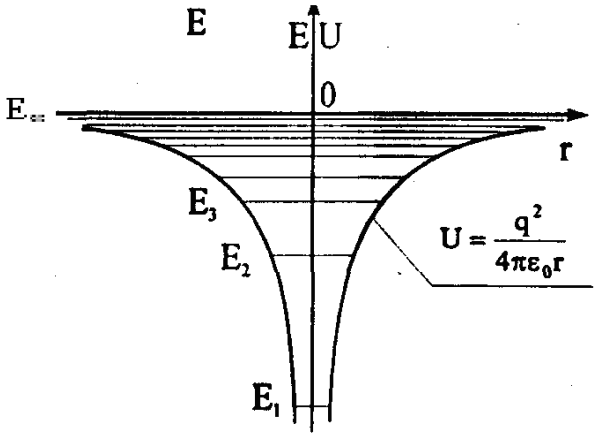


Fig. 1.2.5. The potential energy U of En and energy levels of an electron in a hydrogen atom

In addition to energy, each electron state is characterized by defined values of the orbital angular momentum vector L_o. According to quantum mechanical view, the value of L_o is given by the so-called *orbital quantum number* l = 0, 1, 2, ... , (n-1). As a result, the orbital angular momentum magnitude and direction can take the following set of values:

$$L_0 = \hbar\sqrt{l(l+1)}. \tag{1.2.26}$$

The orientation of the orbital angular momentum is usually defined in relation to some selected direction in space, for example, with respect to the direction of the external magnetic field vector H. In quantum mechanics, for the projection of L_o onto the field, the relation

$$L_{0z} = \hbar m_l \tag{1.2.27}$$

is valid, where the values m_l = 0, ±1, ±2, …, ±l are called *the orbital magnetic quantum number*.

It follows that each state of the electron in the hydrogen atom is determined by the wave function with a particular set of quantum numbers (n, l, m_l). It is easy to show that for every value of l the number states equals 2 (2l + 1) (with m_l = 0, ±1, ..., ± l; m_s = ±1/2), and for every value of n there will be always $\sum_{l=0}^{l=n-1} 2(2l+1) = 2n^2$ states.

The distribution of the wave functions in every state has its own form. For example, for the electron state with the set of the quantum numbers set (n = 1, l = 0, m_l = 0), the wave function is given by

$$\Psi(r) = \frac{1}{\sqrt{\pi r^3}} e^{-\frac{r}{r_1}} \qquad (1.2.28)$$

and is spherically symmetric. The value $r_1 = \frac{4\pi\varepsilon_0 \mathrm{h}^2}{q^2 m}$ is called *the Bohr radius of the electron* is equal in magnitude and 0,529 nm. In accordance with the physical meaning of the wave function, the probability to find an electron from the nucleus of an atom of hydrogen (proton) for this condition has the maximum at a distance $r = r_1$.

In quantum mechanics, in addition to the orbital angular momentum, electron is also characterized by *spin* or *eigen (mechanical) angular momentum* L_z. given by

$$L_s = \hbar\sqrt{s(s+1)}, \qquad (1.2.29)$$

where *s* is called by *spin quantum number*. The spin projection on the direction of the magnetic field vector *H* (the latter is directed along the axis Z) can take only two values

$$L_{sz} = m_s\hbar = \pm\frac{1}{2}\hbar, \qquad (1.2.29)$$

where $m_s = \pm\frac{1}{2}$ – the *magnetic spin quantum number*.

Therefore, any state of the electron in atom always can be characterized by a certain set of four quantum numbers (n, l, m_l, m_s).

The electro-magnetic moments, which correspond to mechanical and own orbital (spin), are given by the formulas

$$\begin{aligned} M_{0m} &= -\frac{q}{2m}L_0 = -\mu_B\sqrt{l(l+1)}; \\ M_{0o} &= -\frac{q}{2m}L_s = -2\mu_B\sqrt{s(s+1)}, \end{aligned} \qquad (1.2.30)$$

where $\mu_B = \frac{qh}{2m} = 0{,}927\cdot 10^{-23}$ J/T is called the Bohr magneton.

According to quantum mechanics, electrons always obey the Pauli exclusion principle. According to this principle, two electrons in an atom (or free electrons in the crystal) can not be in the same condition, which is characterized by a set of quantum numbers (n, l, m_l, m_s). Therefore, in many-electron atoms, electrons fill the first states with the smallest quantum numbers in the sequence n, l, m_l, m_s.

A set of quantum numbers n, l, m_l, m_s for the electrons in an atom is called the electron configuration of the atom. Note that the observed periodicity in the periodic table of the elements is determined by the nature of the electronic states filling with different values of the principal quantum number n in the electron shells of atoms.

1.2.2. Introduction to quantum statistics

In this section we shortly describe properties of the micro particles ensembles whose behavior is controlled by the relations and principles of quantum (wave) mechanics. The latter, in particular, means that the collective movement of free electrons (electron gas) or phonons (system of atomic vibrating modes) in solids does not obey the laws of classical statistics, which is valid for the ideal gas particles. The laws of the statistical distribution of quantum particle by states (velocity, momentum, and energy) are subject to either Fermi-Dirac or Bose-Einstein quantum statistics, not Maxwell-Boltzmann distribution.

Quantum ensemble of micro particles. The statistical description of quantum particles ensembles is fundamentally different from the classical to the following reasons. First, as shown above, the energy and momentum of the micro particles in the limited space region (quantum well), always takes a discrete set of values (quantized). Second, the wave properties of classical particles can be neglected, as their de Broglie wavelength λ at normal temperatures is smaller than the characteristic spatial parameters of the particles. In the case of quantum particles, we should describe the wave function behavior, which determines the micro particles "quantum state" due to a full set of the quantum numbers that determine their dynamic parameters (speed, momentum, energy). If the potential energy of interaction between particles ensemble is much less than the kinetic energy, then the ensemble in its properties will be close to an ideal gas. In this case, the particles can be assumed quasi-free, and their wave functions will be flat or spherical waves.

Third, the most important principle used in the statistical description of quantum ensembles, is *the principle of identity* (*indistinguishability*) of quantum particles: the swapping of two micro particles having the same mass, charge, spin, and in the states with wave functions Ψ_I and Ψ_j, does not change the state of the ensemble. Following the principle of the micro particles identity leads to several important consequences:

1. If we denote $|\Psi(r)|^2$ as the probability density to detect micro particles around the point r, according to the theory of probability the result of the, the probability density for two independent events to find a given particle

near the points r_1 or r_2 equals to the sum of the probabilities ($|\Psi(r_1)|^2 + |\Psi(r_2)|^2$).

2. For the same reason, in case of two different (i.e. non-identical) quantum particles 1 and 2 the probability density to detect near the point *r* either particle "1" or "2" is also equal ($|\Psi_1(r)|^2 + |\Psi_2(r)|^2$).
3. If the given quantum particle is involved in several independent processes, each of which is described by a wave function Ψ_i, then the probability density to describe the whole process, equals to the product of the probability densities for every of the processes $\prod_i |\Psi_i|^2$.
4. In the ensemble of *n* non-interacting quantum particles the resultant wave function (probability amplitude), which describes a system of non-interacting particles, is equal to the product of the wave functions of every particle of the system $\Psi(r_1, r_2, ..., r_n) = \prod_i \Psi_i(r_i)$, where r_i is radius-vector characterizing the position of the *i*-th particle in space.

Fermions and bosons. All quantum particles are divided on *fermions* and *bosons*, depending on their functions of distribution by quantum states in the ensemble. The quantum state of every micro particle is uniquely determined by the form of its wave function (including spin), which, in turn, uniquely identifies the dynamic parameters of the particle through the quantum numbers. If the state of a classical particle is defined with possible assignment of three coordinates and three velocity projections (momentum), then in case of a quantum object its state is defined by a set of four quantum numbers.

Quantum-mechanical studies show that identical Bose particles can not be considered as independent. As a result, in the case of identical bosons, the probability to detect a given quantum state of one of them in the presence of another is doubled compared with the case where the state is not occupied. In other words, there is some interaction between identical bosons, which was called *the exchange interaction*. This kind of interaction has no analogue in classical physics, and characterizes the "statistical attraction" of bosons. In the presence of a certain state of *n* bosons the probability of occupation of this state by other bosons increases in (*n* + 1) times. Particles with zero or integer spin – photons and phonons – are bosons.

Fermions obey the Pauli principle, that is, in a given quantum state at any point at the same time you never can find two completely identical (with the same spins) fermions. By analogy with the "statistical attraction" of bosons this can be regarded as a kind of "statistical repulsion" of fermions. Fermions are particles with half-integer spin, such as electrons.

Further consideration will be mainly related to an ideal gas of identical particles when their force interaction can be neglected, and the effect of the exchange interaction depends on the state of the gas as a system. Specificity of ensembles of fermions and bosons is displayed only in the case of a finite number of possible states in the ensemble. Otherwise, in case of infinite number of possible states (values of coordinate, momentum, and energy) and a finite number of particles, there is also an infinite set of states for every particle. This eliminates the possibility of "hitting" of two micro particles in a single quantum state, and therefore the fermionic and bosonic groups will behave the same way. Such ensembles of particles obey classical Maxwell-Boltzmann statistics and are called non-degenerate.

If number of states in the ensemble is comparable to the number of its constituent quantum particles, such an ensemble is called *degenerate*. In such ensembles of quantum particles there is an effect of *the Pauli exclusion principle*, which leads to degenerate different statistical patterns of behavior of bosons and fermions. As an example, free electrons gas in a metal at low temperatures is the degenerate gas..

The phase space and density of states. Note that even in the case of non-degenerate ensemble of fermions or bosons, the latters are quantum particles, because their energy and momentum are quantized due to the restrictions given by the relation uncertainties. Quantization of energy and momentum for micro particles in the quantum ensemble is usually described in a six-dimensional phase space which contains three coordinates x, y, z and three projections of momentum on the coordinate axes p_x, p_y, p_z.

In the classical case the state of a particle is represented by a point in the phase space because all coordinates and projections of momentums can be estimated accurately simultaneously. The examples of phase trajectories for three main kinds of motion of classical particle along the x-axis are shown In Fig. 1.2.6:

- motion with a constant speed is characterized by a horizontal line 1 $x = v_x t$, where $v = \text{const}(t)$ and $p_x = \text{const}(x)$;
- motion with constant acceleration $a(v_{t=0} = 0)$ is characterized by parabolic curve 2 $x = at^2/2$, where $v_x = at : p_x = m\sqrt{2a}.\sqrt{x}$;

- harmonic motion under exposion of quasi-elastic force $F = -kx$ is characterized by an elliptical phase trajectory $\frac{p_x^2}{2mE} + \frac{x^2}{2E/k} = 1$, where $\frac{p_x^2}{2m} + \frac{kx^2}{2} = E$ is the total vibrational energy.

If in the vicinity of any point of the phase space phase coordinates of particles are varied in limits x, $x + \Delta x$, ... , p_x, $p_x + \Delta p_x$, … $p_z + \Delta p_z$, then the value of $\Delta\phi = \Delta x \cdot \Delta y \cdot \Delta z \cdot \Delta p_x \cdot \Delta p_y \cdot \Delta p_z$ is called *the phase space cell*. The volume of this cell is equal to $\Delta\phi = \Delta\phi_v \cdot \Delta\phi_p$, where $\Delta\phi_v = \Delta x \cdot \Delta y \cdot \Delta z$ is its volume in the coordinate space, and $\Delta\phi_p = \Delta p_x \cdot \Delta p_y \cdot \Delta p_z$ – cell volume in momentum space.

The difference between classical and quantum objects is manifested primarily in a finite number of states in which may be a quantum particle, moving in a limited region of the phase space. Finite number of states follows directly from the uncertainty relation (1.2.4). Indeed, when one-dimensional motion of particle along x axis, two states with momentums p_1 and p_2 on the region with the length Δx can be distinguished if only $p_1 - p_2 = \Delta p$ is at least equal to $h/\Delta x$.

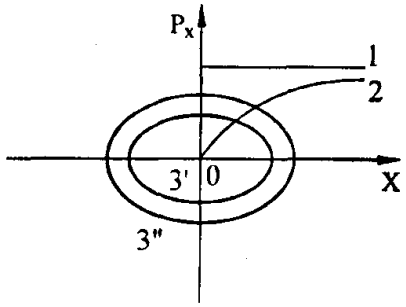


Fig. 1.2.6. The phase trajectories in two-dimensional phase space for the material point moving with constant velocity (*1*), constant acceleration (*2*) or at harmonic motion with different values of the potential energy (curves 3', 3")

This means that, in contrast to the classical case, a single quantum state occupies in the phase space not a point but a certain volume with the edge h. This volum is called *the phase unit cell*. For example, in case of two-dimensional phase space, the h value determines the unit cell area $\Delta p_x \Delta x = h$ (shaded rectangle in Fig. 1.2.7a). Then number of possible quantum states for micro particle is equal to $Z = \Delta\phi/\Delta x \cdot \Delta p_x$.

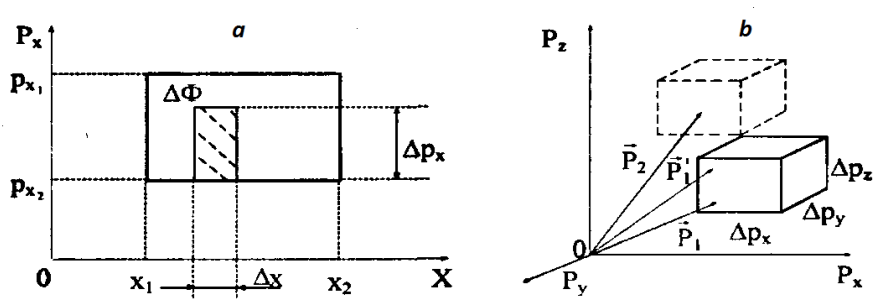


Fig. 1.2.7. Elementary cell for two-dimensional (*a*) or three-dimensional (*b*) phase space

For the three-dimensional space momentum uncertainty is given by $\Delta p_x = \frac{h}{\Delta x}; \Delta p_y = \frac{h}{\Delta y}; \Delta p_z = \frac{h}{\Delta z}.$

Multiplying these uncertainties of momentum components, we can obtain the uncertainty in determining the momentum $\Delta p = \Delta p_x \Delta p_y \Delta p_z = \frac{h^3}{V_0}$, i.e. the value of the unit cell volume in the momentum space, attributable to a single quantum state (see Fig. 1.2.7b). Estimating the volume of single unit cell, one can introduce the concept of density of states in momentum space $g(p)$ as a number of states p_i for the unit p-range.

Let's express the equation $g(p)$ for a quantum particle moving freely in a volume V_o. To do this, we select the volume $4\pi p^2 dp$ of the spherical layer of thickness dp, enclosed between the spheres with radii p and $p + dp$ in momentum space (see shaded region in Fig. 1.2.8). Divided this volume on the volume of the unit cell in momentum space $\frac{h^3}{V_0}$, we get the number of possible states in the momentum range from p to $p + dp$

$$Z(p) = \frac{4\pi p^2 dp}{\frac{h^3}{V_0}}. \qquad (1.2.31)$$

Dividing $Z(p)$ on $V_0 dp$ we can obtain an expression for the density of states $g(p)$ for the crystal of unit volume:

$$g(p) = \frac{4\pi p^2}{h^3}. \qquad (1.2.32)$$

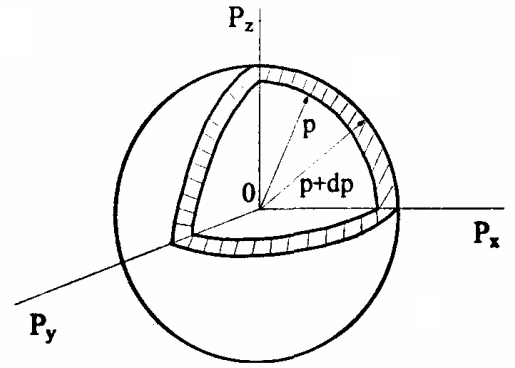


Fig. 1.2.8. To the estimation of density of states $g(p)$ in the momentum space

To obtain the density of states in the energy space $g(E)$, we take into account that for quasi-free micro particles the relation between energy and momentum of the form $E = \frac{p^2}{2m}$, is valid. Taking also into account that $g(p)dp = g(E)dE$, $p^2 = 2mE$ and $dp = \frac{1}{2}(2m)^{1/2} E^{-1/2} dE$ we obtain the following expression for the number of states per the unit energy interval around energy E and for the crystal of unit volume:

$$g(\mathrm{E}) = \frac{2\pi}{h^3}(2m)^{\frac{3}{2}} E^{\frac{1}{2}}. \qquad (1.2.33)$$

The quantum particle distribution function. One of the main tasks of statistics is to determine the number of particles $dn(E)$ in the ensemble, which are being in thermodynamic equilibrium with environment and have an energy between E and $E + dE$. It is obvious that this value $dn(E)$ in the energy interval dE is equal to $g(E)dE$, divided on the average number of particles in every of these states. To determine the latter the particle distribution function $f(E)$ is introduced which determines the probability that a particle has energy E. Then the number of particles in the ensemble, having an energy between E and $E + dE$, is given by the expression

$$\mathrm{dn(E)} = \mathrm{g(E)f(E)dE}, \qquad (1.2.34)$$

where the expression for the density of states $g(E)$ is given by (1.2.33).

As it was stated above, the nature of the filling of quantum states by fermions and bosons are different. The latter means the difference in their distribution functions: for bosons it is a function of the Bose-Einstein ($f_{\mathrm{B\text{-}E}}$), for fermions – the Fermi-Dirac function ($f_{\mathrm{F\text{-}D}}$):

$$f_{F-D} = \frac{1}{e^{(E-\mu)/kT} + 1}$$
$$f_{B-E} = \frac{1}{e^{(E-\mu)/kT} - 1} \qquad (1.2.35)$$

Here μ is chemical potential of particles ensemble in equilibrium state. The distribution functions for fermions (F-D), bosons (B-E) and the ideal gas particles (M-B) by energies is shown in Fig. 1.2.9. As is seen from figure, at energies of particles corresponding to $(E - \mu) > 3\ kT$, a unit in the denominator in (1.2.35) can be neglected, and both quantum distributions reduces to the classical Maxwell-Boltzmann function.

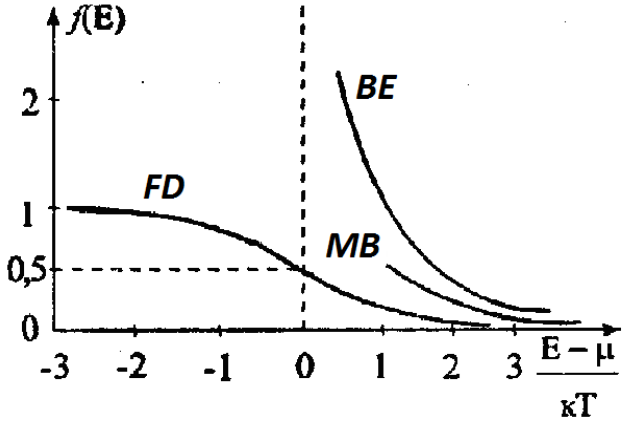


Fig. 1.2.9. Kind of different particle energy distribution functions for the Fermi-Dirac statistics (F-D), the Bose-Einstein (B-E) and the Maxwell-Boltzmann (M-B) at a finite temperature

According to the Bose-Einstein function, at low temperatures a very large number of bosons will occupy the lowest energy level ($E \to 0$), i.e. curve B-E in Fig. 1.2.9 goes up dramatically. For fermions, in this case, in every state cannot be more than two particles.

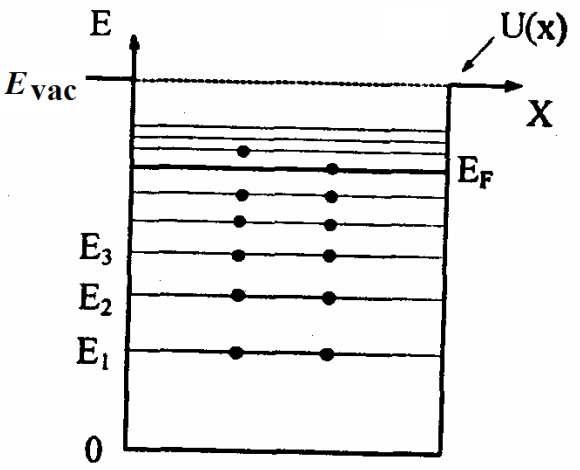


Fig. 1.2.10. Filling the energy levels by electrons in the metal

The application of Fermi-Dirac statistics to the electron gas in metallic solid. According to the Sommerfeld model, quasi-free electrons in metallic crystal can be considered as independent, interacting particles. In this case, in accordance with Fig. 1.2.10, the metal can be represented as a rectangular potential well *U(x)* filled with the gas of quasi-free electrons.

Let us consider the electrons in unit volume of the metallic crystal, which consists of *N* atoms. Since the electrons are fermions, then, at the zero absolute temperature, they will consistently fill the lowest quantum states (energy levels. Therefore, in accordance with the Fermi-Dirac distribution, they can occupy not more than by two electrons with opposite spins per level of the *N* levels. In this case, the "last" electron, occupying the *N*-th level, takes energy, which will determine the maximum kinetic energy of the electrons in the metal at $T = 0$. This energy, which is counted out from the bottom of the well in Fig. 1.2.10, defines the boundary between filled and empty states. It is called *Fermi energy* (or *level*) $E_F(0)$. The Fermi level E_F for the electrons plays the role of the chemical potential μ per particle in the electron ensemble, so that the function of the Fermi-Dirac distribution can now be written as follows:

$$f(E) = \frac{1}{e^{(E-E_F)/kT} + 1}. \qquad (1.2.36)$$

From (1.2.36) it follows that at absolute zero for $E < E_F$ probability of levels occupancy $f_{F\text{-}D} = 1$, and for $E > E_F$ is equal to $f_{F\text{-}D} = 0$. This means that at $T = 0$ K, all the states (levels), lying below the Fermi level, are fully occupied by electrons, whereas above E_F are free (see Fig. 1.2.10). That means that the

Fermi-Dirac distribution (1.2.36) has a step-like shape at $T = 0$ K, see Fig. 1.2.11. At any finite temperature $T \neq 0$ is easy to show that $f_{F\text{-}D} = 0.5$ for $E = E_F$. It reflects statistical significance of the Fermi level: at any temperature is equal to the probability of filling 0.5.

E_F expression for $T = 0$ can be easily found from the condition that the total number n of the metallic crystal of the unit volume can be obtained by integrating the relations (1.2.34) by total range of energy

$$\int_0^{E_F} g(E) \cdot f_{F-D}(E) dE = n. \qquad (1.2.37)$$

Substituting $g(E)$ and $f_{F\text{-}D} = 1$ into the expression (1.2.37), it is easy to obtain the formulae for the Fermi level at the absolute zero temperature

$$E_F(0) = \frac{h^2}{2m}\left(\frac{3n}{8\pi}\right)^{2/3}. \qquad (1.2.38)$$

As can be seen, at $T = 0$, the Fermi energy (chemical potential per electron) is uniquely determined by the concentration of electrons n in the potential well (metal crystal of the unit volume). When temperature T Increasing, a part of electrons, lying below the Fermi level, get additional energy of the order of k_BT from the crystalline lattice. Such electrons can occupy higher levels (i.e. to move with higher velocities) liberating levels with energies lower than E_F.

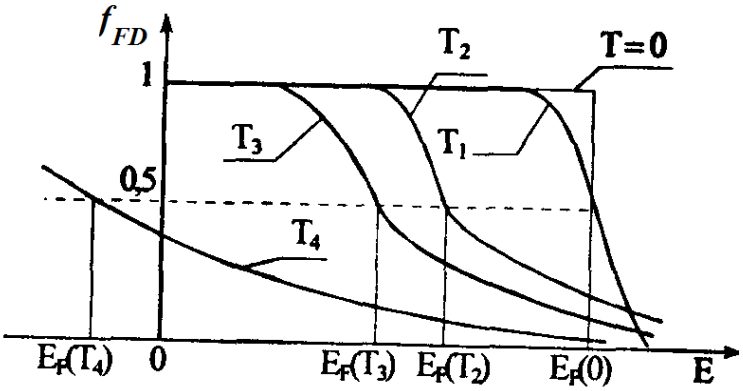


Fig. 1.2.11. Fermi-Dirac distribution at different temperatures $T4 > T3 > T2 > T1 > 0$

In this case, for not too high a temperatures, when the dependence of E_F(T) can be neglected, the distribution function will have the form shown in Fig. 1.2.11 for $T = T_1$. For such temperatures, only a small part of electrons (proportional to the ratio $k_BT/E_F(0)$) changes their energy (jumps to higher energy levels.) The rest of the electrons are completely insensitive to

temperature changes. Just this fact explain that under normal temperatures electrons give almost no contribution to the heat capacity of the crystal.

At high temperatures and/or low electron densities, when (E - E_F) >> k_BT, de Broglie wavelength of electrons is much smaller than the distance between them, so the influence of the Pauli principle on the $f_{F\text{-}D}$ function can be neglected. As a result, the electron gas becomes degenerate and the expression for $f_{F\text{-}D}$ is transformed into the classical Maxwell-Boltzmann relation:

$$f_{F-D} \approx e^{\frac{E_F}{kT}} \cdot e^{\frac{E}{kT}} \to 1 \qquad (1.2.39)$$

(electrons behave as a classical ideal gas, see curve foe $T = T_4$ in Fig. 1.2.11).

In quantum statistics, we can introduce the concept of degeneracy temperature, which is defined as the temperature of the electrons at the Fermi energy $T_F = E_F/k_B$, so that at $T > T_F$ electron gas becomes non-degenerate. Estimates show that for the metal value of T_F is a few tens of thousands of Kelvins, which is much higher than the melting point of a crystal. This means that the free electron gas in metals is always degenerated. We summarize the main results of discussed above using Fermi statistics to describe the electron gas in metallic and semiconducting crystals.

Free electrons, forming electron gas in a metal (or semiconductor) have properties which are different from the ideal gas properties. Therefore, the laws of the statistical distribution of these particles are also different: the ideal gas obeying Maxwell-Boltzmann's statistics, and the electron gas – quantum Fermi-Dirac statistics. To clarify the differences between these statistics, resume and compare assumptions on which they are built:

1. Maxwell - Boltzmann statistics deals with particles whose motion is strictly obeys the laws of classical mechanics. The status of any such particle is uniquely determined by its coordinates x, y, z and components of the momentum p_x, p_y, p_z. In so doing, coordinates and momentums can vary continuously. Therefore some states, infinitely small differ from other in coordinates, momentum and energy, are possible. Such states are considered as different in classical statistics.

The electrons, forming the electron gas, have wave properties, so that their motion is described by the Schrödinger wave equation. The consequence of this is the fact that the energy and some other characteristics of electrons in solids are quantized: they can take only *well-defined values*. Each such value corresponds to the *determined quantum state* of an electron in solids.

The presence of wave properties of the electron excludes the possibility to distinguish between the two states x, y, z, p_x, p_y, p_z, and $x+dx$, $y+dy$, $z+dz$, p_x+dp_x, p_y+dp_y,, p_z+dp_z if the product $dxdydzdp_xdp_ydp_z$ is less than the h^3 (the uncertainty principle)

$$dxdydzdp_x dp_y dp_z < h^3, \quad (1.2.40)$$

where h is the Planck constant. As follows from (1.3.31), the various elements of the phase space $\Delta\phi = dxdydzdp_x dp_y dp_z$ will meet different quantum states of an electron only if the size of these elements is not less than h^3. Therefore, in the quantum statistics the volume

$$\Delta\phi = h^3. \quad (1.2.41)$$

is accepted as an unit cell in the 6-dimensional phase space. When considering free electrons we assume that their potential energy is the same at all points in the metal, so that their distribution is uniform in the whole volume V_o. In this case, instead of the 6-dimensional phase space x, y, z, p_x, p_y, p_z, we can use 3-dimensional space of momentums splitting it into unit cells with the volume

$$\Delta\phi_P = h^3/V_o. \quad (1.2.42)$$

Every cell corresponds to a single quantum state, distinguishable from other states. Thus, the first difference of quantum statistics from Maxwell-Boltzmann one consists in the method of dividing the phase space into unit cells. Classical statistics does not make any restrictions on the size of the cells, suggesting that they may be arbitrarily small. Quantum statistics also believes that the physical sense only have a cell size not less than h^3 for the 6-dimensional phase space and h^3/V_o for three-dimensional momentum space.

2. *Electrons obey the Pauli exclusion principle.* According to this principle in the solids, in every quantum state with energy E can be no more than 2 electrons, which differ from each other with the spin direction. This means that each cell h^3/V_o of momentum space, corresponding to a given quantum state, can hold no more than 2 electrons.

3. *Maxwell-Boltzmann statistics individualizes particles,* assuming that they can be (at least fundamentally) distinguished. Therefore swap of two particles in the ideal gas, which are in different states, according to classical statistics leads to the new microstate of the system. At the same time, Fermi-Dirac quantum statistics considers all particles to be identical (indistinguishable). Therefore, a permutation of their places does not result in a new microstate.

1.2.3. Laue equations (interference condition)

In Section 1.1.6 we introduced the concept for vectors of reciprocal space lattice (their modules have the same dimension as the wave vector) and set the relationship between reciprocal vectors and Miller indices for the direct lattice planes was deduced. The reciprocal lattice concept makes it possible to establish a number of important relations for description of wave diffraction in crystals. In particular, Max von Laue received the equation, which makes it possible to determine the position of the interference maxima occurring at scattering of waves (X-ray, electron, elastic waves of atomic displacements, etc.) on the

crystal lattice. In doing so, this equation does not use the known Wolfe-Bragg assumption (see S. 1.1.5) concerning mirror-like character of X-ray waves reflection from atomic planes. Derivation of the Laue equation is based on the following assumptions:

The crystal consists of a set of identical centers which are placed at the spatial lattice sites and scatter the radiation (X-Ray, electron and elastic waves). Every center can scatter elastically the incident radiation (waves) in all directions. The sharp peaks of intensity, which take place due to scattering of monochromatic radiation with wavelength λ by these centers, are observed only in those directions (for those angles θ), for which the rays, scattered by all lattice sites, are enhanced owing to the interference.

Laue equation can be derived from the condition that the path difference of the rays scattered from two different centers P_1 and P_2 (Fig. 1.2.12) equals to a whole number $m\lambda$ of wavelengths. In accordance with Fig. 1.2.12, this condition can be written as equation

$$\Delta = \mathrm{P_1A} + \mathrm{P_1B} = \left(\vec{d}\left(\vec{n} - \vec{n}'\right)\right) = m\lambda, \qquad (1.2.43)$$

where $\vec{d}$ is the radius-vector, which characterizes the position of the centers P_1 and P_2 relative to each other, $\vec{n}$ and $\vec{n}'$ – the unit vectors showing the propagation direction (the direction of the wave vectors) of the incident and reflected waves, m – an integer (the order of reflection).

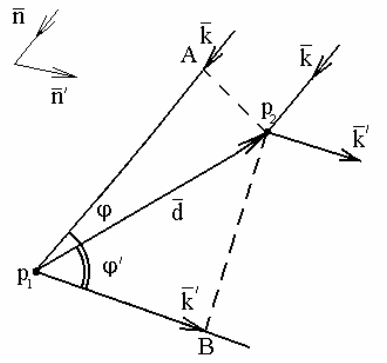


Fig. 1.2.12. For the derivation of the Laue formula

Given that the wave vectors ($\vec{k}$, $\vec{k}'$) and the unit vectors ($\vec{n}$, $\vec{n}'$) for the incident and scattered waves respectively are related as $\vec{k} = \frac{2\pi\vec{n}}{\lambda}$ and $\vec{k}' = \frac{2\pi\vec{n}}{\lambda'}$, formula (1.2.43) can be easily converted to a relation

$$\left(\vec{d}\,\Delta\vec{k}\right) = 2\pi m, \tag{1.2.44}$$

where $\Delta\vec{k} = \vec{k} - \vec{k}'$ is called *the scattering vector*.

Taking into account relations (1.1.8) and (1.1.9) for the reciprocal lattice vectors, equation (1.2.44) can be easily reduced to an equation

$$\Delta\vec{k} = \vec{G}^{*}_{hkl}, \tag{1.2.45}$$

which is called *interference Laue equation*. Fig. 1.2.12 shows that (1.2.45) can be easily reduced to the known Wolfe-Bragg condition (1.1.5).

Thus, according to Laue, diffraction peak, arising from the change of incident wave vector on the value of reciprocal lattice vector G^{*}_{hkl} at the wave scattering by a crystal lattice, is equivalent to Bragg reflection from a family of direct lattice atomic planes {*hkl*} normal to the scattering vector $\Delta\vec{k}$.

1.2.4. Atomic oscillation in solids

The nature of the atom oscillations in crystals is determined by the attraction and repulsion forces which links them. The energies of atomic attraction U_{att} and repulsion U_{rep} are differently dependent on the distances r between atoms (dot-dashed lines in Fig. 1.1.13) in accordance with relations

$$U_{\text{attr}} = -\frac{C_1}{r^m}; \quad U_{\text{rep}} = \frac{C_2}{r^n} \tag{1.2.46}$$

(C_1 and C_2, m and n are constants, characterizing the kind of interaction). So the total energy of interatomic interaction $U(r) = U_{\text{att}} + U_{\text{rep}}$ depends on the interatomic distance and can be represented by a curve with a minimum (solid curve in Fig. 1.2.13. The value of $U_0 = U(r_0)$ at the minimum point of the curve $U(r)$ is none other than *the bonding energy*. If this energy is high enough, the system of atoms forms a stable solid state structure when the atoms are at some distance r_0 from each other. Such bonded atoms have possibility only to fluctuate around their equilibrium positions. Formally, these oscillating atoms in the crystal can be presented as balls linked with elastic springs (Fig. 1.2.1). The presence of these spring-like bonds can transfer oscillations from one atom to another. As a result, at finite temperature in such a system interconnected, collective vibrational motion of the atoms is established, which is equivalent to propagating through the crystal of elastic waves due to atomic displacements. Such waves are called *normal oscillations* of the crystal lattice.

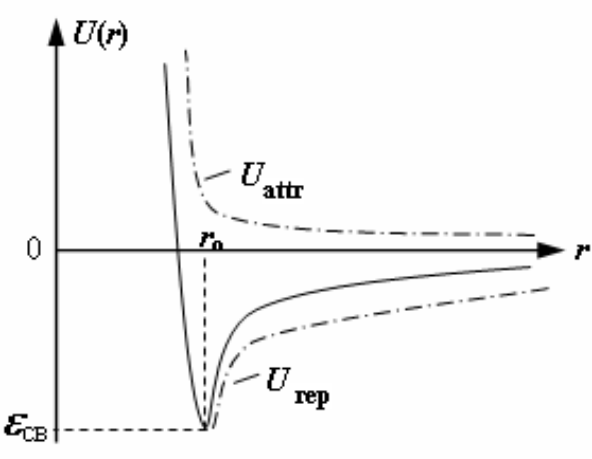


Fig. 1.2.13. The dependence interaction energies $U(r)$ of the atoms on the interatomic distance r: ro is atomic position in the equilibrium state; $U_{att}(r)$ – the energy of attraction; U_{rep} (r) – the repulsion energy; $\varepsilon_{св}$ – chemical bond energy of the interacting atoms

Consider the characteristics of these elastic waves for the one-dimensional chain of atoms (of the same kind) as the model of crystal.

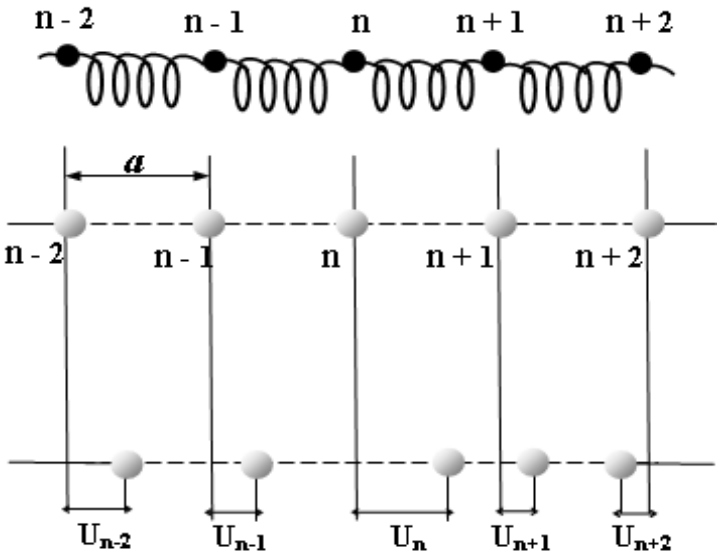


Fig. 1.2.14. The one-dimensional chain consisting of atoms of one kind

The dispersion law for the elastic displacement waves in the one-atom chain. Let us consider the atomic vibrations in some frequency range on the example of a linear chain of N atoms with the length $L = Na$ (a is the lattice parameter), shown in Fig. 1.2.14. We determine the force, which acts on every atom from neighboring ones, using two simplifying assumptions – the pairing interaction and elasticity (right of Hooke's law). Pairing interaction principle means that every atom interacts with only two the nearest neighboring atoms. In this case, the force acting on the n-th atom is determined by relation

$$\begin{aligned} F_n &= F_{n+1} + F_{n-1} = -\mu\left(u_n - u_{n+1}\right) - \mu\left(u_n - u_{n-1}\right) \\ &= \mu\left(u_{n+1} + u_{n-1} - 2u_n\right), \end{aligned} \tag{1.2.47}$$

where μ is elastic coupling constant in Hooke's law, and u_n – displacement of the n-th atom from its equilibrium position during oscillations. Hence, the motion equation for the n-th atom has the form

$$m\frac{d^2u_n}{dt^2} = \mu\left(u_{n+1} + u_{n-1} - 2u_n\right), \tag{1.2.48}$$

where m is atomic mass. The solution of equation (1.2.48) will seek as a flat elastic wave

$$u_n = A_n \exp\left(i\left(\vec{k}n\vec{a} - \omega t\right)\right), \tag{1.2.49}$$

where $x = na$. Then, we obtain

$$F_n = m\frac{d^2u_n}{dt^2} = -m\omega^2 u_n. \tag{1.2.50}$$

or

$$-m\omega^2 u_n = \mu\left(u_{n+1} + u_{n-1} - 2u_n\right). \tag{1.2.51}$$

Substitution of (1.2.49) into (1.2.51) gives the expression

$$\omega^2 = \frac{\mu}{m}\left(2 - e^{+ika} - e^{-ika}\right) = 2\frac{\mu}{m}\left[1 - \cos\left(ka\right)\right] = 4\frac{\mu}{m}\sin^2\frac{ka}{2} \tag{1.2.52}$$

which can be converted to a so-called dispersion law for the one-atom chain

$$\omega = \pm 2\sqrt{\frac{\mu}{m}}\sin\frac{ka}{2}, \tag{1.2.53}$$

which expresses the relationship between the oscillation frequency and the wave number $k = 2\pi/\lambda$ (or wavelength $\omega = 2\pi\upsilon/\lambda$, where υ – elasticc waves propagation velocity). Graphically, the dispersion law in the harmonic (elastic) approximation has the form shown in Fig. 1.2.15.

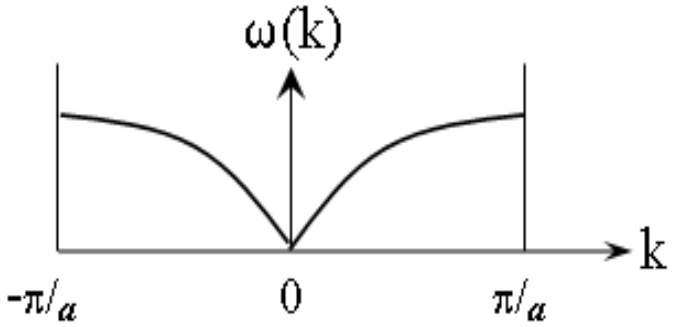


Fig. 1.2.15. The dispersion law for the one-dimensional atomic chain

Analysis of the dispersion relation (1.2.53) shows that it has the fundamental properties such as periodicity:

$$\omega\left(\vec{k}\right) = \omega\left(\vec{k}'\right) \tag{1.2.54}$$

where $\vec{k}' = \vec{k} \pm n\vec{G}^*$; and the lack of new oscillations outside the first Brillouin zone

$$-\pi/a \le k \le +\pi/a. \tag{1.2.55}$$

The relation (1.2.55) automatically means limiting of the frequencies spectrum

$$0 \le \omega \le \omega_{\max} \tag{1.2.56}$$

or wavelengths

$$2a \le \lambda \le L\ (\infty) \tag{1.2.57}$$

of elastic waves in the crystal.

In the so-called *long-wavelength limit*, when $\lambda \to L\ (\infty)$ (or $k \to 0$, $\omega \to 0$), the dispersion law becomes linear:

$$\omega \approx \pm 2\sqrt{\frac{\mu}{m}}\frac{ka}{2} = \pm ka\sqrt{\frac{\mu}{m}} = v_0 k. \tag{1.2.58}$$

In this case, *the phase* and *group velocities* matches, and elastic waves dispersion, as a phenomenon, is missing. In other words, for small k or large λ values, the dispersion law $\omega(k)$ for elastic waves in a crystal, corresponding to the initial parts of the dispersion curve in Fig. 1.2.15, is a straight line. This means the independence of the phase velocity v_0 of elastic waves on the wavelength (the wave vector), as it should be carried out for a continuous media.

The sign ± after the equality sign in (1.2.58) means the presence of waves that propagate along the chain in two opposite directions.

The so-called *short-wavelength limit* ($\lambda \to 2a$, $k \to \pm\pi/a$, $\omega \to \omega_{max}$) of (1.2.53) gives the following relations for the phase and group velocities

$$\left.\begin{aligned} v_f &= \frac{\omega}{k} = \frac{2\sqrt{\frac{\mu}{m}}\sin\frac{ka}{2}}{k} \left| \frac{a}{a} = \frac{a\sqrt{\frac{\mu}{m}}\sin\frac{ka}{2}}{\frac{ka}{2}} = \frac{v_0 \sin\frac{ka}{2}}{\frac{ka}{2}} \to \frac{2v_0}{\pi} \right. \\ v_{gr} &= \frac{d\omega}{dk} = 2\sqrt{\frac{\mu}{m}} \cdot \frac{a}{2}\cos\frac{ka}{2} = v_0 \cos\frac{ka}{2} \to 0 \end{aligned}\right\}. \quad (1.2.59)$$

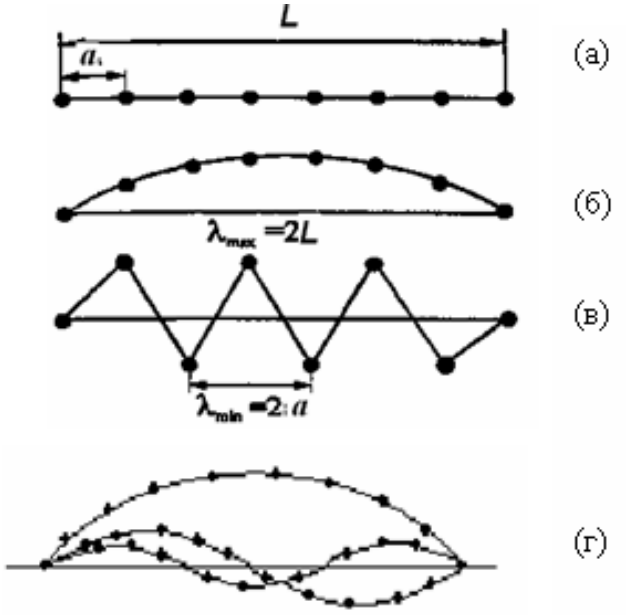


Fig. 1.2.16. Oscillations in monatomic chain (*a*) in one phase in the long-wave limit (*b*), in opposite phases in the short-wave limit (*c*) and at intermediate wavelengths (*d*)

This means that the waves with small λ (large *k*) values display dispersion phenomenon (depending the speed of elastic waves on λ or *k*): decrease of the phase velocity v_p with ω growth so that the dispersion law ω(*k*) becomes nonlinear – frequency growth is slowing down with *k* increase. As seen in Fig. 1.2.16c and from relation (1.2.58) for $k \to \pi/a$ (the edge of the Brillouin zone) the length of the shortest (standing) waves equals $\lambda \to \lambda_{min} = 2a$.

Consider a stable standing wave pattern that occurs in the atomic chain of finite length $L = Na$, fixed at the ends of the chain (Fig 1.2.16a). In this case, oscillation with the smallest frequency ω_{min} or the maximum wavelength $\lambda_{max} = 2L$ (Fig. 1.2.16b)

$$\omega_{min} = \frac{2\pi v}{\lambda_{max}} \tag{1.2.60}$$

occurs. At the same time, oscillations with the maximal possible frequency ω_{max} represent the standing wave with the minimal wavelength $\lambda_{min} = 2a$ (Fig. 1.2.16d). In this case, we obtain:

$$\omega_{max} = \frac{2\pi v}{\lambda_{min}} = \frac{\pi v}{a} = N\frac{\pi v}{l} = N\omega_{min} \tag{1.2.61}$$

For intermediate wavelengths the waves propagated are presented in Fig. 1.2.16d.

The above means that in the linear chain of N atoms N normal oscillations with frequencies ω_{min}, $2\omega_{min}$,..., $N\omega_{min}$ or wavelengths $2a,4a,\ldots,$ $2Na$ are possible. As the number of atoms N is large, so there are a lot of the normal vibrations, and the frequency difference between neighboring vibrations (adjacent frequencies) is very small. Therefore, the spectrum of frequencies of normal vibrations in crystals can be considered as quasi-continuous.

It follows from the above relations, the minimal elastic vibration frequency in a crystal is determined by its size L. Thus, for elastic waves in copper ($v = 3{,}5\cdot10^5$ cm/s) with the sample size $L = 100$ cm minimal frequency $\omega_{min} = 1{,}1\cdot10^4$ s^{-1}. The maximal frequency is limited by the interatomic distance, so that for copper, with $a = 3{,}6\cdot10^{-8}$ cm, $\omega_{max} = 3\cdot10^{13}$ c^{-1}. Thus, the range of the normal vibrations in the crystal lattice includes both acoustic (in the range $\omega = 10^2 - 10^5$ s^{-1}) and ultrasonic (with $\omega \sim 10^{12}$ s^{-1}) waves.

The dispersion law of the elastic displacement waves in one-dimensional diatomic chain. Let us define the range of normal vibration frequencies through lattice parameter a for the linear chain consisting of two sorts of atoms with different masses, shown in Fig. 1.2.17.

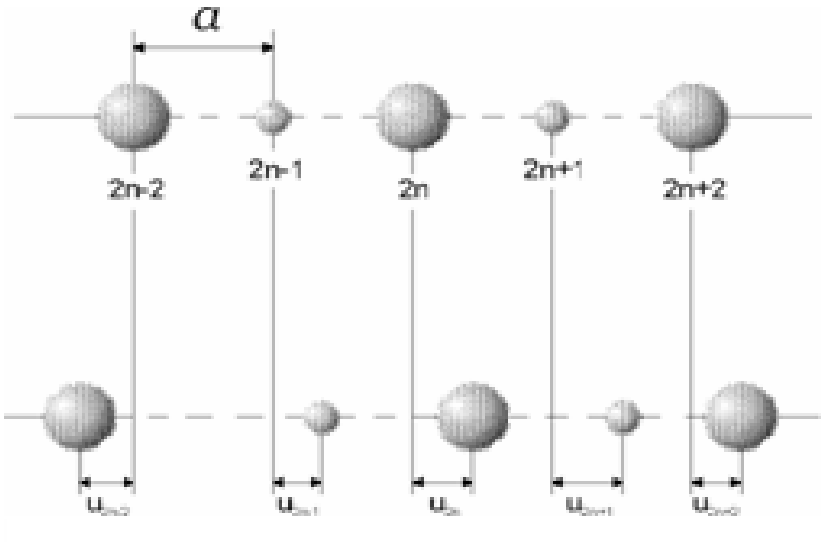


Fig. 1.2.17. The one-dimensional chain consisting of two sorts of atoms

Determining the force acting on every atom is also based on the principle of pairing interaction and the fairness of Hooke's law (elasticity). In this case, we have two equations of the type (1.2.48) and (1.2.49), separately for every sort of atoms. Therefore, the solutions (for two types of elastic waves) in such a crystal can be presented as

$$\begin{aligned} u_{2n} &= Ae^{i[2nka-\omega t]} \\ u_{2n+1} &= Be^{i[(2n+1)ka-\omega t]} \quad x_{2n} = 2na \quad x_{2n+1} = (2n+1)a \end{aligned} \tag{1.2.62}$$

Substituting these solutions in the corresponding motion equations, we obtain

$$\left.\begin{aligned} -\omega^2 m u_{2n} &= m\frac{d^2u_{2n}}{dt^2} = \mu\left(u_{2n+1}+u_{2n-1}-2u_{2n}\right) \\ -\omega^2 M u_{2n+1} &= M\frac{d^2u_{2n+1}}{dt^2} = \mu\left(u_{2n}+u_{2n+2}-2u_{2n+1}\right) \end{aligned}\right\} \tag{1.2.63}$$

which implies

$$\begin{cases} -\omega^2 mAe^{i[2nka-\omega t]} = \mu Be^{i[(2n+1)ka-\omega t]} + \mu Be^{i[(2n-1)ka-\omega t]} - 2\mu Ae^{i[2nka-\omega t]} \\ -\omega^2 MBe^{i[(2n+1)ka-\omega t]} = \mu Ae^{i[2nka-\omega t]} + \mu Ae^{i[(2n+2)ka-\omega t]} - 2\mu Be^{i[(2n+1)ka-\omega t]} \end{cases} \tag{1.2.64}$$

Eliminating the unknown coefficients A and B from (1.2.64), we obtain the following relation for elastic displacement waves of two kinds:

$$\left(2\mu - M\omega^2\right)\left(2\mu - m\omega^2\right) = 4\mu\cos^2 ka. \qquad (1.2.65)$$

Solving the equation (1.2.65), we obtain the dispersion law for the diatomic linear chain in the form of a two-valued function

$$\omega = \pm\sqrt{\frac{\mu}{\overline{m}}\left(1 \pm \sqrt{1 - \frac{4\overline{m}\sin^2 ka}{M+m}}\right)}, \quad \overline{m} = \frac{Mm}{M+m}. \qquad (1.2.66)$$

Graphically, this law is shown in Fig. 1.2.18. The analysis of the dispersion law shows that it has the same fundamental properties as in singl-atomic linear chain: periodicity: $\omega\left(\vec{k}\right) = \omega\left(\vec{k}'\right)$, where $\vec{k}' = \vec{k} \pm n\vec{G}^*$, and the lack of new oscillations outside the first Brillouin zone

$$-\pi/2a \le k \le +\pi/2a. \qquad (1.2.67)$$

Moreover, as is seen in Fig. 1.2.18, this dispersion law is characterised by the presence of two branches – acoustic 1 (low-frequency) and optical 2 (high frequency) 2.

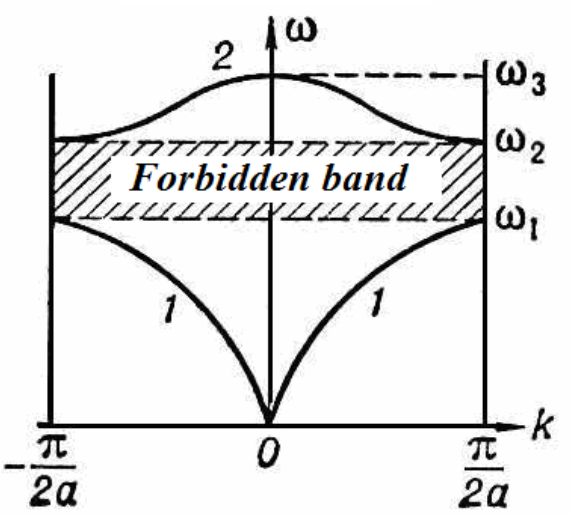


Fig. 1.2.18. The dispersion law for the one-dimensional diatomic chain: *1* – acoustic mode, *2* – optical mode

From relation (1.2.66) it follows that for the acoustic branch 1 (bottom in Fig. 1.2.18), the dispersion is as follows:

$$\omega^2 = \frac{\mu}{\overline{m}}\left[1 - \sqrt{1 - \frac{4\overline{m}\sin^2 ka}{M+m}}\right]. \qquad (1.2.68)$$

In *the long-wavelength limit* ($k \to 0$, $\omega \to 0$, $\lambda \to L(\infty)$) dispersion is linearized for this frequency range:

$$\omega^2 = \frac{\mu}{m}\left[1 - \sqrt{1 - \frac{4\bar{m}\sin^2 ka}{M+m}}\right] \approx \frac{\mu}{m}\left[1 - \sqrt{1 - \frac{4\bar{m}(ka)^2}{M+m}}\right] \to v_{02}k, \quad (1.2.69)$$

where the group velocity is

$$v_{02} = \sqrt{\frac{2\mu a^2}{m+M}} \quad (1.2.70)$$

In the *short-wavelength limit* ($k \to \pm\pi/2a$, $\lambda \to 4a$, $\omega \to \omega_1$) limiting frequency ω_1 is achieved, and the phase and group velocities are expressed by the relations

$$v_f = \frac{\omega}{k} \to \frac{\sqrt{\frac{2\mu}{M}}}{\frac{\pi}{2a}} = \sqrt{\frac{8\mu a^2}{\pi^2 M}} << \frac{2v_0}{\pi}, \quad \text{where } v_0 = \sqrt{\frac{\mu a^2}{m}} \quad (1.2.71a)$$

$$v_{gr} = \frac{\partial\omega}{\partial k} \to 0. \quad (1.2.71a)$$

For the *optical mode* 2 dispersion relation is as follows:

$$\omega^2 = \frac{\mu}{m}\left[1 + \sqrt{1 - \frac{4\bar{m}\sin^2 ka}{M+m}}\right]. \quad (1.2.72)$$

In the *long-wavelength limit* ($k \to 0$, $\omega \to \omega_3$), this branch of the dispersion law tends to saturate

$$\omega \to \omega_3 = \sqrt{\frac{2\mu}{m}}, \quad (1.2.73)$$

so the phase and group velocities are of the form

$$v_{gr} = \frac{d\omega}{dk} \to 0 \quad (1.2.74a)$$

$$v_f = \frac{\omega}{k} \to \frac{\omega_3}{0} \to \infty, \quad (1.2.74b)$$

and the ratio of oscillation amplitudes for heavy and light atoms is

$$\frac{B}{A} = \frac{2\mu - m\omega_3^2}{2\mu\cos ka} \to \frac{2\mu - m\frac{2\mu}{m}}{2\mu} = 1 - \frac{m}{m} = -\frac{m}{M}. \quad (1.2.75)$$

In the *short-wavelength limit* ($k \to \pm \pi/2a$, $\omega \to \omega_2$) the limiting frequency is

$$\omega \to \omega_2 = \sqrt{\frac{2\mu}{m}}, \tag{1.2.76}$$

so the phase and group velocities are of the form

$$v_f = \frac{\omega}{k} \to \frac{\omega_2}{\pi/2a} = \sqrt{\frac{8a^2\mu}{\pi^2 m}} \tag{1.2.77}$$

$$v_{gr} = \frac{d\omega}{dk} \to \frac{\omega_3}{k} \to 0, \tag{1.2.78}$$

and the ratio of of oscillation amplitudes for heavy and light atoms is

$$\frac{B}{A} \to 0. \tag{1.2.79}$$

Stable patterns of standing waves in a diatomic linear chain of finite length with their fixed ends for the acoustic and optical branches are shown in Fig. 1.2.19.

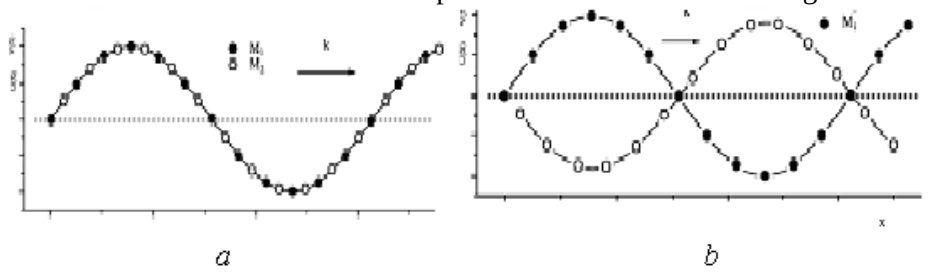


Fig. 1.2.19. Oscillations of the atoms in one-dimensional diatomic chain. Oscillations of dissimilar atoms in-phase mode (*a*) for the acoustic branch and in antiphase (*b*) for the optical branch

The dispersion law for elastic displacement waves in a three-dimensional lattice. If the three-dimensional crystal lattice consists of one type atoms, the total number of oscillating freedom degrees for every atom in a crystal is equal to three (because the oscillations can be carried along three mutually perpendicular axes) – two transversal and one longitudinal. Therefore, if the number of atoms in the crystal is *N*, it can be excited 3*N* normal oscillation modes with a particular frequency ω. This corresponds to the three branches of the dispersion law for acoustic waves (as, for examples, the 3 bottom curves for the two transversal and one longitudinal oscillation modes in Fig. 1.2.20). Note that in an anisotropic medium frequency depends not only on the magnitude of *k*, but also the wave propagation direction.

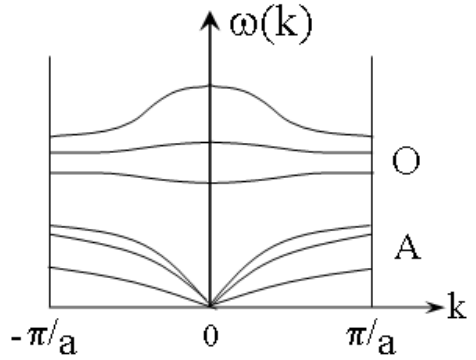


Fig. 1.2.20. Dispersion curves for the three-dimensional crystal

If the crystal consists of *s* sorts of atoms, the total number of vibrational modes is 3*s*. Three of these modes form the acoustic branches of oscillations (three lower curves A of the dispersion law in Fig. 1.2.20) and the remaining (3*s*-3) will correspond to the so-called optical vibrations (the top three of the dispersion curves in Fig. 1.2.20 for the diatomic lattice with $s = 2$).

1.2.5. Concept of phonons

As shown in Section 1.2.1, the wave-particle duality of quantum particles (atoms) makes possible to write the relationship between the characteristics of the wave and particle properties as follows:

$$\varepsilon = \hbar\omega \qquad (1.2.80)$$
$$\vec{p} = \hbar\vec{k}$$

In such a notation, left part of de Broglie relation presents corpuscular characteristics of wave-particle (its kinetic energy *E* and momentum *p*), and the right part – wave characteristics (its frequency ω and wave vector $\vec{k}$). With respect to the atomic oscillations in a crystal, such wave-particles (or, more correctly, *quasi-particles*) have the meaning of a sound quanta and are called *phonons*. The movement of these waves-particles in crystals is described by the Schrödinger equation and obeys the laws of quantum mechanics.

To describe the elastic displacement waves in a crystal as quasi-particles, we should present the crystal as a potential box filled with *phonon gas*. In this case, the collisions (interactions) of phonons should be conformed to the relevant laws of energy and momentum conservation. These laws can be easily obtained from the Laue equation (1.2.44) or (1.2.45), which holds not only for X-ray scattering, but for any scattering (including elastic waves) in the crystals. Laue

equation can be written as the selection rules for the wave vectors giving the interference maxima in the scattering of rays on the direct lattice

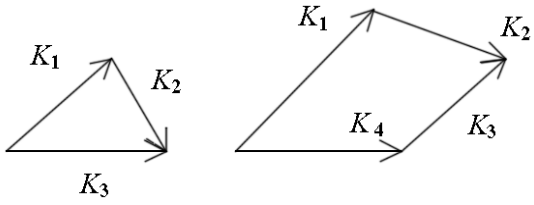
$$\vec{k}' = \vec{k} + \vec{G}^*_{hkl}. \qquad (1.2.81)$$

Taking into account the de Broglie relation (1.2.80), equation (1.2.81) can be regarded as the form of momentum conservation law in the crystal

$$\vec{p}' = \vec{p} + \hbar\vec{G}^*_{hkl}, \qquad (1.2.82)$$

where $\vec{G}^*_{hkl}$ is the reciprocal lattice vector.

Fulfilment of relations (1.2.81) and (1.2.82) means that in the reciprocal space (space of wave vectors), there is a merely definite set of the wave vectors *k* (or momentums *p*) of the incident waves (X-ray, elastic, electronic) which can be scattered by the crystal in accordance with the Laue equation (the selection rules). This region of *k* and *p* values in the reciprocal space is called a *Brillouin zone*. In other words, the Brillouin zone is a region of the reciprocal lattice,



a) *b*)

Fig. 1.2.21. Schemes of normal phonon collisions of types $\vec{k}_1 + \vec{k}_2 \to \vec{k}_3$ (*a*) and $\vec{k}_1 + \vec{k}_2 \to \vec{k}_3 + \vec{k}_4$ (*b*), when the conservation law for phonon momentum s performed which includes the whole set of wave vectors (momentums) of the falling on the crystal rays, for which there is a plane (hkl) of the direct lattice, giving a mirror reflection of waves with these wave vectors and leading to interference maxima

The theory shows that there are the so-called *normal phonons collisions*, for which $\vec{G}^*_{hkl} = 0$. Normal collisions are of two types

$$\vec{k}_1 + \vec{k}_2 \to \vec{k}_3 \text{ and } \vec{k}_1 + \vec{k}_2 \to \vec{k}_3 + \vec{k}_4 \qquad (1.2.83)$$

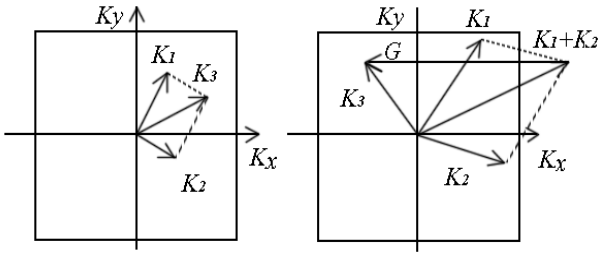
(see Fig. 1.2.21). Relations (1.2.83) show that the total momentum of the phonons at normal collisions $\vec{k}_1 + \vec{k}_2$ is conserved. As a result, the movement direction of the phonons, which is conditioned the heat transfer direction also remains the same.

In the crystals collision of phonons like $\vec{k}_1 + \vec{k}_2 \to \vec{G}^*_{hkl}$ are also possible, where $\vec{G}^*_{hkl} \neq 0$. In this case, the law of momentum conservation in the normal form is not satisfied. Such collisions are called *phonon collisions with a flip*. Peierls have shown that the conservation laws for momentum for such phonons, in contrast to the (1.2.83) law (in Fig. 1.2.22a), the phonon collisions with a flip will have a different form (Fig 1.2.22b):

$$\hbar\omega_1 + \hbar\omega_2 = \hbar\omega_3 \qquad (1.2.84)$$

$$\hbar\vec{k}_1 + \hbar\vec{k}_2 = \hbar\vec{k}_3 + \hbar\vec{G}^*_{hkl}. \qquad (1.2.85)$$

The (1.2.85) coincides with the above selection rules (1.2.81) and (1.2.82). As can be seen from the expressions (1.2.85), at the interaction of phonons their number can be non-conserved.



a) *b*)

Fig. 1.2.22. The scheme of the normal collisions (*a*) and collisions with a flip, when the law of phonons' momentum conservation is not satisfied (*b*)

Furthermore, as seen from Fig. 1.2.22b, in such collisions the phonons' total momentum $\left(\vec{k}_1 + \vec{k}_2\right)$ is outside of the first Brillouin zone. This means that the momentum $\hbar\vec{G}^*_{hkl}$ is added to the momentum of the phonon system as a whole. Vector $\hbar\vec{G}^*_{hkl}$ is called by a *recoil momentum*.

As a result of a collision with a flip (Fig. 1.2.22b), the wave vector $\vec{k}_3$ of the "born" phonon determines the direction of phonon propagation, which differs from the direction of the total phonons' momentum. As will be shown below, such processes, when collisions strongly change the phonons motion direction (an therefore change the direction of thermal energy transfer), have a very strong influence on thermal conductivity of phonon gas (the crystal lattice). Note that, in the case of phonon collisions with a flip, phonons, generally

speaking, does not have the mechanical momentum as the usual material particle. Therefore $\hbar\vec{k}$ is called *a quasi-momentum*.

All this means that every normal phonon mode (every elastic wave of vibrating atoms) with frequency ω can be associated with the quantum harmonic oscillator with a mass of oscillating atoms. The energy of such harmonic oscillator is

$$E_n = \left(n + \frac{1}{2}\right)h\omega, \tag{1.2.86}$$

where n = 1,2,3 ... is oscillator energy level number in Fig. 1.2.23. When quantum oscillator (normal mode) transits from one energy state to another, the energy can be changed merely by an amount $\hbar\omega$. Just as at the transition of an atom from one energy state in the lower state, the emission of a quantum electro-magnetic radiation – photons occurs, in the oscillating crystal we can speak about emission of quasi-particles – phonons with the energy $E_{\text{phon}} = \hbar\omega$ and quasi-momentum $\vec{p} = \hbar\vec{k}$.

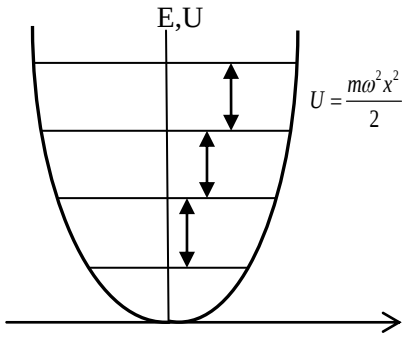


Fig. 1.2.23. The potential energy U and the kinetic energy (energy levels) E of quantum harmonic oscillator

Note that phonons are bosons (particles with a quantum (zero) spin), so their energy distribution is described by the Bose-Einstein distribution (1.2.35a). This function determines the average number of phonons n with frequency ω at temperature T. The higher temperature, the greater number of "born" phonons, and the higher phonon frequency, the less they number is generated at this temperature.

The number of possible oscillations (phonons) in a crystal. As indicated earlier, the number of possible wavelengths (and wave vectors), which can have

the elastic displacement waves propagating through atomic chain of finite length, equals to the number of atoms in it:

$$\lambda = 2a; 4a; ...., 2Na = L \qquad k = \frac{\pi}{Na}; \frac{2\pi}{Na}; \frac{3\pi}{Na}; \ldots \frac{N\pi}{Na} = \frac{\pi}{a} \;\Rightarrow$$

$$\Rightarrow k = (n\pi/Na). \tag{1.2.87}$$

It follows that the volume per one state in the one-dimensional k-space (the distance between the nearest points in the one-dimensional Brillouin zone) is equal $\Delta k = (\pi/Na)$. Hence, the number of states per elementary interval dk of wave numbers in the Brillouin zone equals $\frac{dk}{(\pi / Na)}$. For a crystal of unit volume (in the case of one-dimensional crystal – of unit length), the number of states in dk is

$$\frac{dk}{(\pi / Na)L}. \tag{1.2.88}$$

Then, taking into account that $L = 2Na$, the number of states in k-space, corresponding to an interval dk, equals:

$$\begin{cases} g(k)\,dk = \dfrac{1}{\pi} dk, & |k| \le \pm\dfrac{\pi}{a} \\ g(k)\,dk = 0, & |k| > \pm\dfrac{\pi}{a} \end{cases}, \tag{1.2.89}$$

where the g(k) function is called the density of states in k-space.

Similarly, one can get the *phonon density of states* $g(\omega)$ in frequency space (called *a spectral density of phonons* or *phonon spectrum*). We can obtain $g(\omega)$ using the equation $g(\omega)d\omega = g(k)dk$. This equation is true, because the number of states does not depend on the type of space, in which they are described. Thus, we obtain that

$$g(\omega) = g(k)\left(\frac{d\omega}{dk}\right)^{-1} = \frac{1}{\pi}\left(\frac{d\omega}{dk}\right)^{-1}. \tag{1.2.90}$$

For the three-dimensional isotropic crystal density of states in the Brillouin zone, as it follows from Fig. 1.2.24, can be obtained in spherical coordinates as follows:

$$g(k) = \frac{4\pi k^2}{(2\pi)^3} \Rightarrow\; g(k)dk = \frac{4\pi k^2 dk}{(2\pi)^3}. \tag{1.2.91}$$

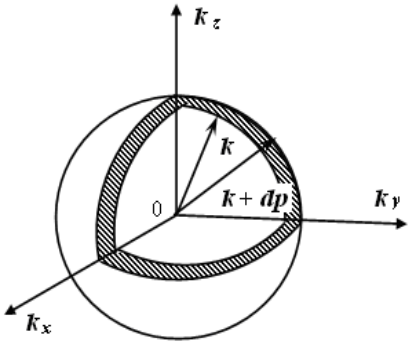


Fig. 1.2.24. Calculation of the states number density

The expression for $g(\omega)$ can be obtained from (1.2.91) in the so-called Debye approximation, when the dispersion law is considered to be linear

$$\omega=\nu_{gr}k = \nu_{ph}k = \nu_o k \, . \qquad (1.2.92)$$

In fact, it is valid only for long-wave phonons excited at low temperatures. In addition, we must note that in the three-dimensional crystal can occur three types of normal modes (two transversal and one longitudinal). In this case, the $g(\omega)d\omega$ value should be increased to three times (we believe in the Debye approximation that the velocities of wave propagation in different directions are the same). So, substituting (2.1.92) in (1.2.90), we obtain the so-called Debye phonon spectrum

$$g(\omega) = g(k)\left(\frac{d\omega}{dk}\right)^{-1} = \frac{3}{8\pi^3}4\pi k^2 v_0^{-1} = \frac{3}{8\pi^3}4\pi\frac{\omega^2}{v_0^2}v_0^{-1} = \frac{3\omega^2}{2\pi^2 v_0^3}, \qquad (1.2.93)$$

with a quadratic dependence on frequency (curve 1 in Fig. 1.2.25).

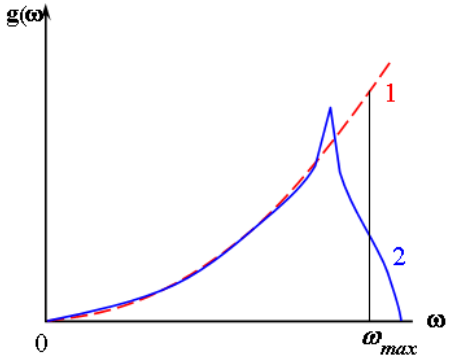


Fig. 1.2.25. States density function for the Debye model (curve 1) and for a real crystal (curve 2)

Note also that for the Debye phonon spectrum (1.2.93), which lies in the frequency range

$$0 \le \omega \le \omega_{\max}, \tag{1.2.94}$$

the following normalising relation should be satisfied

$$\int_0^{\omega_{MAX}} g(\omega) d\omega = 3N. \tag{1.2.95}$$

As can be seen from Fig. 1.2.25, the quadratic (by frequency) Debye phonon spectrum (curve 1) is significantly different from the real phonon spectrum of the crystal (curve 2) in the middle and high frequencies. The maximal frequency $\omega_{\max} = \omega_D$ in the Debye phonon spectrum is called *the characteristic Debye frequency*.

1.3. Electronic dynamics

Crystalline solids greatly vary in their electrical properties. For example, metals are very good conduct electricity and is considered to be conductors. *Conductors* have a high electrical conductivity at a normal temperatures and a positive temperature coefficient of resistivity. At the same time, some crystals practically does not conduct current and are considered to be insulators (or dielectrics). *Insulators* have a very high electrical resistance: their main feature is the ability to polarization and the presentation of the internal electric field in them due to this reason. *Semiconductors* are intermediate conductors by their electrical conductivity between metals and insulators. Besides high dependence of conductivity on type and concentration of impurities, as well as the external energetic impacts (temperature, pressure, light, etc.).

Strong differences in the electrical conductivity between metal, semiconductor and dielectric materials are caused by the features of the distribution of the electrons by energy (energy spectrum) in crystals. This distribution is strongly influenced by the periodic arrangement of atoms in the crystal, forming, in particular, three-dimensional periodic potential in the field of which free electrons move. The nature of this motion of the electrons is also depends very strongly on their interaction with the crystal lattice, and between each other. This section is devoted to the energy spectrum of electrons in crystalline materials and the motion of electrons in a periodic lattice potential, which allow us in the next sections to explain the reasons why all the crystals are divided on electric conductors (metals), poor conductors (insulators) and semiconductors and why they have different electrical properties.

1.3.1. Drude-Lorentz model for free electron gas

The first model describing the electronic properties of crystals (in particular, their electrical conductivity), has been established for metals and was called *the Drude model*. The Drude theory is based on the model of a metallic crystal as a system of N fixed positively charged ions, which form crystalline lattice and are embedded into gas of n_0 free electrons (Fig. 1.3.1).

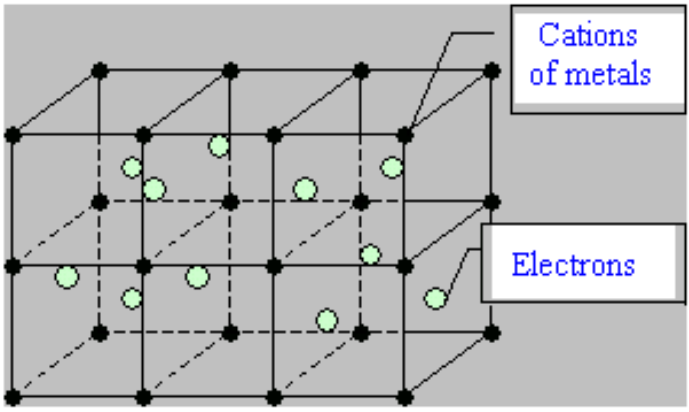


Fig. 1.3.1. Model of the free electron gas in a metallic crystal

The last obey the laws of classical Maxwell-Boltzmann statistics

$$f_0 = 4\pi v^2 \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} e^{-\frac{mv^2}{2kT}}. \qquad (1.3.1)$$

This has allowed to estimate the average thermal velocity of the electrons

$$v_{\text{mean}} = \sqrt{\frac{3kT_0}{m}} \sim 10^7 \ \text{cm/s} \qquad (1.3.2)$$

and their root-mean-square thermal velocity

$$v^2_{\text{mean}} = < v^2_{\text{T}} > = \frac{\int_0^\infty v^2 f_0 dv}{\int_0^\infty f_0 dv} = \frac{3kT}{m}. \qquad (1.3.3)$$

The chaotic and ordered motion of electrons. The Drude model considers that the electrical resistance of the metallic crystals as a result of the scattering of electrons, moving under the influence of an applied electric field, by the fixed lattice ions (Fig. 1.3.2). Thus the Drude theory is based on the following

assumptions at the description of this electrons random motion in a crystal in the absence of an electric field:

1. In the interval between two successive collisions with ions, every electron moves by a straight line with the classical average velocity (1.3.2). This means that any type of Coulomb interactions with ions and other electrons is absent (this is called *the approximation of independent free electrons*).
2. Collisions between electrons and ions are treated as instant random events, suddenly changing the electron speed from some mean value (1.3.2) to zero.
3. It is assumed that per unit of time, the electron experiences a collision with the probability $1/\tau$, where τ is the mean free time of an electron motion between two successive collisions.
4. It is also assumed that electrons come into equilibrium with the crystal only due to their collisions with the ions the lattice.

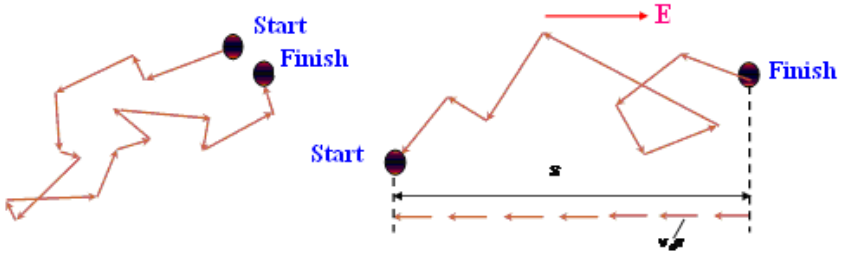


Fig. 1.3.2. Chaotic motion (*a*) and the ordered drift (*b*) of the of electrons in a crystal

These assumptions mean that in the absence of an external electric field, the electrons move randomly in space and do not have a preferred direction of movement. This is called *the chaotic (random) motion* (Fig. 1.3.2a). When applying an external electric field to the crystal, the ordered movement of free electrons is superimposed on the chaotic motion. This second component is called *the drift of electrons* in an electric field (Fig. 1.3.2b). It is the electron drift provides electrical conductivity of crystals.

The electrical conductivity. Consider a group of electrons n_0, moving in an electric field E since the moment on $t = 0$ of its switching-off. As far as, according to our assumption 3, the probability of an electron collision with an ion is equal to $1/\tau$, the number of electrons, which did not experience a collision for time t, is equal

$$n_t = n_0 e^{-\frac{t}{\tau}}. \qquad (1.3.4)$$

Hence the rate of n_t decrease is equal to

$$\frac{dn}{dt} = -\frac{n_0}{\tau} e^{-\frac{t}{\tau}} = -\frac{n_t}{\tau}. \qquad (1.3.5)$$

According to the classical Drude theory of conductivity, an electron moving in the electrictric field with strength *E*, is subjected to the effect of the force $F = eE$, which gives it the acceleration a = F/m = eE/m, where m – mass of a free electron. According to the Drude model, the electron, being accelerated, can not increase its velocity indefinitely due to collisions with ions and other electrons in the crystal lattice. The scattering of electrons by ions lead to the fact that the velocity to be gained in the direction of the electric field, practically falls to zero after each collision. As a result of such collisions, cycles "acceleration – scattering" for the electrons are repeated many times.

The average distance λ, which electrons pass from the collision to collision, is called *the mean free path*. Since the electron travels this distance for time τ, the latter is called *the mean free path time*. Thus the velocity vector, which electron gains in the electric field for the time t, is not more than

$$\vec{v}_t = -\frac{e\vec{E}t}{m} = \vec{a}t\,.$$

This is an additional increase in the speed of electrons (due to the action of an electric field), which did not experience collisions with ions for the time t. It follows that the shift vector (drift) of the electron in the direction of the electric field is

$$\vec{x}_t = -\frac{e\vec{E}t^2}{2m} = \frac{\vec{a}t^2}{2} \qquad (1.3.6)$$

(this drift is superimposed on the random thermal motion of electron, Fig. 1.3.2b).

Average total path (the total shift) in the direction of the electric field vector *E* for each of n_t electrons, which have not undergone collisions with ions for the time *t*, is equal to

$$\begin{aligned} < x_t > &= \int_0^n x_t dn = \int_0^\infty \vec{x}_t \left(\frac{dn}{dt}\right) dt = -\frac{e\vec{E}n_0}{2m\tau} \int_0^\infty t^2 e^{-\frac{t}{\tau}} dt = \\ &= -\frac{e\vec{E}n_0\tau^2}{m} \int_0^\infty \frac{1}{2} y^2 e^{-y} dy = -\frac{e\vec{E}n_0\tau^2}{m} = \vec{v}_D \tau n_0. \end{aligned} \qquad (1.3.7)$$

The average velocity v_D, which carriers gain under the influence of an electric field (it is called *the drift velocity*) for uniformly accelerated motion from the quiescent state is equal to half of the maximal velocity $v_D = 0{,}5eE\tau/m$. A more rigorous derivation, taking into account the distribution of free electrons in the crystal by the energies, leads to the expression:

$$\vec{v}_D = -\frac{e\vec{E}\tau}{m}. \tag{1.3.8}$$

According to the Drude model, for the loss of velocity, which electron gains in the electric field, only one collision is enough. So it gives

$$\tau = \frac{\lambda}{v}, \tag{1.3.9}$$

where $v = v_T + v_D$ is full velocity of the electron, which is a sum of the thermal v_T and drift v_D components, where usually $v_T >> v_D$. Note that in the classical Drude model the mean free path of electrons λ is independent on temperature and is given by

$$\lambda \approx \frac{1}{\pi R^2 N}, \tag{1.3.10}$$

where N is the number of centers of elastic electrons scattering (equal to the concentration of atoms in the crystal), and πR^2 is the cross section of electron scattering on a fixed atom with radius R.

The total density of the electron flow (current density) moving along the electric field can be expressed as the product of free electron charge on its drift velocity

$$\vec{j} = ne\vec{v}_D = \frac{ne^2\tau}{m}\vec{E} = \sigma\vec{E}. \tag{1.3.11}$$

This expression is called Ohm's law in differential form. Hence, the ratio front of the field vectors in (1.3.10), called the specific electric conductivity of the Drude electron gas will be equal

$$\sigma = \frac{ne^2\tau}{m}. \tag{1.3.12a}$$

The magnitude of the drift velocity in the electric field of unit strength

$$\mu = \frac{e\tau}{m} = \frac{\vec{v}_D}{\vec{E}} \tag{1.3.13}$$

is called *the electron mobility*. From here one can get a second relation for conductivity of free electrons gas resulting from the Drude model:

$$\sigma = ne\mu\,. \tag{1.3.12b}$$

Substituting (1.3.2) to (1.3.10) or (1.3.12a), we can obtain the temperature dependence of the conductivity for the Drude classical electron gas in the form

$$\sigma = \frac{ne^2}{m}\frac{\lambda}{\sqrt{\frac{3kT}{m}}} = \frac{ne^2\lambda}{\sqrt{3kTm}}. \tag{1.3.14}$$

As will be shown later, this kind of dependence $\sigma(T)$ does not coincide with the experimental results.

The heat capacity and thermal conductivity of the classical electron gas in the Drude model. According to the classical Maxwell-Boltzmann statistics, which holds true for the gas of n free electrons in the Drude model, the energy ($k_B T/2$) should fall at every electron degree of freedom. This means that the heat capacity, as the derivative of the electron gas kinetic energy by temperature, must be equal

$$C_e = \frac{\delta U}{\delta T} = \frac{3}{2} k_B n, \tag{1.3.15}$$

where k_B is Boltzmann constant.

The most impressive success of the Drude model was the explanation of the empirical Wiedemann-Franz law. This law states that the ratio κ/σ of thermal κ to the electrical conductivity σ for the most metals is directly proportional to temperature, in doing so the proportionality coefficient is the same, with sufficient accuracy, for all metals.

We can calculate the thermal conductivity of the electron gas, assuming that the major part of the heat flow in the metal is transferred by conduction electrons (therefore metals conduct heat much better, than insulators!) Taking into account that electron gas obey classical statistics, we can get thermal conductivity of an ideal electron gas in form $\kappa_e \approx \frac{1}{3} < v_T >^2 \tau C_e$, where C_e is specific heat of the classical electron gas in the metal. Substituting (1.3.10) and (1.3.15) to this relation, we obtain

$$\kappa_e = \frac{1}{2} nk\tau \frac{3kT}{m} = \frac{3nk^2}{2m} \tau T. \tag{1.3.16}$$

Hence it is easy to derive the relation

$$L = \frac{\kappa_e}{\sigma T} = \frac{\frac{3}{2}\frac{nk^2\tau T}{m}}{\frac{ne^2\tau}{m} T} = \frac{3}{2}\left(\frac{k}{e}\right)^2 = const, \tag{1.3.17}$$

which became known as the Wiedemann-Franz law. It is easy to calculate that the constant L in this equation, called the Lorentz number, is equal to $1{,}11 \cdot 10^{-8}$ W/Ω·K^2.

Boltzmann kinetic equation. The discrepancy between the calculated and experimental curves $\sigma(T)$ for metals required to transform the Drude model by taking into account the influence of the electric field on the distribution function f of the electrons by velocities (energies). According to Lorentz, the rate of

change of *f* after the electric field switching on can be considered as the sum of two terms

$$\frac{\partial f}{\partial t} = \left(\frac{\partial f}{\partial t}\right)_{\text{full}} = \left(\frac{\partial f}{\partial t}\right)_{\text{field}} + \left(\frac{\partial f}{\partial t}\right)_{\text{collision}}, \tag{1.3.17}$$

where contribution

$$\left(\frac{\partial f}{\partial t}\right)_{\text{field}} \approx -\frac{e\vec{E}}{m}\frac{\partial f}{\partial \vec{v}} \tag{1.3.18}$$

is called by *the drift term*, and contribution

$$\left(\frac{\partial f}{\partial t}\right)_{\text{collision}} \approx \frac{f - f_o}{\tau_r} \tag{1.3.19}$$

is *collisional term.* The parameter τ_r in this case is called *the relaxation* time *of electrons:*this is the time during which the electronic system comes into equilibrium with the lattice by collisions between electrons and ions. The value of f_0 is the unperturbed function of the Maxwell-Boltzmann (1.3.1) in the absence of an electric field.

For the stationary case ($t > \tau_r$) relation (1.3.17) becomes the well-known Boltzmann kinetic equation

$$\frac{\partial f}{\partial t} = \frac{e\vec{E}}{m}\frac{\partial f}{\partial \vec{v}} + \frac{f_0 - f}{\tau_r} = 0, \tag{1.3.20}$$

whence

$$\frac{e\vec{E}}{m}\frac{\partial f_0}{\partial \vec{v}} = -\frac{f_0 - f}{\tau_r}. \tag{1.3.21}$$

From this equation we can get the expression for the Boltzmann distribution function for the stationary state, perturbed by the electric field,

$$f \cong f_0 + \frac{e\vec{E}}{m}\frac{\partial f_0}{\partial \vec{v}}\tau_r. \tag{1.3.22}$$

Here we consider that the shape of the distribution function in the electric field is not changed, i.e. it is assumed that

$$\frac{\partial f}{\partial \vec{v}} = \frac{\partial f_0}{\partial \vec{v}}. \tag{1.3.23}$$

In the improved Drude-Lorentz model electron relaxation time is assumed to be

$$\tau_r = A < v_T >^j. \tag{1.3.24}$$

Here A is the mean free path length λ, if j = -1, which corresponds to the elastic mechanism of electron scattering on neutral (uncharged) ions. For elastic scattering of electrons on the fixed ions of the crystal lattice the A coincides with the expression (1.3.10) and then the relaxation time depends on the velocity, but not on its direction:

$$\tau_r = \frac{\lambda}{< v_T >} = \tau. \tag{1.3.25}$$

Substitution of (1.3.22) in the definition of the average velocity of the electron (by the mean-value theorem) provides an additional contribution to the electron velocity in the direction of the electric field $<v_T> = <v_T> + <v_D>$, which determines the electric current. Here

$$< \vec{v}_D >= \frac{e\tau_r \vec{E}}{m}. \tag{1.3.26}$$

In the direction of the electric field (along axis x) the current density is

$$J_x = -\iiint e v_x f dv_x dv_y dv_z, \tag{1.3.27}$$

where the distribution function f is given by (1.3.22).

Since $J_x = \sigma E_x$ and the integral (1.3.27) comprising f_0, is equal to zero, we obtain

$$J_x = \sigma E_x = -\int\int\int \left(\frac{E_x e^2}{m} \right) \tau_r v_x \left(\frac{\partial f}{\partial v} \right) dv_x dv_y dv_z. \tag{1.3.28}$$

Using (1.3.25), we get relation,

$$\sigma = \frac{4ne^2\lambda}{3(2\pi mkT)^{1/2}} = \left(\frac{3\pi}{8} \right)^{1/2} \frac{ne^2\tau}{m}, \tag{1.3.29}$$

which is almost identical with the Drude formula (1.3.12a).

Hall effect. As can be seen from the Drude and Drude-Lorentz relations, to calculate the conductivity, we need to know the concentration and mobility of electrons. To estimate the electron density, Hall effect is generally used.

If we put conductor with a current density j (drift velocity $\vec{v}_D$) along x axis in a magnetic field $H \perp j$ (z direction), the electric field $\vec{E}_H$ (Fig. 1.3.3) appears in y direction (normally to H and $\vec{v}_D$). This field is called *the Hall field*. Try to calculate the $\vec{E}_H$.

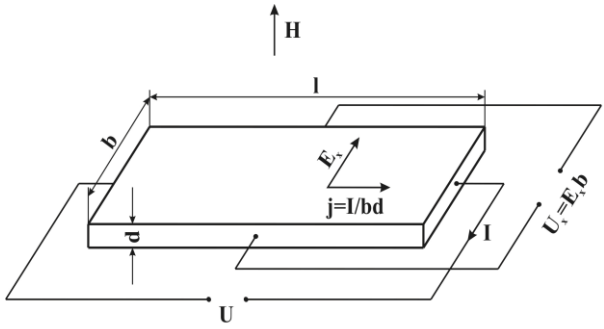


Fig. 1.3.3. The scheme of the Hall effect measurement

At equilibrium state, the impact of the Hall field on electron is balanced by the Lorentz force $e\vec{E}_H = e[\vec{v}_D \vec{H}]$, also acting on the electron. This relation means that in the case of $H \perp j$ electrons, accumulating on the conductor faces in the direction y (see Fig. 1.3.3) under the influence of Lorentz force, will create an electric field which counteracts a charge accumulation. In equilibrium and with $H \perp j$, this field will be equal to

$$E_H = v_D H = \frac{j}{ne} H = R_H jH, \tag{1.3.30}$$

where

$$R_H = -\frac{1}{ne} \tag{1.3.31}$$

is called *the Hall constant*. Thus, by measuring the Hall field, one can use the formula (1.3.31) to determine the concentration of electrons in the metal experimentally. As is seen, the Hall coefficient for electron gas must be negative and do not depend on any parameters except the carrier concentration.

To measure the Hall effect (see Fig. 1.3.3), the rectangular metallic plate of length l, which is significantly greater the width b and thickness d, is placed perpendicularly to the magnetic field H and through the plate an electric current $I = jbd$ is passed. The Hall voltage $V_H = E_H b = R_H Hj/d$ is measured between two electrodes situated in the middle of the side faces. By measuring the Hall constant R_H and conductivity σ, we can estimate according to (1.3.12b) and (1.3.31) not only the concentration n, but also to calculate the electron mobility

$$\mu_H = \sigma R_H. \tag{1.3.32}$$

The difficulties of the classical theory of electrical conductivity. The Drude-Lorentz model for free-electron gas, despite the apparent simplicity, allowed to explain Ohm's law (1.3.11) and to understand qualitatively some of the

experimental data concerning the binding energy, electrical conductivity (for alkali metals), thermal conductivity, Wiedemann-Franz law, etc. However, when comparing with the experimental data, significant discrepancies between theory and experiment were revealed.

Drude theory allows to estimate the values of the relaxation times, mean free path and mobility of the electrons from measurements of conductivity and Hall effect. It is easy to estimate that $\tau \approx 10^{-14}$-10^{-15} s and the mean free path $\lambda \approx 0.1$-1 nm (for $v_T \sim 10^7$ cm/s). However, experiments have shown that for a very pure metals λ value can reach 1 cm when the temperature lowering, which is quite unclear in terms of the Drude theory. In addition, the real temperature dependences of conductivity in metals were quite different (see Fig. 1.3.4), than the Drude (1.3.14) or the Drude-Lorentz (1.3.29) formulae give.

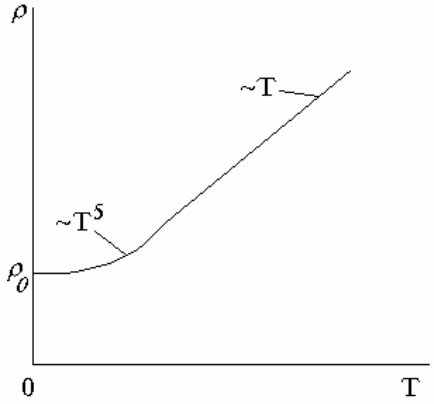


Fig. 1.3.4. The experimental temperature dependence of the conductivity in a metal

Experimental data on heat capacity C_e also differ substantially from the predictions of the Drude model: real $C_e \sim T$ whereas the Drude electronic capacity is constant, according to (1.3.15). Moreover C_e values for room temperature was by two orders of magnitude less than was predicted by Drude model for metals.

The measurements of the Hall effect for very pure substances at low temperatures showed that the experimental results for concentrations n of electrons in for alkaline (monovalent) metals are close to the Drude value of n_0, corresponding to one electron per atom. In the noble metals (as monovalent) value $n/n_0 = (1.3\text{-}1.5)$. At the same time, for the divalent Be and Mg R_H was found to be positive (!) and $n/n_0 = (0.2\text{-}0.4)$.

Thus, it has been shown that the Drude model can not explain some properties of metallic crystals, in particular, the difference between the

measured values of the electron density in metals of different groups of the periodic table and a positive sign of R_H for some metals (in this model, R_H shoulf have only negative sign!). The Drude model also gives the wrong temperature dependence of the electrical conductivity and heat capacity of free electrons gas. Moreover, this model could not explain, in principle, the division of materials on metals and dielectrics.

1.3.2. Quantum theory of free electrons in metals (Zommerfeld model)

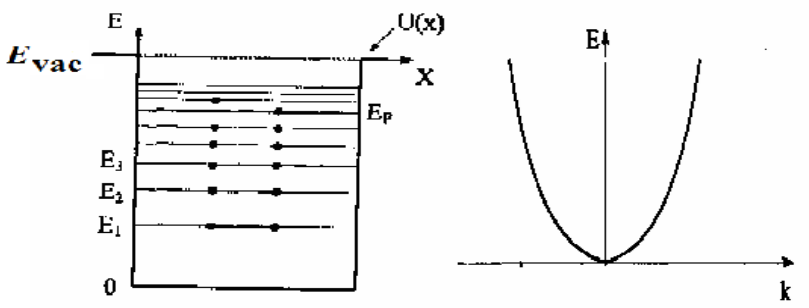
The disadvantages of the classical Drude-Lorentz model were overcome by the quantum theory of free electrons in metals, developed by Sommerfeld. The Sommerfeld model uses the idea of quantization of the kinetic energy of free electrons, the Pauli exclusion principle and quantum Fermi-Dirac statistics.

The dispersion relation of the electrons in the crystal (the dependence of the electron energy on the wave vector). According to de Broglie's hypothesis (see above) and collectivized movement of free electrons in the internal electric field, produced by ions of the crystal lattice, in the weak-coupling approximation can be considered as movement of almost free particles. In this case, the expression for the kinetic energy of a free electron having a quasi-momentum p and mass m, can be written as

$$\varepsilon = \frac{p^2}{2m} = \frac{h^2 k^2}{2m}, \qquad (1.3.33)$$

where $k = 2\pi/\lambda$ is the wave number. In this case, according to the Sommerfeld model, the metallic crystal can be replaced by a kind of rectangular potential box (Fig. 1.3.5a), in which free electrons can be considered as noninteracting particles of an ideal gas having a parabolic dispersion law (Fig. 1.3.5b). In this case, the electrons have discrete energy spectrum and occupy energy levels (Fig. 1.3.5a) not more than two electrons to the level (due to the Pauli principle). The highest energy level that electrons in a metal can take at $T = 0$ K is called the Fermi level (see Fig. 1.3.5a).

In other words, the movement of the free electron can be interpret as propagation of plane wave with the velocity $v = \hbar k/m$ (see below). Plot of the $\varepsilon(k)$ function has the same form as the $\varepsilon(p)$ curve. At the same time, by virtue of parabolic dispersion law for free electrons in a metal (Fig. 1.3.5b), the distribution of electron states in k-space is characterized by a spherical symmetry (Fig. 1.3.6).



a) b)

Fig. 1.3.5. The system of energy levels (*a*) and parabolic dispersion law (*b*) of the conduction electrons in the metal

The Fermi level of electrons in metals. According to the Sommerfeld model, at $T = 0$ K, all the n electrons in a metal tend to occupy the states with the lowest values of the energy E subject to the Pauli exclusion principle (no more than two electrons with opposite spins per state). In such a case, due to squared-like electron dispersion law in a metallic crystal, in the absence of an external electric field, all occupied electron states in k-space will be inside a sphere of radius k_F (see Fig. 1.3.6). The surface of the ball is called *the Fermi surface*, and the highest energy of electrons, as noted above, is *the Fermi energy*.

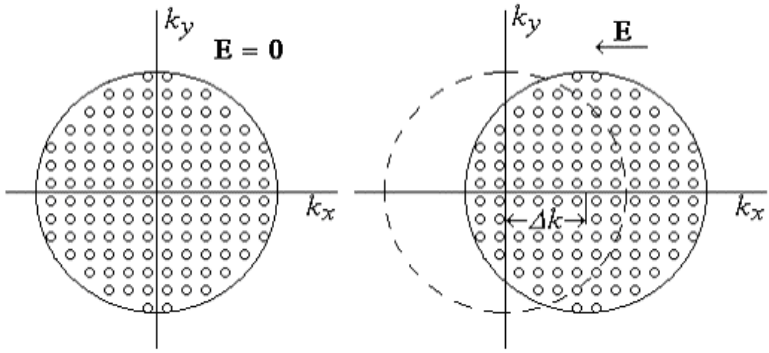


Fig. 1.3.6. Changing the distribution of electron states of a conductor in the Brillouin zone when exposed to an electric field E

When an external electric field E is applied along the k_x, distribution of electrons by the states, shown in Fig. 1.3.6, is shifted on some distance δk_x. Obviously, after a sufficiently long time, the velocity of the electrons and displacement of the electron distribution in Fig. 1.3.6 can become very large.

However, according to the Drude-Lorentz model (see Section 1.3.1) for time $t > \tau_r$ collisions between electrons and lattice ions will result in the distribution to the steady state (the right part of Fig. 1.3.6).

It is easy to show that the Fermi energy is dependent on the concentration of free electrons n and at T = 0 is given by formula

$$\varepsilon_F = \frac{h^2\left(3\pi^2 n\right)^{2/3}}{2m}, \qquad (1.3.34)$$

which implies that $k_F \sim n^3$. As the temperature increases, the probability of electrons states (k values) occupying given by the Fermi-Dirac function is changed in accordance with Fig. 1.3.7. It should be noted that for all metals at all temperatures, including their melting temperature, the Fermi energy is 50-200 times greater than the value of thermal energy kT. Therefore, the electron gas in metals should be considered as highly degenerate electron Fermi gas.

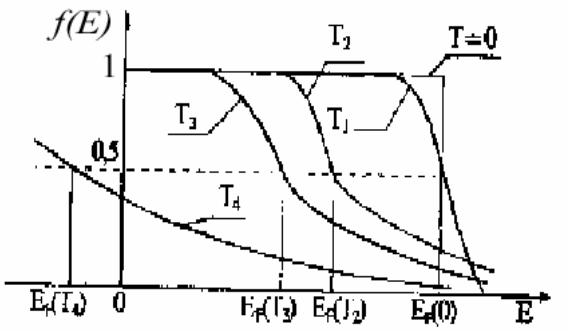


Fig. 1.3.7. The filling state function for the electron Fermi gas at different temperatures $T_4 > T_3 > T_2 > T_1 > 0$

The density of states and its dependence on energy. To determine the number of electrons with energies in a given interval (ε, ε + dε), we need, except for the distribution function f(ε), to know the density of states g(ε). The described in Section 1.2.2 function g(ε) in the form of (1.2.33), specifies the number of levels, which are attributable to unit energy (density of levels).

Extending the theory of the electron gas on the case of arbitrary temperatures and using the well-known normalization condition

$$n = \int_0^\infty f_{F-D}(\varepsilon)\, g(\varepsilon)\, d\varepsilon = (2s+1)\frac{2\pi}{h^3}(2m)^{3/2}\int_0^\infty \varepsilon^{1/2} f_{F-D}(\varepsilon)\, d\varepsilon, \qquad (1.3.35)$$

we can determine the temperature dependence of the Fermi level. The calculations show that the Fermi energy increases with temperature

$$\varepsilon_F(T) = \varepsilon_F(0)\left[1 - \left(\frac{\pi^2}{12}\right)\left(\frac{kT}{\varepsilon_F(0)}\right)^2\right], \qquad (1.3.36)$$

where the Fermi energy $\varepsilon_F(0)$ at T = 0 K is given by (1.3.34). The weak dependence of the Fermi level on temperature indicates that the temperature increase leads only to a slight ($\sim kT$) smearing of the Fermi-Dirac distribution function (see Fig. 1.3.7).

The main content of the Sommerfeld model. Thus the essence of the Sommerfeld model is that the free electrons in the metal can be considered noninteracting particles of an ideal gas in a squared-like potential well (Fig. 1.3.5a). Since electrons are fermions, they will consistently fill the energy levels up to the Fermi level, which will determine the maximal kinetic energy of the electrons in the metal at T = 0. Thus, the Fermi level $\varepsilon_F(0)$ in a metal defines the boundary between filled and unfilled (empty) states.`

As a result, at the absolute zero function has a stepped form (Fig. 1.3.7), while when increasing the temperature, $f_{\text{F-D}}$ is spraded by the width of $2kT$ (Fig 1.3.7). This means that the electrons, lying below the Fermi level, jump to higher, empty energy levels due to thermal excitation, freeing states below ε_F. Moreover, the main part of the electrons is completely insensitive to the very substantial changes in temperature. In doing so, the value of the Fermi energy is uniquely determined by the concentration of electrons. Dependence $\varepsilon_F(0)$ of n is non-linear, because at the growth of $g(\varepsilon)$ more and more electrons can be located at higher energy levels in the energy interval $d\varepsilon$.

Such an approach to the description of the electron gas in metals has allowed to eliminate a part of shortcomings for classical Drude and Drude-Lorentz models. In particular, they have described the reasons of lack of electron contribution to the heat capacity of the metallic crystal at temperatures of the order and above the Debye temperature.

The main content of the Sommerfeld model. One of the mostly impressive successes of the Sommerfeld model is the explaination of thermoelectronic emission in metals at high temperatures. This phenomenon can be explained using Fig. 1.3.7 and 1.3.8. The x axis in Fig. 1.3.8 is perpendicular to the surface of a metallic crystal, on which there is a potential step with the height φ with respect to the Fermi energy ε_F. With the temperature growth, an increasing number of electrons n will have energy (velocity v_x), sufficient to yield a metal. Obviously, the condition when electron leaves the metallic surface in the direction x:

$$v_x \geq \sqrt{\frac{2(\varphi + \varepsilon_F)}{m}}, \qquad (1.3.42)$$

where is called *the work function*.

It is easy to show that the Fermi-Dirac function for the electron velocity distribution at high temperatures (when the energy of the electrons are much higher than the Fermi energy at T = 0 K) becomes the classical Maxwell-Boltzmann distribution (see curve for T_4 in Fig. 1.3.7).

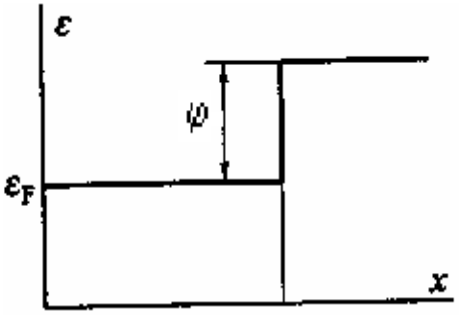


Fig. 1.3.8. The model used when considering the thermionic emission from the metal surface(φ is work function)

According to the Sommerfeld model, current density of electrons, which are capable for emission along the x axis and have the velocities in the range of (v_x, $v_x + dv_x$), is determined by the Richardson-Dushman law

$$j = AT^2\left(1 - r^2\right)\exp\left(-\frac{\varphi}{kT}\right), \qquad (1.3.43)$$

where $A = \frac{emk^2}{2\pi^2\hbar^3} \approx 120\frac{A}{cm^2K^2}$, r is the reflection coefficient of electrons with the velocity that satisfies (1.3.42) on the potential step back into the metal from its surface. This coefficient is small and does not depend on temperature, but, as follows form the quantum mechanics, is not zero.

The values for work function φ for metals lie in the range of 1.8 eV (Cs) to 5.3 eV (Pt), and for some oxides and other compounds they are considerably less. Work function depends not only on temperature, but also very strongly on the surface state (the above values are given for sufficiently clean surfaces). For single crystals work function to some extent depends on the orientation of the surface relative to the crystal axes.

Metallic conductivity in the Sommerfeld model. The aforementioned drawbacks of classical models in the description of the electrical conductivity of

metals have been eliminated through the use of the Sommerfeld model based on the quantum Fermi-Dirac function for the electron energy distribution. According to the Sommerfeld model, with increasing temperature the Fermi distribution is "smeared" only slightly (Fig 1.3.7), so that the main part of electrons is not involved in the formation of the electrical conductivity (and heat capaity) of metallic crystal.

Using the Fermi-Dirac statistics instead of the Maxwell-Boltzmann in the kinetic Boltzmann equation, it is possible to obtain a quantum expression for the conductivity. As it turned out, in this case, we receive the same formula as was in the Drude-Lorentz model

$$\sigma = \frac{ne^2\tau_r}{m}. \qquad (1.3.44)$$

However, the characteristic time τ_r is called *the relaxation time*, and has a completely different meaning. It is given by relation

$$\tau_r = \frac{\lambda(\varepsilon_F)}{v(\varepsilon_F)}. \qquad (1.3.45)$$

Since only applies to electrons (with the concentration on the order $kTn/E_F(0)$), which have an energy around the Fermi energy at T = 0 K. Then (3.1.44) implies that

$$\sigma = \frac{ne^2\lambda(\varepsilon_F)}{mv(\varepsilon_F)} = \frac{ne^2\lambda(\varepsilon_F)}{\sqrt{2m\varepsilon_F}}, \qquad (1.3.49)$$

where the Fermi velocity $v(\varepsilon_F) = \sqrt{\frac{2\varepsilon_F}{m}} \approx 10^8\ cm/s$ is much higher than the thermal velocity of the classical electron gas. This approach allowed Gryunayzen theoretically obtain the temperature dependence of the conductivity $\sigma \sim T^n$, where n = 3÷5, which is close to the experimental ones (see section 4 below).

Advantages and disadvantages of the Sommerfeld model. Sommerfeld model has allowed to remove some inconsistencies and explain the shortcomings of the classical models of the electron gas. In particular, it has explained why not all free electrons may be involved in the energy transfer and contribute to the electric current. It also allowed to calculate correctly the laws of thermionic emission and electrical conductivity for metallic crystals. However, this model has not been able still to explain the reasons for the existence of insulators, semiconductors and even more semiconductors. In particular, it remained unclear fundamentally different temperature dependence of the electrical conductivity in metals (power-like) and dielectrics (exponential), a positive sign of the Hall effect in some crystalline materials, as well as the high sensitivity of

semiconductors to external influences (temperature, radiation, light, magnetic and electric fields, etc.).

1.3.3. Zone model of solids

Tight-binding approximation. Consider what happens to the energy of the electrons in the atoms, which at first are at large distances r, and then draw closer to a distance r_0, which corresponds to the equilibrium interatomic distances in the crystal (~10^{-8} cm). In the remote (non-interacting) atoms the total energy of the electrons takes only a number of distinct values of E_1, E_2 , E_3, ... (Fig. 1.3.9a). When $r = r_0$, potential energy curves $U(r)$ in the spaces between the atoms are overlapped, so that potential barriers, separating the electrons in adjacent atoms, drop (Fig. 1.3.9b). As a result, the electrons, being on the highest level (E_3 in this figure), can migrate freely from atom to atom, and therefore belong the entire crystal that corresponds to Sommerfeld.

If the crystal contains N atoms, then all electrons being on the outer levels in isolated atoms could be in the same state in the crystal. However, the Pauli principle forbids the presence of more than two electrons at one energy level. Therefore, at the formation of the crystal the energies of these electrons should be changed due to the electron interaction. In the simplest case, the interaction between the N electrons on these external levels (level E_3 in this case in Fig. 1.3.9*a*) should result in their splitting onto N close sub-levels, every of which may be occupied by only two electrons with opposite spins. The same applies to higher (excited) levels E_4, E_5, …, E_∞. Fig. 1.3.10 schematically shows how discrete levels of N isolated atoms (being at $r >> 10^{-8}$ cm) expand in the zone (band) with r decreasing. In the simplest case N sub-levels appears in every allowed band, having energy minima at $r = r_0$.

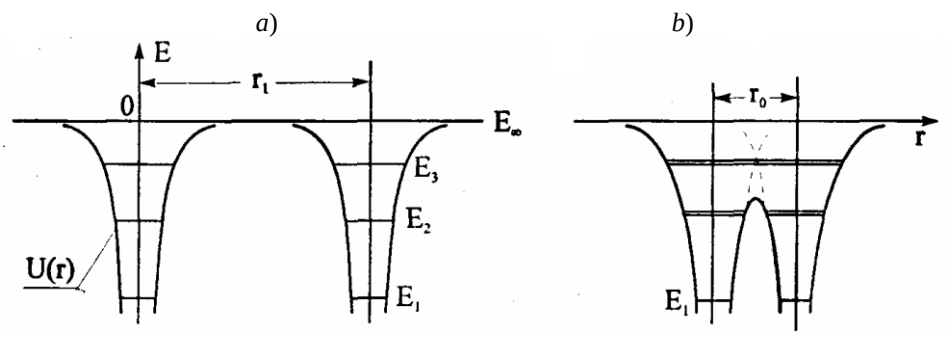


Fig. 1.3.9. The energy of the electrons in the isolated (non-interacting) atoms (*a*) and atoms in the crystal lattice (*b*)

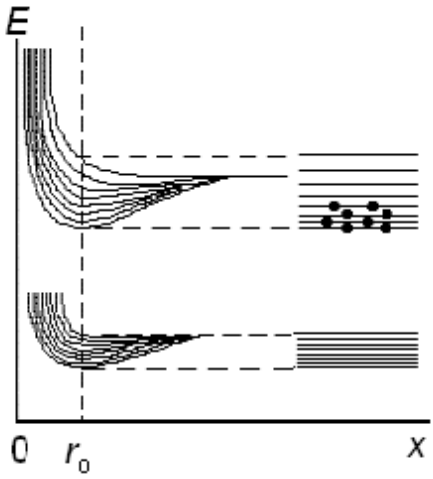


Fig. 1.3.10. Scheme of energy bands creation in the crystal from atomic energy levels when atoms approach equilibrium interatomic distance r_o

For the valence electrons the width of the allowed energy bands is of a few electron volts $\Delta E \sim h/\tau \sim 1$ eV (close to the energy of interatomic bonds). It follows that for typical value $N \approx 10^{22}$ cm^{-3} for concentration of atoms in crystals, the interval between levels constitutes $(\Delta E/N) \sim 10^{-22}$ eV. In other words, the levels in the allowed band are so close that even at low temperatures thermal energy of the electrons kT is much greater than this value, so that this band can be considered as a zone of *quasi-continuous* values of the allowed energies.

As a result, to move through the crystal can not only electrons with energy E_3, but the electrons, which are located at the levels of E_2, which are separated in the crystal low potential barrier (see Fig. 1.3.11b). As a result, due to the tunneling effect or the lattice thermal energy kT, these electrons are able to overcome these barriers with the width of about 10^{-8} cm, and also become common throughout the crystal. Therefore, in this case, a range of allowed energies, although with a smaller width, is also appears.

Electrons, located at lower energy levels (like E_1 in Fig. 1.3.9b), are separated from those of the electrons in the neighbor atoms much more high and wide potential barriers, so they remain localized on their atoms. This means that the deep level E_1 is not splitted and does not create the allowed band for the electron states.Therefore, if the levels E_2 of the electrons in the atom were filled with electrons, in the crystal these electrons form allowed band of free electrons. If a high-energy levels E_3 in isolated atoms were free of electrons

under normal conditions, in the crystal such levels form the empty band (free of electrons).

The mentioned approach, when the energy spectrum of electrons in a crystal is based on the presence of energy levels in isolated atoms, is called *the strong-coupling approximation*. It illustrates well (at least qualitatively) the general regularities of the energy bands formation at coming together of isolated atoms when crystal lattice is formed, Fig. 1.3.10.

Thus, as follows from the strong-coupling approximation, the kinetic energy of the electrons in the crystal presents a set of wide empty and filled with electrons bands, which are separated by forbidden zones (the electrons with such kinetic energy are absent!). Such a form of energy spectrum of electrons is called *band (zone) energy spectrum*. The band spectrum (see Fig. 1.3.10) is common to all solids and determins many of the properties of crystals (electrical, magnetic, optical, etc.).

An electron in a periodic field of the crystal. Single-electron adiabatic approximation. As the crystal is composed of about 10^{23} cm^{-3} atoms and valence electrons, the most complete data on the electronic subsystem of the crystal can be obtained by solving the full Schrödinger equation with taking into accounts of all kinds of interactions: every electron interacts with every atom and every electron. Because it is difficult, we need to simplify the task. One of the simplifications in the calculation of the electron energy spectrum suggests that the electronic and lattice subsystems move independently of each other, that is without exchange by energy (*adiabatic approximation*). Another simplification assumes that behavior of all electrons is the same in different unit cells, and the action of the ions and the other electrons on a given electron can be replaced by an impact of some periodic potential field (*a single-electron approximation*).

A single-electron Schrödinger equation (1.2.9) for electron moving in the field of ions for the stationary case ($t > \tau_r$), will be the following:

$$-\frac{h^2}{2m}\frac{d^2\Psi(r)}{dr^2} + V(r)\Psi(r) = E\Psi(r), \qquad (1.3.50)$$

where the potential of the crystal lattice, in which free electron is moved, is periodic $V(x) = V(x + a)$. Therefore, the solution of the type (1.2.10) for free electron will be in the form of a plane wave de Broglie

$$\Psi_k(x) = U_k(x)\cdot \exp(ik\cdot x), \qquad (1.3.51)$$

where

$$\Psi_k(x) = \Psi_k(x + Na),\ U_k(x) = U_k(x + Na). \qquad (1.3.52)$$

Note that factor $U_k(r)$ accounts for the effect of periodic crystal field on the electron and reflects the fact that the probability to find an electron in a particular region of the crystal is repeated from one cell to another.

Solution of the stationary Schrödinger equation for the Kronig-Penney model. In addition to a strong coupling approximation another approach was proposed by Kronig and Penny. It shows that the presence of the band structure of the energy spectrum for electrons in a crystal is a fundamental consequence of the translational symmetry of the crystal lattice. For simplicity, in this approach, the case of one-dimensional lattice with periodic potential

$$V(x) = V(x+a) = V(x+2a) = ... \qquad (1.3.53)$$

which is a combination of an infinite number of potential wells of width b and potential barriers of height V_0 and width of $a = a - b$ (Fig. 1.3.11).

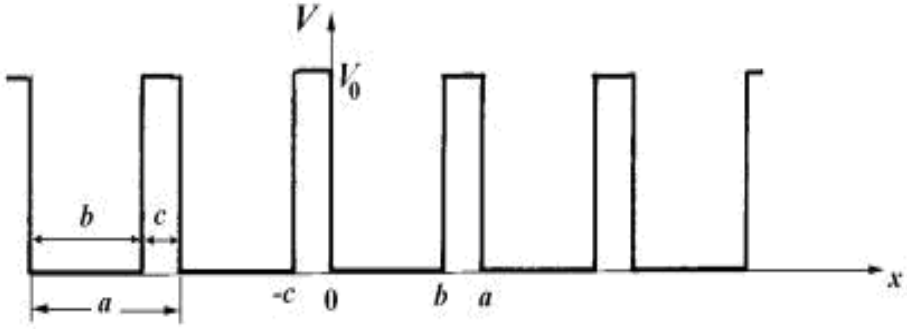


Fig. 1.3.11. The change of potential energy of an electron in a one-dimensional single-atom chain with recurrent rectangular potential wells in the model of the Kronig-Penney

Note here that the lattice period is $a = b + c$. In this case, the potential energy of the electron in the equation (1.3.50) is given by the relation

$$\begin{cases} V(x) = 0 & \text{when} \quad 0 < x < b \\ V(x) = V_0 & \text{when} \quad b < x < a = b + c \end{cases}, \qquad (1.3.54)$$

and the electron in the well is at the energy level $E < V_0$. The height of the potential barrier for electron is equal $\Delta V = V_0 - E$, whereas the thickness of the barrier is c.

If the velocity of an electron in the chain is equal to v, it realizes v/b approaches to the barrier per unit time at the motion in the well along the chain. Then, the frequency of transitions of n electrons from one well to another (from one atom to another) is proportional to the probability of tunneling through a potential barrier. Using relations of Section 1.2.2, the probability of tunneling through a rectangular potential barrier (transition frequency) is exponentially dependent on the width of the potential barrier and its height

$$\nu = \frac{v}{b} \exp\left(-\frac{2h}{c} \sqrt{2m\Delta V} \right). \qquad (1.3.55)$$

Evaluation by formula (1.3.55) indicates that, when the potential barrier width $c \sim 10^{-8}$ cm, its height of about $V_0 \sim 10$ eV (close to the ionization potential of the isolated atom), velocities of electron in the atom $v \sim 10^8$ cm/s and radius Bohr orbit $b \sim 10^{-8}$ cm, the time during which an electron is in a particular lattice site equals merely only $\tau = 1/\nu \sim 10^{-15}$ seconds. In other words, the electrons of the outer atomic shells in the crystal are not localized near a specific site in the lattice, but move across the crystal at a velocity $v \sim 10^{-8}/10^{-15} \sim 10^7$ cm/s.

For the electrons of the inner atomic shells the potential barrier is wider and higher, and the probability of tunneling is much smaller than for the valence electrons. Consequently, electrons from substantially deeper atomic levels are localized in lattice sites.

For a selected Kronig-Penney potential, the general solution of Schrödinger equation will be in the form of a plane wave with an modulated amplitude like (1.3.51). Therefore the general solution of the Schrödinger equation for the electrons in one-dimensional Kronig-Penney model will look like:

$$\frac{\beta^2-\alpha^2}{2\alpha\beta} sh(\beta c)\sin(\alpha a) + ch(\beta c)\cos(\alpha a) = \cos\big(k(b+c)\big) \qquad (1.3.56)$$

where $\alpha = \sqrt{\frac{2m\varepsilon}{\hbar^2}}$ and $\beta = \sqrt{\frac{2m(V_0-\varepsilon)}{\hbar^2}}$. The expression (1.3.56) can be greatly simplified by assuming that the width of the barrier tends to zero $c \to 0$, and its height – to infinity $V_0 \to \infty$, but in a way that their product remains constant ($V_0 c \approx$ const). Under these conditions the expression (1.3.56) is transformed to

$$p\frac{\sin(\alpha a)}{\alpha a} + \cos(\alpha a) = \cos(ka), \qquad (1.3.57)$$

where the parameter p characterizes the "power" of the potential barriers that separate regions with zero potential

$$p = \frac{mabV_0}{\hbar^2}. \qquad (1.3.58)$$

Since parameter α defines kinetic energy E of the electron and k is its wave vector, the expression (3.1.58) is actually presents dispersion law $E(k)$ for an electron in a crystal lattice. When $p \to 0$ (*weak-coupling approximation*), we can get from (3.1.58) that $\alpha a = ka$, whence

$$\varepsilon = \frac{\hbar^2 k^2}{2m}, \qquad (1.3.59)$$

which corresponds to the dispersion law for free electrons in the Sommerfeld model (see above).

On the contrary, if $p \to \infty$ (*approximation of absolutely bound electrons*), the energy of the electrons is independent of *k*. As seen from the equation (3.1.58), allowed energy values will have only those electrons for which $\sin(bc)= 0$, i.e. $\alpha a = n\pi$. Hence the energy of the electrons is given by relation

$$\varepsilon = \frac{\hbar^2\pi^2}{2ma^2}n^2, \tag{1.3.60}$$

where *n* is the integer. The expression (3.1.60) corresponds to the solution of the Schrödinger equation for a particle, which is in a one-dimensional potential well of width *c* and the walls with infinite height. The energy levels of a particle in such a well are discrete that corresponds to the case of isolated atoms.

The analysis shows that if the value of *p* is finite ($p >> 1$), the electron will overcome potential barriers easier and easier if its energy (parameter α) significantly increases compared with V_0 and there will come a time when the electron becomes like a free. In the intermediate case (*strong coupling approximation*), the energy of the electrons is characterized by the dispersion relation:

$$\begin{aligned}\varepsilon_n &= \frac{\hbar^2 n^2\pi^2}{2ma^2}\left\lfloor 1+\frac{2}{p}\left((-1)^n\cos(ka)-1\right)\right\rfloor = \\ &\frac{\hbar^2 n^2\pi^2}{2ma^2}\left(1-\frac{2}{p}\right)+\frac{2\hbar^2 n^2\pi^2}{2pma^2}(-1)^n\cos(ka).\end{aligned} \tag{1.3.61}$$

As follows from this relation, the dispersion law for the electrons is multi-valued, and consists of *n* allowed energy bands (zones). By choosing different values of $n = 1,2,3...$, we obtain the a whole set of dispersion laws for different energy bands:

$$\begin{cases}\varepsilon_0(k) = 0 \\ \varepsilon_1(k) = \dfrac{\hbar^2\pi^2}{2ma^2}\left(1-\dfrac{2}{p}\right)-\dfrac{\hbar^2\pi^2}{pma^2}\cos(ka) \\ \varepsilon_2(k) = \dfrac{4\hbar^2\pi^2}{2ma^2}\left(1-\dfrac{2}{p}\right)+\dfrac{4\hbar^2\pi^2}{pma^2}\cos(ka) \\ \varepsilon_3(k) = \dfrac{9\hbar^2\pi^2}{2ma^2}\left(1-\dfrac{2}{p}\right)-\dfrac{9\hbar^2\pi^2}{pma^2}\cos(ka) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\end{cases} \tag{1.3.62}$$

Thus, using the approach of the Kronig-Penney, we have shown that the energy spectrum of electrons in a periodic potential field which is due to the translational symmetry of the lattice, really has a band structure. Dependence of the electrons energies the crystal on the wave vector (quasi-momentum) for

different types of allowed bands (1.3.62) is schematically shown in Fig. 1.3.12. Dot-dashed parabola in this figure corresponds to the limiting case of the energy of free electron with a quadratic dispersion law (1.3.59). For the electron in crystal this parabola is replaced by σ-shaped portions of sinusoid which are separated from each other by discontinuities in the energy spectrum at the boundaries of Brillouin zone at $k = \pm\pi n/a$, where $n = 1, 2, 3, \ldots$. These discontinuities just are those gaps between allowed bands, which were mentioned above.

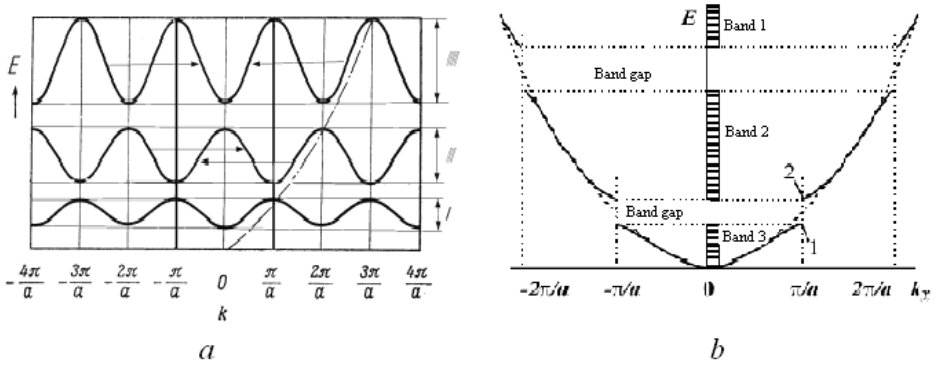


Fig. 1.3.12. The Kronig-Penney $E(k)$ dependences for different allowed energy bands for an electron in a one-dimensional lattice in the representation of extended zones (*a*) and the generalized dispersion law (*b*)

As is seen from Fig. 1.3.12, the parts of sine curves in different allowed energy bands I, II, III may belong to different Brillouin zones. The ranges of allowed values of the energy (allowed bands) I, II, III are separated by intervals of the forbidden values (energy gaps).

Since the electron states in all Brillouin zones are physically equivalent, we can use the procedure of shift (umklupp process) of the dispersion relation branches on the vector $G = 2\pi/a$ of reciprocal lattice. Using this procedure we can reduce the dispersion laws of different Brillouin zones only to the first Brillouin zone. Horizontal arrows in Fig. 1.3.12a indicate the directions of these shifts. The generalized dispersion law in Fig. 1.3.12*b* has a form of the "disrupted" parabola, showing the dependence of $E(k)$ for a free electron in the first Brillouin zone $-\frac{\pi}{a} < k \le \frac{\pi}{a}$. Such an image is called *the representation of the reduced Brillouin zone*.

So, we have proved, using Kronig-Penney model, that the electron energy spectrum is divided into allowed zones (bands) and forbidden zones (gaps).

The presence of gaps E_g in the energy spectrum means the lack of electrons with such kinetic energies.

Filling the energy bands by electrons. Dividing crystals on metals, dielectrics and semiconductors. As noted earlier in this chapter, all crystalline solids can be divided into metals, dielectrics and semiconductors primarily by the values of conductivity. For typical metals, at room temerature this value is of $10^8...10^6$ $(\Omega \cdot m)^{-1}$. For very good insulators electrical conductivity does not exceed 10^{-11} $(\Omega \cdot m)^{-1}$. Crystals with intermediate values of electrical conductivity are usually related to semiconductors.

However, as noted above, insulators, semiconductors and metals are different from each other not only by the resistivity values. In metals and dielectrics a fundamentally different types of the temperature dependence of the electrical conductivity are observed (see Fig. 1.3.13). Furthermore, semiconductors are also characterized by a very high sensitivity to various external impacts (temperature, light, magnetic and electric fields, etc.).

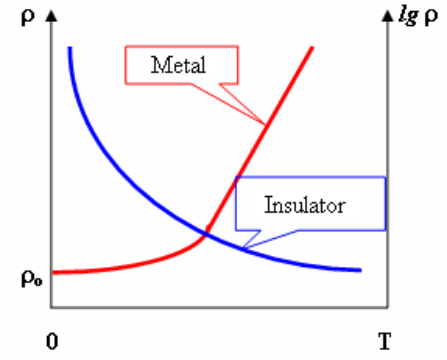


Fig. 1.3.13. The temperature dependences of the electrical resistance in metals and dielectrics

The division of crystalline solids on metals, semiconductors and insulators, as well as marked differences between their conductivities can be explained qualitatively on the basis of the foregoing model of electron energy spectrum. As follows from the model, such large differences in the electrical properties of solids are related to the structure and the degree of filling of energy bands by the electrons.

The number of electrons in a crystal, of course, depends both on number of atoms N and number of electrons per atom. The electrons in the crystal tend to occupy the lowest energy levels. However, the Pauli principle forbids to be more than two electrons at every energy level. Therefore, in the crystal the lowermost energy levels in the allowed bands are primarily filled. As a result,

some (the lowest) bands becomes completely filled, whereas the uppermost are filled either partially or remain completely free of electrons (Fig. 3.1.14).

Band with the highest energies, which is completely filled with electrons, is called *the valence band*. Next, higher band, which can be either partially filled (Fig. 1.3.14*a*) or fully unfilled with electrons (see Fig. 1.3.14*b, c*), is called *the conduction band.*

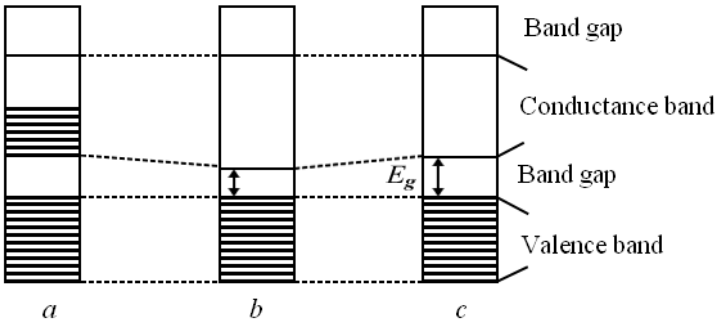


Fig. 1.3.14. Energy bands in the crystal and a scheme of their filling in the case of metal (*a*), dielectrics (*b*) and the semi-conductors (*c*)

So, it follows from the above, the structure of the energy bands (primarily, the value of band gap E_g) of the crystal and the nature of the filling of allowed zones by electrons render a decisive impact on the value of its electrical conductivity. If the magnitude of the electric field is not higher than 10^4 V/m, then at a distance equal to the mean free path (usually ~ 10^{-8} m), the electron in the crystal acquire energy of about 10^{-4} eV. It is clear that these values allow the electron to pass from one level to another but only within the same energy band. To move between the bands it needs the energy of the order of band gap E_g.

These considerations lead to the conclusion that the metallic crystal should have partially filled conduction band in the energy spectrum (Fig. 1.3.14a) to be highly conductive. If the valence band in the crystal is completely filled with electrons and the conduction band is empty (Fig. 1.3.14b, c), weak external electric field will not be able to throw electrons over the band gap from the valence to the conduction zone. Therefore, application of such field will not lead to an electric current in such a crystal and it behaves as dielectric or semiconductor. In this case, dielectrics crystals are characterized by a relatively wide band gap (Fig. 1.3.14c) with typical E_g > 3 eV. Thus, for diamond E_g = 5.2 eV, for boron nitride 4.6 eV and for alumina ~ 7 eV. In typical semiconductors bandgap is significantly less (Fig. 1.3.14b) and not more than 3 eV. For example, in germanium E_g = 0.66 eV, for silicon 1.12 eV, while the indium antimonide has E_g = 0.17 eV.

1.3.4. Electron dynamics in periodic lattice

Features of the electrons motion by the crystal are not only due to the external electric field, but also owing to their interaction with the crystal lattice. In general, the motion of the electron can be described by Newton's second law $\vec{F} + \vec{F}_{in} = m\frac{d\vec{v}}{dt}$, where electron is subjected to the impact of two forces – from the external electric field and from the inner periodic lattice field. However, as it turns out, the free motion of a single electron in a crystal can be described by the same equation of Newton, but which takes into account only the influence of the external force

$$\vec{F} = m^* \frac{d\vec{v}}{dt}. \tag{1.3.63}$$

This description is called *the effective mass approximation*. In this approximation, the electron in a crystal is considered as a quasi-particle described by a wave function whose energy and velocity depends on the wave vector through the dispersion law, but the mass is not equal to the mass of a free electron in a vacuum. As noted above, the absolutely free electron is described by a monochromatic de Broglie wave and is not localized anywhere. Real electron in the crystal is necessary to compare the group of de Broglie waves with different frequencies ω and wave vectors $\vec{k}$. Center of this group is moving in space with the group velocity, which corresponds to the velocity of the electron energy transfer. For one dimensional case, it is expressed as

$$v_{gr} = \frac{dx}{dt} = \frac{d\omega}{dk} = \frac{1}{\hbar}\frac{dE}{dk}. \tag{1.3.64}$$

The increase in the electron energy *dE* under the impact of an external force *F* equals an elementary work *dA*, which accomplishes this force (such as an electric field) over the electron for infinitely small time interval *dt*:

$$dE = dA = Fdx = Fv_{gr}dt. \tag{1.3.65}$$

Substituting the expression (1.3.64) for the group velocity in (1.3.65), we obtain

$$dE = \frac{F}{\hbar}\frac{dE}{dk}dt. \tag{1.3.66}$$

Hence

$$\frac{d}{dt}(\hbar k) = F. \tag{1.3.67}$$

Extending this result to three dimensions case, we obtain the vector equality

$$\frac{d}{dt}\left(\hbar\vec{k}\right) = \vec{F}. \tag{1.3.68}$$

As seen from this equation, the value $\hbar\vec{k}$ for electron in a crystal varies with time under the influence of external force in the same way as the particle momentum in classical mechanics $(dP/dt) = F$. Despite this, as noted in the case of phonons, $\hbar\vec{k}$ can not be identified with the momentum of an electron in a crystal, since the vector components are defined accurately within constant terms of the form $(2\pi/a)n_i$ (here a is lattice parameter, n_i = 1, 2, 3, ...). However, within the first Brillouin zone, the quasi-momentum $\hbar\vec{k}$ has all the properties of the momentum.

We now calculate the acceleration a, acquired by an electron as a quasi-particle under the influence of an external force F. For the one-dimensional task

$$a = \frac{dv}{dt} = \frac{d}{dt}\left(\frac{1}{\hbar}\frac{dE}{dk}\right) = \frac{1}{\hbar}\frac{d^2E}{dk^2}\frac{dk}{dt}. \tag{1.3.69}$$

At the calculation of the acceleration, we take into account that the electron energy is a function of time $E = E\big(k(t)\big)$. Given that $\frac{dk}{dt} = \frac{F}{\hbar}$, we get

$$\frac{d\vec{v}}{dt} = \frac{1}{\hbar^2}\frac{d^2E}{d\vec{k}^2}\cdot\vec{F} \tag{1.3.70}$$

or

$$\vec{F} = \hbar^2\frac{1}{\left(d^2E\big/d\vec{k}^2\right)}\frac{d\vec{v}}{dt} = m^*\frac{\partial v}{\partial t}. \tag{1.3.71}$$

Comparing the expression (1.3.71) with Newton's second law, it is easy to see that the electron in a crystal moves under the influence of an external force such as a free electron in a vacuum would move under the impact of the same forces, if it has mass

$$m^* = \frac{\hbar^2}{\left(d^2E\big/d\vec{k}^2\right)}. \tag{1.3.72}$$

The value of m^* in (3.1.72) is called *the effective mass* of an electron in a crystal. Strictly speaking, the effective electron mass is irrelevant to real mass of the electron. It is *characteristic of the electrons in the crystal as a whole*.

By introducing the concept of effective mass, we describe real electron in a crystal as some new free quasiparticle having only two physical parameters of the real electron – its charge and spin. All other parameters – the quasi-momentum, effective mass, kinetic energy, etc. – are determined by interactiin of electron with the crystal lattice. Such quasi-particle should be called as *quasi-electron*, to emphasize its difference from the real electron in a crystal.

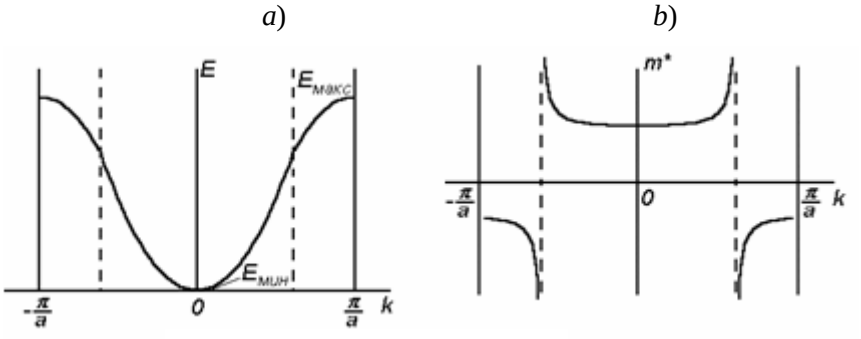


Fig. 1.3.15. The dispersion law (*a*) and dependence of the effective mass on the wave number (*b*) for an electron in a crystal

Features of the electron effective mass are connected with the type dispersion relation of the electron in the crystal (Fig. 1.3.15a). The dependence of the electron effective mass on its wave vector is shown in Fig. 1.3.15b. For electrons, which are located at the bottom of the energy band (in the center of the first Brillouin zone), the dispersion relation can be approximately described by the parabolic law

$$E = E_{\min} + Ck^2, \qquad C > 0. \tag{1.3.73}$$

Since the second derivative in this case equals $\frac{d^2E}{dk^2} = 2C > 0$, therefore, the effective mass is positive and practically independent on the energy of the electron (Fig. 1.3.15a). Behavior of such electrons in an external electric field is qualitatively similar to the free electrons in a vacuum: they are accelerated by an external electric field. The difference between these electrons and the free ones is that their effective mass may differ substantially from the free electron mass. For many metals, in which the concentration of electrons in the partially filled band is small, and they are located near its bottom, the conduction electrons behave similarly. Moreover, if these electrons are weakly bound to the crystal, their effective mass is merely slightly different from the real electron rest mass.

For electrons that are at the top of the energy band (near the edge of the first Brillouin zone), the dispersion relation can be loosely described by the inverted parabola

$$E = E_{\max} - D\left(\frac{\pi}{a} - k\right)^2, \qquad D > 0, \tag{1.3.74}$$

so that the effective mass becomes negative (Fig 1.3.15b). Such behavior of the electron effective mass can be explained as follows: during its movement by the crystal the electron has not only a kinetic energy E_k of its translation motion, but also the potential energy U due to its interaction with the crystal lattice. Therefore, part of the work A of external force can go into kinetic energy and change it to the value of ΔE_k, while another part to the potential energy ΔU:

$$A = \Delta E_K + \Delta U \tag{1.3.75}$$

If at the motion of an electron, not only the whole work of the external force turns into the potential energy, but also some part of its kinetic energy ($\Delta E_к < 0$), then its velocity will decrease. In this case the electron behaves as a particle with a negative effective mass (or with a positive charge!). In the case where all of the work of an external force turns into potential energy ($\Delta E_к = 0$), the increment velocity and kinetic energy does not occur and electron behaves as a particle with an infinitely large effective mass (Fig. 1.3.15b). Infinitely large effective mass of the electron corresponds to the inflection points of the dispersion curve (in Fig. 1.3.15a they are indicated by dashed lines). Note that dispersion law for the free electron is

$$\varepsilon = \frac{\hbar^2 k^2}{2m_0}, \tag{1.3.76}$$

whereas for the electron in a crystal in the effective mass approximation, when taking into account (1.3.73), it will be of the form

$$\varepsilon = \frac{\hbar^2 k^2}{2m^*}. \tag{1.3.77}$$

1.4. Fundamentals of semiconductor physics

As was shown in the previous Chapter, separation of defect-free crystalline materials on metallic, dielectric and semiconducting is determined by the structure of the allowed energy bands and the character of their filling with electrons. Therefore, the most of properties of semiconductor crystals also depend on the value of the forbidden band gap E_g, their doping (type and concentration of impurities) and damages of crystalline lattice.

This section is devoted to the analysis of main factors which influences electric conductivity and optical properties of semiconductor crystals. We shall study dependence of their properties on type and concentration of impurities and defects, incorporated into the crystal, and external impacts (temperature, electric fields, electro-magnetic irradiation, etc.). Also the distribution of equilibrium carriers (electrons and holes) by energies and their interaction between each other and with the crystalline lattice will be also discussed.

1.4.1. Impurities and defects in semiconductors

(Electrical conductivity of intrinsic and doped semiconductors)

As shown in Section 1.3, carrier transport (conductivity $\sigma = ne\mu$) over semiconductor crystals is determined by charge carriers concentration n and their mobility μ. Therefore, the temperature dependence of σ will be determined not only by the charge carriers statistics (which determines their concentration), but also the mechanisms of their scattering. Thus, to solve the problem of the conductivity of the crystal it is necessary to find the concentration of electrons and holes in the crystal, their mobility and (if necessary) their temperature dependence.

Mobility of charge carriers in semiconductors. As follows from the conductivity quantum theory (see Section 1.3), unlike metals, not only vibrations of the atoms (phonons in Zommerfeld model) and damages in crystal lattice (impurities, dislocations, grain boundaries) but also electrical fields of charged defects should have a significant impact on conductivity in real semiconductors. Let us first consider how the change in temperature affects the scattering of charge carriers (electrons and holes) at their motion in a semiconductor crystal, i.e. the temperature dependence of mobility $\mu(T)$.

Phonon contribution to the mobility. As was discussed above for the case of metals, at high temperatures $T >> \theta_D$ (when the total phonon spectrum is excited), the mean free path of electrons at their scattering on acoustic phonons is inversely proportional to temperature $l \sim T^{-1}$. Since the thermal velocity $v = (3k_BT/m)^{1/2} \sim T^{1/2}$, we obtain for this contribution to the mobility of non-degenerate carriers (e.g., free electrons in the C-band of semiconductor or holes in V-band)

$$\mu_{ph} \sim l/v \sim \left(m^*\right)^{-\frac{5}{2}} T^{-1} \cdot T^{-1/2} \sim \left(m^*\right)^{-\frac{5}{2}} \cdot T^{-3/2}, \qquad (1.4.1a)$$

while for degenerate semiconductors

$$\mu_{ph} \sim T^{-1}. \qquad (1.4.1b)$$

Ionized impurities contribution to the mobility. At $T > 0$ K a part of the impurity atoms can be ionized. The problem of electron scattering by charged point defects in the semiconductor crystal lattice in many ways similar to that which has been considered in the study of Rutherford scattering of α-particles on metallic foil. The charged ion of impurity generates around itself electrical field, which changes the direction of movement of flying charge carriers. The deviation from the original direction of their motion is more perceptible for carriers moving with low velocities: the higher their velocity v, the faster they pass by ion, without having to deviate at high angle. The solution of Rutherford gives that in this case for a non-degenerate gas of charge carriers, scattered by

charged impurity ions, the temperature dependence of mobility is given by relation

$$\mu_{ion} \sim v^3 \approx v_m{}^3 \sim N_{ion} \cdot T^{3/2}, \quad (1.4.2)$$

while for a degenerate gas (e.g., metal) mobility is virtually independent of temperature:

$$\mu_{ion} \sim v_F{}^3 \approx \text{const}. \quad (1.4.3)$$

Note, that mobility decreases with increase of the effective mass, but increases with temperature. As follows from (1.4.2), the increase of mobility with temperature is due to the increase of charge carriers velocity. Scattering by the impurity ions is the dominant mechanism at low temperatures where the vibrations of lattice are not significant.

Neutral impurities contribution to the mobility. If neutral impurities are the scattering centers (this is observed at low temperatures, when the carriers, excited to the allowed band, return back to impurity centers), mobility μ_o of charge carriers is determined only by their concentration N_{imp} and does not depend on the temperature:

$$\mu_o \sim (1/N_{ion}) \quad (1.4.4)$$

If the main mechanism of scattering of charge carriers is related with dislocations, mobility does not depend on the temperature and is inversely proportional to their density N_D $\mu_D \sim (1/N_D)$. In polar semiconductors, such as GaAs, carrier scattering by optical phonons is possible. For this type of scattering $\mu_{DO} \sim T^{1/2}$. In real conditions several scattering mechanisms usually coexist simultaneously. The resulting value of mobility can be found from consideration that the total number of scattering events per unit time z equals to the sum of scattering events caused by each of these factors. The mean free path of charge carriers and their mobility will be inversely proportional to the number of scattering events for each channel. Thus, since $\mu_i \sim \frac{1}{z_i}$, the resultant mobility can be represented as $\mu^{-1} = \left(\sum_i \frac{1}{\mu_i} \right)$.

In such semiconductors as germanium and silicon two main mechanisms of scattering are the scattering by acoustic phonons and ionized impurities. Thus $\frac{1}{\mu} = \frac{1}{\mu_T} + \frac{1}{\mu_i} = aT^{3/2} + bT^{-3/2}$, where a and b are some coefficients that do not depend on T. Then $\mu = \frac{1}{aT^{3/2} + bT^{-3/2}} = \frac{T\sqrt{T}}{aT^3 + b}$.

Fig. 1.4.1 presents the temperature dependence of the mobility for non-degenerate semiconductor. As seen, $\mu(T)$ curve passes through a maximum: at high temperatures acoustic phonon scattering with $\mu \sim T^{-3/2}$ prevails (the right branch in Fig. 1.4.1), while the scattering by ionized impurities, where $\mu \sim T^{3/2}$, dominates at low temperatures (the left branch in Fig. 1.4.1). The peak position on the $\mu(T)$ curve depends on the type of semiconductor material and, particularly, on the impurity concentration. Generally it corresponds to temperatures that are considerably below room temperature.

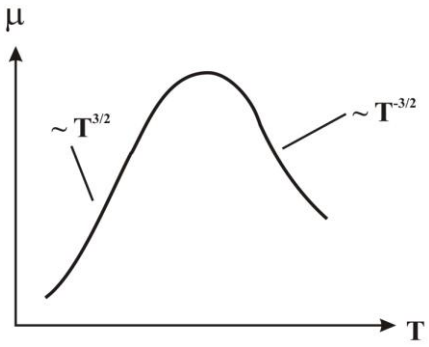


Fig.1.4.1. Temperature dependence of the charge carrier mobility in semiconductor

From the foregoing, it is clear that charge carriers mobility changes rather weakly with temperature by fractionally-power law and cannot provide a strong dependence of the electrical conductivity of a semiconductor on temperature observed in experiments. Moreover, carrier mobility decreases with temperature at high temperatures, while the electrical conductivity strongly increases. All this support the fact that carrier concentration, but not mobility, has determining influence on the electrical conductivity of semiconductors and its temperature dependence. Below we consider the main factors that affect to the concentration of electrons and holes in semiconductors.

Defect-free (intrinsic) and doped semiconductors. Let us consider the crystal at $T = 0$ K with the valence band completely filled with electrons (Fig. 1.4.2a). In the case, when heating, part of electrons overcomes the band-gap Eg and can be transferred from the fully filled valence band (V-band) into the conduction band (C-band) as the result of interaction with the oscillating atoms of the crystal lattice (interaction with phonons), see Fig. 1.4.2b. As a result, the C-band states, occupied by electrons, appear, and the same number of empty states (holes) appears in the V band. This process is called *generation of electron-hole pairs* in the process of *band-to-band transitions*. Just such

substances exhibit semiconducting properties, and defect-free semiconductors themselves are called *intrinsic*.

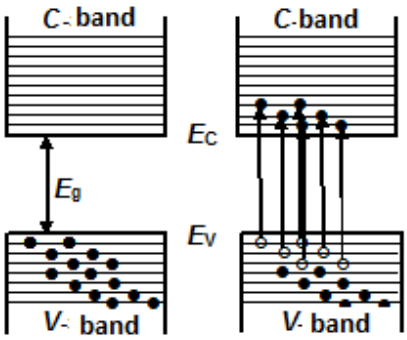


Fig. 1.4.2. Filling the C and V bands by electrons in the defect-free (ideal) semiconductor at T = 0 K (*a*) and at finite temperature (*b*). Black circles are electrons and white circles – holes

In intrinsic semiconductors, when applying the external electric field, the electrons are rearranged by states (move up by the energy levels), as shown in Fig. 1.4.2b, in both conduction and valence bands. Moreover, if the electrons in *C*-band behave as a negatively charged particles with a positive effective mass, the holes in V-band act either as particles with a negative effective mass or as a positively charged particles with a positive effective mass. Furthermore, since there are exponentially few electrons in conduction band and holes in valence band, the defect-free semiconductor will conduct electric current worse than metal but better than dielectric.

As mentioned, the considered in Fig. 1.4.2 band structure can be realized only in the case of perfectly pure and defect-free (intrinsic) semiconductor crystals. Conductivity of intrinsic semiconductor is called the *intrinsic conductivity*. In real crystalline semiconductors, impurities and other defects are always present, some of which have a substantial influence on their conductivity. For example, the addition to silicon of boron atoms in amount of one atom per 10^5 silicon atoms (10^{-3} at.%) results in the increase of its conductivity at room temperature in 1000 times. Semiconductors containing impurities are called *impurity (extrinsic) semiconductors*, and their electrical conductivity is called, respectively, the *impurity conductivity*.

Let us consider the changes that different types of impurities create in band structure and conductivity of semiconductor (for example, like silicon or germanium) at T = 0 K and when heating. In pure silicon crystal, each atom uses four valence electrons for the formation of covalent bonds with four the

nearest neighbors (Fig. 1.4.3a). Suppose that the pentavalent impurity atoms (e.g., phosphorus) are introduced into silicon crystal lattice. They are located in lattice sites (Fig. 1.4.3b) substituting the silicon atoms. Then, as in pure silicon, 4 electrons of each phosphorus atom form 4 covalent bonds with 4 nearest neighboring silicon atoms but the fifth is not involved in the chemical bond. As a result, the extra electron is connected with the phosphorus atom much weaker than other 4 electrons bounded to silicon atoms. This means that at $T = 0$ K, this fifth electron is said to be localized on the impurity.

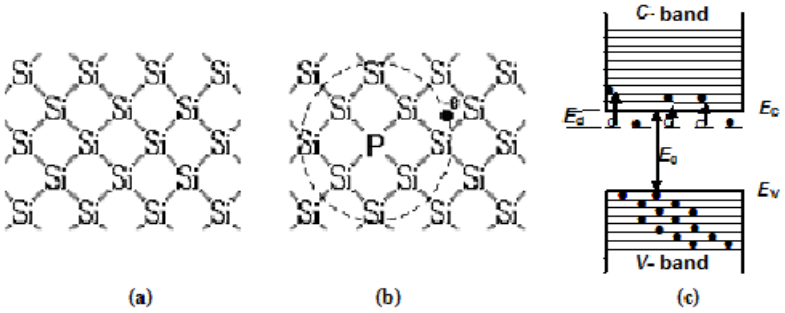


Fig. 1.4.3. Scheme of formation of electronic states in the semiconductor silicon: *a*) the covalent bonds in pure silicon; *b*) silicon with an impurity atom of phosphorus; *c*) band structure of silicon with donor impurities

Such behavior of the phosphorus atom means that it will cause the appearance of energy levels in the band gap of silicon at the depth E_d below the bottom of C-band (Fig. 1.4.2c). Since these levels belong to electrons localized near the impurity atoms, they are represented in the band diagram by the dashed lines.

It will be easy to tear this extra electron away from the phosphorus atom (to make it free) in case of heating of the crystal, permitting its free motion through the crystal when an electric field is applied. In terms of band structure, this means that the electron can move up, getting energy from phonons, from the impurity level to the empty levels of C band (leaving localized hole at the impurity level in the gap). Such an electron can participate in the creation of an electric current when an external electric field is applied. Such an impurity which gives electrons to the C-band is called *donor impurity,* and semiconductors with such impurities are called *electronic semiconductors* or *n-type semiconductors*. E_d energy that must be expended to move the electron from the center of the donor type impurity into the conduction band is called the *ionization energy of donor impurity*. The donor center becomes positively charged after electron separation, since the hole is localized at it.

Atoms of the fifth group in the Mendeleev periodic table, such as phosphorus (P), arsenic (As) and antimony (Sb), are the most common donor impurities in silicon and germanium crystals. Impurity atoms from the third group of the periodic table, such as boron (B), aluminum (Al), gallium (Ga) and indium (In) behave itself differently in silicon and germanium. For example, the substitution of one silicon atom in the silicon lattice by boron atom leads to the fact, that covalent bond of one of the 4 silicon atoms closest to the boron atom remains unfilled. This bond can be restored if one electron from the silicon-silicon chemical bond passes to the boron atom forming electron vacancy, or hole (see Fig. 1.4.4a). In the band diagram, this corresponds to the appearance of local boron impurity levels in the band gap of silicon near the top of the V band (Fig. 1.4.4b). At T = 0 K, this level is free, but electrons from the V-band can fill it if the crystal is heated. The holes formed in the valence band are the carriers of electric current in this type of impurity semiconductors.

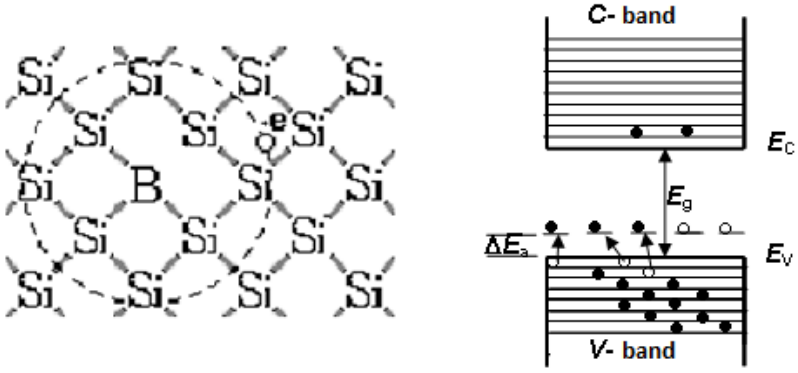


Fig. 1.4.4. Scheme of formation of the hole states and conductivity of the semiconductor silicon: *a* – silicon with boron impurity atom, *b* – band structure of silicon with acceptor impurities

Impurities capturing electrons from the valence band of semiconductors are called the *acceptor impurities*, and the energy levels of these impurities are *acceptor levels*. Difference between the acceptor level energy and the energy of V-band top is called the *ionization energy of acceptor impurity* E_a. Semiconductors containing acceptor impurities are called *hole semiconductors* or *p-type semiconductors*. After electron transfer from the silicon to acceptor center (formation of a free hole in the valence band), the latter becomes negatively charged.

The binding energy of impurities in a semiconductor crystal (ionization energy of the donor or acceptor) can be estimated on the basis of a simple model, similar to the Bohr model of the electron in hydrogen atom. So these impurities are called *hydrogen-like impurities*. The calculations show that the ionization energy of a hydrogen-like donor (acceptor) in eV can be expressed by the relation

$$E_d = \frac{e^4 m^*}{2\left(4\pi\varepsilon\varepsilon_0\hbar\right)^2} = E_d = 13{,}53\frac{m_n^*}{m\varepsilon^2}, \qquad (1.4.5)$$

where ionization energy of free hydrogen atom is 13.6 eV, ε is the dielectric constant, ε_o – electrical constant, m and m^* – mass of electron in vacuum and effective mass of electron in semiconductor crystal, correspondingly.

Electrical conductivity of intrinsic and impurity semiconductors. According to the band structure of intrinsic semiconductor (see Fig. 1.4.2), its conductivity depends on the concentration of generated electrons and holes (moving in electric field in the opposite directions) and can be expressed by the following equation:

$$\sigma = \sigma_n + \sigma_p = ne\mu_n + pe\mu_p. \qquad (1.4.6a)$$

Here σ_n and σ_p are contributions to the conductivity of electrons and holes, n and p –their concentrations, μ_n and μ_p – mobility of electrons and holes, respectively. So, intrinsic conductivity in (1.4.11a) consist of two components – electron (first contribution) and hole (second contribution).

Since in the intrinsic semiconductor electron and hole concentrations are the same (since $n = p = n_i = p_i$, where n_i and p_i are intrinsic concentrations of electrons and holes), the conductivity of intrinsic semiconductor will be equal to

$$\sigma = n_i e(\mu_n + \mu_p). \qquad (1.4.6b)$$

Conductivity of impurity semiconductors is determined by the concentration of charge carriers in corresponding allowed band (electrons in C-band or holes in V-band) and their mobility. In doing so, carrier concentration depends on the level of doping (impurity concentration) and temperature.

The main source of electrons in *n*-type semiconductor at low temperatures are donor impurities. When heating of impurity semiconductor (by thermal excitation), electrons pass from the donor levels in band-gap into the C band, so that the conductivity is defined by

$$\sigma_n = ne\mu_n. \qquad (1.4.7a)$$

At very high temperatures, when the electrons are excited into the C band due to their transition from the V band, the electronic semiconductor behaves like an intrinsic semiconductor (see equation (1.4.6)).

For the hole semiconductor, the following relation is true

$$\sigma_p = pe\mu_p. \qquad (1.4.7b)$$

1.4.2. Statistics of carriers in intrinsic and doped semiconductors

Electrical neutrality of semiconductors. As noted in the preceding section, the impurity conductivity of semiconductors is caused by formation of some free electrons or holes in C- or V-band, respectively, due to the influence of temperature. Concentration of generated free charge carriers is determined by the balance between impurity-to-band and band-to-impurity transitions of electrons. First process is called *generation of carriers* and their return back – *carriers recombination*.

At steady process of generation and recombination of electrons in the *n*-type semiconductor their equilibrium concentration on donors (in this case, the donor atoms are electrically neutral) will be equal

$$N_d^0 = n_d = N_d - N_d^+ = N_d - p_d, \qquad (1.4.8)$$

where: N_d is total concentration of donors (donor levels), n_d – electron density at the donors, N_d^+ – concentration of charged donors (which is equal to concentration of holes p_d localized on them).

In equilibrium state, the concentration of positive charges (localized holes and positively charged donors) and negative charges (free electrons and negatively charged acceptors) must be equal. Only in this case, the semiconductor is electrically neutral.

The electrical neutrality condition for in n-type semiconductor at low temperatures, when $kT << E_d$, and no band-to-band transitions with the formation of electron-hole pairs (as in the intrinsic semiconductor), will have the form

$$n = N_d^+ = p_d. \qquad (1.4.9)$$

The electrical neutrality condition of for electronic semiconductor at high temperatures ($kT >> E_d$) is equal

$$n = p_d + p, \qquad (1.4.10)$$

where

$$p_d = N_d^+ = N_d. \qquad (1.4.11)$$

Equilibrium concentration of holes localized on acceptors in *p*-type semiconductor in steady generation-recombination process equals

$$N_a^o = p_a = N_a - N_a^- = N_a - n_a, \qquad (1.4.12)$$

where N_a is the concentration of acceptor levels, p_a – the holes concentration at acceptor levels, N_a^- – concentration of charged acceptors (which is equal to the concentration of electrons n_a localized on them).

Then the electrical neutrality condition for hole semiconductor at low temperatures ($kT << E_a$) is

$$p = N_a. \tag{1.4.13}$$

The condition of electrical neutrality for hole semiconductor at high temperatures ($kT >> E_a$) is

$$p = n_a + n. \tag{1.4.14}$$

Electroneutrality condition in a semiconductor doped with both types of impurities (such a semiconductor is called compensated):

$$p + p_d - n - n_a = 0. \tag{1.4.15}$$

The condition of electrical neutrality in an intrinsic semiconductor:

$$p_i = n_i \text{ or } \quad p_i - n_i = 0. \tag{1.4.16}$$

For the equilibrium concentrations of electrons and holes in any semiconductor, the following relation is true

$$p_i^2 = n_i^2 = np, \tag{1.4.17}$$

which is called the *mass action law*. The assumption that degree of filling of the energy levels by charge carriers is much less than unity was used in the derivation of this law. This gas of carriers is called *non-degenerate gas*, and semiconductors are called *non-degenerate semiconductors*, respectively.

Equilibrium concentration of charge carriers in defect-free semiconductor crystals. Let us first analyze the temperature dependence of concentration of intrinsic carriers, which will allow to describe temperature dependence of electrical conductivity in pure (defect-free) semiconductors. Since, there is no localized levels in the band gap, electrons can occupy levels in C-band only due to band-band transitions (see Fig. 1.4.2b). So their concentration in C-band will be determined by the known relation

$$n = N_C \exp\left(-\frac{\Delta E_g}{2k_B T}\right), \tag{1.4.18}$$

i.e. near the bottom of C-band electrons behave almost as an ideal gas, for which the kinetic energy is $E = p^2/2m$. Thus, as was shown in Section 1.4 the same concentration of free holes

$$p = N_V \exp\left(-\frac{E_g - \varepsilon_F}{2k_B T}\right) \tag{1.4.19}$$

is generated, in V-band. Here N_C and N_V are effective densities of states in C- and V-bands, correspondingly.

Multiplying the equations (1.4.18) and (1.4.19), we get another form of the law of mass action:

$$np = n_i^2 = N_C N_V \exp\left(-\frac{E_g}{kT}\right) = 4\left(\frac{kT}{2\pi^2\hbar^2}\right)^3 \left(m_p^* m_n^*\right)^{3/2} \exp\left(-\frac{E_g}{kT}\right) \qquad (1.4.20)$$

The Fermi level in an intrinsic semiconductor. Using the equality $n = p$ and relations (1.4.31) and (1.4.34), we obtain:

$$N_C \exp\left(\frac{\varepsilon_F}{kT}\right) = N_V \exp\left(-\frac{E_g + \varepsilon_F}{kT}\right). \qquad (1.4.21)$$

Hence

$$\varepsilon_F = -\frac{E_g}{2} + \frac{kT}{2}\ln\frac{N_V}{N_C} = -\frac{E_g}{2} + \frac{3}{4}kT\ln\frac{m_p^*}{m_n^*}. \qquad (1.4.22)$$

This relation means that at $T = 0$ Fermi level of an intrinsic semiconductor is located exactly in the middle of the band gap, if the effective masses of electrons and holes are the same (since for $m_p^* = m_n^*$, $\ln\left(\frac{m_p^*}{m_n^*}\right) = 0$): $\varepsilon_F(0) = -\frac{E_g}{2}$.

Substituting (1.4.22) to one of the relations (1.4.18) and (1.4.19), we obtain an expression for intrinsic carrier concentration in semiconductor

$$n_i = p_i = \sqrt{N_C N_V}\exp\left(-\frac{E_g}{2kT}\right). \qquad (1.4.23)$$

Temperature dependence of Fermi level in an intrinsic semiconductor is determined by the second term in equation (1.4.22). If the effective mass of holes in the V band is more than effective mass of electron in the C band, Fermi level is shifted with temperature closer to the bottom of the C band. In the opposite case, Fermi level moves toward the V band (Fig. 1.4.5).

For most semiconductors, hole effective mass is not much greater than the electron effective mass, and the shift of the Fermi level with temperature is insignificant. However, for example, InSb has the ratio $\frac{m_p^*}{m_n^*} \approx 10$, and band gap is small (E_g = 0.17 eV). Therefore its Fermi level is in the conduction band at

T > 450 K. At this temperature, the semiconductor passes into degenerate state and behaves like a metal.

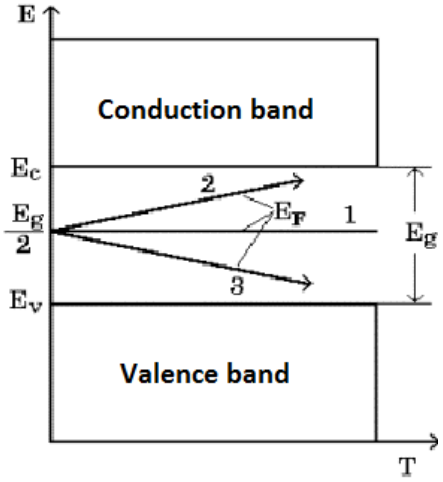


Fig. 1.4.5. Schematic view of temperature dependence of the Fermi level position in an intrinsic semiconductor: $1 - m_p^* = m_n^*$; $2 - m_p^* > m_n^*$; $3 - m_p^* < m_n^*$

Electrons and holes concentrations in doped semiconductors. All the most interesting properties of semiconductor crystals which allow to use them in devices manifest themselves in the case of doped semiconductors. Impurities, even in relatively low concentrations, strongly modify electrical, optical and other properties of semiconductors.

To estimate the equilibrium concentrations of carriers in bands in the presence of donor and acceptor levels, we have to known positions of the band edges E_V and E_C, the ionization energies (positions of impurities levels) for acceptors E_a and donors E_d, donor N_d and acceptor N_a concentrations, effective masses of electrons m_n and holes m_p in corresponding bands and temperature T.

Calculations allow us to find the carrier concentrations n and p in doped semiconductors, as well as the position of Fermi level separately (see below). However, the expressions for them, in general, are complex and sometimes have an unclear physical meaning. So, we consider only a few special cases.

If impurity levels (for definiteness by donor impurity, see Fig. 1.4.3b) correspond to E_d of about few hundredths of eV, the impurity is strongly ionized even at room temperature, and at higher temperatures it is ionized completely. In the last case, n value is almost equal to the impurity concentration N_d.

Formula (1.4.18), obtained above for intrinsic semiconductors, is true for doped semiconductor at high temperatures, when transitions band-to-band are take place. At low temperatures, part of electrons (n_d) occupy donor levels and the electron concentration satisfies the relation

$$n = \sqrt{\frac{N_d N_C}{2}} \exp\left(-\frac{E_d}{2kT}\right). \qquad (1.4.24a)$$

Similarly, for the hole-type semiconductor at low temperatures

$$p = \sqrt{\frac{N_a N_V}{2}} \exp\left(-\frac{E_a}{2kT}\right). \qquad (1.4.24b)$$

Sometimes crystals contain both donor and acceptor impurities with concentrations N_d and N_a. In this case, electrons should pass ("fall down") from donor to acceptor levels. So, if $N_d > N_a$, only $(N_d - N_a)$ donors can participate in creation of the electronic conductivity. Such semiconductors are called *partially compensated*. In this case, concentration of dominating charge carriers (electrons) at $E_d/2kT >> 1$ and not very low concentration of compensating impurity N_a looks as follows:

$$n = \frac{N_d - N_a}{2N_a} N_C \exp\left(-\frac{E_d}{kT}\right). \qquad (1.4.25)$$

Fermi level in doped semiconductor. The Fermi level position in doped semiconductors can also be found from the condition of electrical neutrality of the crystal (1.4.10). For electronic semiconductor this condition can be re-written as

$$n = p_d + p = p + (N_d - n_d), \qquad (1.4.26)$$

where N_d is concentration of donor levels, n_d – electron concentration at the donor levels. Concentration of electrons in V band is equal to the sum of concentrations of holes in the valence band and concentration of positively charged donor ions (the latter, obviously, is equal to $N_d - n_d$). So, concentration of electrons at donor level can be calculated by multiplying the concentration of these levels N_d on the function of Fermi-Dirac:

$$n_d = \frac{N_d}{1 + \exp\left(\frac{E_C - \varepsilon_F - E_d}{kT}\right)}, \qquad (1.4.27)$$

where E_d is activation energy of donor levels.

Substitution of concentrations of electrons (1.4.18) and holes (1.4.19) as well as charged donor levels (1.4.27) into the electrical neutrality condition (1.4.26) leads to the following equation with respect to Fermi level ε_F for non-degenerate electron gas:

$$N_C \exp\left(\frac{\varepsilon_F - E_C}{kT}\right) - N_V \exp\left(\frac{E_C - E_g - \varepsilon_F}{kT}\right) + \\ + \frac{N_d}{\exp\left(\frac{E_C - \varepsilon_F - E_d}{kT}\right)} = N_d \qquad (1.4.28)$$

Usually the equation (4.1.28) is not solved in general case because of its complexity. So below we are limited to reviewing special cases, for example, low or high temperatures:

$$\begin{cases} \varepsilon_F(T) = -\frac{E_d}{2} + \frac{kT}{2}\ln\frac{N_d}{2N_C} & \text{for} \quad kT << E_d \\ \varepsilon_F(T) = -\frac{kT}{2}\ln\frac{N_d}{2N_C} & \text{for} \quad kT >> E_d \end{cases} \qquad (1.4.29)\ (1.4.30)$$

From the equation (4.1.29) follows that, at absolute zero temperature, Fermi energy of electronic semiconductor is located exactly in the middle between the bottom of the conduction band and the position of donor levels. In so doing, temperature dependence of Fermi level is determined by the second term in the equation (4.1.29), which changes its sign with temperature. So, firstly, Fermi level moves toward the C band bottom with temperature increasing, and then (when impurity semiconductor is heated to high temperatures) – to the V band, as in intrinsic semiconductor (Fig. 1.4.6a). The free carriers, excited from impurity levels, are called *the majority carriers*, while carriers with opposite sign generated due to band-band transitions are *the minority carriers*. Similarly, we can obtain the expressions for temperature dependences of Fermi level in p-type semiconductors:

$$\begin{cases} \varepsilon_F(T) = -\frac{E_a}{2} + \frac{kT}{2}\ln\frac{N_a}{2N_V} & \text{for} \quad kT << E_a \\ \varepsilon_F(T) = -\frac{kT}{2}\ln\frac{N_a}{2N_V} & \text{for} \quad kT >> E_a \end{cases} \qquad (1.4.31)\ (1.4.32)$$

Schematic view of the temperature dependences of Fermi level in doped semiconductors is schematically shown in Fig. 1.4.6.

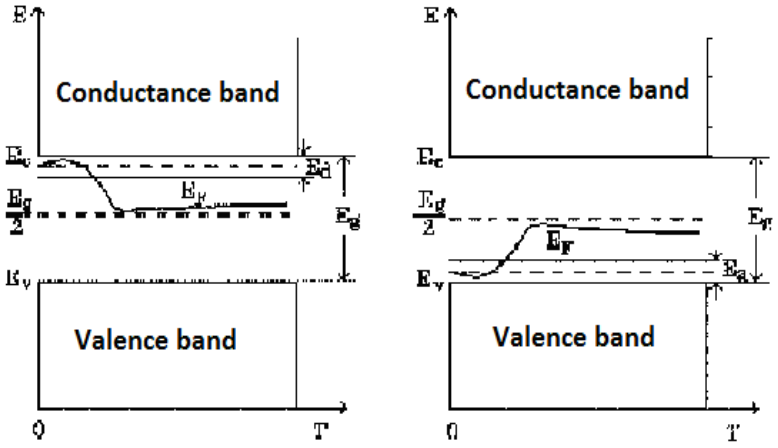


Fig. 1.4.6. Schematic view of temperature dependence of the Fermi level position in n-(*a*) and p-type (*b*) semiconductors

1.4.3. Estimation of gap energies, impurity ionization energies, concentration and mobility of carriers in semiconductors

Note that temperature dependence of carrier concentration in the intrinsic and doped (extrinsic) semiconductors are obtained from measurements of the Hall constant R_x, which is inversely proportional to the concentration of charge carriers

$$R_x = \frac{r}{ne} \text{ or } R_x = \frac{r}{pe}. \tag{1.4.33}$$

If we plot the $n(T)$ dependencies in Arrhenius scale, we obtain the curves shown schematically in Fig. 1.4.7. Here straight line 1 will give the energy of impurities ionization (because its slope is equal to E_d), and the line 3 – the band gap E_g in accordance with Eqs. (1.4.23) and (1.4.24).

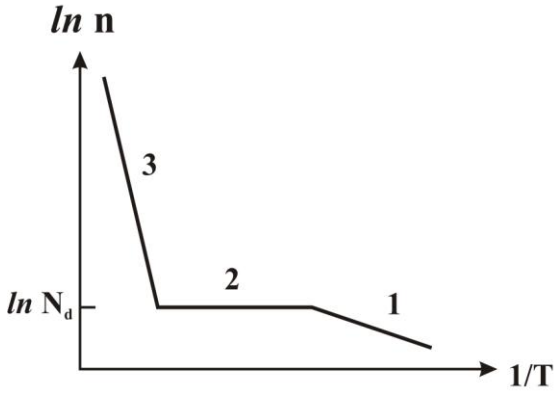


Fig. 1.4.7. Schematic view of temperature dependence of electron concentration in the C-band for the n-type semiconductor:*1* – impurity-band transitions, *2* – depletion region of impurities, *3* – band-band transitions

At the intermediate temperature range, when all donors are fully ionized, so that $n = N_d$ and the band-band transitions are still impossible, $n(T)$ dependence is saturated (see horizontal part 2 of the curve in Fig. 1.4.7). This temperature range is called the region of *impurities depletion*.

If semiconductor contains acceptor impurities, then impurity-band transitions dominate at low temperatures and hole concentration p in the V-band is given by the Eq. (1.4.24b), where energy ionization E_a of acceptor levels is used instead of E_d and is counted from the top of V-band. The function $p(T)$ has the same shape as the dependence $n(T)$ in Fig. 1.4.7. The horizontal part 2 of these curve is given now by concentration of acceptors N_a.

In the case of Ge and Si, E_g values are 0.68 and 1.2 eV, respectively, and a typical depths of the donor and acceptor levels lies lie at 0.01-0.05 eV, so at room temperature these impurities are completely ionized. Typically, the concentration of impurity atoms in semiconductors, do not exceed the 10^{16}-10^{17} cm^{-3}, which is much higher than the electron density, which can create inter-band transitions at the same temperature (it will be 10^{13} cm^{-3} in the case of silicon at $T = 300$ K).

If the impurity concentration is high enough, impurity atoms begin to interact with each other because of the overlapping of their wave functions. In this case, there is a possibility for tunneling ("hopping") of the charge carriers between the impurities. As a result, the discrete impurities levels begin to form the so-called *impurity band* in the band gap, in which the carriers can move by jumps, although with significantly lower mobility than in the area of allowed states.

At a sufficiently high impurity concentrations (approximately 10^{17} cm^{-3} for germanium and 10^{18} cm^{-3} for silicon), the impurity region may merge with the allowed band, so that the bandgap decreases. In such a heavily doped semiconductor, boundaries of these bands becomes blurred, so that at the edges of the allowed bands the so-called "*tails of the localized states*" are appeared. Besides, the random variations in the impurities concentration by the crystal can result to the formation of the so-called random *large-scale potential relief*.

Let us compare the experimentally observed temperature dependence of conductivity $\sigma(T) = n(T) \cdot e \cdot \mu(T)$ in semiconductors with the theoretical relations for $n(T)$ and $\mu(T)$ (see above). Schematic image of the experimental temperature dependence of the resistivity for extrinsic semiconductor is shown in Fig. 1.4.8.

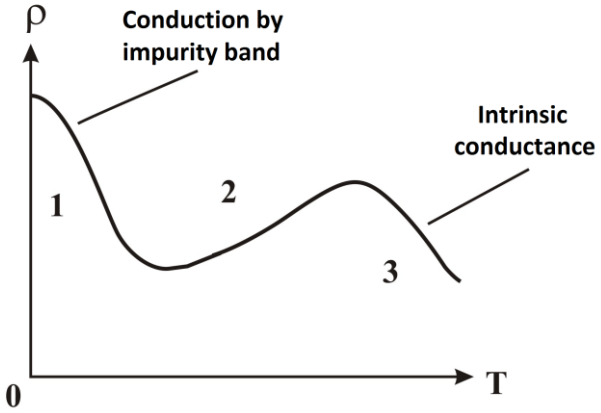


Fig. 1.4.8. Temperature dependence of the electrical conductivity for extrinsic semiconductor

Analysis shows, that with the temperature increasing when $k_B T << E_g$, the conductivity must grow exponentially with the growth of the carriers number due to their transitions between the impurities levels and the allowed area. In other words, in the Arrhenius scale of this dependence is linearized (similar to part 1 on Fig. 1.4.7 above), allowing to specify the value of the ionization energy of the impurity $E_d(E_a)$.

Note that the role of the temperature dependence of mobility is low, until all the impurities will not be ionized. At the intermediate temperature region, where impurities depletion occurs and $n(T) \approx$ const, the carrier mobility, as mentioned above, decreases with increasing temperature (due to phonon scattering) and the semiconductor resistance will increase slightly (conductivity decreases) with temperature growth (part 2 of the curve in Fig. 1.4.8).

With the further increase in temperature, when the condition $kT >> E_g$ will be approached, transitions of electrons from the V-band to the C-band begin to play the growing role. As a result, the concentration of charge carriers will grow exponentially according to Eq. (1.4.38) as for intrinsic semiconductors. Then, according to the calculations, the temperature dependence of intrinsic conduction (resistance) in Arrhenius coordinates will also be linearized, and the slope of the straight line 3 in Fig. 1.5.7 will correspond to the band gap E_g.

1.5. Generation and recombination of charge carriers

(non-equilibrium processes in semiconductors)

1.5.1. Basic concepts and definitions for the physics of non-equilibrium processes in semiconductors

In the previous section 1.3, we discussed semiconductor crystals containing equilibrium electrons and holes. Concentration of such electrons and holes are uniquely determined by the position of Fermi level. If we know this position we can always determine distribution of *the equilibrium charge carriers* in the C band, in the V band and by impurity levels in the band gap. Thus, Fermi level (Fermi energy), as the normalization parameter, describes the system of charged carriers as a whole.

The process of generation of free electrons and/or holes in semiconductor crystal requires energy to overcome the energy gap E_g between the allowed bands (for intrinsic semiconductor) or the ionization energy (E_a or E_d) between allowed band and the local levels in the forbidden band (for doped semiconductor). Under equilibrium conditions, this energy is taken from the crystal thermal energy. Free charge carriers appearing due to thermal generation and being in thermal equilibrium with the lattice are called an *equilibrium charge carriers (ECC).*

Simultaneously with the generation of free charge carriers, the reverse process takes place (due to collisions of electrons and/or holes with the crystal lattice ions and defects), which is called *recombination*. In this process the electrons in C-band return back to empty states of the V-band or on local levels in the gap resulting in the disappearance of electron-hole pairs or neutralization of impurities. Processes of generation and recombination are fully balanced at thermodynamic equilibrium.

It is known that various external impacts (electric and magnetic fields, pressure, irradiations, etc.) disturb the thermodynamic equilibrium of electron and ion subsystems in the crystal. However, this disturbance does not always lead to the change in the charge carriers concentration. For example, applying

not very high electric field to the homogeneous crystal does not change carrier concentration in comparison with those in the equilibrium state. Indeed, the energy distribution of charge carriers in the crystal while the electric current flows differs from the equilibrium distribution given by the Fermi function $f(E,T)$, see above in Section 1.3.

However, in addition to the equilibrium charge carriers formed due to thermal generation, charge carriers can occur due to such external impacts as lighting of a crystal, its bombardment by electrons and other high-energy particles, the current flow through the semiconductors-semiconductor or metal-semiconductor junctions, application of strong electric fields, and some other reasons.

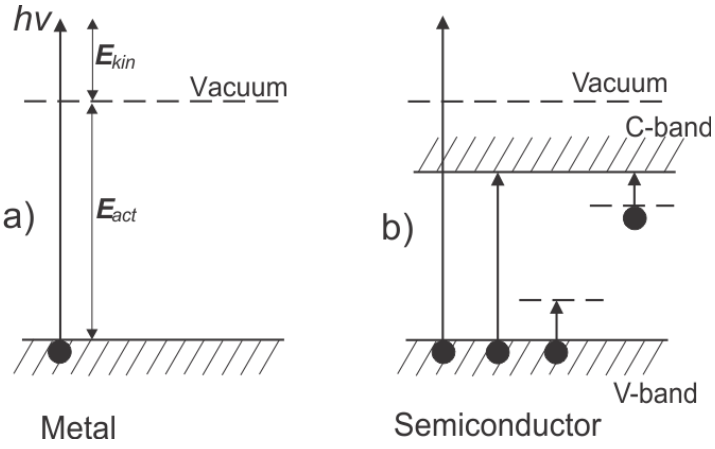


Fig. 1.5.1 Schemes of free electrons and/or holes generation in a metal (a) or semiconductor (b) due to internal or external impacts

Let us consider the case of lighting of a semiconductor crystal by monochromatic light with quantum energy $h\nu > E_g$. In 1905, Albert Einstein formulated the equation of the *external photoelectric effect* (photoemission)

$$h\nu = E_{act} + \frac{m\nu^2}{2} = E_{act} + E_{kin}, \qquad (1.5.1)$$

where E_{act} is the work function of an electron in a vacuum. Metals have the lowest work function values (1.1-2.5 eV).

However, for semiconductors, there is another possibility to make free the excited electron by means of its transition from the V- to C-band if $h\nu > E_{act}$ (Fig. 1.5.1, *b*). In this case electron becomes free within the crystal so that this process is called the *internal photoelectric effect*. It is possible to write an

equation similar to (1.5.1) for the internal photoelectric effect. The role of the activation energy can be played by the E_g value, the ionization energy of donors or acceptors, and capture of an electron by empty donor or transition of the electron from the acceptor to the C-band. For all such processes is characteristic that $h\nu > E_{act}$.

An important result of the above description is the fact that added external energy is spent when atoms are ionized due to the process of optical generation of free charge carriers, and the excessive energy is carried away with the free carriers. Thus, we have a clear breaking of thermal equilibrium between the electronic subsystem and the crystal lattice: the energy of the electrons increases, while the mean energy of the lattice remains unchanged. Such free charge carriers which appears as the result of external impacts in the crystal (not due to thermal generation) are called *non-equilibrium charge carriers (NECC)*.

Obviously, thermal equilibrium between the electron and lattice subsystems must be restored after lighting switching-off. It should be noted that, as a rule, the concentration of generated NECCs is substantially less than the total concentration of electrons, and the latter is less than the concentration of ions. Besides, the mass of an electron is almost 2000 times smaller than the mass of the ion. Thus, the energy stored by the NECCs is small compared with the thermal energy of the lattice. As a result, the relaxation (establishment of equilibrium between the lattice and electron subsystems) is reduced to the fact that the NECCs recombine but the temperature of the crystal lattice remains unchanged. Take this account, one can assume that application and removal of excitation changes the NECC concentration, but does not change the ECC concentration. Thus, the total concentration of free charge carriers is equal to the simple sum of equilibrium and non-equilibrium carriers:

$$\begin{aligned} n &= n_0 + \Delta n, \\ p &= p_0 + \Delta p. \end{aligned} \tag{1.5.2}$$

We shall often use the concept of *excitation level*, which is determined as the ratio of the equilibrium and non-equilibrium carriers concentrations. Let us define the condition $\Delta n << n_0$ as low excitation level (LEL) and $\Delta n >> n_0$ – as high excitation level (HEL).

The statistics of equilibrium concentrations of electrons and holes was described in Section 1.4.2. For non-degenerate carriers in a crystal, obeying Maxwell-Boltzmann statistics,

$$n_0 = N_C \exp\left(\frac{\varepsilon_F}{kT}\right) \tag{1.5.3}$$

and

$$p_0 = N_V \exp\left(-\frac{E_g + \varepsilon_F}{kT}\right). \tag{1.5.4}$$

Obviously, Fermi level does not reflect the concentration of electrons and holes in the crystal in case of NECCs generation. Moreover, the concentrations of non-equilibrium electrons and holes are not connected together. Indeed, we can change freely and independently the concentration of non-equilibrium electrons and holes by changing excitation character, so that a single Fermi level can not determine *n* and *p* values simultaneously.

To describe this case, William Shockley introduced the concept of the so-called *quasi-Fermi levels*, which are different for non-equilibrium electrons (φ_n) and holes (φ_p) and are independent from each other. They were formally introduced in such a way, that the concentrations of electrons and holes completely coincide with Eqs. (1.5.3) and (1.5.4). Then, we can determine the quasi-Fermi levels in form of the following equations

$$n = N_C \exp\left(\frac{\varphi_n}{kT}\right), \tag{1.5.5}$$

$$p = N_V \exp\left(-\frac{E_g + \varphi_p}{kT}\right). \tag{1.5.6}$$

Note that, in contrast to formulas for the ECCs (1.5.3) and (1.5.4), in Eqs. (1.5.5) and (1.5.6) the positions of quasi-Fermi levels are determined separately by concentrations of electrons and holes.

However, one should remember that the Fermi level and the quasi-Fermi levels are interconnected. To find this relationship, we obtain the expressions which contain the Fermi level and the quasi-Fermi levels simultaneously. From (1.5.5) and (1.5.6), excluding N_C and N_V, one can obtain:

$$\varphi_n = \varepsilon_F + kT \ln\frac{n}{n_0}, \tag{1.5.7}$$

$$\varphi_p = \varepsilon_F - kT \ln\frac{p}{p_0}. \tag{1.5.8}$$

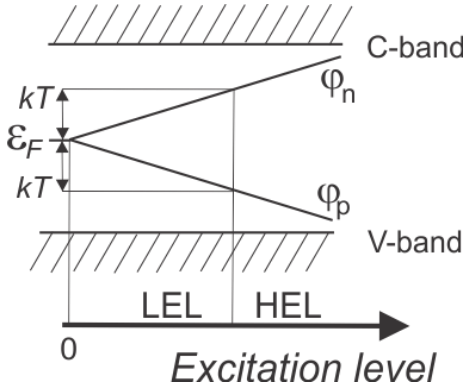


Fig. 1.5.2. Shifts of quasi-Fermi levels for electrons and holes relative to the equilibrium Fermi level

As can be seen, the quasi-Fermi level for electrons is shifted up and for holes is shifted down with respect to the Fermi level in the unexcited crystal (see, Fig. 1.5.2). This shift value for the quasi-Fermi levels can be related with the excitation level. Indeed

$$\varphi_n = \varepsilon_F + kT \ln \frac{n}{n_0} = \varepsilon_F + kT \ln \frac{n_0 + \Delta n}{n_0} = \varepsilon_F + kT \ln \left(1 + \frac{\Delta n}{n_0} \right).$$

Expanding the logarithm in the series $\left| \ln \left(1 + \frac{\Delta n}{n_0} \right) \approx \frac{\Delta n}{n_0} \right|$ results in

$$\varphi_n \approx \varepsilon_F + kT \frac{\Delta n}{n_0}. \qquad (1.5.9)$$

The analysis of this expression shows that for low excitation levels, when $\Delta n << n_0$, one has $|\varphi_n - \varphi| < kT$, whereas for high excitation levels $|\varphi_n - \varphi| > kT$.

1.5.2. Stationary and non-stationary processes. The lifetimes for NECCs

We have introduced above the concepts of *generation* and *recombination* of charge carriers, which describe the processes of the charge carriers appearance and disappearance. It is important to understand in this connection, that the concentrations of NECCs and ECCs are only a consequence of these two processes. Let us introduce the following definitions:

- *rate of generation* of charge carriers g [$cm^{-3}s^{-1}$] as the number of carriers generated per unit time in unit volume of the material;

- *rate of recombination* of charge carriers as the number of charge carriers r [$cm^{-3}s^{-1}$] recombined per unit time in unit volume of the material.

It is necessary to make a remark. Generally speaking, it is necessary to distinguish two processes of generation: *thermal generation* g_0 (i.e. thermodynamically equilibrium) and *inequilibrium generation* g. The total generation rate is

$$G = g_0 + g. \tag{1.5.10}$$

The same could be said for the total recombination rate:

$$R = r_0 + r. \tag{1.5.11}$$

However, in experiments, the generation rate can be separated on equilibrium and non-equilibrium contributions (because we can switch off the non-equilibrium generation). At the same time, the recombination rate cannot be separated in such a way because one cannot discriminate equilibrium or non-equilibrium charge carriers.

After introducing the concepts of generation and recombination rates of charge carriers, one can now write the equations of balance between these processes, i.e. the equations of the generation-recombination processes:

$$\begin{aligned} G_n - R_n &= g_n + g_{0n} - r_n - r_{0n} = g_n - r_n = \frac{dn}{dt}; \\ G_p - R_p &= g_p + g_{0p} - r_p - r_{0p} = g_p - r_p = \frac{dp}{dt}. \end{aligned} \tag{1.5.12}$$

In this equation, since the generation and recombination are related only to non-equilibrium processes, it is more correct to write the derivatives (rates) only for NECCs (although this is not so principal), i.e.

$$\begin{aligned} \frac{dn}{dt} &= \frac{d\Delta n}{dt} = g_n - r_n; \\ \frac{dp}{dt} &= \frac{d\Delta p}{dt} = g_p - r_p. \end{aligned} \tag{1.5.13}$$

At the beginning, from general considerations, let us try to determine the value of r. Let a non-equilibrium electron is generated in the crystal at the moment $t = 0$. It will move in the crystal until it meets a hole and an act of recombination will take place. In general physics there is the formula for the length, which the electron passes until its collisions (interaction) with another particle (hole)

$$S_n = \frac{1}{\pi r^2 p}. \tag{1.5.14}$$

As is seen, it is inversely proportional to the concentration p of holes and their cross sections. In our case, the cross-sectional area will be replaced by the "interaction cross section" q. This is a broader interpretation of this parameter because it takes into account the interaction determined by the Coulomb attraction of the electron and hole and the fact that a moving electron is not a material point, because it has a finite size. Therefore, we shall understand the "interaction cross section" as a certain spatial region in which an electron can recombine with a hole, hence

$$S_n = \frac{1}{q_p p}. \tag{1.5.15}$$

On the other hand, the path passed by an electron before the interaction with a hole is the product of the velocity of an electron on the average time between the generation and recombination of free electron, i.e. $S_n = v_n \tau_n$.

From a comparison of the last two expressions one can obtain $v_n \tau_n = \frac{1}{q_p p}$ or

$$\tau_n = \frac{1}{p q_p v_n} = \frac{1}{\gamma_n p}. \tag{1.5.16}$$

The average time between the acts of electron generation and recombination is called the *electron lifetime* τ_n. A similar analysis can be made for the holes, so that

$$\tau_p = \frac{1}{n q_n v_p} = \frac{1}{\gamma_p n}. \tag{1.5.17}$$

In Eqs. (1.5.16) and (1.5.17), γ_n and γ_p are called *the recombination coefficients* for electrons and holes, respectively.

Let us the crystal with several kinds of carriers of one sign, for example, holes of k types: free holes, holes localized on impurities of type A, holes localized on impurities of type B. In this case, there are the alternative processes for recombination of a free electron, moving through the crystal: with a free hole, with a bound hole on A-type or with a bound hole on B-type impurities. Then the electron lifetime relative to recombination with a hole of k-type is:

$$\tau_{nk} = \frac{1}{p_k q_{pk} v_n}.$$

Effective lifetime, measured in the experiment, which takes into account the possibility of electrons recombination with all types of holes can be found from the following considerations. If τ_{nk} is the time between the generation and recombination of an electron with a hole of k-type, then $1/\tau_{nk}$ value determines the number of recombination events for the electron per unit time (i.e. one

electron could recombine $1/\tau_{nk}$ times for the unit time). If there are several possible recombination channels, the total number of recombination events will equal to the sum of recombination processes in each channel. Then the effective lifetime including all the channels will take form $\sum_k p_k q_{pk} v_n = \sum_k \tau_{nk}^{-1} = \left(\tau_{eff}\right)^{-1}$.

In other words,

$$\frac{1}{\tau_{eff}} = \sum_k \frac{1}{\tau_{nk}}, \tag{1.5.18}$$

i.e., inverse times for each process should be summarized when considering the complex processes of recombination.

We can express the recombination rate r, appearing in Eq. (1.5.4), through the parameter τ. Indeed, if $1/\tau_n$ defines the number of recombination events per unit time and per electron, the total number of recombination events per unit volume per unit time is n/τ_n. Thus, Eq. (1.5.13) can be transformed into

$$\begin{aligned} \frac{dn}{dt} &= G_n - R_n = G_n - \frac{n}{\tau_n} = G_n - \gamma_n np; \\ \frac{dp}{dt} &= G_p - R_p = G_p - \frac{p}{\tau_p} = G_p - \gamma_p np. \end{aligned} \tag{1.5.19}$$

The recombination term γnp in Eq. (1.5.19) for the balance between electrons and holes, which describes the rate of recombination as the product of the concentrations of electrons and holes, can be interpreted as follows: only one electron and one hole are involved in the elementary recombination act.

Let us obtain one more modification of the recombination term in Eq. (1.5.19), taking into account that we cannot directly specify the recombination rate of NECCs only. However, we know the recombination rate for all carriers as γnp contribution. Besides, we know that in equilibrium state $r_0 = \gamma n_0 p_0$

Hence $r = R - r_0 = \gamma np - \gamma n_0 p_0$, so that Eq. (1.5.13) will get the form

$$\begin{aligned} \frac{d\Delta n}{dt} &= g_n - \gamma_n (np - n_0 p_0); \\ \frac{d\Delta p}{dt} &= g_p - \gamma_p (np - n_0 p_0). \end{aligned} \tag{1.5.20}$$

If trapping centers are lack (opposite case we discuss later) and NECCs in the crystal are generated only due to the band-band transitions, their recombination is also carried out only due to the band-band transitions. Then, at any time the neutrality condition $\Delta n = \Delta p$ will be satisfied and the Eq. (1.5.20) is transformed into relation

$$\frac{d\Delta n}{dt} = g_n - \gamma_n\left[(n_0+\Delta n)(p_0+\Delta p) - n_0 p_0)\right] =$$
$$= g_n - \gamma_n\left[n_0 p_0 + n_0\Delta n + p_0\Delta n + (\Delta n)^2 - n_0 p_0\right] = \quad (1.5.21)$$
$$= g_n - \gamma_n \Delta n(n_0 + p_0 + \Delta n).$$

Similarly,

$$\frac{d\Delta p}{dt} = g_p - \gamma_p \Delta n(n_0 + p_0 + \Delta n). \quad (1.5.22)$$

1.5.3. The relaxation of non-equilibrium conductivity

In this section, we explore some characteristics of such a fundamental concept as the lifetime of NECCs, to discuss the photoconductivity phenomenon as an example. Additional conductivity, arising in the crystal under light irradiation with photon energies sufficient for the appearance of NECCs, is called *the photoconductivity*. These NECCs may have an initial kinetic energy much greater than the energy of ECCs (~ kT). If we irradiate a silicon crystal with green light ($\lambda_{green} \sim 0.55$ μm), then $h\nu_{green} = \frac{1.24}{\lambda_{green}} \approx 2.2\,\text{eV}$, i.e. twice more than E_g value in Si. Thus the optically generated NCCs will have the kinetic energy of the order of 1.1 eV, while the equilibrium carriers will have the kinetic energy of about $kT = 26\,\text{meV}$. This means that the excess energy of the NCCs is higher than 1 eV, if compare with equilibrium carriers. Estimations show that this energy is dissipated within about 1,000 collisions which occur during ~10^{-10} s. The lifetime value for NCCs, as we will show below, is much higher (~ 10^{-8} -10^{-4} s). Hence, the generated hot NECCs relax rapidly to the bottom of the allowed bands and become indistinguishable from the ECCs.

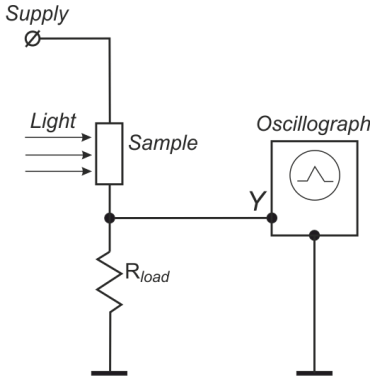


Fig. 1.5.3. The experimental scheme to determine lifetimes of carriers by the photoconductivity decay

Let us consider the electrical circuit presented in Fig. 1.5.3, which allows to determine lifetimes of carriers by decay of photoconductivity excited by lighting. In the absence of light the dark current

$$I_{dark} = \frac{U}{R_{sample}^{dark} + R_{load}} \tag{1.5.23}$$

flows through the circuit with full resistance (R_{sample} + R_{load}), where U is supplied voltage, R_{sample}^{dark} – dark electric resistance of the sample, R_{load} – load resistance. Resistance of the sample in the dark is

$$R_{sample}^{dark} = \frac{l}{S\sigma_0} = \frac{l}{Se(n_0\mu_n + p_0\mu_p)}, \tag{1.5.24}$$

where S is the cross section area of the sample, l – length.

When the light is switch on, the resistance of the sample

$$R_{sample}^{light} = \frac{l}{S(\sigma_0 + \Delta\sigma)} = \frac{l}{Se\left[(n_0 + \Delta n)\mu_n + (p_0 + \Delta p)\mu_p\right]} \tag{1.5.25}$$

is reduced due to the appearance of additional NECCs. Light current in the circuit with the resistance (R_{sample} + R_{load}) increases and becomes equal to

$$I_{light} = \frac{U}{R_{sample}^{light} + R_{load}}. \tag{1.5.26}$$

The load voltage drop $V = IR_{load}$ is applied for the input terminals of the Y-sweep of the oscilloscope. Current change in the circuit when switching of the light (with light flux modulation) will lead to a change in the input oscilloscope voltage.

Let us set a rectangular modulation of the light flux (light will be instantly switched on and off at regular intervals). In order to get the voltage drop V at the oscilloscope input to be proportional to the NECCs concentration, i.e. $\Delta V = V_{light} - V_{dark} \sim \Delta n$, it is necessary to fulfill certain conditions. Indeed,

$$\Delta V = V_{light} - V_{dark} = (I_{light} - I_{dark})R_{load} =$$

$$= \left(\frac{1}{R_{sample}^{light} + R_{load}} - \frac{1}{R_{sample}^{dark} + R_{load}}\right) UR_{load} =$$

$$= \left(\frac{R_{sample}^{dark} - R_{sample}^{light}}{(R_{sample}^{light} + R_{load})(R_{sample}^{dark} + R_{load})}\right) UR_{load}$$

If the condition

$$R_{load} << R_{sample}^{light}, R_{sample}^{darh} \tag{1.5.27}$$

is fulfilled, then

$$\Delta v \approx \frac{R_{sample}^{dark} - R_{sample}^{light}}{R_{sample}^{light} \cdot R_{sample}^{dark}} = \frac{\dfrac{l}{S\sigma_0} - \dfrac{l}{S(\sigma_0 + \Delta\sigma)}}{\dfrac{l}{S(\sigma_0 + \Delta\sigma)} \cdot \dfrac{l}{S\sigma_0)}} = \frac{S}{l}\Delta\sigma \sim \Delta\sigma \tag{1.5.28}$$

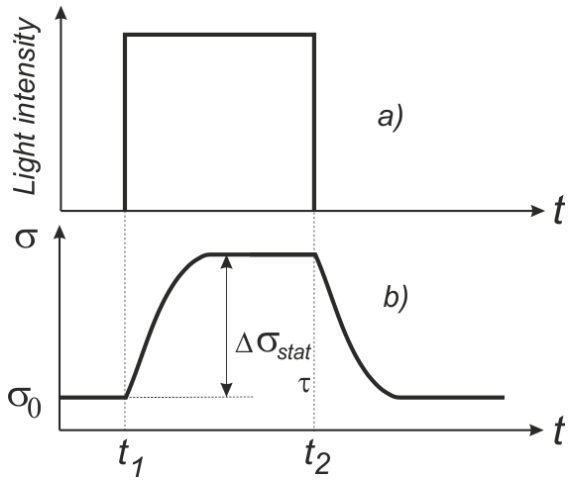


Fig. 1.5.4. Time dependences of optical excitation (*a*) and conductivity (*b*)

Let us consider the processes of growth and decay of photocurrent (Fig. 1.5.4, b) when the light is switching (Fig. 1.5.4, a). Until time t_1, there is only the dark conductivity of the sample, so that $\Delta n = 0$. After switching on the light, the carrier concentration will increase gradually due to the generation process. However, the concentration of NECCs cannot increase infinitely, reaching maximal value (*the stationary concentration*) in the steady state, which is determined by the light intensity (and therefore the *g* value).

To find the form of the time dependence (kinetics) of NECCs concentration, it is necessary to solve the generation-recombination equations with some initial and boundary conditions. We assume that in our case $h\nu > E_g$ and concequently the absorption of a photon leads to the generation of both electrons and holes, i.e. $\Delta n = \Delta p$. We need to solve the generation-recombination equation

$\frac{d\Delta n}{dt} = g - \frac{\Delta n}{\tau_n}$ for the carriers of certain sign, for example, for electrons for two limiting excitation levels.

In the case of low-level excitation, $\tau_n = \frac{1}{\gamma_n(p_0 + \Delta p)} \approx \frac{1}{\gamma_n p_0} = const$ and does not depend on the excitation level (NECCs concentration). In this case, the generation-recombination equation is a homogeneous linear differential equation. Its solution splits into two separate solutions with the increasing and decreasing branches.

For the increasing branch ($g \neq 0$) we obtain relation

$$\Delta n = g\tau(1 - e^{-\frac{t}{\tau}}), \tag{1.5.29}$$

which shows that, if $t \to \infty$, then $\Delta n \to g\tau = \Delta n_{stat}$. For the decreasing branch ($g = 0$)

$$\Delta n = g\tau e^{-\frac{t}{\tau}}. \tag{1.5.30}$$

Thus, an important result in this case is that the NECC lifetime in such process is constant. It is obvious that the measurement of photoconductivity kinetics allows to estimate the τ value. Indeed, from the analysis of the decreasing branch of the photoconductivity signal detected on oscilloscope (see Fig. 1.5.4b), we can conclude: if the condition (1.5.31) is fulfilled, the voltage signal V on an oscilloscope will be proportional to the conductivity σ, which, in turn, is proportional to the concentration of charge carriers, because $\sigma = en\mu$.

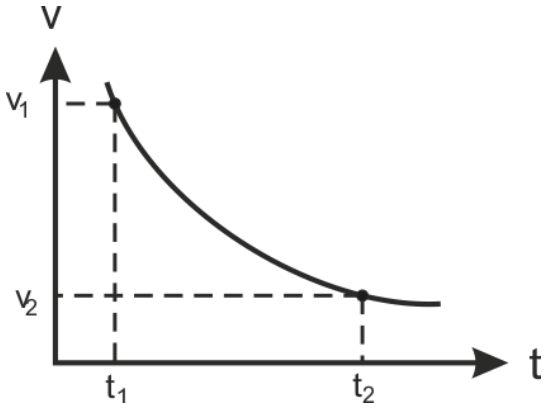


Fig. 1.5.5. The scheme for estimation of the carriers lifetime using the curve of photoconductivity decay

Thus, selecting two points with $V_1(t_1)$ and $V_2(t_2)$ coordinates on the decay curve of the photoconductivity signal in Fig. 1.5.5, one can write: $V_1 \sim n_1 = A\exp(-\frac{t_1}{\tau});$ where A is the proportionality factor. Then, dividing the $V_2 \sim n_2 = A\exp(-\frac{t_2}{\tau}),$
first equation by the second, we eliminate the A and obtain relation

$$\frac{V_1}{V_2} = \exp\left(-\frac{t_2 - t_1}{\tau}\right), \tag{1.5.31}$$

whence

$$\tau = \frac{t_2 - t_1}{\ln\frac{V_1}{V_2}}. \tag{1.5.32}$$

Thus, to obtain the correct τ value, we need to measure correctly the time interval between the points t_1 and t_2 and to find the ratio of volt signals for these points. In the case of high-level excitation, $\tau_n = \frac{1}{\gamma_n p} = \frac{1}{\gamma_n(p_0 + \Delta p)} \approx \frac{1}{\gamma_n \Delta p}$ is no longer a constant since Δp depends on the level of excitation (intensity of light). Let us write the generation-recombination equation for electrons as follows $\frac{d\Delta n}{dt} = g - \gamma_n np$. If we consider the NECCs generation at the lighting as the interband process, then $\Delta n = \Delta p$, so that

$$\frac{d\Delta n}{dt} = g - \gamma_n(\Delta n)^2, \tag{1.5.33}$$

i.e. recombination is proportional to the square of the NECC concentration. This is called *a quadratic recombination*.

The solution of equation (1.5.33) has the form

$$\Delta n = \sqrt{\frac{g}{\gamma}} th(t\sqrt{\gamma_n g}) \tag{1.5.34}$$

for the increasing branch, where hyperbolic tangent is $\left| th(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \right|$.

For the decreasing branch we have

$$\Delta n = \sqrt{\frac{g}{\gamma_n}} \frac{1}{t\sqrt{\gamma_n g} + 1}. \tag{1.5.35}$$

It is easy to see, that for $t \to \infty$ the hyperbolic tangent becomes equal to 1 and $\Delta n_{stat} = \sqrt{\frac{g}{\gamma}}$.

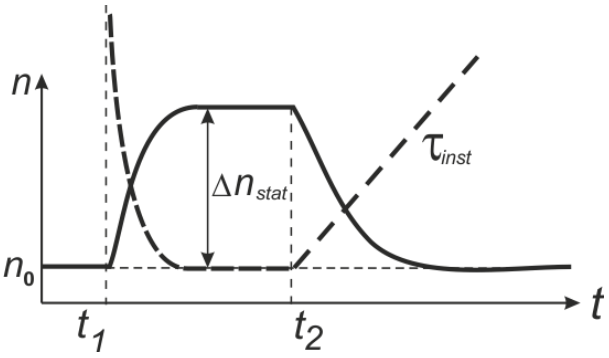


Fig. 1.5.6. The change of instant lifetime and photoconductivity for the case of squared recombination

If we introduce an instant lifetime and define it as $\tau_{inst} = \frac{1}{\gamma_n \Delta p}$, the curves of instant lifetime will have the form shown in Fig. 1.5.6. Let us introduce a more general definition of instant lifetime.

We assume that the differential equations $\frac{dn}{dt} = g_n - \frac{n}{\tau_n}$ and $\frac{dp}{dt} = g_p - \frac{p}{\tau_p}$ have a sense if τ_n and τ_p are instant lifetimes of electrons and holes. Then

$$\tau_{inst} = \frac{\Delta n}{g - \frac{dn}{dt}}. \qquad (1.5.36)$$

It is easy to show that in the case of a low-excitation level, such as for increasing branches, $\Delta n = g\tau(1 - e^{-\frac{t}{\tau}})$ we obtain relation $\frac{d(\Delta n)}{d\tau} = g\tau e^{-\frac{t}{\tau}} \cdot \frac{1}{\tau} = g e^{-\frac{t}{\tau}}$.

Then $\tau_{inst} = \frac{\Delta n}{g - \frac{d\Delta n}{dt}} = \frac{g\tau(1 - e^{-\frac{t}{\tau}})}{g - g e^{-\frac{t}{\tau}}} = \tau$.

Stationary lifetimes can be easily calculated, if we set $\frac{d\Delta n}{dt} = 0$. Then $\tau_{stat} = \frac{\Delta n_{stat}}{g}$, so that for the case of linear recombination we obtain $\tau_{stat} = \tau_{inst} = const$ and in the case of quadratic recombination

$$\tau_{stat} = \frac{\Delta n_{stat}}{g} = \frac{\sqrt{\frac{g}{\gamma}}}{g} = \frac{1}{\sqrt{\gamma g}} \sim \frac{1}{\sqrt{g}}.$$

In the case of linear recombination, the NECC concentration is proportional to the rate of NECC generation $\Delta n_{stat} = g\tau$, while in the case of quadratic recombination $\Delta n_{stat} = \sqrt{\frac{g}{\gamma}} \sim \sqrt{g}$.

This allows to determine the type of recombination processes in the crystal from the shape of NECC concentration or photocurrent dependences on the exitation level (intensity of light).

1.6. Recombination processes in semiconductors

1.6.1. Concept of recombination of non-equilibrium charge carriers

As was mentioned in Section 1.1.4, free charge carriers in semiconductors arise from the processes of generation which can be equilibrium and non-equilibrium. Equilibrium generation is a process in which electrons and holes are formed due to thermal transitions over the band gap or from local impurity levels to the allowed bands. Recombination is the reverse process of carrier generation. The recombination rate R is measured in the same units as the rate of generation, i.e. [$cm^{-3} \cdot s^{-1}$].

In addition to the equilibrium (thermal) generation of free electrons and holes, there are other ways of their occurrence: optical generation, impact ionization, injection, etc. All these methods differ primarily in sources of energy required to create free electrons and holes. Similarly, the main question that determines the type of recombination process is where the energy of recombining electron-hole pair will be transferred to, i.e. basic laws describing the recombination process (and generation, too) are the laws of conservation of energy and momentum.

1.6.2. Band-to-band radiative recombination

Radiative recombination mechanism is based on the fact that energy and momentum of recombining pair transfers to emitted photon.

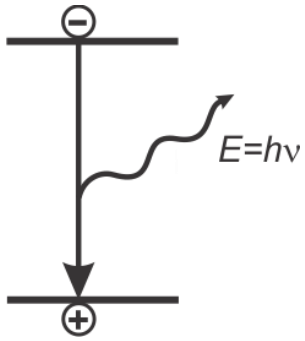


Fig. 1.6.1 Scheme of radiative recombination

Radiative recombination rate is proportional to the concentration of electrons and holes, i.e.

$$R_{rad} = \gamma_{rad} np,$$

where γ_{rad} is radiative recombination coefficient. At thermodynamic equilibrium, when there are only thermal generation and recombination transitions equal to each other, we have

$$R_{0rad} = \gamma_{rad} n_0 p_0.$$

Then

$$\gamma_{rad} = \frac{R_0}{n_0 p_0}.$$

As has been shown previously in (1.5.39)

$$\tau_n = \frac{\Delta n}{g_n - \frac{dn}{dt}} \qquad \text{and} \qquad \tau_p = \frac{\Delta p}{g_p - \frac{dp}{dt}}.$$

And since

$$\frac{d\Delta n}{dt} = g_n - \gamma_{rad} \Delta n (np - n_0 p_0)$$

and

$$\frac{d\Delta p}{dt} = g_p - \gamma_{rad} \Delta p (np - n_0 p_0),$$

then

$$\tau_n = \tau_p = \frac{1}{\gamma_{rad}(n_0 + p_0 + \Delta n)} = \frac{n_0 p_0}{R_0(n_0 + p_0 + \Delta n)}. \qquad (1.6.1)$$

It is of interest to find the dependence of the charge carriers lifetime on the doping level of crystal. At high temperature, when the impurity is ionized, doping level determines concentration of equilibrium electrons (in the case of donor impurity) or holes (in the case of acceptor impurity). At low excitation level, when the concentration of non-equilibrium carriers is less than the equilibrium carriers concentration, the dependence of charge carriers lifetime on doping level becomes $\tau = f(n_0)$ or $\tau = f(p_0)$. At low excitation level relation (1.6.1) will have the form

$$\tau_n = \tau_p = \frac{1}{\gamma_{rad}(n_0 + p_0)} = \frac{n_0 p_0}{R_0(n_0 + p_0)}. \qquad (1.6.2)$$

For intrinsic semiconductor, when $n_0 = p_0 = n_i$, lifetime is

$$\tau_i = \frac{1}{\gamma_{rad} 2n_i}.$$

We should take into account that

$$\tau_n = \tau_p = \frac{2n_i\tau_i}{n_0 + p_0},$$

or

$$\frac{\tau}{\tau_i} = \frac{2n_i}{n_0 + p_0} = \frac{2}{\frac{n_0}{n_i} + \frac{p_0}{n_i}}.$$

Let us denote $\frac{n_0}{n_i} = x$ (dimensionless), then $\frac{p_0}{n_i} = \frac{p_0 n_i}{n_i^2} = \frac{\not{p}_0 n_i}{\not{p}_0 n_0} = \frac{n_i}{n_0} = \frac{1}{x}$. In such dimensionless units $\frac{\tau}{\tau_i} = \frac{2}{x + \frac{1}{x}} = \frac{2x}{x^2 + 1}$.

Graph of this function in a double logarithmic scale has the form presented in Fig. 1.6.2.

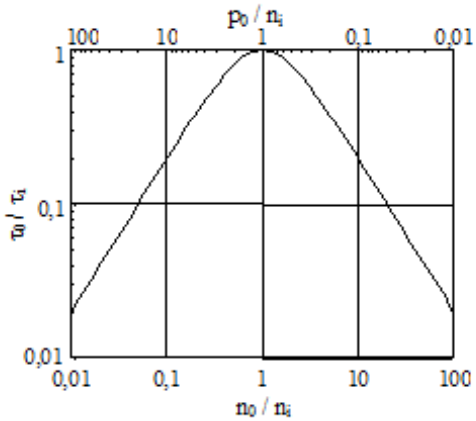


Fig. 1.6.2. Dependence of charge carriers lifetime on concentration of equilibrium electrons or holes (in dimensionless units)

Let us note that x values are plotted on abscissa axis x, i.e. n_0 / n_i, but p_0 / p_i values are simultaneously plotted in the opposite direction because $p_0 / p_i = 1/x$. Let us draw some conclusions:

The maximum lifetime takes place in intrinsic (undoped) semiconductor. Increasing the concentration of impurity in k times leads to the same decrease of NCC lifetime in k times relatively to the undoped crystal. At high excitation level $\Delta n > n_0 , p_0$. In this case

$$\tau_n = \tau_p = \frac{1}{\gamma_{rad} \Delta n} .$$

Thus, in the case of high excitation level, lifetime τ does not depend on the doping level of crystal and is inversely proportional to the excitation level. When considering the processes of radiative recombination, we still do not take into account the law of momentum conservation which has a significant influence on the absorption of light. Most probably the same strong effect will be observed for processes reverse to absorption of light – light generation processes, i.e. at radiative recombination. Similar analysis can be performed by dividing all the crystals into indirect- and direct-band ones. It is also necessary to include in consideration the processes involving phonons, etc. However, there is the other way that takes into account the fact that the processes of radiative recombination are inverse to optical generation of NCC. Let us use for this the principle of detailed balance whose essence is that in a system in thermodynamic equilibrium the transition rates for the forward and inverse processes are equal and balanced in detail. Based on this, we can write

$$r_0(\nu)d\nu = g_0(\nu)d\nu = \alpha\beta I_0(\nu)d\nu. \qquad (1.6.3)$$

According to the principle of detailed balance, we must compare radiative band-to-band transition with similar band-to-band transition at absorption. The quantum yield of the photoionization process for latter is $\beta = 1$. One should take equilibrium radiation, i.e. blackbody radiation at a chosen temperature, as equilibrium exciting light flux I_0. Thus, we can write

$$r_0(\nu)d\nu = \alpha(\nu)\frac{2\pi h\nu^3}{c^2}\frac{1}{\exp\left(\frac{h\nu}{kT}\right)+1}d\nu. \qquad (1.6.4)$$

Given the fact that this is a quantum radiation flux, let us divide it by energy of one quantum $h\nu$. We also consider that the radiation propagates not in vacuum, but in crystal with refractive index $\tilde{n}$.

As a result, we obtain:

$$r_0(\nu)d\nu = \frac{2\pi\tilde{n}^2\nu^2}{c^2}\frac{\alpha(\nu)}{\exp\left(\frac{h\nu}{kT}\right)+1}d\nu. \qquad (1.6.5)$$

The resulting expression gives the rate of equilibrium radiative recombination and simultaneously the *spectral dependence* of the band-to-band equilibrium recombination radiation. It should be noted that this expression is valid for both direct-gap and for the indirect-gap crystals.

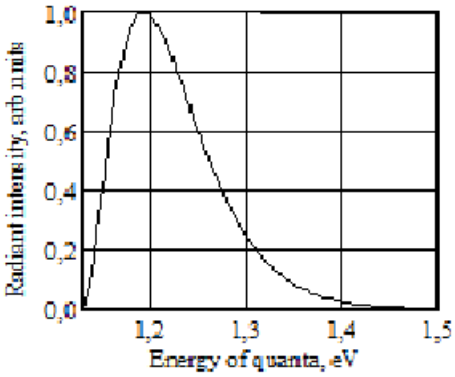


Fig. 1.6.3. Spectral dependence of band-to-band radiative recombination in silicon at T = 300 K

We take into account that the laws of conservation of energy and momentum are realized through the absorption coefficient $\alpha(\nu)$. Thus, knowledge of the spectral dependence $\alpha(\nu)$ allows us to calculate the spectral dependence $r_0(\nu)$. A characteristic feature of the spectrum of radiative band-to-band recombination is the fact that the spectrum is located in the region $h\nu > E_g$. However, in this area, the more $\alpha(\nu)$, the more light absorption band-to-band (photon energy) occurs. As the result, the radiation spectrum may be distorted due to re-absorption of light. The calculated spectrum of radiative band-to-band recombination of silicon, without taking re-absorption into account, is shown in Fig. 1.6.3. The total radiative recombination rate can be found by integrating (1.6.5) over the entire spectrum of radiation

$$r_0 = \frac{2\pi}{c^2}\int_0^\infty \frac{\alpha(\nu)\tilde{n}^2\nu^2}{\exp\left(\frac{h\nu}{kT}\right)+1}d\nu = \frac{2\pi}{c^2}J, \tag{1.6.6}$$

where

$$J = \int_0^\infty \frac{\alpha(\nu)\tilde{n}^2\nu^2}{\exp\left(\frac{h\nu}{kT}\right)+1}d\nu. \tag{1.6.7}$$

The resulting equation (1.4.72) is called the relation of van Rusbrek-Shockley. Since $r_0 = \gamma_{rad} n_0 p_0 = \gamma_{rad} n_i^2$, then

$$\gamma_{rad} = \frac{r_0}{n_i^2} = \frac{2\pi}{c^2 n_i^2}J. \tag{1.6.8}$$

In the non-equilibrium case $r = \gamma_{rad} np$, so that

$$r = \frac{2\pi np}{c^2 n_i^2}J. \tag{1.6.9}$$

Using the relations for the rate of radiative recombination let us write previously obtained dependence for the lifetime. For low excitation level we have

$$\tau_n = \tau_p = \frac{1}{\gamma_{rad}(n_0 + p_0)} = \frac{c^2 n_i^2}{2\pi J(n_0 + p_0)}. \tag{1.6.10}$$

In so doing, for material with *n*-type conductivity $(n_0 >> p_0)$

$$\tau_n = \tau_p = \frac{c^2 n_0 p_0}{2\pi J(n_0 + p_0)} = \frac{c^2 p_0}{2\pi J}, \tag{1.6.11}$$

and for material with *p*-type conductivity $(p_0 >> n_0)$

$$\tau_n = \tau_p = \frac{c^2 n_0 p_0}{2\pi J (n_0 + p_0)} = \frac{c^2 n_0}{2\pi J}. \qquad (1.6.12)$$

We can make an important conclusion from these relations: the lifetime of charge carriers relative to the radiative recombination is determined by *minority* carriers. For high excitation level we have

$$\tau_n = \tau_p = \frac{1}{\gamma_{rad}(n_0 + p_0 + \Delta n)} = \frac{c^2 n_i^2}{2\pi J \Delta n}. \qquad (1.6.13)$$

Indeed, as was shown earlier, the lifetime at high excitation level is independent of the concentration of equilibrium electrons and holes and, hence, the doping level, but it depends on excitation level (inversely proportional).

1.6.3. Exciton radiative recombination

This recombination mechanism is of great theoretical and practical importance, especially for indirect-band crystals. There are 2 ways of excitons formation:

- direct excitation to exciton state, when exciton is generated as a result of a photon absorption (Fig. 1.6.4, a);
- through allowed bands, when as a result absorption of a photon with $h\nu \geq E_g$ non-equilibrium electron and hole are generated which bound then (if the temperature is sufficiently low) into the exciton (Fig. 1.6.4, b).

Further, electron and hole composing exciton can recombine radiatively. Regardless of the mechanism of excitons generation, their radiative recombination can be described by the ratio of van Rusbrek-Shockley (1.6.6).

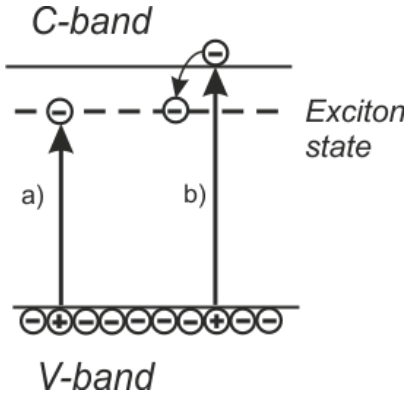


Fig. 1.6.4. Two processes for excitons formation

In principle, when integrating from 0 to ∞, we include the region of exciton radiative recombination (RR) into consideration but $\alpha(\nu)$ dependences for the exciton and band-to-band RR are different. In addition, when calculating exciton RR, one must consider the fact that exciton is a free particle having, therefore, a certain velocity (kinetic energy). So, one should take into account the excitons velocity distribution. Based on the above and considering that for the exciton absorption $\alpha \sim (h\nu)^{1/2}$ or $\alpha \sim (h\nu)^{3/2}$, one can propose the following semi-empirical formula describing the spectrum of exciton radiation involving one or several phonons:

$$I(h\nu) = Aw^{1/2} \exp(-w), \tag{1.6.14}$$

where

$$w = \frac{h\nu - E_g + E_{phon}}{kT}.$$

Expression (1.6.14) describes the contour of line of exciton radiation for multi-phonons processes in indirect-band semiconductors. The contour of exciton radiation for single-phonon processes is described by:

$$I(h\nu) = I_{\max} w^{3/2} \exp(-w). \tag{1.6.15}$$

Sign + before E_{phon} means that there are processes with emission of phonons. Processes with absorption of phonons are unlikely because exciton radiative recombination takes place only at low temperatures (at high temperatures excitons is decopmposed into free electrons and holes).

Exciton emission spectrum for Si at temperature T = 20 K (liquid hydrogen) and T = 77 K (liquid nitrogen) is shown in Fig. 1.6.5.

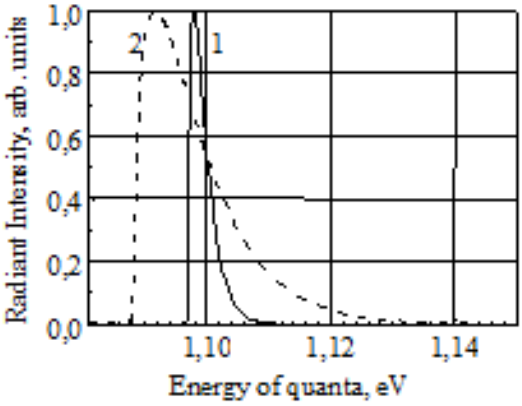


Fig. 1.6.5. Spectral dependence of excitons radiative recombination in silicon at T = 20 K (*1*) and T = 77 K (*2*)

At low temperatures (T < 20 K) excitons are bounded with unionized impurities, thermal defects, radiation defects, dislocations, etc. Being in a bound state, they can recombine radiatively. Since each bound is characterized by its specific energy, the emission spectrum of bound excitons contains information about the presence in the crystal of corresponding centers.

1.6.4. Band-to-band impact recombination (Auger recombination)

The entity of Auger recombination consists that recombining electron and hole transfer their energy (and momentum) to free charge carriers, electron or hole. One should however take into account the momentum conservation law according to which the recoil electron (or hole) may appear far from the dispersion curve limits, i.e. where a very small probability of their finding.

However, one should consider the fact that sometimes the probabilities of Auger processes for electrons and holes become strongly different. Very often, this situation is realized in indirect-band crystals. For example, in crystalline Si probability of Auger recombination involving recoil electrons is rather significant but, at the same time, the probability of a similar process involving the recoil hole is close to zero. For comparison, in Ge, both processes have essentially zero probability.

For radiative recombination we noted above that its rate is proportional to the concentration of electrons and holes $R_{irr}=\gamma_{rad}np$. From a microscopic point of view this can be interpreted as follows: in one elementary event of recombination one electron and one hole are involved.

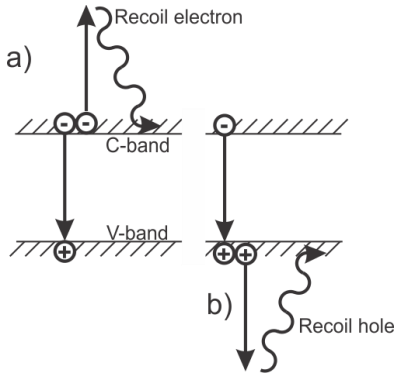


Fig. 1.6.6 The schematic diagram of impact recombination process. Recoil particles – electron (*a*), hole (*b*)

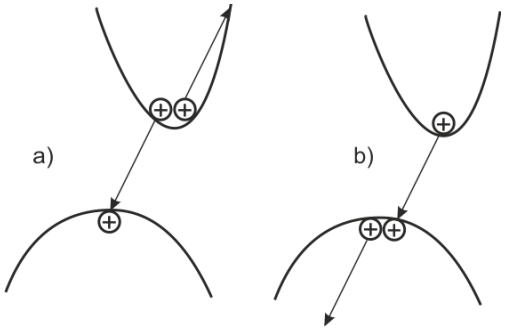


Fig. 1.6.7. Auger recombination in E-k space for high (*a*) and low (*b*) probabilities of the process with recoil electron

Since in the elementary event of Auger recombination two electrons and a hole or two holes and electron are involved, we can write:

$$R_{Auger} = \eta_n nnp + \eta_p npp. \tag{1.6.16}$$

Coefficients η_n and η_p define probabilities of processes involving recoil electrons and recoil holes, respectively. Then the generation-recombination equation will look as follows:

$$\begin{aligned}
&\frac{dn}{dt} = g - \eta_n n^2 p + \eta_p np^2 = \\
&= g - [\eta_n (n_0 + \Delta n)^2 (p_0 + \Delta p) + \eta_p (n_0 + \Delta n)(p_0 + \Delta p)^2] = \\
&= g - [\eta_n (n_0^2 p_0 + n_0^2 \Delta p + 2n_0 \Delta n p_0 + 2n_0 \Delta n \Delta p + \Delta n^2 p_0 + \Delta n^2 \Delta p) + \\
&+ \eta_p (p_0^2 n_0 + p_0^2 \Delta n + 2p_0 \Delta p n_0 + 2p_0 \Delta p \Delta n + \Delta p^2 n_0 + \Delta p^2 \Delta n)].
\end{aligned}$$

Let us group the terms in powers Δn and Δp.

$$\frac{dn}{dt} = g - R \begin{cases}
\Delta n(\eta_p p_0^2 + 2\eta_n n_0 p_0) + \\
+\Delta p(\eta_n n_0^2 + 2\eta_p p_0 n_0) + \\
+2\Delta n \Delta p(\eta_n n_0 + \eta_p p_0) + \\
+\Delta n^2 \eta_n p_0 + \\
+\Delta p^2 \eta_p n_0 + \\
+\Delta n^2 \Delta p \eta_n + \\
+\Delta p^2 \Delta n \eta_n.
\end{cases} \tag{1.6.17}$$

Note that lifetime of electrons and holes will be in accordance with (1.5.38). Let us analyze this expression for the lifetime in the case of low and high levels of excitation.

For low excitation level when $\Delta n << n_0$ and $\Delta p << p_0$. When expressing R in (1.6.17), let us limit oneself by terms of the first order with respect to Δn and Δp. In this case we obtain

$$\tau_n = \frac{1}{\underbrace{(\eta_p p_0^2 + 2\eta_n n_0 p_0)}_{A} + \frac{\Delta p}{\Delta n}\underbrace{(\eta_n n_0^2 + 2\eta_p p_0 n_0)}_{B}} = \frac{1}{A + \frac{\Delta p}{\Delta n} B} \qquad (1.6.18)$$

and

$$\tau_p = \frac{1}{\frac{\Delta n}{\Delta p}(\eta_p p_0^2 + 2\eta_n n_0 p_0) + (\eta_n n_0^2 + 2\eta_p p_0 n_0)} = \frac{1}{\frac{\Delta n}{\Delta p} A + B}. \qquad (1.6.19)$$

It is easy to find the relation between τ_n and τ_p:

$$\tau_n = \frac{\frac{\Delta n}{\Delta p}}{\frac{\Delta n}{\Delta p} A + \frac{\Delta n}{\Delta p}\frac{\Delta p}{\Delta n} B} = \frac{\Delta n}{\Delta p}\tau_p, \quad \text{or} \qquad \tau_n \Delta p = \tau_p \Delta n.$$

Thus, if $\Delta p = \Delta n$, then $\tau_n = \tau_p$. Indeed, although for one recombination event two electrons and a hole or two holes and an electron are necessary but in the result of one recombination event one electron and one hole disappears. Thus, if equal amounts of electrons and holes are generated (band-to-band generation) and Auger recombination takes place, the lifetimes of electrons and holes are equal to each other.

Let us find the dependence of NCC lifetime on doping level just as we did for the radiative recombination. We assume that $\tau_n = \tau_p$, i.e. $\Delta p = \Delta n$. Then,

$$\tau = \frac{1}{\eta_n n_0^2 + \eta_p p_0^2 + 2n_0 p_0(\eta_n + \eta_p)} \qquad (1.6.20)$$

Initially, for simplicity, we assume that the coefficients of Auger recombination of electrons and holes are equal to each other, i.e. $\eta_n = \eta_p$. In this case, the expression (1.4.78) will look as follows $\tau = \frac{1}{\eta(n_0^2 + p_0^2 + 4n_i^2)}$.

For intrinsic semiconductor crystal it adds up to $\tau_i = \frac{1}{6\eta n_i^2}$.

Then $$\frac{\tau}{\tau_i} = \frac{6n_i^2}{n_0^2 + p_0^2 + 4n_i^2} = \frac{6}{\frac{n_0^2}{n_i^2} + \frac{p_0^2}{n_i^2} + 4} = \frac{6}{x^2 + \frac{1}{x^2} + 4} = \frac{6x^2}{x^4 + 4x^2 + 1}.$$

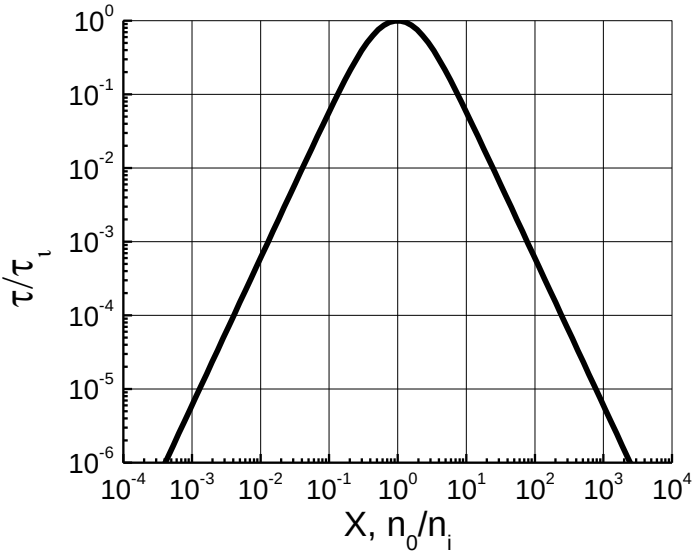


Fig. 1.6.8. Dependence of charge carriers lifetime on concentration of equilibrium electrons or holes (in dimensionless units)

Here the replacement $\frac{n_0}{n_i} = x$ and $\frac{p_0}{n_i} = \frac{1}{x}$ was performed.

If the coefficients of Auger recombination are very different, for example, η_n has a real value other than zero and $\eta_p = 0$, the expression (1.6.20) will look as follows:

$$\tau = \frac{1}{\eta_n (n_0^2 + 2n_i^2)}.$$

Then

$$\tau_i = \frac{1}{\eta_n (n_i^2 + 2n_i^2)} = \frac{1}{3\eta_n n_i^2}.$$

Hence,

$$\frac{\tau}{\tau_i} = \frac{3n_i^2}{n_0^2 + 2n_i^2} = \frac{3}{\frac{n_0^2}{n_i^2} + 2} = \frac{3}{x^2 + 2}.$$

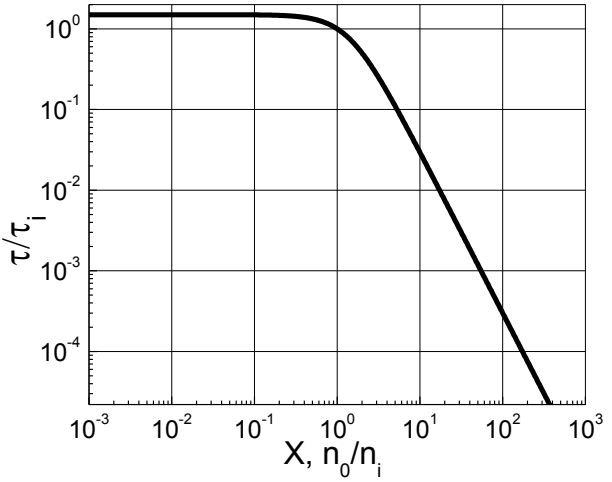


Fig. 1.6.9. Dependence of charge carriers lifetime on concentration of equilibrium electrons or holes if $\Delta n \neq \Delta p$

View of this function of the lifetime on the doping level is presented in Fig. 1.6.9. We can see that this dependence looks close to the same dependence for radiative recombination. The main similarity is that the maximal lifetime takes place in the intrinsic crystal and it decreases with the increase of doping level. However, unlike the previous case, this decrease is much sharper, because it is quadratic with respect to the dopant concentration. Since the formula (1.6.20) does not contain terms and, then lifetime does not depend on the level of excitation, i.e. this type of recombination is a linear recombination.

For high excitation level when relations $\Delta n >> n_0$ and $\Delta p >> p_0$ are realized. In R expression let us leave the terms with the highest powers of Δn and Δp. We obtain:

$$\tau_n = \frac{\Delta n}{\Delta n^2 \Delta p \eta_n + \Delta p^2 \Delta n \eta_p} = \frac{1}{\Delta n \Delta p \eta_n + \Delta p^2 \eta_p},$$

$$\tau_p = \frac{\Delta p}{\Delta n^2 \Delta p \eta_n + \Delta p^2 \Delta n \eta_p} = \frac{1}{\Delta n^2 \eta_n + \Delta p \Delta n \eta_p}.$$

If $\Delta n = \Delta p$, then

$$\tau_n = \tau_p = \frac{1}{\Delta n^2 (\eta_n + \eta_p)},$$

i.e. the lifetime τ depends quadratic on Δn. Let us remind that in the case of radiative recombination similar dependence was linear.

Analyzing the radiative recombination we managed simply to find the explicit form of the radiative recombination coefficient γ_{rad}. This is possible due to the use of the principle of detailed balance for the absorption and emission of radiation by crystal under equilibrium conditions, i.e., in fact, we solve the inverse problem. To find η_n and η_p, it is necessary to solve direct problem. Its solution is extremely difficult because of necessity of precise knowledge of the band structure (dispersion curve) of a concrete crystal.

Remark. The probability of impact recombination is always lower than the radiative one because the impact recombination is three-particle interaction, while zero-phonon radiative recombination is two-particle interaction. However, if the pair of particles is already bound by any interaction (e.g., in exciton), the probability of the Auger process increases dramatically.

1.6.5. Recombination through simple local centers

Formulation of the problem. We have considered two mechanisms of NCC recombination: band-to-band RR and band-to-band Auger recombination. Calculation of τ for these recombination mechanisms shows that for the semiconductor crystal having the band gap E_g of the order of 1-2 eV lifetimes should be ~ 0.1-1 s. At the same time, measured experimentally τ values are rarely more than 10^{-3} s. This fact indicates that both considered recombination mechanisms are not determinative in real crystals.

As established, such mechanisms in the vast majority of semiconductors are the mechanism of recombination through impurity and defect centers. Elementary event of such recombination consists of two stages: in the beginning, carrier of the same sign is captured by the center, and then the carrier of the opposite sign is captured. Besides, each of stages may be radiative or Auger process. In addition, in this case non-radiative processes with energy transfer to phonons become possible. Suchprocesses are impossible in the case of direct band-to-band transitions because one phonon is not able to take the energy of one or a few eV order.

In the case of recombination through local center, emission of a large number of phonons is possible. M.Lax pointed the possibility of such a process out. Its essence is that the local center described by hydrogen-like model has, in

addition to the ground state in the band gap a significant number of excited states which are usually empty. Charge carrier from the allowed band occupies one of these states, generating or absorbing *one* phonon. Further, this carrier goes down by the "ladder" of allowed states emitting (or absorbing) one phonon at each transition (see Fig. 1.6.10).

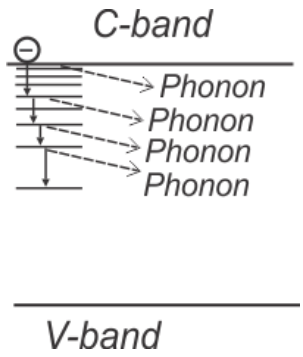


Fig. 1.6.10. The cascade mechanism of energy transmission by M.Lax

Usually simple impurity centers are in one of two states: a neutral or singly charged –positively for donor and negatively for acceptor. However, there are the impurities which may have multiple charge states (amphoteric), such as gold or copper in silicon or germanium. We will consider the simplest case of singly charged impurities. Such a center will be characterized by two coefficients of capture – for holes and electrons. Naturally, these coefficients will be different at least because one carrier is captured by a neutral center and the other – by charged one, in doing so Coulomb potential additionally is "working".

Let us consider semiconductor crystal having one type of recombination centers in the band gap (sometimes in the literature, such centers are called *traps*). We assume (to fix the idea) that recombination centers are of donor type and their energy level is located in the upper half of the band gap. We also assume that the concentration of recombination centers is M, and the number of electrons localized at these centers is m. If the centers are singly charged, it is always $m \leq M$. Let excitation of NCC in crystal is realized by band-to-band regime (process 1 in Fig. 1.6.11).

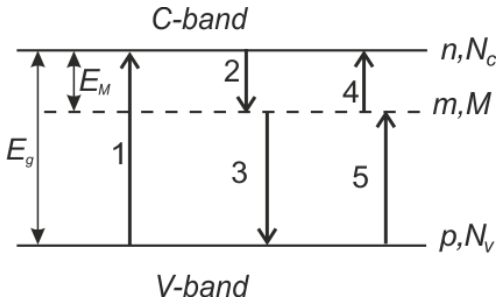


Fig. 1.6.11. Electronic processes for recombination over simple local centers

Concentration of free electrons in the conduction band is n, the concentration of holes in the valence band is p, the densities of allowed states in C-and V-bands are N_C and N_V. NCC recombination occurs in two stages: electron is captured by the center (process 2) with subsequent recombination of this electron with a hole from the valence band (process 3). Since we do not indicate the exact position of the level of recombination centers in the band gap, therefore it can be close enough to the C-or V-bands. In this case, the processes of reverse thermal emission (release) of carriers captured by centers have a high probability. Let us introduce these processes (4 and 5) into consideration. Rate of the processes 2 and 3 can be easily found from the following reassons. Rate of the process 2 is proportional to the concentration n of electrons n in the conduction band and the number M-m of free places for electrons on the centers, i.e. $n(M-m)$. Similarly, the rate of the process 3 is proportional to $m \cdot p$. In a different way such relations can be interpreted as follows: for realization of the process 2 *one electron* and *one unoccupied place on the center* are required.

For the processes 4 and 5 the case is somewhat more complicated. On the one hand, using the same algorithm, we can say that the process 4 will be proportional to $m \cdot N_C$, and the process 5 to $p \cdot (M-m)$. However, one should take into account the fact that the processes 4 and 5 are realized with energy absorption. Based on the above we write the processes 1 – 5 as follows:

g;

$\gamma_n n(M-m)$;

$\gamma_p p\, m$;

$\alpha_n m$;

$\alpha_p (M-m)$.

Let us derive the coefficients α_n and α_p. In equilibrium state, transitions 2 and 4 are equal to each other

$$\alpha_n m_0 = \gamma_n n_0 (M - m_0);$$
$$\alpha_n m_0 - \gamma_n n_0 M + \gamma_n n_0 m_0 = 0;$$

so that

$$m_0 = \frac{\gamma_n n_0 M}{\alpha_n + \gamma_n n_0} = \frac{M}{\frac{\alpha_n}{\gamma_n n_0} + 1}. \qquad (1.6.21)$$

On the other hand, according to the Fermi statistics for electrons concentration at the level we can write:

$$m_0 = f_{F-D} M = \frac{M}{\exp\left(\frac{E_M - \varepsilon_F}{kT}\right) + 1}. \qquad (1.6.22)$$

From a comparison of expressions for m_0 we obtain:

$$\alpha_n = \gamma_n n_0 \exp\left(\frac{E_M - \varepsilon_F}{kT}\right) = \gamma_n N_C \exp\left(\frac{\varepsilon_F}{kT}\right) \exp\left(\frac{E_M - \varepsilon_F}{kT}\right) =$$
$$= \gamma_n N_C \exp\left(\frac{E_M}{kT}\right). \qquad (1.6.23)$$

Thus, the transition 4 can be written as $\gamma_n m N_C \exp\left(\frac{E_M}{kT}\right)$.

We really have the product mN_C but an additional term $\exp\left(\frac{E_M}{kT}\right)$ appeared. Its meaning is that the transition 4 comes "up" and additional energy originating from thermal energy of crystal is necessary for its realization. The greater the distance from the recombination level to the bottom of the conduction band is, the smaller the probability of getting this energy. This is expressed in an additional term (exponent power is negative hence the exponent is less than 1). The deeper recombination level in the band gap is, the closer exponential factor to 0. Similarly, for the transition 5 we have

$$\alpha_p (M - m_0) = \gamma_p p_0 m_0;$$
$$\alpha_p M - \alpha_p m_0 - \gamma_p p_0 m_0 = 0;$$
$$m_0 = \frac{\alpha_p M}{\alpha_n + \gamma_p p_0} = \frac{M}{1 + \frac{\gamma_p p_0}{\alpha_p}} \qquad (1.6.24)$$

From a comparison of the expression for m_0 with the expression derived from Fermi-Dirac statistics (1.6.22), we obtain:

$$\frac{\gamma_p p_0}{\alpha_p} = \exp\left(\frac{E_M - \varepsilon_F}{kT}\right),$$

whence

$$\alpha_p = \gamma_p p_0 \exp\left(\frac{\varepsilon_F - E_M}{kT}\right) = \gamma_p N_V \exp\left(\frac{E_g - \varepsilon_F}{kT}\right)\exp\left(\frac{\varepsilon_F - E_M}{kT}\right) = \\ = \gamma_p N_V \exp\left(\frac{E_g - E_M}{kT}\right). \tag{1.6.25}$$

Thus the transition 5 can be written as

$$\gamma_p (M - m) P_V \exp\left(\frac{E_g - E_M}{kT}\right)$$

We introduce the notations:

$$N_C \exp\left(\frac{E_M}{kT}\right) = N_{CM} \text{ and } N_V \exp\left(\frac{E_g - E_M}{kT}\right) = N_{VM}. \tag{1.6.26}$$

Parameters N_{CM} and N_{VM} are called reduced densities of states in C-and V-bands, respectively, relatively to transitions from the M level. Thus, the transitions 1 - 5 in the final form will look as:

g;

$\gamma_n n(M-m)$;

$\gamma_p p\, m$;

$\gamma_n n N_{CM}$;

$\gamma_p (M-m) N_{VM}$.

The system of kinetic equations describing these processes, taking into account the obtained expressions for the transition rates, will look as follows:

$$\frac{dn}{dt} = g_n - \gamma_n n(M - m) + \gamma_n m N_{CM}; \tag{1.6.27}$$

$$\frac{dm}{dt} = \gamma_n n(M - m) - \gamma_n m N_{CM} + \gamma_p m p - \gamma_p (M - m) N_{VM}; \tag{1.6.28}$$

$$\frac{dp}{dt} = g_p - \gamma_p m p + \gamma_p (M - m) N_{VM}. \tag{1.6.29}$$

The solution of this system is not possible because, when generating band-to-band, i.e. $g_n = g_p$, equation (1.4.86) for electrons on the centers is a linear combination of (1.4.85) and (1.4.87) equations. Electroneutrality equation for non-equilibrium charge carriers can be taken as missing equation

$$\Delta n + \Delta m = \Delta p. \tag{1.6.30}$$

Thus, definition of $g(t)$ function (in the case of optical generation $g(t) = \alpha\beta I(t)$), in principle, allows us to find $n(t)$, $p(t)$, $m(t)$ dependences but the general solution of this system faces a number of mathematical difficulties because it is necessary to solve a system of nonlinear differential equations.

Shockley-Read-Hall model. In this model (SRH), a number of simplifications were accepted which allowed not only to solve the problem of recombination through local centers but also to get a number of obvious representations that extend a particular solution proposed by the authors of model.

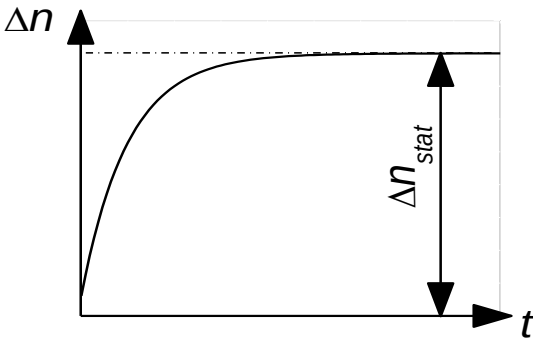


Fig. 1.6.12. Attainment of stationary concentration of ICC

The first simplification accepted in this model concerns the fact that a stationary process is considered. Stationary process takes place if rate of NCC generation g is not changed with time. Some time after such generation turning on, system will come in the equilibrium state, i.e. generation rate will be equal to the rate of recombination and

$$\frac{dn}{dt} = \frac{dp}{dt} = 0 .$$

Therefore

$$\frac{d}{dt}(\Delta n) = g - \frac{\Delta n}{\tau} = 0$$

and

$$\Delta n_{stat} = g\tau_{stat} .$$

In the stationary case, the equations (1.4.85)-(1.4.87) must be equated to zero. From equation (1.4.879) we obtain:

$m = \dfrac{g + \gamma_p M N_{VM}}{\gamma_p p + \gamma_p N_{VM}}$; and $M - m = \dfrac{\gamma_p pM - g}{\gamma_p p + \gamma_p N_{VM}}$.

Let us substitute these expressions into equation (1.4.85):

$$g - \gamma_n n \frac{\gamma_p pM - g}{\gamma_p p + \gamma_p P_{VM}} + \gamma_n N_{CM} \frac{g + \gamma_p M P_{VM}}{\gamma_p p + \gamma_p P_{VM}} = 0,$$

$$g\gamma_p p + g\gamma_p P_{VM} - \gamma_n \gamma_p npM + g\gamma_n n + g\gamma_n N_{CM} + \gamma_n \gamma_p M N_{CM} P_{VM} = 0,$$

$$g = \frac{M\gamma_n\gamma_p (np - N_{CM} P_{VM})}{\gamma_n (n + N_{CM}) + \gamma_p (p + P_{VM})}. \qquad (1.6.30)$$

The second and very important factor that simplifies SRH model is the fact that low concentration of recombination centers is accepted. This requires some explanation. If we assume the low concentration M of recombination centers, then m will be small, too, because $m \le M$. In this case, we can assume that any filling of recombination centers practically does not change the concentration of carriers in the allowed bands. Mathematically this can be written in such a way:

$m << n$, and then $\Delta m << \Delta n$.

Then the equation (1.6.30) will be:

$$\Delta n = \Delta p \qquad (1.6.31)$$

This, in turn, highly simplifies the model and the resulting solution. Indeed, if $\Delta n = \Delta p$, then $\tau_n = \dfrac{\Delta n}{g} = \dfrac{\Delta p}{g} = \tau_p$, i.e. lifetimes of electrons and holes are the same (we denote them just as τ). Using (1.6.11) we obtain

$$\tau = \frac{\Delta n}{g} = \Delta n \frac{\gamma_n (n + N_{CM}) + \gamma_p (p + N_{VM})}{M\gamma_n\gamma_p \underbrace{(np - N_{CM} N_{VM})}_{A}}. \qquad (1.6.32)$$

Let us find the value A.

$$np = (n_0 + \Delta n)(p_0 + \Delta n) = n_0 p_0 + \Delta n(n_0 + p_0 + \Delta n),$$

$$n_0 p_0 = N_C \exp\left(\frac{\varepsilon_F}{kT}\right) N_V \exp\left(\frac{E_g - \varepsilon_F}{kT}\right) = N_C N_V \exp\left(\frac{E_g}{kT}\right),$$

$$N_{CM} N_{VM} = N_C \exp\left(\frac{E_M}{kT}\right) N_V \exp\left(\frac{E_g - E_M}{kT}\right) = N_C N_V \exp\left(\frac{E_g}{kT}\right).$$

Thus,

$$A = (np - N_{CM} N_{VM}) = \Delta n(n_0 + p_0 + \Delta n).$$

Then

$$\tau = \frac{n_0 + N_{CM} + \Delta n}{M\gamma_p(n_0 + p_0 + \Delta n)} + \frac{p_0 + N_{VM} + \Delta n}{M\gamma_n(n_0 + p_0 + \Delta n)}.$$

Values $\frac{1}{\gamma_p M}$ and $\frac{1}{\gamma_n M}$ have the dimension of time. Let us denote them as τ_{n0} and τ_{p0}, respectively. Then

$$\tau = \tau_{p0}\frac{n_0 + N_{CM} + \Delta n}{(n_0 + p_0 + \Delta n)} + \tau_{n0}\frac{p_0 + N_{VM} + \Delta n}{(n_0 + p_0 + \Delta n)}. \qquad (1.6.33)$$

The resulting formula is called Shockley-Read equation. It determines the lifetime of NCC in the crystal in conditions of stationary excitation and low concentration of recombination centers. Let us analyze this formula for some particular cases.

A. Low excitation level (LEL)

LEL is characterized by the fact that $\Delta n < n_0$, p_0. Under these conditions, Shockley-Read formula is transformed into

$$\tau = \tau_{p0}\frac{n_0 + N_{CM}}{(n_0 + p_0)} + \tau_{n0}\frac{p_0 + N_{VM}}{(n_0 + p_0)}. \qquad (1.6.34)$$

Let us find the dependence of τ on the degree of crystal doping. In this case, we assume that the degree of doping is determined by the position of the Fermi level which is not connected with recombination centers (more precisely, connected not only with the recombination centers, but also with other impurity centers). In other words, in the crystal position of the recombination centers is only defined, but the Fermi level, regardless of this, may occupy any position in the band gap which is determined by the type (donor or acceptor) and the depth of the energy levels of all impurity centers, including recombination ones. For definiteness, we consider a crystal having levels of recombination centers located in the upper half of the band gap.

Terms in the formula (1.4.93) have the form:

$$
\begin{aligned}
n_0 &= N_C \exp\left(\frac{\varepsilon_F}{kT}\right);\\
p_0 &= N_V \exp\left(\frac{E_g - \varepsilon_F}{kT}\right);\\
N_{CM} &= N_C \exp\left(\frac{E_M}{kT}\right);\\
N_{VM} &= N_V \exp\left(\frac{E_g - E_M}{kT}\right).
\end{aligned}
\qquad (1.6.35)
$$

Note that the structure of these terms is similar: factors N_C or P_V and exponential factor with negative the exponent power everywhere is less than 1. Thus, when comparing the terms in (1.4.94), exponent powers which are determined by one or another energy gap should be compared. Where this gap is larger, the exponential term will be less and, consequently, the terms in series n_0, p_0, N_{CM} and P_{VM} are less. We assume that the coefficients N_C and P_V are close to the order of values.

1. If the Fermi level is in the upper half of the band gap and lies above the recombination level (Fig. 1.6.13, a), the following relations are realized:

$n_0 > N_{CM} > N_{VM} > p_0$.

Leaving only the leading terms in numerator and denominator of the formula (1.4.93), we obtain:

$$\tau = \tau_{p0}. \qquad (1.6.36)$$

Thus, while the Fermi level lies above the recombination level, lifetime of NCC is constant and does not dependent on its position. This is explained by the fact that all levels of recombination centers are occupied by electrons. There are also a lot of electrons in the conduction band. Holes appearing as a result of the generation quickly recombine with electrons on these centers. So the lifetime is determined by hole lifetime relatively to recombination centers completely filled with electrons.

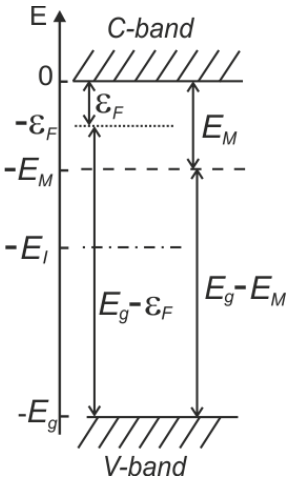


Fig. 1.6.13. The schematic diagram of terms for recombination over local centers. Situation when Fermi level is in the upper half of the band gap and lies above the recombination level

2. If the Fermi level lies in the upper half of the band gap but is below the level of recombination centers (Fig. 1.6.14), the following relations is realized:

$N_{CM} > n_0 > p_0 > N_{VM}$.

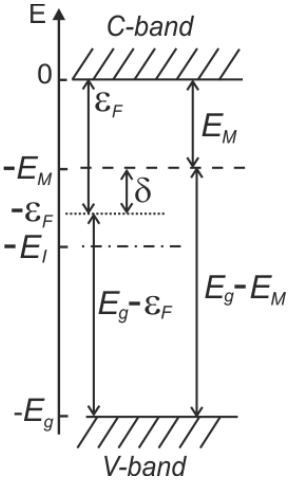


Fig. 1.6.14. The schematic diagram of terms for recombination over local centers. Situation when Fermi level lies in the upper half of the band gap but below the level of recombination centers

Formula (1.6.34) takes the form:

$$\tau = \tau_{p0}\frac{N_{CM}}{n_0} = \tau_{p0}\frac{N_C \exp\left(\dfrac{E_M}{kT}\right)}{N_C \exp\left(\dfrac{\varepsilon_F}{kT}\right)} = \tau_{p0}\exp\left(\frac{\overbrace{E_M - \varepsilon_F}^{\delta}}{kT}\right). \qquad (1.6.37)$$

The numerator of the exponent $\delta = E_M - \varepsilon_F$ is positive and increases with the movement of the Fermi level to the middle of band gap.

$$\ln\tau = \ln\tau_{p0} + \frac{\delta}{kT}.$$

Lifetime of NCC also increases with increasing δ.

Crystal still refers to the n-type conductivity, however, filling the recombination centers by electrons is low because the Fermi level lies below the level of recombination centers. While lowering of the Fermi level, filling centers by electrons decreases and hence the lifetime increases.

3. Let us the Fermi level in the crystal of p-type is lowered to the bottom of band gap but the energy gap $|E_g - \varphi| > |E_M|$ (Fig. 1.6.15). In this case relations between the parameters of the formula (1.4.93) take the form: $N_{CM} > p_0 > n_0 > P_{VM}$ and formula (1.4.93) is transformed into

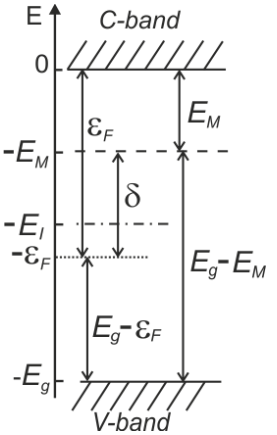


Fig. 1.6.15. The schematic diagram of terms for recombination over local centers. Situation for p-type crystal and when $|E_g - \varphi| > |E_M|$

$$\tau = \tau_{p0}\frac{N_{CM}}{p_0} = \tau_{p0}\frac{N_C \exp\left(\dfrac{E_M}{kT}\right)}{P_V \exp\left(\dfrac{E_g - \varepsilon_F}{kT}\right)} =$$

$$= \tau_{p0}\frac{N_C}{N_V}\exp\left(\frac{E_M - E_g + \varepsilon_F}{kT}\right) = \tag{1.6.38}$$

$$= \tau_{p0}\frac{N_C}{N_V}\exp\left(\frac{-(E_g - E_M - \varepsilon_F}{kT}\right) = \tau_{p0}\frac{N_C}{N_V}\exp\left(\frac{\delta}{kT}\right)$$

In contrast to the previous case, while the Fermi level moving to the V-band, δ value decreases but remains positive. Consequently, the lifetime of the NCC decreases:

$$\ln\tau = \ln\tau_{p0} + \frac{\delta}{kT}.$$

4. If the Fermi level approaches the top of V-band (Fig. 1.6.16), relations between the parameters of the formula (1.4.93) become equal to

$$p_0 > N_{CM} > N_{VM} > n_0,$$

and Shockley-Read formula (1.6.34) takes the form

$$\tau = \tau_{n0}.$$

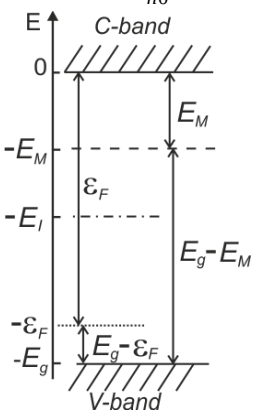


Fig. 1.6.16. The schematic diagram of terms for recombination over local centers. Situation when Fermi level approaches the top of V-band

The Fermi level is near the V-band and recombination levels located in the upper half of band gap are empty, there are very few electrons in C-band and a lot of holes in V-band. However, the recombination in this case is defined not

by them, but by free electrons. As soon as free electron appears in C-band, it is captured on the center and recombines with a hole in V-band.

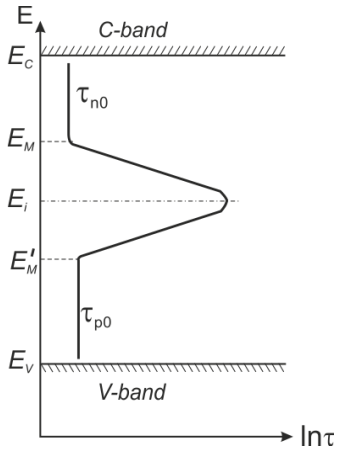


Fig. 1.6.17. Dependence of carriers lifetime on position of Fermi level

The resulting graph of $\ln\tau = f(\varphi)$ dependence is shown in the Fig. 1.6.17. As can be seen from this figure, the general shape of the curve is similar to the corresponding curves when the radiative and Auger recombinations occur. Common is that the maximal lifetime of NCC takes place for intrinsic or close to intrinsic crystal. However, in contrast to the previous cases, in recombination through the local centers, when the Fermi level lies near the allowed bands, i.e. when the conductivity of the crystal has a distinctly impurity character, the lifetime becomes constant independently on the Fermi level position. If recombination centers are located in the lower part of band gap, the dependence $\tau = f(\varphi)$ is the same, it is only necessary to permute τ_{n0} and τ_{p0}.

Let us consider the temperature dependence of the lifetime in the case of low excitation. Here we follow the same approach, assuming that the Fermi level shifts with temperature change. Its position determines the lifetime. For definiteness, we assume that the crystal possesses *n*-type conductivity having recombination centers in the upper half of the band gap. Again, for definiteness, we assume that the recombination centers and centers providing donor conductivity type are different centers, and recombination centers lie deeper then donors.

Remember that temperature dependence of concentration of equilibrium charge carriers (electrons in our case) has three characteristic regions *A, B, C* (see. Fig. 1.6.18a).

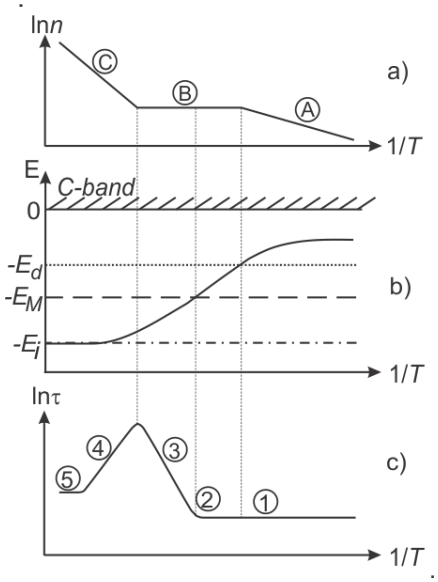


Fig. 1.6.18. Temperature dependence of: *a*) concentration of a free electrons; *b*) position of Fermi level; *c*) lifetime of charge carriers

Region *A* characterizes temperature ionization of impurity, so that the Fermi level lies above the donor levels, and relations between the coefficients in the formula (1.4.93) will be:

$n_0 > N_{CM} > N_{VM} > p_0$.

In this case, formula (1.6.36) gives

$$\tau = \tau_{p0}.$$

This corresponds to region 1 of $\ln\tau = f(1/T)$ dependence in Fig. 1.6.18c.

Region *B* in Fig. 1.6.18a characterizes complete ionization of the impurities, so that n_0 = const. Until the Fermi level is above the level of recombination centers (see., Fig. 1.6.18b), the relation $n_0 > N_{CM} > N_{VM} > p_0$ is fulfilled, and still $\tau = \tau_{p0}$ (range 2 in $\ln\tau = f(1/T)$ curve in Fig. 1.6.18c).

When the Fermi level is lowered below the recombination level (see., Fig. 1.6.15,b), the relation between the coefficients in the formula (1.6.34) will be

$$N_{CM} > n_0 > p_0 > N_{VM},$$

and the formula (1.6.34) after discarding the small terms (and taking into account the fact that $n_0 = const$) becomes equal to

$$\tau = \tau_{p0}\frac{N_{CM}}{n_0} = \tau_{p0}\frac{2\dfrac{(2\pi m_n^* kT)^{3/2}}{h^3}\exp\left(\dfrac{E_M}{kT}\right)}{n_0} = \tag{1.6.39}$$

$$= const \cdot T^{3/2}\exp\left(\frac{E_M}{kT}\right).$$

In this expression, two factors are temperature dependent, and both increase with temperature growth (of course, the exponential term is decisive). There is a region of lifetime growth with temperature increase on the $\ln\tau = f(1/T)$ dependence (range 3 in Fig. 1.6.18c).

For temperature region *C* in Fig. 1.6.18a the Fermi level approaches the middle of the band gap (see, in Fig. 1.6.18b), so that thermal band-to-band transitions begin to occur and

$$n_0 = p_0 = n_i = \sqrt{N_C N_V}\exp\left(\frac{E_g}{2kT}\right). \tag{1.6.40}$$

In this case, formula (1.6.34) will have the form:

$$\tau = \tau_{p0}\frac{n_i + N_{CM}}{2n_i} + \tau_{n0}\frac{n_i + N_{VM}}{2n_i} = \frac{\tau_{p0}}{2}\left[1 + \sqrt{\frac{N_C}{N_V}}\exp\left(\frac{E_M - \dfrac{E_g}{2}}{kT}\right)\right] + \tag{1.6.41}$$

$$+\frac{\tau_{n0}}{2}\left[1 + \sqrt{\frac{N_V}{N_C}}\exp\left(\frac{\dfrac{E_g}{2} - E_M}{kT}\right)\right].$$

If the recombination level is located in the upper half of the band gap, then $E_M - \dfrac{E_g}{2} > 0$, i.e. the first exponent in (1.6.41) is positive, and the second exponent is negative. In so doing, the second term $\approx \dfrac{\tau_{n0}}{2}$. While $E_M - (E_g/2) < kT$, the first exponential contribution decreases with temperature increasing, and we have region 4 on the curve $\ln\tau = f(1/T)$ in Fig. 1.6.18c. However, if the increase in temperature will lead to the $E_M - (E_g/2) > kT$, then both exponents will tend to unity (see Fig 1.6.19) and then

$$\tau = \frac{\tau_{p0}}{2}\left(1 + \sqrt{\frac{N_C}{N_V}}\right) + \frac{\tau_{n0}}{2}\left(1 + \sqrt{\frac{N_V}{N_C}}\right). \tag{1.6.42}$$

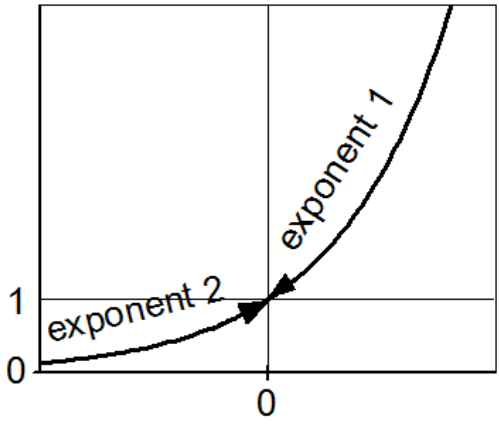


Fig. 1.6.19. Exponential contributions to NCC lifetime in formulae (1.6.41)

If the recombination level is located in the upper half of the band gap, then $E_M - \frac{E_g}{2} > 0$, i.e. the first exponent in (1.6.41) is positive, and the second exponent is negative. In so doing, the second term $\approx \frac{\tau_{n0}}{2}$. While $E_M - (E_g/2) < kT$, the first exponential contribution decreases with temperature increasing, and we have region 4 on the curve $\ln\tau = f(1/T)$ in Fig. 1.6.18c. However, if the increase in temperature will lead to the $E_M - (E_g/2) > kT$, then both exponents will tend to unity (see Fig 1.6.19) and then

$$\tau = \frac{\tau_{p0}}{2}\left(1+\sqrt{\frac{N_C}{N_V}}\right)+\frac{\tau_{n0}}{2}\left(1+\sqrt{\frac{N_V}{N_C}}\right). \qquad (1.6.42)$$

If the values of N_C and N_V are of the same order, then

$$\tau = \frac{\tau_{p0}}{2}(1+1)+\frac{\tau_{n0}}{2}(1+1) = \tau_{p0}+\tau_{n0}. \qquad (1.6.43)$$

Thus, at high temperatures $\ln\tau = f(1/T)$ curve in Fig. 1.6.18c can reach the range 5 where $\ln\tau \neq f(1/T)$. But the latter is not always possible. If the level is deep enough, the temperature of outlet to the range 5 may be too high (above the melting point of the crystal). For example, the recombination level of gold in silicon has a level is by 0.35 eV away of the edge of the allowed band. In this case $\left|E_M - \frac{E_g}{2}\right| = |0.35 - 0.55| = 0.2eV$, i.e in this case, $kT >>$ 0.2 eV. Therefore,

$T = \frac{0.2eV}{8.6 \cdot 10^{-5} eV \cdot K} \approx 2500K$, that is significantly higher than the melting point of silicon (1410 ºC).

B. High excitation level (HEL).

At HEL we have

$$\Delta n = \Delta p \sim n_o, p_o, N_{VM}, N_{CM}. \qquad (1.6.44)$$

In this case the Shockley-Read formula (1.4.92) provides for the lifetime:

$$\tau = \tau_{n0} + \tau_{p0} = \frac{1}{M\gamma_n} + \frac{1}{M\gamma_p} = \frac{1}{M} \cdot \frac{\gamma_n + \gamma_p}{\gamma_n \cdot \gamma_p}. \qquad (1.6.45)$$

As we can see, in this case, the lifetime is independent of the position of the Fermi level, temperature and concentration of charge carriers (equilibrium and non-equilibrium). The latter conclusion is essentially distinguishes the case of recombination through local centers from other types of the above discussed band-to-band recombination. As was shown above, for radiative and Auger recombination life times are $\tau \sim 1/\Delta n$ and $\tau \sim 1/(\Delta n)^2$, respectively. Independence of τ from Δn at HEL indicates that the number of recombining carriers is determined not by number of carriers but by the "capacity" of recombination centers and the number of these centers M.

The present case is quite often realized in pure semiconductors and insulators with deep recombination centers. In such crystals Δn, n_0, p_0, N_{CM}, N_{VM} are small and even in the case of not very high levels of lighting or injection, relation (1.6.45) can be satisfied.

The consideration of transition from LEL to HEL, for example by increasing the optical lighting of crystal or by increase of injection of NCC in crystal, is of great interest. In this case, as can be seen from Fig. 1.6.20, when increasing the excitation level, the increase of lifetime τ for heavily doped crystal and decrease of τ for the crystal close to intrinsic is observed.

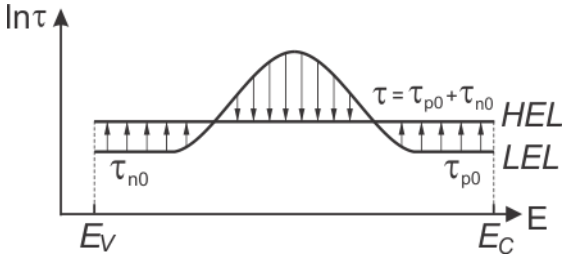


Fig. 1.6.20. Change of charge carrier lifetime for transition from LEL to HEL depending on position of Fermi level

Similar situation is observed also with temperature change. As can be seen from the Fig 1.6.21, at low temperatures the transition from LEL to HEL increases lifetime τ. At high temperatures, when intrinsic conductivity starts, the increase of excitation level leads to reduction of τ.

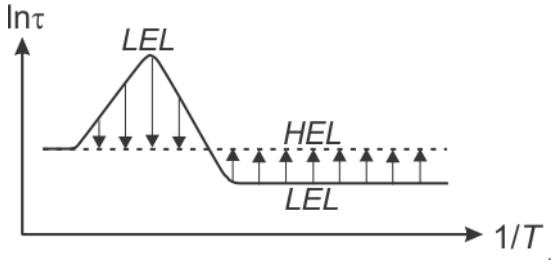


Fig. 1.6.21. Change of charge carrier lifetime for transition from LEL to HEL depending on temperature change

Trapping and recombination centers. When considering the recombination processes over local centers we operated with two types of transitions: direct (down, with the release of energy) and inverse (up, with energy absorption). For example, for electrons it will be

$$\frac{dn}{dt} = g_n - \underbrace{\gamma_n n(M-m)}_{\text{direct transition}} + \underbrace{\gamma_n m N_{CM}}_{\text{reverse transition}} .$$

Indeed, a single recombination event through the local level consists of two stages:

- capture of carrier of one sign on the center, and
- finishing recombination event with the carrier of the other sign.

But this second process has an alternative possibility of thermal emission of carrier back into the allowed band (3) in Fig. 1.6.22.

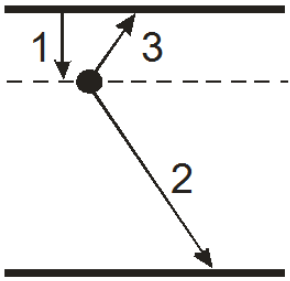


Fig. 1.6.22. alternative ways for electron, captured on local level (*1*):
2 – a recombination with a hole of V-band; *3* – return thermal emission in C-band

It is obvious that the increase in the rate of direct transitions leads to a reduction of the lifetime of NCC, and the increase of the inverse transitions rate leads to an increase of τ. This follows from the definition of "NCC lifetime":

$$\tau_n = \frac{\Delta n}{g_n} = \frac{\Delta n}{\gamma_n n(M-m) - \gamma_n m N_{CM}}.$$

In this case, the higher the probability of inverse (thermal) transitions is, the closer corresponding recombination level to the allowed band. For shallow level and at sufficiently high temperatures the probability of thermal transitions to the allowed band can be so high that the recombination (more precisely, its final stage) may stop at all. Centers of this type are called *trapping (attachment) centers*.

Let us introduce the coefficient of recombination probability (for the carrier captured on the center) equals to the ratio of direct and inverse processes rates:

$$K_n = \frac{\gamma_p m p}{\gamma_n m N_{CM}} = \frac{\gamma_p p}{\gamma_n N_{CM}}. \qquad (1.6.46)$$

Similarly,

$$K_p = \frac{\gamma_n n}{\gamma_p N_{VM}}. \qquad (1.6.47)$$

Centers for which the coefficient of the recombination probability is greater than 1 will be called recombination centers, and the corresponding levels in the band gap are recombination levels. Conversely, centers with K < 1 will be called the trapping centers and their corresponding levels are the trapping levels. Boundary separating the regions of trapping and recombination levels is called *demarcation line* (denoted as E_{Dn}). Similar reasoning can be done for trapping centers of holes, when we can also introduce the concept of demarcation line (E_{Dp}).

Let us find the position of demarcation lines in the band gap. As follows from the definition of demarcation line, its position corresponds to K = 1. Let us substitute E_{Dn} instead of E_M in (1.6.46).

$$K_n = \frac{\gamma_p p}{\gamma_n N_{CM}} = \frac{\gamma_p N_V}{\gamma_n N_C} \exp\left(\frac{E_g - \varphi_p - E_{Dn}}{kT} \right) = 1.$$

After taking the logarithm we obtain

$$E_{Dn} = E_g - \varphi_p - kT \ln \frac{\gamma_n N_C}{\gamma_p N_V}.$$

As a result, for E_{Dp} we obtain:

$$E_{Dp} = E_g - \varphi_n + kT \ln \frac{\gamma_p N_V}{\gamma_n N_C}.$$

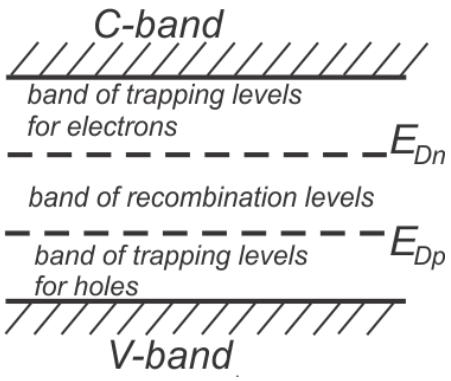


Fig.1.6.23. Division of the forbidden band into band of recombination levels and band of trapping levels

Thus, the demarcation line for electrons is located at the same distance from C-band, at which quasi-Fermi level for holes is away from the V-band bottom (with accuracy up to the term $kT \ln \frac{\gamma_n N_C}{\gamma_p N_V}$). Similarly, the demarcation line for holes is located at the distance from V-band equal to the distance between C-band and quasi-Fermi level for electrons (again, with accuracy up to the term $kT \ln \frac{\gamma_p N_V}{\gamma_n N_C}$). Thus, the demarcation lines divide the band gap into three regions: the region of trapping levels for electrons, the region of trapping levels for holes and recombination levels region (see, Fig. 1.6.23). It should be noted that this differentiation is determined by specific levels definition and, moreover, depends on the temperature.

Relaxation of non-equilibrium conductivity during recombination through local centers. In the standard SRH model stationary case is considered, when

$$\frac{d(\Delta n)}{dt} = \frac{d(\Delta p)}{dt} = 0.$$

Let us extend the applicability limits of the model. According to the conditions of SRH model, we accept $\Delta n = \Delta p$ and also assume that

$$\frac{d(\Delta n)}{dt} = \frac{d(\Delta p)}{dt}.$$

It is not obvious, but there is a strict mathematical proof of this state. Then, equating equation (1.6.27) and (1.6.29) to each other we obtain

$$\gamma_n m N_{CM} - \gamma_n n(M-m) + \gamma_p pm - \gamma_p (M-m) N_{VM} = 0 .$$

Hence we obtain:

$$m = \frac{\gamma_n nM + \gamma_p M N_{VM}}{\gamma_n (n + N_{CM}) + \gamma_p (p + N_{VM})},$$

and

$$M - m =$$

$$= \frac{\cancel{\gamma_n nM} + \gamma_n N_{CM} M + \gamma_p pM + \cancel{\gamma_p N_{VM} M} - \cancel{\gamma_n nM} - \cancel{\gamma_p N_{VM} M}}{\gamma_n (n + N_{CM}) + \gamma_p (p + N_{VM})} =$$

$$= \frac{\gamma_n N_{CM} M + \gamma_p pM}{\gamma_n (n + N_{CM}) + \gamma_p (p + N_{VM})}.$$

We substitute obtained values of m and (M-m) into (1.6.27)

$$\frac{dn}{dt} = g - \frac{-\gamma_n n(\gamma_n N_{CM} M + \gamma_p pM) + \gamma_n N_{CM} (\gamma_n nM + \gamma_p N_{VM} M)}{\gamma_n (n + N_{CM}) + \gamma_p (p + N_{VM})} =$$

$$= g - \frac{\gamma_n \gamma_p M \overbrace{(N_{CM} N_{VM} - np)}^{A}}{\gamma_n (n + N_{CM}) + \gamma_p (p + N_{VM})} ,$$

to find A (when assuming $\Delta n = \Delta p$):

$$N_{CM} N_{VM} - np = N_C \exp\left(\frac{E_M}{kT}\right) N_V \exp\left(\frac{E_g - E_M}{kT}\right) -$$

$$- \quad n_0 p_0 \quad - n_0 \Delta n - p_0 \Delta n - (\Delta n)^2 =$$

$$= N_C N_V \exp\left(\frac{E_g}{kT}\right)$$

$$= N_C N_V \exp\left(\frac{E_g}{kT}\right) - N_C N_V \exp\left(\frac{E_g}{kT}\right) - \Delta n (n_0 + p_0 + \Delta n)$$

Then

$$\frac{dn}{dt} = g - \frac{\gamma_n \gamma_p M (n_o + p_o + \Delta n)}{\gamma_n (n + N_{CM}) + \gamma_p (p + N_{VM})} \Delta n = g - \frac{\Delta n}{\tau_{stat}},$$

where τ_{stat} – is stationary lifetime obtained earlier (see equation (1.6.33)). At the same time, obtained expression corresponds to the expression describing linear recombination (see (1.5.32) and (1.5.33)). This allows us to use all the methods

of experimental τ finding which were discussed earlier. At the same time, it should be noted that the obtained dependence will take place only in the case of absence of charge carriers trapping.

1.7. Diffusion and drift of non-equilibrium charge carriers

1.7.1. Charge carriers transfer in a stationary non-equilibrium state

As was shown in Sections 1.3 and 1.4.2, heating of the crystal leads to generation of free charge carriers, which move chaotically in the crystal lattice. Application of electric field to the crystal results in the appearance of ordered contribution to the movement of charge carriers, i.e. occurrence of electric current. The directed carrier transport, in contrast to their random motion under thermal impact, requires to create certain conditions. These conditions usually lead to the fact that the system cannot be considered as equilibrium.

Corresponding electric current, arising at electric field application, is described by Eq. (1.3.11) and called *drift current*. In semiconductors, ordered movement of charge carriers may be also due to the presence of spatial gradient of their concentration. In this case, a *diffusion current* appears.

Drift current. Suppose that a sufficiently weak electric field with a strength E is applied to semiconductor sample. In this case, the directed movement of free charge carriers along the applied field is observed due to appearance of additional non-zero *average drift velocity* of carriers. The drift contribution to the carriers velocity was described by Eq. (1.3.26) in Section 1.3, which can be transformed for semiconductors in the relations

$$v_n = \frac{e\tau_n}{m_n^*} E = \mu_n E \qquad (1.7.1a)$$

for the drift of electrons, and

$$v_p = \frac{e\tau_p}{m_p^*} E = \mu_p E \qquad (1.7.1b)$$

for the drift of holes. Here τ_n and τ_p are the relaxation times of electrons and holes (or the times between two consecutive scattering events); μ_n and μ_p – electron and hole mobilities and m^*_n and m^*_p – effective masses of electrons and holes.

Taking into account the Eq. (1.3.11), we can define the density of the electron and hole drift currents as

$$J_n^{drift} = env_n = en\mu_n E, \qquad (1.7.2a)$$

and

$$J_p^{drift} = env_p = en\mu_p E \tag{1.7.2b}$$

with appropriate conductivities

$$\sigma_n = \frac{J_n^{drift}}{E} = en\mu_n \tag{1.7.3a}$$

and

$$\sigma_p = \frac{J_p^{drift}}{E} = en\mu_p \tag{1.7.3b}$$

for electrons and holes, correspondingly.

Electron and hole drift velocities are opposite in sign, whereas their current densities will have the same directions with respect to the electric field vector *E* (Fig. 1.7.1). Therefore, when drift currents of electrons and holes summarizing, we can come to the formulation of the Ohm law for semiconductors in the following form

$$J^{drift} = J_n^{drift} + J_p^{drift} = (n\mu_n + p\mu_p)eE = \sigma E. \tag{1.7.4}$$

Note that for doped semiconductors one contribution to the full current relates to majority carriers, while another one – to minority carriers. In doing so, in such crystals drift current of majority carriers is always much larger than the drift current of minority carriers.

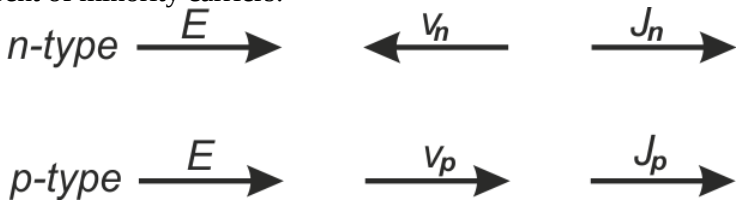


Fig. 1.7.1. Electrons and holes drift velocities, as well as their current densities in semiconductor under exposion of electric field E

In spite of the fact that the occurrence of current is associated with the disturbance of thermodynamic equilibrium in the crystal (by electric field), we shall, however, assume that in stationary conditions the distribution of majority carriers is almost equilibrium. This takes place in case of weak current densities.

Diffusion current. Let us consider that distribution of charge carriers is spatially non-uniform in the crystal due to inhomogeneous generation and other reasons. Then, even without applied external electric field, the directed transfer of carriers will take place from the region with their high concentration to the region where their concentration is low, in addition to their chaotic motion under thermal impact. This ordered carriers flow is called the *diffusion current*.

Density of carriers' diffusion current is proportional to the gradient of their concentration and has a form

$$J_n^{diff} = eD_n \frac{dn}{dx} \tag{1.7.5a}$$

for electrons, and

$$J_p^{diff} = -eD_p \frac{dp}{dx} \tag{1.7.5b}$$

for holes. Here D_n and D_p are diffusion coefficients of electrons and holes, respectively.

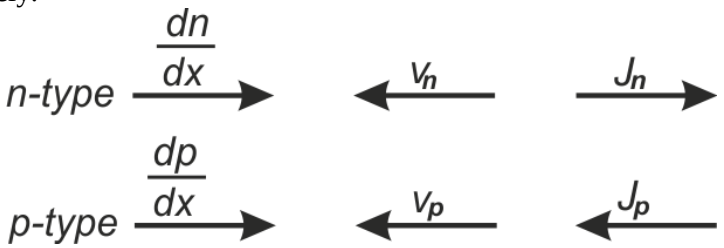


Fig. 1.7.2. Electrons and holes drift velocities, as well as their current densities in semiconductor under exposion of the "buried" fild due to gradient of carriers concentration

It should be noted that, oppositely to the case of carriers drift, the diffusion currents of electrons and holes have the opposite directions, while drift velocities have the same directions. This is due to the fact that the direction of electrons and holes motion is the same, but they have different signs of charge.

The phenomena of diffusion and drift of free charge carriers can arise simultaneously. The resulting current for one-dimensional model equals

$$J_n = J_n^{dr} + J_n^{diff} = e\mu_n nE + eD_n \frac{dn}{dx} \tag{1.7.6a}$$

for electrons, and

$$J_p = J_p^{dr} + J_p^{diff} = e\mu_p pE - eD_p \frac{dp}{dx} \tag{1.7.6b}$$

for holes.

It should be expected, that there is some connection between mobility μ_n, μ_p related to the process of drift, and parameters D_n , D_p which characterize diffusion process, because mobility and diffusivity are determined by the scattering mechanisms of charge carriers.

Indeed, in the isolated sample which is not subjected to external forces, any deviation of the charge carriers concentration from uniform distribution causes a

diffusion current tending to balance the arose concentration gradient. On the other hand, any redistribution of charge in semiconductor will bring occurrence the internal electric fields causing the drift current. In the absence of external electrical circuits (load), these processes are balanced, i.e.:

$$J_{no}^{diff} = -J_{no}^{drift}. \tag{1.7.7}$$

Consequently, in the one-dimensional case diffusion equation for electron semiconductor is

$$D_n \frac{dn}{dx} = -\mu_n n\mathcal{E}, \tag{1.7.8}$$

where $\mathcal{E} = -\frac{d\psi}{dx}$ is the strength of the internal (buried) electric field and ψ is electrostatic potential.

In expressions (1.7.6) – (1.7.8) for currents we used n instead of equilibrium concentration n_0. The fact is that, although we believe that the system is in equilibrium with the environment, but within the system we consider just the case when deviation from equilibrium state occurs due to some fluctuations. Since the potential energy $U = -e\psi$ (in our notation, $e > 0$) of the electrons has changed as a result of application of the electric field, we have the right to write

$$n = N_C e^{\frac{\varepsilon_F - U}{kT}} = n_0 e^{-\frac{U}{kT}}. \tag{1.7.9}$$

From (1.7.9) we find

$$\frac{dn}{dx} = n_0 e^{\frac{e\psi}{kT}} \frac{e}{kT} \frac{d\psi}{dx}. \tag{1.7.10}$$

Substituting n and $\frac{dn}{dx}$ in equation (1.7.8), we obtain

$$D_n n_0 e^{\frac{e\psi}{kT}} \frac{e}{kT} \frac{d\psi}{dx} = \mu_n n_0 e^{\frac{e\psi}{kT}} \frac{d\psi}{dx};$$

$$D_n \frac{e}{kT} = \mu_n.$$

Hence we obtain the known relation, derived by Einstein in the theory of Brownian motion, which is also applicable for electrons and holes:

$$\left.\begin{aligned} \frac{\mu_n}{D_n} &= \frac{e}{kT} \\ \frac{\mu_p}{D_p} &= \frac{e}{kT} \end{aligned}\right| \text{Einstein relations for charge carriers} \tag{1.7.11}$$

These relations connect the diffusion coefficients of charge carriers with their mobilities in their thermodynamic equilibrium. However, they are also applicable in some cases for non-equilibrium processes.

1.7.2. Electronic processes in crystals with gradient of charge carriers concentration.

The continuity equation for NECC

In previous chapters, when considering the generation-recombination processes in crystals, we did not take into account the spatial distribution of NECC suggesting that the observed processes evolve equally in the whole volume of the considered crystal. It is clear that in the most cases it is not true. Let us remember the process of optical absorption which is characterized by the curve of light penetration into medium with the corresponding profile $g = g(x)$ of NECC generation.

Let us introduce the spatial coordinates to discuss non-equilibrium processes in crystal. First, we consider one-dimensional case. We also assume that the concentrations of free electrons and holes will be not only a functions of time but also the functions of the coordinates: $n = n(x,t)$ and $p = p(x,t)$.

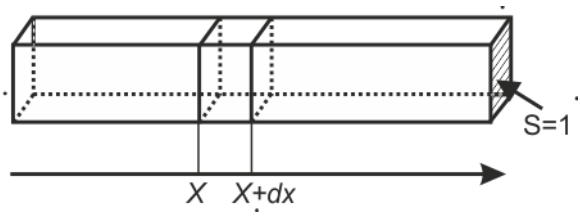


Fig. 1.7.3. One-dimensional model of diffusion and drift in a crystal

When at time t the concentration of electrons was $n = n(x, t)$ and at time $t + dt$ their concentration is equal $n = n(x, t + dt)$, so that the change in the concentration (the number in a unit volume) becomes

$$n(x,t+dt)dx - n(x,t)dx = \frac{dn}{\partial t}dxdt .$$

Changing the concentration of carriers in the selected volume per selected period of time may be due to processes of NECC generation and recombination discussed earlier and additionally as a result of their migration in the crystal.

We assume that the change in the number of NECC in the selected volume per time dt is:

the result of the generation $gdxdt$;

the result of the recombination $rdxdt = \frac{\Delta n}{\tau_n} dxdt$,

and similarly $\frac{\Delta p}{\tau_p} dxdt$;

the result of the migration the carriers through the crystal.

Let $I_n(x,t)$ is the number of electrons passing through a unit area at the point x and $I_n(x + dx,t)$ is their number passing through a unit area at the point $x + dx$. Then the change of the number of carriers in the volume dx per time dt will be:

$$I_n(x,t) - I_n(x+dx,t) = -\frac{\partial I_n}{\partial x} dxdt .$$

Complete change of the number of electrons in the selected volume will be:

$$\frac{\partial n}{\partial t} dxdt = gdxdt - \frac{\Delta n}{\tau_n} dxdt - \frac{\partial I_n}{\partial x} dxdt .$$

Hence:

$$\frac{\partial n}{\partial t} = g - \frac{\Delta n}{\tau_n} - \frac{\partial I_n}{\partial x} .$$

A similar expression can be written for holes:

$$\frac{\partial p}{\partial t} = g - \frac{\Delta p}{\tau_p} - \frac{\partial I_p}{\partial x} .$$

Taking into account the expressions for the current density $j_n = - eI_n$ and $j_p = eI_p$ we can write

$$\left.\begin{aligned} \frac{\partial n}{\partial t} &= g + \frac{1}{e}\frac{\partial j_n}{\partial x} - \frac{\Delta n}{\tau_n} \\ \frac{\partial p}{\partial t} &= g - \frac{1}{e}\frac{\partial j_p}{\partial x} - \frac{\Delta p}{\tau_n} \end{aligned}\right\} \qquad (1.7.12)$$

For the three-dimensional case this equation looks like

$$\left.\begin{aligned} \frac{\partial n}{\partial t} &= g + \frac{1}{e}\underbrace{\left(\frac{\partial j_n}{\partial x} + \frac{\partial j_n}{\partial y} + \frac{\partial j_n}{\partial z}\right)}_{div(j_n)} - \frac{\Delta n}{\tau_n} \\ \frac{\partial p}{\partial t} &= g - \frac{1}{e}\underbrace{\left(\frac{\partial j_p}{\partial x} + \frac{\partial j_p}{\partial y} + \frac{\partial j_p}{\partial z}\right)}_{div(j_p)} - \frac{\Delta p}{\tau_p} \end{aligned}\right\} \qquad (1.7.13)$$

These equations are called the *continuity equations* because they reflect condition of charge carriers conservation. At the same time, these equations do not contain any information about the reasons of an electric current appearance. In section 1.7.1 it has been shown that there are two such reasons: *diffusion* and *drift*.

The diffusion of particles is going due to their concentration gradient. Diffusion is not a consequence of the influence of any forces. It is a purely statistical phenomenon. Particles pass from the region of space where their concentration is larger to the region with a lower concentration. This redistribution appears as diffusion current which is always proportional to the gradient of particles concentration

$$\begin{aligned} I_n &= -D_n \frac{dn}{dx}; \\ I_p &= -D_p \frac{dp}{dx}, \end{aligned} \tag{1.7.14}$$

where D_n and D_p are the diffusivities of electrons and holes, respectively, in the crystal.Diffusion current of electrons and holes can be written using the term of the electric current density as

$$\left.\begin{aligned} j_n^{diff} &= eD_n \frac{dn}{dx}; \\ j_p^{diff} &= -eD_p \frac{dp}{dx}. \end{aligned}\right\} \tag{1.7.15}$$

for one-dimensional case, or

$$\left.\begin{aligned} \vec{j}_n^{diff} &= eD_n \overrightarrow{grad}(n); \\ \vec{j}_p^{diff} &= -eD_p \overrightarrow{grad}(p). \end{aligned}\right\} \tag{1.7.16}$$

for three-dimensional case.

If the crystal is placed in an external electric field the drift current appears. The density of this drift current is equal

$$\left.\begin{aligned} \vec{j}_n^{drift} &= en\mu_n \vec{E}, \\ \vec{j}_p^{drift} &= ep\mu_p \vec{E}. \end{aligned}\right\} \tag{1.7.17}$$

It should be noted that both components in the expression for the drift current have the same sign. This is due to the fact that μ_n and μ_p have different signs by definition. Total current through the sample is the sum of the diffusion and drift currents:

$$j_n = j_n^{diff} + j_n^{drift} \tag{1.7.18}$$

Thus, a complete solution for the distribution of charge carriers in the crystal is possible when the corresponding equations for the density currents are joined with the continuity equations for electrons and holes.

1.7.3. Diffusion and drift of NECC in the case of unipolar conductivity

We shall assume that majority charge carriers are uniformly distributed throughout the sample volume in the crystals with a pronounced p-type (for concreteness) conductance. The movement of carriers takes place in auniform electric field. In this case we can consider only the equations for the minority carriers for the description of NECC motion assuming that the change in the concentration and spatial distribution of the majority carriers is small.

Let us consider the problem when a long thin crystal with p-type conductivity is extended along the X axis (Fig. 1.7.4). At the point $X = 0$ the crystal is illuminated by a narrow strip of light from fundamental absorption region. We consider LEL case with the condition $\Delta n = \Delta p << p_o$.

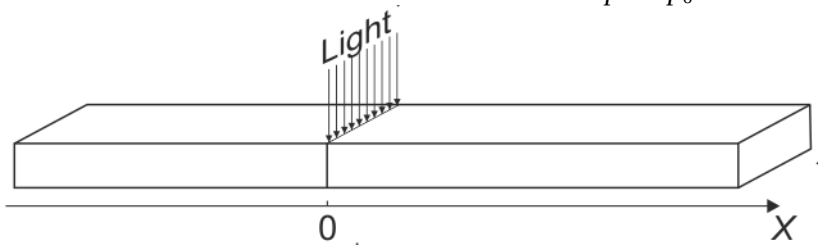


Fig. 1.7.4. One-dimensional model of diffusion and drift of charge carriers under optical excitation in a point $X = 0$

We will look for $\Delta n = f(x)$. Generation of NECC is absent in the crystal regions with $x \neq 0$, i.e. in these areas $g = 0$. The continuity equation for these areas in the stationary case will be of the type:

$$\frac{\partial n}{\partial t} = \frac{1}{e}\frac{\partial j_n}{\partial x} - \frac{\Delta n}{\tau_n} = 0 .$$

The total electron current is

$$j_n = j_n^{drift} + j_n^{diff} = en\mu_n E + eD_n \frac{dn}{dx} .$$

Let us find the value of $\frac{\partial j_n}{\partial x}$:

$$\frac{dj_n}{dx} = e\mu_m E\frac{dn}{dx} + eD_n\frac{d^2n}{dx^2}.$$

Note that an electric field applied to the crystal is constant and $\frac{\partial E}{\partial x} = 0$ due to the homogeneity of the majority carriers distribution in it.

Before solving the complete problem, we consider two limiting cases: *E* is lack and *E* is strong.

1. No external electric field. In this case, the electron current is determined only by the diffusion component. The continuity equation is of the type:

$$D_n\frac{d^2(\Delta n)}{dx^2} - \frac{\Delta n}{\tau_n} = 0 \text{ or } \frac{d^2(\Delta n)}{dx^2} - \frac{\Delta n}{D_n\tau_n} = 0. \qquad (1.7.19)$$

The solution of this equation is

$$\Delta n(x) = C_1 e^{\frac{x}{\sqrt{D_n\tau_n}}} + C_2 e^{\frac{-x}{\sqrt{D_n\tau_n}}}. \qquad (1.7.20)$$

Constants C_1 and C_2 are found from the following considerations. First, our solution is not continuous, as we have removed the point $x = 0$. Second, in the physical sense, function $\Delta n = f(x)$ should decrease with increasing distance from the point $x = 0$. So, the solution of the type $\Delta n(x) = Ce^{\frac{x}{\sqrt{D_n\tau_n}}}$ cannot exist for $x > 0$. Accordingly, the solution can be expressed like this:

$$\begin{aligned} \Delta n &= C_1 e^{\frac{x}{\sqrt{D_n\tau_n}}} \Rightarrow (x < 0) \\ \Delta n &= C_2 e^{\frac{-x}{\sqrt{D_n\tau_n}}} \Rightarrow (x > 0) \end{aligned} \qquad (1.7.21)$$

We also take into account the fact that if $x \to 0$ then $\Delta n \to C_1$ or $\Delta n \to C_2$. From this condition we conclude that $C_1 = C_2 = \Delta n(0)$.

Let us also remember that the optical generation of NECC takes place only in the $x = 0$ region. Thus, the general solution of the equation (1.7.19) can be written as:

$$\Delta n(x) = \Delta n(0)\exp\left(\frac{-|x|}{\sqrt{D_n\tau_n}}\right) \qquad (1.7.22)$$

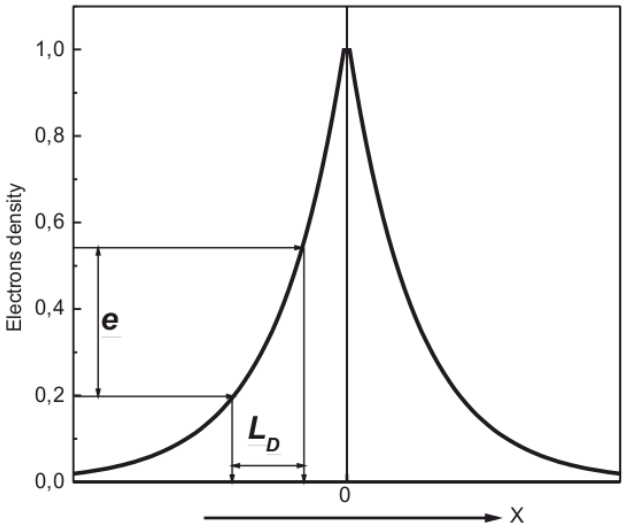


Fig. 1.7.5. Distribution of electrons in the sample (for the scheme in Fig. 7.1.4)) at presence merely diffusions of charge carriers

The value

$$L_{Dn} = \sqrt{D_n \tau_n} \tag{1.7.23}$$

has the dimension of length and is called the length of diffusion for electrons. It determines the distance at which excess NECC concentration decreases in *e* times. At room temperature in silicon for electrons

$$D_n = \frac{kT}{e}\mu_n = \frac{0.026eV}{1e} \cdot 1350 \frac{cm^2}{V \cdot s} \approx 35 \frac{cm^2}{s},$$

and for holes:

$$D_p = \frac{kT}{e}\mu_p = \frac{0.026eV}{1e} \cdot 470 \frac{cm^2}{V \cdot s} \approx 12 \frac{cm^2}{s}.$$

It follows that if the lifetime τ is equal, for example, 1 μs then the diffusion length $L_D = \sqrt{D\tau}$ will have values of about $6 \cdot 10^{-3}$ cm = 60 μm for electrons and $3.5 \cdot 10^{-3}$ cm = 35 μm for holes.

2. Strong electric field. Let us consider the case of the influence of the drift component of the electric current on charge carriers movement in the one-dimension sample. Strictly speaking, this approach has no physical meaning because, in contrast to the drift current, that we can turn on or off we cannot turn off the diffusion of charge carriers. Some real plausibility can be achieved if we assume that the drift component of the current is so much more than

diffusion component that the latter can be neglected. We assume in this case that the electron current is determined only by the drift component. The continuity equation for this case takes the form

$$\mu_n E \frac{d(\Delta n)}{dx} - \frac{\Delta n}{\tau_n} = 0 \text{ or } \frac{d(\Delta n)}{dx} - \frac{\Delta n}{\mu_n \tau_n E} = 0. \quad (1.7.24)$$

The solution of this equation is:

$$\Delta n(x) = \Delta n(0) \exp\left(\frac{-x}{\mu_n \tau_n E} \right) \quad (1.7.25)$$

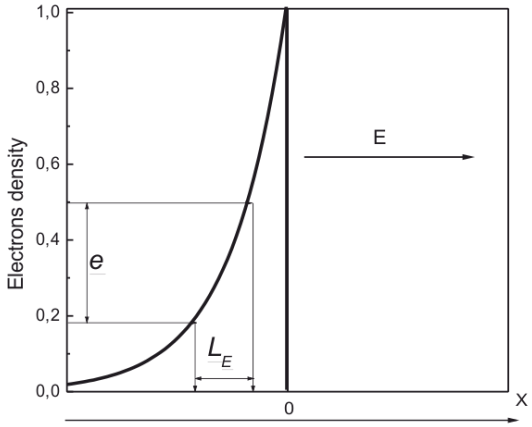


Fig. 1.7.6. Distribution electrons in the sample (the scheme in Fig. 1.7.4) at the presence merely the drift of charge carriers

As in the previous case, the denominator of the exponent has the dimension of length. Let us call it the length of the drift for electrons:

$$L_{En} = \mu_n \tau_n E. \quad (1.7.26)$$

It is the distance, passed by NECC under impact of an electric field *E*, when concentration of electrons is decreased due to the recombination in *e* times. Note that the length of the drift has no such a determinative importance as the diffusion length because the former depends on the magnitude of the electric field.

The obtained dependence indicates that, as the result of the drift, the carriers (in this case – the holes) are shifted with respect to the generation region toward one end of the sample. The concentration decrease is an exponential function of this displacement. However, drift displacement is asymmetric in contrast to

a diffusion displacement. Under an electric field holes move along the field and the electrons – against the field.

Let us now consider the general problem of diffusion and drift of NECC. In this case,

$$\frac{dj_n}{dx} = e\mu_n E \frac{d(\Delta n)}{dx} + eD_n \frac{d^2(\Delta n)}{dx^2}. \tag{1.7.27}$$

Substituting this expression into the continuity equation we obtain:

$$D_n \frac{d^2(\Delta n)}{dx^2} + \mu_n E \frac{d(\Delta n)}{dx^2} - \frac{\Delta n}{\tau_n} = 0. \tag{1.7.28}$$

Let us transform the equation to the form:

$$\begin{aligned} &\frac{d^2(\Delta n)}{dx^2} + \frac{\mu_n E}{D_n}\frac{\tau_n}{\tau_n}\frac{d(\Delta n)}{dx^2} - \frac{\Delta n}{D_n \tau_n} = \\ &= \frac{d^2(\Delta n)}{dx^2} + \frac{L_E}{L_D^2}\frac{d(\Delta n)}{dx^2} - \frac{\Delta n}{L_D^2}0 \end{aligned} \tag{1.7.29}$$

This equation is of the type *y" + py' + qy = 0,* where $p = \frac{L_E}{L_D^2}$ and $q = -\frac{1}{L_D^2}$.

The solution of this equation will be $\Delta n(x) = C_1 e^{r_1 x} + C_2 e^{r_2 x}$, where r_1 and r_2 are:

$$r_{1,2} = -\frac{L_E}{L_D^2} \pm \sqrt{\frac{L_E^2}{L_D^4} + \frac{1}{L_D^2}} = \frac{-L_E \pm \sqrt{L_E^2 + L_D^2}}{L_D^2}.$$

Following the same considerations as for derivation of the formula (1.7.25), for decrease of $\Delta n(x)$ function with increasing distance from the excitation region, as well as from the conditions of the discontinuity of the $\Delta n(x)$ function at $x = 0$, we obtain:

$$\begin{aligned} &\Delta n(x) = \Delta n(0) \exp\left(\frac{-L_E - \sqrt{L_E^2 + L_D^2}}{L_D^2} x \right) \Rightarrow x > 0; \\ &\Delta n(x) = \Delta n(0) \exp\left(\frac{-L_E + \sqrt{L_E^2 + L_D^2}}{L_D^2} x \right) \Rightarrow x < 0. \end{aligned} \tag{1.7.30}$$

This formula is transformed into (1.7.22) without an electric field.

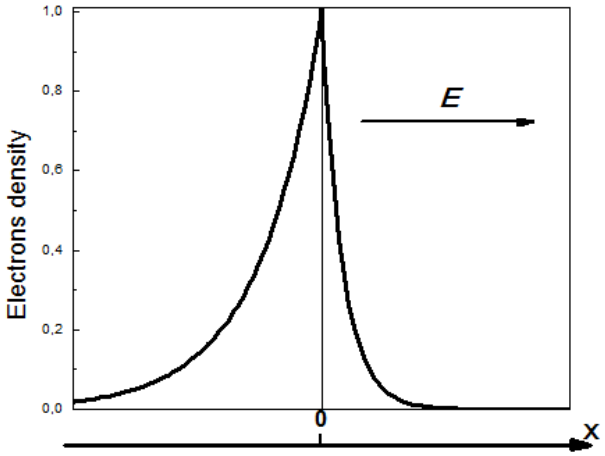


Fig. 1.7.7. Distribution of electrons in the sample (the scheme in Fig. 7.1.4) at presence of both diffusion and drift of charge carriers

As is seen in Fig. 1.7.7, in the presence of strong electric field our problem gives an asymmetric distribution of carriers with respect to the excitation region. The greater the electric field, the greater is the difference between left and right branches of $\Delta n(x)$. The proposed scheme of the experiment and carried out calculation is quite concrete, but the results have general character. In particular, the diffusion length L_D and drift length L_E do not depend on the geometry of the experiment.

1.7.4. Diffusion and drift of NECC in the case of bipolar conductivity

In the previous section we discussed the processes of NECC diffusion and drift in a crystal which has a pronounced impurity conductance. In this section we will consider these processes in the crystal close to the intrinsic, i.e. in a crystal with a small (and a little different) concentration of electrons and holes. We choose the same experimental geometry as in the case of unipolar conductivity, when a narrow strip of long thin crystal is illuminated. Then

$g \neq 0$ at $x = 0$;

$g = 0$ at $x \neq 0$.

When considering a crystal with intrinsic conductivity, it is necessary to analyze the motion of NECC of both signs. In this case,

$$j = j_n + j_p = e(n\mu_n + p\mu_p)E + e\left(D_n \frac{dn}{dx} - D_p \frac{dp}{dx} \right) \qquad (1.7.31)$$

The process of diffusion in such crystal is a diffusion of electron-hole pairs and involves their Coulomb interaction. At the same time, the carriers drift is carried out separately, since electrons and holes move in opposite directions in the electric field. Collective motion of electron-hole pairs will be characterized by effective values of diffusion coefficient D and drift mobility $\mu_D = \frac{e}{kT} D$.

These values are called bipolar diffusion coefficient and bipolar diffusion mobility. To describe the collective motion of the electrons and holes, we need to solve two continuity equations and two equations for the current densities of electrons and holes:

$$\left| \begin{array}{l} \dfrac{\partial n}{\partial t} = g + \dfrac{1}{e}\dfrac{\partial j_n}{\partial x} - \dfrac{\Delta n}{\tau_n} \\ \dfrac{\partial p}{\partial t} = g - \dfrac{1}{e}\dfrac{\partial j_p}{\partial x} - \dfrac{\Delta p}{\tau_p} \\ j_n = en\mu_n E + eD_n \dfrac{dn}{dx} \\ j_p = ep\mu_p E - eD_p \dfrac{dp}{dx} \end{array} \right| \qquad (1.7.32)$$

Here we shall solve the stationary problem, assuming that $\frac{\partial n}{\partial t} = \frac{\partial p}{\partial t} = 0$. The solution is looked for in the areas with $x \neq 0$, where $g = 0$.

Let us differentiate j_n and j_p with respect to x. It is necessary to consider the following. We have concluded that in the case of low equilibrium concentration of charge carriers and excitation of an equal number of non-equilibrium electrons and holes (due to band-to-band transitions) the movement of carriers with the same sign will cause the motion of carriers with the opposite sign due to the Coulomb interaction. In this case, if the mobility of electrons and holes are not equal to each other (usually, $\mu_n > \mu_p$), holes will lag behind electrons that cause the occurrence of the internal electric field gradient. This field will "speed up" lagging holes and "brake" runaway electrons. Thus, the resulting electric field causing NECC drift will not be homogeneous, i.e. $E \neq const$. Then

$$\frac{\partial j_n}{\partial x} = e\mu_n E\frac{\partial n}{\partial x} + e\mu_n n\frac{\partial E}{\partial x} + eD_n\frac{\partial^2 n}{\partial x^2}$$
$$\frac{\partial j_p}{\partial x} = e\mu_p E\frac{\partial p}{\partial x} + e\mu_p p\frac{\partial E}{\partial x} - eD_p\frac{\partial^2 p}{\partial x^2}. \qquad (1.7.33)$$
$$\sigma_n \qquad \sigma_p$$

It is necessary to find $\frac{\partial E}{\partial x}$. Since $E = -\frac{d\psi}{dx}$, then

$$\frac{\partial E}{\partial x} = -\frac{\partial^2\psi}{\partial x^2} = \frac{\rho}{\varepsilon\varepsilon_0} = -\frac{e(\Delta n - \Delta p)}{\varepsilon\varepsilon_0}.$$

Substituting this into the equation of continuity, we get

$$D_n\frac{d^2 n}{dx^2} + \mu_n E\frac{dn}{dx} - \sigma_n\frac{\Delta n - \Delta p}{\varepsilon\varepsilon_0} - \frac{\Delta n}{\tau_n} = 0;$$
$$D_p\frac{d^2 p}{dx^2} - \mu_p E\frac{dp}{dx} + \sigma_p\frac{\Delta n - \Delta p}{\varepsilon\varepsilon_0} - \frac{\Delta p}{\tau_p} = 0. \qquad (1.7.34)$$

We assumed $\Delta n = \Delta p$ at $x = 0$ and believe that any change of Δn or Δp quickly leads to the same change of NECCs of opposite sign. In this case, however, the term

$$\frac{\Delta n - \Delta p}{\varepsilon\varepsilon_0}.$$

is undefined. We resolve this uncertainty as follows. Let us multiply the first equation (1.7.34) on σ_p, the second on σ_n and add them to each other. Members with uncertainty will be cancelled because they have opposite signs. We obtain:

$$\sigma_p D_n\frac{d^2 n}{dx^2} + \sigma_p\mu_n E\frac{dn}{dx} - \sigma_p\frac{\Delta n}{\tau_n} + \sigma_n D_p\frac{d^2 p}{dx^2} -$$
$$-\sigma_n\mu_p E\frac{dp}{dx} - \sigma_n\frac{\Delta p}{\tau_p} = 0. \qquad (1.7.35)$$

Now, since we got rid of the uncertainty, we can assume that $\Delta n = \Delta p$ at any point and therefore

$$\frac{d^2 n}{dx^2} = \frac{d^2 p}{dx^2} = \frac{d^2(\Delta n)}{dx^2}$$

and

$$\frac{dn}{dx} = \frac{dp}{dx} = \frac{d(\Delta n)}{dx}.$$

Besides, we believe $\tau_n = \tau_p = \tau$. Given this:

$$\frac{\sigma_p D_n + \sigma_n D_p}{\sigma_n + \sigma_p} \frac{d^2(\Delta n)}{dx^2} + \frac{\sigma_p \mu_n - \sigma_n \mu_p}{\sigma_n + \sigma_p} \frac{d(\Delta n)}{dx} - \frac{\Delta n}{\tau} = 0 \qquad (1.7.36)$$

The resulting equation will be identical in form with (1.7.28), which describes the unipolar conductance, if we put:

$$D = \frac{\sigma_p D_n + \sigma_n D_p}{\sigma_n + \sigma_p} = \frac{ep\mu_p D_n + en\mu_n D_p}{e\mu_n n + e\mu_p p} = \left| \underbrace{\mu_p D_n = \mu_n D_p}_{Einstein\_relations} \right| = \frac{n+p}{\frac{n}{D_p} + \frac{p}{D_n}}; \qquad (1.7.37)$$

$$\mu_E = \frac{\sigma_p \mu_n - \sigma_n \mu_p}{\sigma_n + \sigma_p} = \frac{e\mu_p p\mu_n - e\mu_n n\mu_p}{e\mu_n n + e\mu_p p} = \frac{p-n}{\frac{n}{\mu_p} + \frac{p}{\mu_n}}. \qquad (1.7.38)$$

The value of D is called bipolar diffusion coefficient and μ_E is the bipolar drift mobility. There is another physical quantity, called bipolar diffusion mobility. It can be obtained as follows. Sometimes, it can be used mobility, instead of the diffusion coefficient, expressing it from the Einstein relations:

$$D_n = \frac{kT}{e}\mu_n \text{ and } D_p = \frac{kT}{e}\mu_p.$$

Believing $\mu = \frac{e}{kT} D$ and multiplying (1.7.37) by $\frac{e}{kT}$, we obtain:

$$D \cdot \frac{e}{kT} = \frac{ep\mu_p D_n + en\mu_n D_p}{e\mu_n n + e\mu_p p} \cdot \frac{e}{kT} = \frac{ep\mu_p D_n \cdot \frac{\mu_n}{D_n} + en\mu_n D_p \cdot \frac{\mu_p}{D_p}}{e\mu_n n + e\mu_p p} =$$

$$= \frac{ep\mu_p \mu_n + en\mu_n \mu_p}{e\mu_n n + e\mu_p p} = \frac{n+p}{\frac{n}{\mu_p} + \frac{p}{\mu_n}} = \mu_D$$

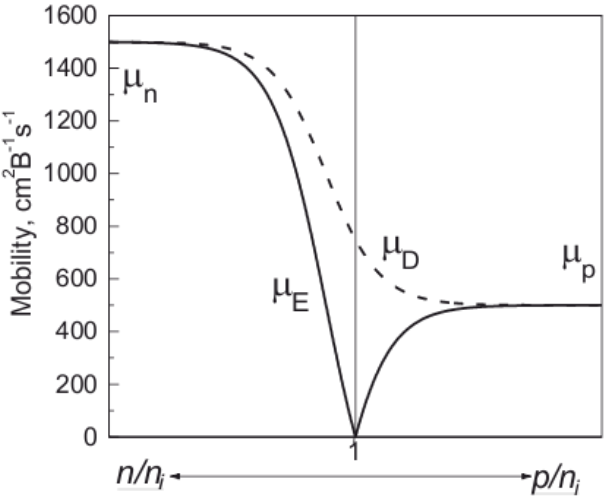


Fig. 1.7.8. Dependence of bipolar diffusion mobility (1.7.39) and bipolar drift mobility (1.7.38) on relative carrier concentration

As we can see, the bipolar diffusion mobility differs from bipolar drift mobility by a sign in the numerator. The physical sense of this difference between the diffusion and drift bipolar mobilities is that the former describes the diffusion process in the intrinsic (or close to intrinsic) crystal and the latter describes the process of drift in the electric field.

It is easy to see that in intrinsic crystal, where $n = p$, bipolar drift mobility $\mu_E = 0$ and bipolar diffusion mobility will be of the type

$$\mu_D = \frac{n + p}{\frac{n}{\mu_p} + \frac{p}{\mu_n}} = \frac{2\mu_n \mu_p}{\mu_n + \mu_p}. \qquad (1.7.39)$$

In the heavily doped crystal μ_E and μ_D coincide with the corresponding values of mobilities μ_n or μ_p. For silicon, the dependence of μ_E and μ_D on the doping level is shown in Fig. 1.7.8.

Thus, in a crystal with a conductivity close to the intrinsic, processes of diffusion and drift differ substantially from similar processes in crystals with a high concentration of carriers with the same sign. In the case of *bipolar diffusion*, carriers move in pairs coupled by the Coulomb interaction. Their velocity for strictly intrinsic crystal (and, therefore, the diffusion mobility) is the geometric mean value of the velocities of the electron and hole. *Bipolar drift* in the intrinsic crystal is equal to zero. This is because the quantity of electrons

precisely equals to the quantity of holes. They are coupled by the Coulomb interaction forming electrically neutral complexes that are not involved in the electrical conductivity. Further solving the problem of diffusion and drift of NECCs in a bipolar conductivity is similar to that in the previous paragraph.

1.8. Properties of crystalline solids

Energy-efficient materials are quite varied by both the properties and functions. However, for convenience they can be conditionally divided into two major classes – constructional and special. The special materials include materials with specific functional properties (mechanical, electric, magnetic, conducting, insulating, optic, etc.), and in particular, crystalline materials with speciefic behavior when exposed to temperature as well as electrical and magnetic fields, irradiation, mechanical impacts, etc.

1.8.1. Electrical properties of metallic crystals

By magnitude of the electrical resistivity at ambient temperature (Fig. 1.10.1) and its behavior in electric field materials is usually divided into conductor, semiconductor and dielectric.

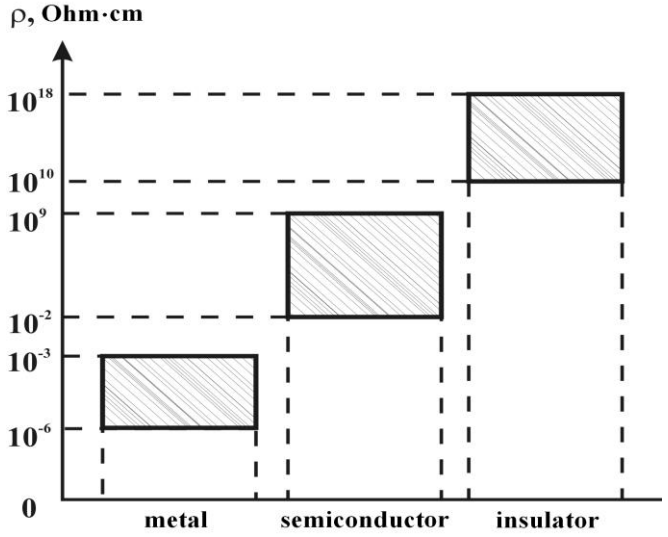


Fig. 1.10.1. Ranges of resistivity at room temperature for different types of materials

At room temperature the metallic conductors are characterized by high electrical conductivity (low resistivity). In dielectrics, except for a very high resistivity, the main feature is the ability to polarization and the possibility of

the presence of internal electric field in them. Semiconductors are intermediate by the values of the conductivity of materials between dielectrics and conductors. And particularly, strong dependence of conductivity on the concentration and type of impurities and defects, and, in most cases, the external impacts (temperature, radiation, light, etc.) are characteristic for them. The specified gradation of crystals on the conductivity is a logical consequence of the zone nature of the electrons energy spectrum in crystals, and the difference in filling of V-band by electrons.

By the nature of the response to the impact of a magnetic field, it is accepted to classify crystalline materials as diamagnetic, paramagnetic and ferromagnetic. According to functional properties of magnetic materials, they are usually divided into the non-magnetic (diamagnetic and paramagnetic) and ferromagnetic, and the last – on the soft and hard ferromagnets.

A special type of response to the action of the electric and magnetic fields (as well as temperature) have superconducting materials, which in the normal (non-superconducting) state can relate to metals and to semiconductors or insulators. Below certain temperatures such substances pass into a state where the electric resistance becomes zero and the magnetic flux is ejected from the sample (ideal diamagnetism).

Conducting materials. Discussed in section 1.3 the classical and quantum representations of the energy spectrum of electrons in crystalline solids, allowed to get similar by form but different by the physical meaning equations for the conductivity of free electron gas. As shown in quantum conductivity models, the formula (1.3.45) includes the electron relaxation time (1.3.45), rather than the mean free time of electrons (1.3.9), as is in the Drude-Lorentz model. Note also that this relaxation time applies only to those electrons which lie around the Fermi level.

Note that in the Sommerfeld model, when deriving these relations, we did not use any assumptions concerning the mechanisms of energy exchange between electrons, accelerated by the electric field, and the atoms of the crystal lattice, defects, and other electrons. This did not allow to determine the temperature dependence of the conductivity of metallic crystals in the framework of this model. This problem was solved by Gruneisen on the basis of Sommerfeld model by estimating mean free path in assumption that scattering of electrons by vibrating atoms and fixed defects (impurity atoms) are the only mechanism for the energy exchange between electrons and crystal lattice. In doing so, it is assumed that the atoms and lattice defects scatter elastically and independently moving electrons. Then, in accordance with the Sommerfeld model, the collision frequency (probability of elastic scattering) of an electron with defects and oscillating atoms consists of two components:

$$\frac{1}{\lambda} = \frac{1}{\lambda_{def}} + \frac{1}{\lambda_T}. \qquad (1.8.1)$$

The first term in this equation is independent of temperature. The second, temperature-dependent contribution is determined by the scattering of electrons by phonons (vibrating atoms) and depends on their concentration.

From (1.8.1) for the resistivity of metal crystals with defects and impurities follows the so-called Matthiessen's rule:

$$\rho = \frac{1}{\sigma} = \frac{1}{\sigma_{def}} + \frac{1}{\sigma_T} = \rho_o + \rho_T(T). \qquad (1.8.2)$$

In this rule, the temperature-dependent term of the resistivity obeys the known GruÈneisen relation

$$\rho_T(T) = C\left(\frac{T}{\theta_D}\right)^5 \int_0^x \frac{x^3 dx}{(e^x - 1)(1 - e^{-x})}, \qquad (1.8.3)$$

where $x = \theta_D/T$ and C is constant. konstanta. Here θ_D is Debye characteristic temperature, G(0) – the matrix element of the electrons scattering by phonons, k_F – Fermi wave vector for the electron (the radius of the Fermi sphere), q_D – Debye phonon momentum (momentum of phonons with a characteristic Debye frequency). The reasons for this temperature dependence of the resistivity of the type (1.8.6) will be discussed below, when considering the mechanisms of electrons scattering in the crystal. Equation (1.8.6) gives the dependence of the type $\rho_{ph} \sim T^n$ with $n = 5$ below the Debye temperature and $\rho_{ph} \sim T$ ($n = 1$) for high temperatures $T >> \theta_D$, that is confirmed by experimental curves in Fig. 1.8.2.

As can be seen from the Fig. 1.8.2a, at very low temperatures, when the influence of thermal vibrations on the scattering of electrons can be neglected, the resistivity of the metal becomes virtually independent of temperature. The limit value ρ_0 for the resistivity of metals with temperature decreasing to absolute zero, is called the *residual electrical resistivity*. This is a very important characteristic sensitive to the concentration of lattice defects.

A comparison of the curves, shown in Fig. 1.8.2b, displays that the contributions to the resistivity of the conductor on the dynamic and static defects can actually be considered as additive. Thus, the residual resistivity of the conductor ρ_0 is usually proportional to the concentration of impurity atoms or defects (for their small contents).

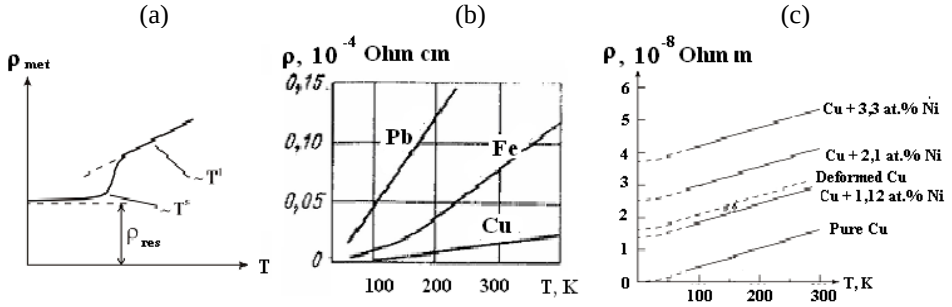


Fig. 1.8.2. The temperature dependences of resistivity by Gryunayzen law (*a*) and the experimental curves for pure metals (*b*) and copper-based alloys with nickel (*c*)

Measurements of Hall effect as a function of temperature allows us to estimate the concentration and mobility of electrons. Their qualitative form for metals is shown in Fig. 1.8.3.

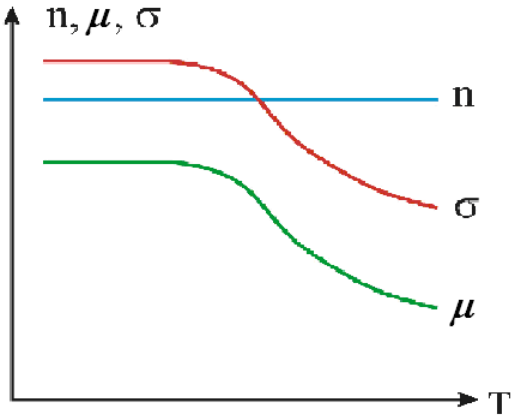


Fig. 1.8.3. Dependences of the carrier concentration n, mobility μ and conductivity σ on the temperature T for a metal

Comparison of the curves in Fig. 1.8.3 shows that the concentration of electrons in metals is not changed with temperature, so that the temperature dependence of the electrical conductivity is determined only by the change of mobility (which is inversely proportional to the relaxation time) of the electrons with the temperature. This means that the temperature dependence of the electrical conductivity of metallic crystals is determined by the mechanisms of electron scattering (or more exactly, by temperature dependence of the electrons relaxation time). Among all the existing mechanisms in the metallic crystals, the

mostly important are electron-phonon, electron-electron and electron-impurity scattering of electrons. In some cases, collisions of electrons with defects like dislocations, surfaces, grain and interphase boundaries may play a significant role.

Scattering by impurity ions in the metallic crystals is most often considered as elastic, and the impurity concentration n_i and the energy, which they are exchanged, are small. In the theory of collisions, their probability is characterized by the *scattering cross section Q*. Since the scattering of electrons by ions is considered to be elastic, the process can be represented as a collision of elastic spheres with a cross section $\sim r_D^2$, where r_D is the so-called *Debye length* of *screening* ($\sim 10^{-8}$ m) of the impurity atom electrostatic field. In this case, the cross-section is equal to $Q_{eff} \sim r_D^2 \sim 10^{-16}$ cm^2, and the mean free path of conduction electrons $l \sim 1/N_i Q_{eff}$. Then, to estimate the conductivity value we can use relation

$$\sigma \approx \frac{n\, e^2 \tau}{m} \approx \frac{n\, e^2 l}{m v_F} \approx \frac{n\, e^2}{m v_F N_i Q_{eff}} \tag{1.8.4}$$

where m is effective mass of electron, n_e – concentration of electrons.

Given that the ratio of concentration of electrons n_e and impurity atoms of i-th sort n_i is always more than unit ($n_e/N_i > 1$) and that $e^2/\hbar v_{mean} \sim 1$, we obtain the following expression for the conductivity

$$\sigma = n\hbar/(mN_i Q_{eff}) , \tag{1.8.5}$$

from which we estimate that $\sigma \approx 10^{-8}/N_i\ \Omega^{-1}\text{cm}^{-1}$. Accordingly, for resistivity we obtain $\rho \approx 10^8 N_i\ \Omega\cdot$cm. A comparison with the characteristic values of resistivity for different metals shows that scattering by impurity atoms can give a significant contribution to the total resistance of the metallic crystal.

As follows from experiments and theory, the contribution of *electron-electron scattering* to the conductivity is equal

$$\sigma \approx \frac{ne^2\tau}{m} \approx \frac{ne^2\hbar E_F}{A(k_B T)^2}, \tag{1.8.6}$$

where A is constant. Assuming that at room temperature $k_B T \approx 10^{-2}$ eV and $E_F \approx 1$ eV, we find that the contribution of the electron-electron scattering processes to the electrical resistivity is very small.

The processes of *electron-phonon scattering* are considered as inelastic. It is also assumed that at the electron-phonon scattering only one phonon can be emitted or absorbed. This means that, when $T >> \theta_D$, contribution of the electrons scattering by phonons to the electrical resistivity results in the following relation for the resistivity

$$\rho \approx m/ne^2\tau \approx mk_B T/ne^2\tau \tag{1.8.7}$$

At low temperatures, $T << \theta_D$, it is necessarily to take into account that actually not all phonons, with wave vectors allowed by the conservation laws (their number is proportional to T^2), can participate in the processes of absorption and emission of electrons. Given that, the relaxation time for the electron-phonon collisions is inversely proportional to T^3 (due to the Debye law for the concentration of phonons at low temperatures,) at $T << \theta_D$ we obtain $1/\tau \sim \rho \sim T^5$. As noted above, for the case of $T << \theta_D$, this law follows from Gruneisen relation (1.8.3).

Applications of electrically conductive materials. In power technologies and energy saving, various types of metallic materials are widely used. They are presented at the scheme in Fig. 1.8.4. These materials include pure metals with high electrical conductivity (low values of resistivity), their alloys and non-metallic compositions with substantially higher values of resistivity. First need to create the most highly conductive (with low Joule losses) electrical wires, cables and other electrical items, and the second – for various heaters (in furnaces), temperature sensors, etc.

The most of applied metallic conductive materials (mostly pure metals) has a conductivity in the range (0.015 ÷ 0.03) μΩ·m. For high-resistive metallic alloys ρ value ranges (0,4 ÷ 2) μΩ·m. Most pure metals and alloys can be used in an ambient air at temperatures not higher than 200°C and some up tu 500°C. Above these temperatures, the oxide films are formed at the conductor surface. These oxides have typically a loose structure, so that the oxygen, penetrating through it, prolongs to oxidize item. Modern control electronics needs in conducting materials which are not oxidized by atmospheric oxygen at temperatures up to 800-1000°C.

To create electrotechnical materials with the lowest electrical resistance, it is necessary to provide the minimal concentration of defects in crystal structure (in particular, impurities). To reduce the density of defects (vacancies, interstitials, dislocations) and lower the ρ values, additional annealing of metals is used. To reduce resistivity of the electrical conductors in high-power systems (transformers, motors, etc.), their cooling is often used, for example, by forced-air cooling, circulation of heat-removal liquids (oil, water) around or within the wire or cable, embedding of wires in liquid nitrogen, etc. Copper and aluminum (and in some cases, high-purity silver) are commonly used as the mostly efficient conductors.

To create conductors with the highest electrical resistance, one should use either metals with the maximal concentration of defects or special alloys. In the latter case, alloys of the metallic elements with very different structure of the electron shells, but well soluble in each other, are used. In this case the crystal

lattice consists of randomly alternating atoms of different types, which leads to strong scattering of the electrons.

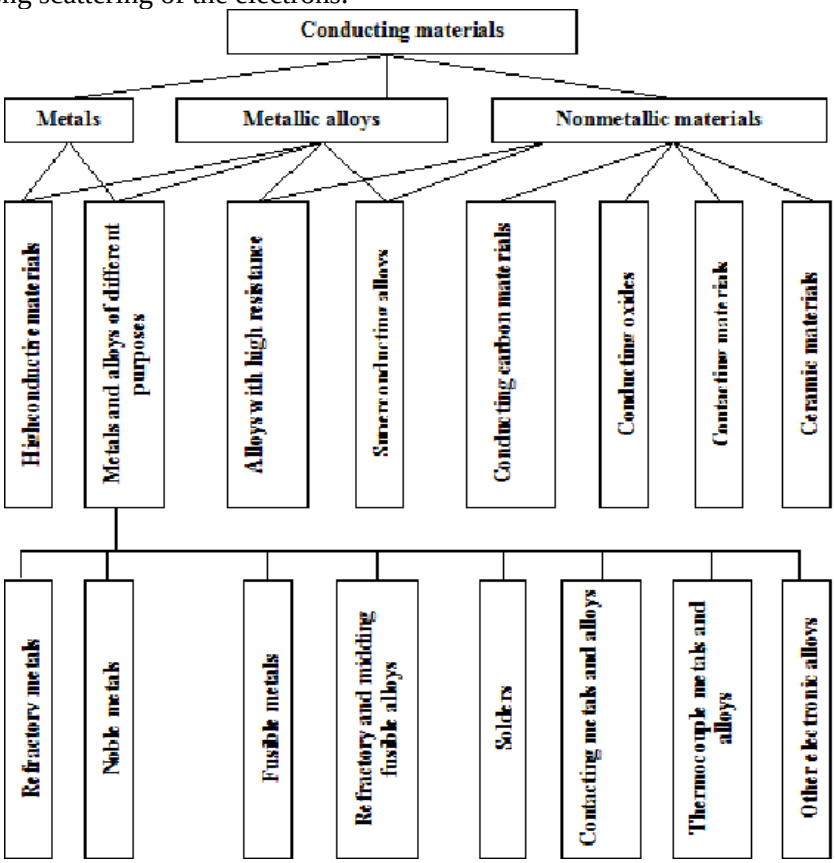


Fig.1.8.4. Types of conductive materials used in power industry

Most often, as materials with the highest electrical resistance (e.g. for the windings in electric furnaces), nickel-chromium alloys (so-called nichrome) are used. They also have a high oxidation resistance up to temperatures of about 1400 K.

1.8.2. Electrical properties of semiconductor crystals

Table 1.8.1 shows main properties of such basic semiconductors as Si, Ge and GaAs

Paramters	Notation	Unit of measure-ment	Si	Ge	GaAs
Density	ρ	g/cm^3	2,33	5,23	-
Permittivity	ε		11,8	16	11
Effective mass at T=300K:	electrons m_n		0,33	0,22	0,07
	holes m_p		0,55	0,39	0,5
Band gap	E_g	eV	1,11	0,67	1,4
Effedtive density of states	N_C (in C-band)	cm^{-3}	$2,8\cdot10^{19}$	$1,04\cdot10^{19}$	-
	N_V (in V-band)	cm^{-3}	$1,02\cdot10^{19}$	$0,61\cdot10^{19}$	-
Concentration of intrinsic carriers	n_i	cm^{-3}	$1,6\cdot10^{10}$	$2,5\cdot10^{13}$	$1,5\cdot10^{6}$
Drift mobility of intrinsic carriers	electrons μ_n	$cm^2/V\cdot s$	1400	3800	11000
	holes μ_p		500	1800	450
Intrinsic resistivity	ρ_i	$\Omega\cdot cm$	$2\cdot10^5$	60	$4\cdot10^8$
Diffusivity	electrons D_n	cm^2/s	36	100	290
	holes D_p		13	45	12
Thermal capacity	C	J/g·K	0,75	0,41	-
Thermal conductivity	λ	Wt/cm·K	1,2	0,5	-
Melting temperature	T_m	K	1593	1233	1653

1.8.3. Properties of crystalline dielectrics

Usually we rank to the dielectric solids such crystals, in which the band gap is greater than 3-5 eV, so that the turnover of the electrons from the V- to the C-band at room temperature can be practically zero, and the concentration of free carriers is negligible. One of the most important properties of dielectrics, except for the very high electrical resistivity, is also their ability to be polarized (possibility of the presence of internal electrostatic field in them).

The electrical conductivity of insulators, in principle, can be described on the basis of the theory of semiconductor crystals. However mobility of electrons and holes in dielectrics may in 100-1000 times less than in wide-gap semiconductors. This is due to the fact that charge carriers in dielectrics are practically in bounded state (due to the strong interaction with the crystal lattice). As a result of strong electron-lattice interaction, free carrier moves along the crystal together with the region of the distorted lattice. Such a carrier, "coated" with the induced positive charge of the polarized lattice, is called

polaron. Thus, in addition to the usual electronic (or hole) conduction mechanism, dielectrics can have *a polaron conduction mechanism*.

Polaron conductivity is characteristic for ionic crystals, where the Coulomb interaction between electrons and ions, forming the crystal lattice, is especially large. Because at the polaron conduction velocity of free electrons in an external electric field is slowed down (due to the strong electron-phonon interaction), we can attribute them the effective mass which is thousands of times greater than the effective mass of free carriers in metals and semiconductors. Due to this reason, the mobility of carriers in dielectrics is a thousand times lower. In some dielectrics ionic conductivity is the dominated, at which not only charge but also the mass is transferred in electric field.

Polarization of dielectrics. Under exposion of an electric field, the particles constituting the insulator (atoms, ions, molecules) are transformed into dipoles, that results in arising of *electric (dipole) moment*. Its total value depends on the dipole moments of the particles and their density (concentration). Dielectric polarization leads, in particular, to reduce the electric field inside the condenser filled with a dielectric, i.e. to increase its capacity.

Electric (dipole) moment per unit volume of the dielectric is called *the dielectric polarizability P*, which is generally equal to the vector sum of the dipole moments of elementary particles forming the dielectric

$$P = \sum_{i=1}^{N} \mathbf{p}_i, \qquad (1.8.9)$$

where N is the volume concentration of the elementary dipoles p_i. In addition to the polarizability P of dielectrics, we introduce for its characterization such macroscopic characteristics as the *intensity of the electric field E* in the dielectric and the *electric induction D*, which are related by

$$D = \varepsilon_0 \varepsilon E \quad \text{or} \quad D = \varepsilon_0 E + P, \qquad (1.8.9)$$

whence

$$P = \varepsilon_0 (\varepsilon - 1) E = \varepsilon_0 \chi E, \qquad (1.8.10)$$

where $\varepsilon_0 = 8{,}85 \cdot 10^{-12}$ Ф/м is *dielectric constant*, ε – *permittivity*, and $\chi = \varepsilon - 1$ is called the *relative dielectric susceptibility*. In isotropic dielectrics, vectors D, E and P have the same direction, so that the dielectric constant and the susceptibility are prime numbers. For the most dielectrics permittivity does not depend on electric field intensity (at least, in the fields to 10^6 V/m).

Polarization of dielectrics in an electric field is due to a number of the following microscopic processes: displacement of the electron shells of atoms and ions, the displacement of cations and anions in relation to each other, the orientation of molecules with a permanent dipole moment by the electric field,

etc. As a result, the polarization can be subdivided into the following three types – electronic, ionic and dipole.

Within each type of polarization, its type is determined by the nature and characteristics of the interaction of (the natural of acting forces) between the particles forming the dielectric. If the forces that aspire to return particles, shifted by the electric field, to the starting position, are the quasi-elastic (linear) character, then we deal with *the elastic polarization*. If the particles (electrons, ions, dipoles) after displacement in the field overcome the potential barrier due to thermal energy, then we deal with *the thermal polarization*.

The most common form of polarization in dielectrics is *elastic electron polarization* that occurs due to displacement of the electron shell of the ion relative to the atom nucleus. This kind of polarization is set for times of the order of 10^{-17}-10^{-16} s. Dipole moment produced by the electron polarization is given by

$$P = (e^2/k)E. \qquad (1.8.11)$$

The coefficient of proportionality between E and P is called *the electronic polarizability*

$$\alpha_e = e^2/k = 4\pi\varepsilon_0 r^3, \qquad (1.8.12)$$

where k characterizes the elasticity of the electron-nucleus system, and r – radius of the electron orbit of a hydrogen-like atom before the application of an electric field.

The values of the electronic polarizability of atoms is of the order of 10^{-40} F·m^2, although the experimental values are always somewhat smaller than is calculated from the formula (1.8.12), especially for complex atoms, when hydrogen-like model is not applicable.

The polarizability increases strongly with the increasing of the electron shells radius, because the interaction between the nucleus and the electron is decreased and the electrons in the electric field are subjected to the greatest displacements.

In dielectrics with ionic chemical bonds the *elastic ion polarization* may occur due to the displacement of the anions and cations relative to each other. This kind of polarization is set for times of the order of 10^{-15}-10^{-14} s. For this kind of polarization *the ionic polarizability* has a form

$$\alpha_i = e^2/k = \frac{4\pi\varepsilon_0(r_a + r_k)}{n-1}, \qquad (1.8.13)$$

where k – a coefficient of elasticity for the anion-cation system, r_a and r_k – the effective radii of the cations and anions, and the exponent n is a repulsive Born potential, describing the potential energy of interaction between anions and cations.

In many molecular dielectrics, molecules have their own dipole moments. Their orientation, in the absence of an external electric field *E*, is determined by the existing internal electric field E_{int} (Fig. 1.8.8). Under the impact of an external electric field, elastic deflection from the equilibrium orientation can result in some preferred orientation of the dipoles, which is called *the elastic dipole polarization*.

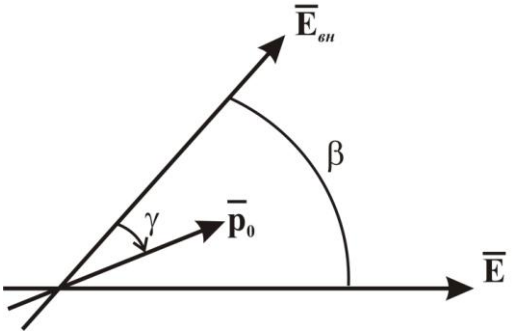


Fig. 1.8.8. Diagram of a dipolar elastic polarization

The polarizability of polar molecules (such as HCl) with elastic bonds

$$\alpha_в = \frac{P_0^2}{|U_0|}\sin^2\beta \qquad (1.8.14)$$

depends on the direction of the electric field (through angle β between the vectors *E* and E_{int}), the electric moment of P_0 and the energy of the dipole (intermolecular bonds) U_0. The contribution of this kind can polarization cause the anisotropy of the permittivity.

The orientation of the dipole moments of the particles (atoms, ions, molecules) is also dependent on their thermal movement which prevents the orienting action of the internal electric field. The polarization of the dielectric in an external electric field is called in this case *the thermal polarization*. For such kind of polarization, polarizability of dielectric without external electric field is zero (because of heat disorientation by atomic vibrations in the crystal lattice). In the external field, it is strongly dependent temperature. Depending on the source of a dipoles, it is accepted to distinguish *thermal*, *thermal ion* and *thermal electron polarization*s.

Due to slow setting (compared with the elastic polarization) of heat polarization, it obeys by law

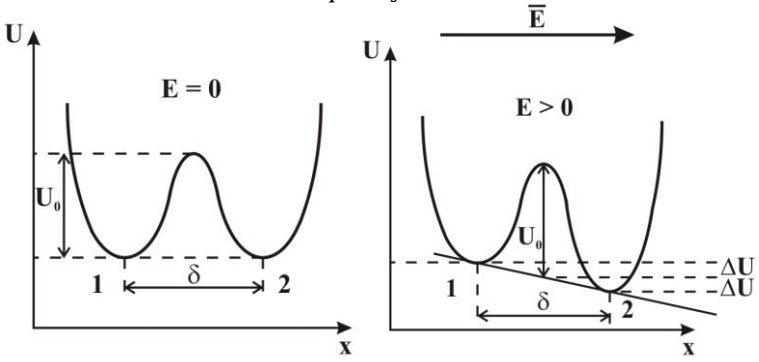
$$P = P_0 \exp(-t/\tau), \qquad (1.8.15)$$

so that this kind of thermal polarization is called *a relaxation polarization*. Here τ is a dipole moment relaxation time (~ 10^{-8}-10^{-4} s).

Ionic thermal polarization is due to changes of the height of the potential barrier in the electric field, that the ion must overcome for transition from one equilibrium position to another. This creates an opportunity for directed (and not random) movement of ions in the damaged regions. Due to probabistical nature of the ions transition from the equilibrium position 1 to position 2 (Fig. 1.10.9a), number of ions, overcoming the barrier in the *x* direction for 1 second, equals to

$$n = \frac{n_0 \nu}{6} \exp\left(-\frac{U_0}{k_B T} \right), \tag{1.8.16}$$

where n_0 – the total number of weakly bounded ions per unit volume, U_0 – the activation energy of the transition (jump) from state 1 to state 2 at zero electric field, ν – the oscillation frequency of the ions.



a) *b*)

Fig. 1.8.9. Diagram of a thermal ionic polarization

When an electric field E is applied along the *x* direction (Fig. 1.8.9b), the ion hopping activation energy (the barrier height) in this direction will decrease on the value of $\Delta U = eE\delta/2$ (e is ion charge, δ – the distance between the ion states 1 and 2). As a result, the number of jumps in the direction of $1 \rightarrow 2$ increases by an amount Δn compared with reverse direction $2 \rightarrow 1$. As a result, in the dielectric asymmetric charge distribution will be established, forming some dipole moment $P = \Delta n e \delta$. In this case, assuming that the $\Delta U << k_B T$, polarizability, calculated for each weakly-bounded ion, will be generated

$$\alpha_{iT} = \frac{P}{n_0 E} = \frac{\Delta n}{n_0} \frac{e\delta}{E}. \tag{1.8.17}$$

Calculation of Δn, depending on the electric field and temperature, leads to the following expression for *the ionic thermal polarizability* after switching on the field

$$\alpha_{iT} = \frac{e^2\delta}{12k_BT}\left(1-e^{-t/\tau}\right), \tag{1.8.18}$$

that for the stationary case (when $t \to \infty$) corresponds to the establishment of a permanent polarization

$$\alpha_{iT} = \frac{e^2\delta}{12k_BT}. \tag{1.8.19}$$

Since the dipole moment $P = e\delta$, occurring, when passing of every ion, is constant and does not depend on the electric field, in contrast to elastic ionic polarization, the polarizability of every ion is inversely proportional to the field, and as a result, the coefficient of *electronic thermal polarization* α_{iT} is not dependent on the electric field. α_{iT} arises in solid dielectrics with lattice defects such as anion vacancies (for example, in the crystals of rutile TiO_2 or alkali halide) for which the electrons can jump to the neighboring ions under the influence of thermal fluctuations, thus overcoming some of the potential barrier.

When the electric field lack, electron hopping between the anion vacancy and the neighboring cations occur randomly and polarization does not arise. When electric field is applied, the induced by the field jumps become significantly directed that leads to the formation of dipole moment with the time constant of the order of 10^{-7}-10^{2} s. The contribution of electronic thermal polarization can be significant even at low concentrations of defects due to the high polarizability of the electrons that are weakly coupled with defects.

Dipole thermal polarization takes place at the presence of a polar molecules in dielectric with weak intermolecular coupling, which are relatively easily able to rotate by an electric field, creating a dipole polarization, depending on the temperature. Unlike thermal ionic polarization, wherein ion transition from one equilibrium state to another takes place due to translational motion, polar molecule changes its state owing to the rotational movement.

By analogy with the ion thermal polarization, one can obtaine the temperature dependence of the polarizability for the relatively weak electric field in the form of

$$\alpha_{dT} = \frac{P_0}{3k_BT}, \tag{1.8.20}$$

where P_0 is electric dipole moment of the molecule.

In strong fields, the dipole moment of the molecule tends to saturation, ceasing to depend on the field with the growth of the latter. Since the polarization of the real dielectric is always a set of the simplest t ofupes polarization, in general, the resultant dipole moment (polarizability) per the unit volume of the dielectric presents the vector sum

$$P = \left(\sum_m n_m \alpha_m \right) E, \tag{1.8.21}$$

where α_m is a polarizability for the m-th polarization type, n_m – the concentration of particles which contribute to the m-th kind of polarization.

The interconnection between the permittivity and polarizability. Let us establish the relation between the dielectric permittivity and polarizability. From (1.8.17) it follows that the permittivity ε, as the mostly macroscopic important parameter of the dielectric, is related to the polarizability P and the electric field *E* by ratio

$$\varepsilon = 1 - P/(\varepsilon_0 E), \tag{1.8.22}$$

where *E* – the average macroscopic field in the dielectric, which is composed of the external and internal fields. Theoretical analysis shows that the interconnection between the permittivity and polarization is determined by the Clausius-Mosotti relationship

$$\frac{\varepsilon - 1}{\varepsilon + 1} = \frac{1}{3\varepsilon_0} \sum_m n_m \alpha_m. \tag{1.8.23}$$

The frequency dependence of the dielectric permittivity. In a constant electric field, all kinds of polarization, specific to the any dielectric substances, have time to steady, so that the contribution to the dielectric permittivity carry in both fast and slow polarization mechanisms. In an alternating electric field, owing to the different times of steady for different mechanisms of polarization in the matter, there is a dispersion of permittivity, i.e. its dependence on the frequency ε(ν). The common form of ε(ν) is shown in Fig. 1.8.10.

In the low frequencies, all kinds of polarization influence on the dielectric permittivity ε_{LF}. However at ν ~ 10^4-10^5 Hz various kinds of the so-called *space-charge polarization* start to "turn off". They are associated with the motion and accumulation of charged particles (electrons , ions) at the interfaces of heterogeneous dielectrics. In the radio-frequency range 10^4-10^{11} Hz, the thermal polarization mechanisms (electronic, ionic and dipole) begin to contribute to the permittivity ε_{RF}. In the infrared region of the spectrum (10^{11}-10^{15} Hz), a delay of elastic and dipole ion polarization happens. In the field of optical frequencies (10^{17}-10^{18} Hz), the ε_{OPT} value is due only to the elastic electron polarizability, since at such high frequencies all the other mechanisms

of polarization does not have time to steady. In the fields above the optical frequencies, polarization is not possible, so that ε = 1.

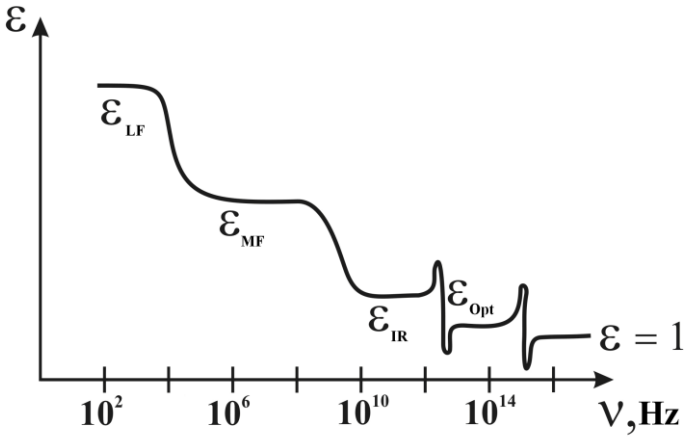


Fig. 1.8.10. The frequency dependence of permittivity

Polarization in non-centrosymmetrical dielectrics. In dielectrics with non-centro-symmetrical structure, is also possible the forced polarization except for the above-mentioned mechanisms of polarization induced by an external electric field. In such a crystals, the dielectric moment occurs under the influence of mechanical stresses (*piezopolarization*), temperature change (*piropolyarization*) or light irradiation (*photopolarization*). In addition, these crystals can have *spontaneous polarization* that occurs without any external influences.

Pezopolyarization occurs due to the displacements of the charged particles along the polar axises. For example, in quartz SiO_2, it appears under impact of mechanical stresses along the direction (axis), where the alternation of anions and cations takes place. Such effect is called by *direct piezoelectric effect*.

Piropolyarization occurs along the polar axis due to reduction of the polarization degree due to heating or its increase owing to cooling of the crystal (*pyroelectric effect*), which has a spontaneous polarization.

Photopolarization occurs when an intense non-uniform illumination of widebandgap crystals (such as lithium niobate $LiNbO_3$) through the redistribution of charge at the generation of free charge carriers due to photoionization of impurities in the dark areas, they pass under the influence of the internal electric field arising in the dark spots and the subsequent capture by defects (traps) in the dark areas.

Conservation of the polarized state in a dielectric for a long time after removal of the external impact was called *electret state*. Dielectrics, having spontaneous polarization, are divided into two classes. For one class, not any electric field can change the direction of the spontaneous polarization of the dielectric until the breakdown. The second class of dielectric (*ferroelectric*s) is characterized by a nonlinear dependence of the polarization of the electric field and the ability to repolarization, which is manifested by the hysteresis loop for $P(E)$ dependence.

1.8.4. Properties of crystalline magnets

Let us study the magnetic properties of crystalline solids, including their response on the application of a magnetic field (though, as will be shown below, there are some substances, where magnetization is present even at zero field). Some magnetic crystals are an important kind of special materials that are used for various purposes, including power technologies and energy saving.

From the standpoint of the nature their response to an external magnetic field, all the known substances can have a magnetic moment directed either along the field (the positive magnetic susceptibility) or against it (negative magnetic susceptibility), see Fig. 1.8.11a. Furthermore, among the materials with positive magnetic susceptibility, the so-called *magnetically ordered substance*s are observed. In such a materials, the magnetic moments of adjacent atoms can be ordered even at the absence of a magnetic field, Fig. 1.8.11b-d.

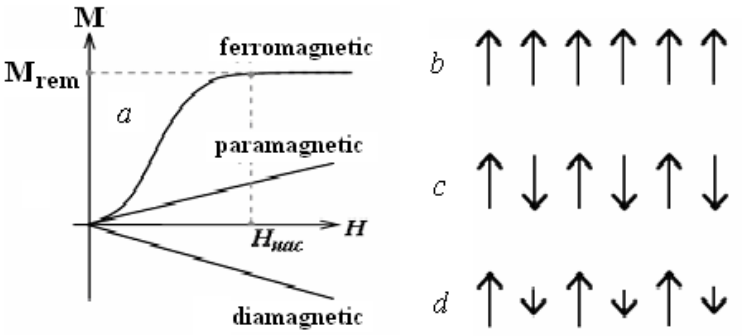


Fig. 1.8.11. Magnetic field dependence of the magnetization M for the diamagnetic, paramagnetic and ferromagnetic (*a*) substances, and the types of the ordered arrangement of spins for a ferromagnetic (*b*) and antiferromagnetic (*c*) and ferromagnetic (*d*) substances

This regularity in the arrangement of the magnetic moments causes features not only in the magnetic properties but also in many other physical properties

such as electrical conductivity, heat capacity, etc. Just magnetically ordered materials play a crucial role in modern technology.

The main magnetic properties of substances. The magnetic properties of materials are characterized by a magnetic moment per unit volume

$$M = \chi H = (1/\mu\mu_o)\chi B, \qquad (1.8.24)$$

which is called *the specific magnetization*. Here, H – a magnetic field vector, B – the vector magnetic field induction, χ – *magnetic susceptibility* of the substance, μ – *relative permeability* of the material, $\mu_o = 4\pi 10^{-7}$ H/m – *magnetic constant (permeability of vacuum*).

The magnetic induction B is related to the magnetic field intensity by relations

$$B = \mu\mu_o H \text{ and } B = \mu_o H + \mu_o M. \qquad (1.8.25)$$

For isotropic materials one can get from these relations the connection between the magnetic susceptibility and the magnetic permeability

$$\chi = \mu - 1. \qquad (1.8.26)$$

For many substances, magnetic susceptibility and permeability are scalar quantities. Substances with $\chi < 0$ (vector M and H are antiparallel) are called *diamagnetics*. Substances with $\chi > 0$ (vector M and H are parallel) are called *paramagnetics*. In zero magnetic field, the magnetization of dia- and paramagnetic substances is zero. The dependence of the magnetization for these two types of magnets on magnetic field at low fields, as a rule, is linear (Fig. 1.8.11a). In the most cases, by modulo, the magnetic susceptibility of paramagnetic materials exceeds the magnetic susceptibility of diamagnetics.

The sources of magnetism in crystals. The diamagnetism is associated with the occurrence of the electron orbital precession in the atoms in a magnetic field. Therefore, it is inherent in all the atoms without exception.

The linear dependence of $M(H)$ for the paramagnetic substances is observed only in weak fields and at high temperatures, coming to saturation at high fields and low temperatures. Group of magnets having a spontaneous magnetization (not equal to zero even at zero magnetic field) is called *ferromagnets*. For them, the $M(H)$ dependence is non-linear function (see Fig. 1.8.11a), and magnetic susceptibility $\chi = dM/dH$ depends on the field (Fig 1.8.12b).

Ferromagnetic as well as paramagnetic materials are magnetized in the H direction, i.e. they have $\chi > 0$. However, for these substances $M(H)$ is saturated in fields much smaller than in the case of paramagnets, and magnetic permeability $\mu = 1 + \chi$ depends on H, passing through a maximum.

The values of the magnetic susceptibility of certain substances at room temperature are as follows: for bismuth $\chi = -1.8 \cdot 10^{-4}$ (diamagnetic) for gold $\chi = -3.7 \cdot 10^{-5}$ (diamagnetic) for germanium $\chi = -8 \cdot 10^{-6}$ (diamagnet) for uranium

$\chi = 4.14 \cdot 10^{-5}$ (paramagnet), for titanium $\chi = 6.1 \cdot 10^{-4}$ (paramagnet), for platinum $\chi = 2.6 \cdot 10^{-6}$ (paramagnet).

The total magnetic moment of a free atom is composed from the own (spin) magnetic moments of electrons M_S and orbital magnetic moments M_L, governed by the motion of electrons around the nucleus (nuclear magnetism can be ignored). It is known that M_S and M_L are linked with the appropriate mechanical moments P_S and P_L by so-called *gyromagnetic ratios*

$$\frac{M_S}{P_S} = -\frac{\mu_o e}{m}; \quad \frac{M_L}{P_L} = -\frac{\mu_o e}{2m}. \tag{1.8.27}$$

Due to the quantization of P_S and P_L, magnetic moments are also quantized. The Bohr magneton $\mu_B = e\hbar/(2m) = 9{,}27 \cdot 10^{-24}$ A·m^2 is called the quantum of magnetic moment.

Total angular momentum of the atom, defined as the vector sum of $P_j = P_L + P_S$, corresponds to the total magnetic moment of the atom M_j, whose projection on the direction of the field H is given by relation $M_j = - m_{jH} g\mu_B$, where m_{jH} is the magnetic quantum number, g – Lande splitting factor. Note that for pure spin magnetism $g = 2$, and for purely orbital magnetism $g = 1$.

In all atoms and ions, which have an even number of electrons, the electron shells are completely filled, so that the resulting spin and orbital magnetic moments are equal to zero, and as a result the total magnetic moment is also zero. Non-zero magnetic moments (diamagnetism) can have only the atoms and ions that have undeveloped internal d-and f-shells or contain an odd number of electrons in the valence shell. Oxygen, in which the number of electrons is even, but the spin moments of the two of them uncompensated is also paramagnetic.

Nature of diamagnetism. Magnetization in diamagnets is induced in the external magnetic field. As a result, a diamagnetic moment of the electron magnetic field rotating in a circular orbit will equal $M_{dia} = (e^2 S/2\pi m)B$.

For the calculation of the diamagnetic susceptibility of the substance, it is necessary to calculate the magnetic moment of many-electron atoms that are composed from the orbital angular momenta of the individual electrons. For atoms with Z electrons which have the mean squared distance $\langle a^2 \rangle = 2/3 \langle r^2 \rangle$ of electrons from the axis of their rotation, passing through the core in parallel to the field, the magnetization of the substance of the N atoms per unit volume is

$$M_{dia} = \frac{NZe^2 \mu_B < r^2 >}{6m} B. \tag{1.8.28}$$

Hence, in accordance with the Eq. (1.8.31), we obtain an expression for the diamagnetic susceptibility of the material unit volume

$$\chi_{dia} = \frac{NZe^2 \mu_B < r^2 >}{6m}. \qquad (1.8.29)$$

As follows from this relation and is in good agreement with experiment, the diamagnetic susceptibility is independent on temperature and increases with the ordinal number of the element in the periodic table according to the law $\chi_{dia} \approx - Z \cdot 10^{-6}$ (for $N = 5 \cdot 10^{22}$ cm^{-3} and $r = 10^{-8}$ cm).

Since the orbital motion of the electrons occurs in all substances, the diamagnetism is common to all substances without exceptions. Often, however, it is covered by stronger magnetic effects like paramagnetism and ferromagnetism. Unlike dielectrics, where all the electrons are bound to their atoms, at the calculation of the diamagnetic susceptibility of metals we need also to take into account *the diamagnetism of the free electron gas.*

Taking into account the quantum properties of free electrons, moving in a magnetic field by circular paths under the influence of Lorentz force, allowed Landau to calculate the diamagnetic susceptibility for free electrons

$$\chi_{dia}^{el} = \frac{N \mu_o \mu_B^2}{2\varepsilon_F}. \qquad (1.8.30)$$

Estimation gives $\chi_{dia}^{el} \sim 10^{-6}$, which is about three times smaller than the paramagnetic susceptibility of the free electron gas.

Nature of substances paramagnetism. As already mentioned, with a classical standpoint, paramagnetism is the occurrence of magnetization under exposion of the magnetic field due to the orientation of the elementary magnetic momenta, which were in substance even in the absence of the field, but were directed randomly due to the thermal vibration of the atoms.

According to the above classification of magnetic materials, the magnetization vector in paramagnetic substances (with a positive paramagnetic susceptibility) is parallel to the magnetic field. This means that the magnetic induction vector in a paramagnetics is higher than in vacuum.

Paramagnetism is presented in atoms and molecules, that have an odd number of electrons, which have uncompensated spin magnetic moment; the free atoms or ions with incompleted d-and f-shells; some of the molecules with an even number of electrons (e.g., O_2 and S_2) due to the non-compensated spins of two electrons; the lattice defects with odd electrons; and also metals due to free electrons gas. In the paramagnetics, susceptibility χ_{para} depends on the temperature following the Curie law

$$\chi_{para} = C_1/T, \qquad (1.8.31)$$

where C_1 is a Curie constant.

Paramagnetism in dielectrics. Thus, as shown above, the physical nature of paramagnetism in dielectrics is due to the orientation of the spin magnetic moments in atoms, ions or molecules in a magnetic field (Fig. 1.5.12).

Paramagnetism in dielectrics can be easily explained by *the classical theory of Langevin*. Langevin theory comes from the fact that the net magnetization (*paramagnetic moment*) of the media, containing N atoms per unit volume, is determined by statistical equilibrium between the orienting impact of the magnetic field and misorienting effect of the thermal motion of the atoms. As a result, the total moment is determined bythe sum of the projections $M_B = M\cos\theta$ onto the field (Fig. 1.8.12).

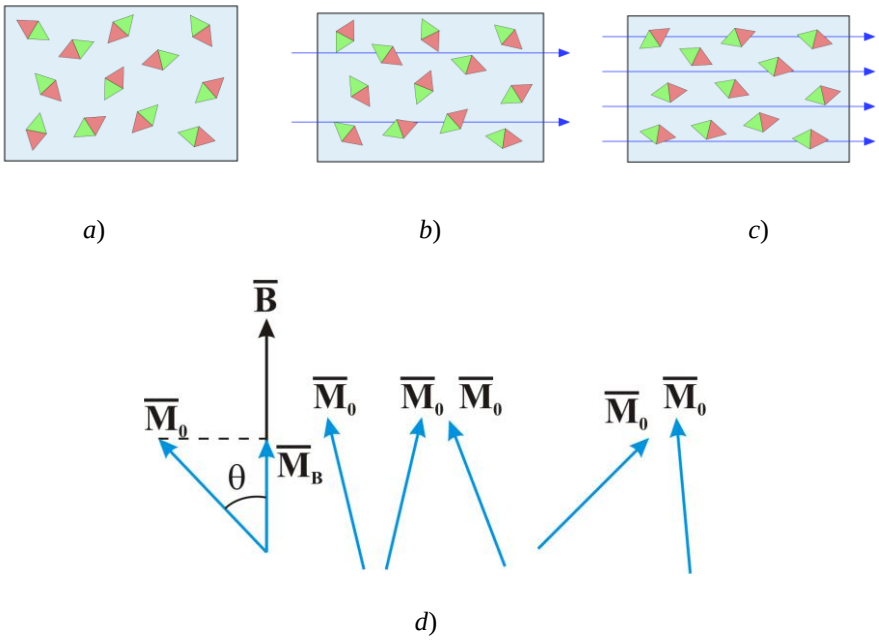


Fig. 1.8.12. The scheme of the spin magnetic moments orientation in atoms in a magnetic field

The classical Langevin theory estimates the average projection of the magnetic moment using the classical statistics methods, calculating the probability of its various orientations θ. This calculation for the N paramagnetic centers per unit volume gives the expression

$$N\langle M_B\rangle = N\langle M\cos\theta\rangle = NM(\mathrm{cth}\beta - 1/\beta) = NM\cdot L(\beta), \qquad (1.10.32)$$

where $\beta = MB/(k_BT)$ is the reduced magnetic energy and $L(\beta) = (\mathrm{cth}\beta - 1/\beta)$ – classical Langevin function. As a result, according to the Langevin model, the

paramagnetic susceptibility of the dielectric in a weak magnetic field and relatively low temperatures is expressed by

$$\chi_{para} = \mu_o N\langle M_B\rangle / = N\mu_o M^2/(3k_B T), \qquad (1.8.33)$$

where $C_1 = N\mu_o M^2/(3k_B)$ is a Curie constant. As you can see, the formula (1.10.40) coincides with the well-known experimental *Curie law* (1.5.38). In strong fields, where $\beta \to \infty$, $\mathrm{ctg}\beta \to 1$ and $L(\beta) \to 1$, the magnetization reaches saturation, which is also in agreement with experiment. Condition $\beta = M_B/(k_B T) << 1$ is almost always satisfied at room temperature T = 300 K, because even in the fields with the induction of the order of 1 Tesla magnetic energy $M_B \sim 10^{-23}$ J, while the lattice thermal energy $k_B T \sim 3 \cdot 10^{-21}$ J.

The quantum nature of paramagnetism. The classical theory of the Langevin is in good agreement with experiments for non-metallic paramagnetic substances, but has a number of shortcomings in principle. In particular, it can not explain a number of experimental results for the paramagnetic metal. The reason for these shortcomings of the classical Langevin theory, first of all, is the injustice of the assumption concerning the possibility of any orientations of atomic magnetic moments relative to the magnetic induction vector. According to the quantum Langevin theory, in a magnetic field orientation of the magnetic moment vector is quantized. In view of this, the average projection of the magnetic moment of the atom on the magnetic induction vector is expressed in terms of the Brillouin function or a generalized Langevin function $B_j(\beta)$ as follows

$$\langle M_j\rangle = g\mu_B j B_j(\beta), \qquad (1.8.34)$$

where $\beta = g\mu_B jB/(k_B T)$ and J is number of moment orientations, and

$$B_j(\beta) = \frac{2j+1}{2j}\mathrm{cth}\frac{2j+1}{2j}\beta - \frac{1}{2j}\mathrm{cth}\frac{\beta}{2j}.$$

Note that for $j \to \infty$ $B_\infty(\beta) = L(\beta)$.

Using (1.10.34), it is easy to obtain the following relation for the magnetic moment of the substance containing N atoms per unit volume

$$M = Ng\mu_B j B_j(\beta). \qquad (1.8.35)$$

If the magnetic moment of atoms is caused by the electron spin only (orbital moment is zero), the number j of its possible orientations in a magnetic field is reduced to two. In this case formula (1.8.35) takes the simpler form:

$$M = Ng\mu_B \frac{1}{2}\mathrm{th}\left(\frac{g\mu_B B}{2k_B T}\right) \qquad (1.8.36)$$

In the case of weak magnetic fields and not too low temperatures (($\beta << 1$) we can get from (1.10.36) the *adjusted Curie law* for paramagnetic susceptibility

$$\chi_{para} = \frac{N\mu_o \mu_B^2 g(j+1)}{3k_B T} \qquad (1.8.37)$$

In strong fields and at very low temperatures, when $\beta \to \infty$ and $B_j(\beta) \to 1$, magnetization, as in the classical case, reaches saturation

$$M = Ngj\mu_B, \qquad (1.8.38)$$

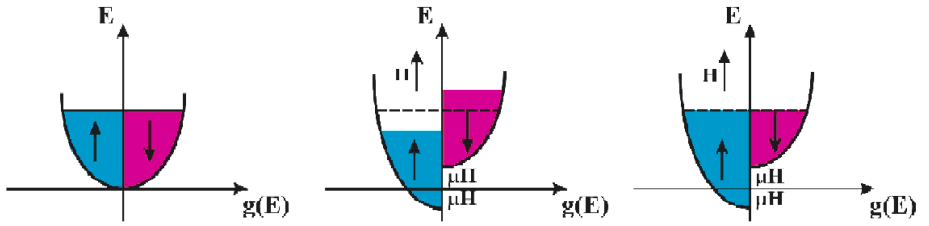
albeit at a lower value of the magnetic moment.

Paramagnetism of the free electron gas. Contribution to the paramagnetic susceptibility of metals and doped semiconductors is introduced not only by ions of the crystal lattice, but also by collectivized conduction electrons. The paramagnetism of the electron gas is associated with the presence of the electron spin magnetic moment equal to the Bohr magneton. Experiments show that the paramagnetic susceptibility of free electrons does not depend on temperature.

Paramagnetism of free electrons in a metal is described by quantum theory of Pauli based on Fermi-Dirac statistics. According to this theory, in the absence of a magnetic field ($B = 0$), the total magnetic moment of the electron gas at $T = 0$ K is zero, since the electrons in accordance with the Pauli principle are located on the energy levels in pairs (with opposite spins), see Fig. 1.10.13. Therefore, before the switching on of magnetic field, the electrons density of states with spin up and spin down can be presented by Eq. (1.3.32), and the dispersion laws for them are similar parabolas (Fig 1.10.13a):

$$g_+(E) = g_-(E) = (1/2)g(E).$$

When magnetic field is applied ($B \neq 0$), electron subband, in which the magnetic moments are oriented along the field, will be shifted along the energy axis E down on the amount of magnetic energy $U = -\mu_B B$, and subband with opposite spins - up to $U = \mu_B B$ (Fig. 1.10.13b). Since both subbands will shift relative to each other by the $2\mu_B B$ (Fig. 1.10.13c). The scheme of the magnetic moment of a free electron Fermi levels should be equalized (to the system of electrons come into equilibrium), then the electrons will flow from the right to the left poluzony, thus changing the direction of the spin (Fig. 1.10.13c). As a result, the magnetic moment of



a) *b*) *c*)

Fig.1.10.13. Scheme of appearance of free electron paramagnetic moment

electron gas, directed along the field, appears, which equal

$$M = \mu_B(g_+ - g_-), \tag{1.8.39}$$

where the number of the electrons, changed the energy, in the first approximation is proportional to the applied field

$(g_+ - g_-) \approx g_+(E_F)\Delta E = g_+(E_F)\mu_B B$.

Substituting into (1.10.39) the previously derived expression for the electron gas density of states, we easily obtain the Pauli paramagnetic susceptibility

$$\chi_{par}^{el} = \frac{3n\mu_o\mu_B^2}{2\varepsilon_F}, \tag{1.8.40}$$

where $n = \int g(E)f(E)dE$ – the concentration of electrons in the metal, ε_F – weakly temperature-dependent Fermi energy and $\mu_B = e\hbar/(2m)$.

Consequently, given the diamagnetism of free electrons as Eq. (1.10.30), the total magnetic susceptibility of the electron gas will be paramagnetic and equals

$$\chi^{el} = \chi_{dia}^{el} + \chi_{para}^{el} = \frac{n\mu_o\mu_B^2}{\varepsilon_F}. \tag{1.8.41}$$

Comparison with experiment for such metals (even without ion paramagnetism), as, for example, copper and bismuth, is complicated, first, the fact that there will inevitably contribute atomic diamagnetism. As we have seen, this contribution is of the same order or even higher than contribution of itinerant electrons. Secondly, although the above calculation of the electron susceptibility is quite accurate, but is based on a free electrons model, i.e. does not take into account the influence of the periodic potential.

Ferromagnetism and the magnetic ordering. In many substances, which are paramagnetic at high temperatures, elementary magnetic moments (spins) are ordered when the temperature is lowered. The critical temperatures, at which the phase transition into ordered state of the spins occurs, are called the Curie or

Neel temperature, depending on the type of ordering. The simplest form of ordering is *ferromagnetic state*, when tends all the magnetic moments of the substance becomes parallel despite the influence their thermal motion. As a result, the occurrence of spontaneous (in the absence of an external magnetic field) magnetization, the material gains very large susceptibility which is strongly dependent on magnetic field (Fig. 1.8.11a). Magnetic ordering, in particular ferromagnetic, is essentially a quantum effect and is, as noted above, one of the major collective effects in solids.

Ferromagnetism was detected only in 9 chemicals – Fe, Co, Ni, Gd, Dy, Tb, Ho, Er, Tm. The only common feature of these metals is that they have unfilled d- and f-electron shells. There are also a variety of chemical compounds (including non-metallic) which have ferromagnetic properties, such as EuO, $CdCr_2Se_4$, ferromagnetic oxoides with spinel structure, etc.

The orderly arrangement of the atomic magnetic moments is electrostatic in nature. It is due to the limitations imposed by the Pauli principle on the shape of the electrons wave functions. As a result, a different relative orientation of the spins is associated with different forms of the electron clouds and different values of the electrostatic (Coulomb) energy. The difference values between the energies for parallel and antiparallel spin orientations (or for the symmetric and asymmetric coordinate parts of the wave function) is called *the exchange energy* U_{ex}. The highest U_{ex} values have atoms with unfilled internal electron shells, e.g. in iron, chromium, manganese, cobalt, nickel, rare earth and transuranic elements.

When of the atomic magnetic moments ordering, significant energy gain is obtained, which is often comparable to the gain in energy of chemical bonds at the rearrangements of atoms. For this reason, the ordering of the atomic magnetic moments can cause a rearrangement of the atoms in substance and even division of the matter with a homogeneous composition onto two phases, one – highly enriched in element that is characterized by strong ordering of the magnetic moments, and another – consisting of "other" items.

Weiss mean-field theory. The phenomenological theory of ferromagnetism, which explains the existence of a spontaneous magnetic moment, was established by Weiss so far ascertaining the true (quantum) nature of the magnetic ordering. This model qualitatively describes many of the properties of ferromagnets, but, instead of considering the Coulomb interaction between the electron shells (and the related exchange integral), it introduces some *exchange magnetic field* B_i, which just provides an ordered arrangement of the magnetic moments.

The *Weiss mean-field model* is based on the Langevin paramagnetism model (see above) with the important additional assumption. He assumed that in addition to the external magnetic field H a hypothetical internal magnetic field

B_i (called *molecular* or *Weiss mean-field*) acts on the elementary magnetic moments in a ferromagnetic material, later. This Weiss field is proportional to the magnetization

$$B_i = \lambda \mu_o M , \tag{1.8.42}$$

where λ is *constant of the molecular field*. This constant depends on the interaction of intrinsic magnetic moments of the atoms in a ferromagnetic material, just creating parallel orientation of the atomic magnetic moments in ferromagnetic crystal in the absence of external magnetic field. Thus, total field, which acts on atom in ferromagnetic material, equals

$$B_{eff} = B_i + \lambda \mu_o M. \tag{1.8.43}$$

Taking into account the previously obtained relation (1.10.44), and replacing the induction vector B in a ferromagnet on B_{eff} from (1.10.49), one can get approximatly for weak magnetic fields and very low temperatures

$$M = \frac{N\mu_B^2 g^2 j(j+1)}{3k_B T}(B + \lambda \mu_0 M). \tag{1.8.44}$$

Solving the equation (1.8.44) with respect to M, we obtain the following expression for the magnetic susceptibility of ferromagnet in the form of the Curie-Weiss law

$$\chi = \frac{N\mu_0\mu_B^2 g^2 j(j+1)}{3k_B\left[T - \frac{N\lambda\mu_0\mu_B^2 g^2 j(j+1)}{3k_B}\right]} = \frac{C_2}{T-\theta} = \frac{C_2}{T - C_2\lambda}. \tag{1.8.45}$$

As can be seen, as the temperature increases, the spontaneous magnetization of a ferromagnet above the Curie temperature θ decreases, so that it behaves like a normal paramagnetic. Introduction of internal molecular field by Weiss enabled one to explain many of the properties of ferromagnets. However, the nature of this field (magnetic or non-magnetic) remained unclear.

The exchange interaction. At first glance, magnetic interaction of the magnetic moments would be a cause of ferromagnetic ordering because the magnetostatic energy of the magnetic dipoles interaction is minimal when they are parallel. However, the dipole-dipole interaction of the magnetic momenta can not explaine other types of magnetic ordering (e.g., antiferromagnetic, and ferrimagnetic). Moreover, the interaction magnitude is too small to ensure a high Curie temperature in such ferromagnetic materials as, for example, iron.

Frenkel and Heisenberg theoretically established that ferromagnetism is due to the strong electrostatic (Coulomb) interaction of the electrons, so that spontaneous magnetization with a parallel orientation of spins becomes energetically favorable state.

Indeed, the energy of the Coulomb interaction between two electrons is $e^2/a \sim 10^{-11}$ erg, which is significantly more than their magnetic energy. In other words, from a classical point of view, even a small fraction of this energy would be enough for the ferromagnetic ordering. However, in classical physics, the magnetic moments of the electrons do not affect its electrostatic energy and thus, the Coulomb interaction can not cause magnetic ordering. Therefore, the nature of ferromagnetism remained unclear until the advent of quantum mechanics.

In quantum mechanics, the energy of the electrostatic (Coulomb) interaction of electrons, as in classical physics, is not dependent on their spins and is determined only by the distance between the electrons. However, the wave functions of the electrons depend on the relative orientation of the spins of electrons, and so the mean energy of the Coulomb interaction is also dependent on this orientation.

The difference values of the average electrostatic energy of the electrons at different mutual orientations of their spins is called *the exchange energy*. Therefore, such an electrostatic interaction in nature, leading to the appearance of the exchange energy in quantum mechanics, has been called *the exchange interaction.*

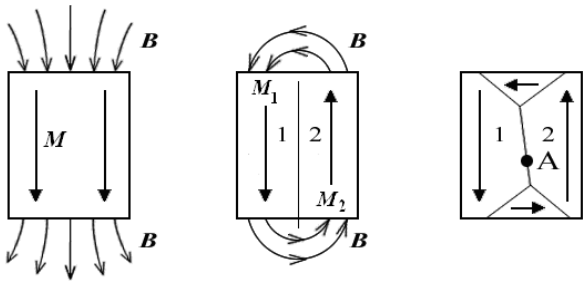
A detailed quantum-mechanical calculations have shown that indeed the resulting interaction energy of electrons includes not only purely Coulomb contribution but still an additional contribution in the form of an exchange of energy E_{ex}, depending on the relative orientation of the spins. When $E_{ex} > 0$, the minimal energy of the spin system corresponds to their parallel orientation (ferromagnetic state). When $E_{ex} < 0$, the minimal energy will be in an antiparallel orientation of the spins that corresponds to the antiferromagnetic state.

The magnetic structures with a strictly parallel and antiparallel arrangement of the spin magnetic moments are called *collinear*. There are also *non-collinear* structure, such as triangular, spiral, etc., which correspond to the ferrimagnetic state with incomplete compensation of the magnetic moment. Structures with antiparallel (antiferromagnetic) orientation of the spins, but their lack of compensation, are realized in the materials called ferrites. The resulting material has a magnetization in several times smaller than in the past.

Domains and magnetic hysteresis. As follows from the above description, the speciefic magnetic properties of ferromagnetic materials are determined by the presence of a large spontaneous magnetization. However, many experiments have shown that the magnetization of ferromagnetic materials in the absence of an external magnetic field is very often turns out to be zero. In such cases ferromagnet is separated into a variety of small regions (typically with a size of several micrometers) with different directions of the magnetization vectors,

which leads to a zero average magnetization of the whole sample. These regions are called *magnetic domains*.

The origin of the domains in a crystal is due to its desire to have the lowerst total free energy. If the crystal has had only one domain, large magnetic field was out of the crystal (Fig. 1.8.14a), giving a significant contribution to the magnetic field energy density equal to $B^2/2\mu\mu_0$. This energy is significantly reduced if the crystal will be devided on two roughly equal domains with the opposite orientation of the M (Fig. 1.8.14b). In the presence of the four domains (Fig. 1.8.14c), it will be even less. Thus, from the viewpoint of reducing the energy of the magnetic field it is beneficial for the force lines of vector B to be "closed" in the crystal, i.e. decomposition of crystal domains occurs.



a) *b*) *c*)

Fig. 1.8.14. The scheme for crystal sharing on ferromagnetic domains

In connection with the formation of the domain structure in ferromagnets, the question arises, how the magnetization vector is rotated at the transition from one domain to another, and what are the shape and dimensions of the domains. The area, in which the rotation of the spins in two neighboring domains occurs, is called *a domain wall*.The shape and dimensions of the domains, in contrast to the wall thickness, essentially depend on the shape and dimensions of the sample.

Landau and Lifshitz considered the sample in the form of an infinite plane-parallel uniaxial ferromagnet perpendicular to the easy axis of magnetization with the domain structure shown in Fig. 1.510.15. Minimizing the energy of such a structure, consisting of a domain wall energy and the energy of the magnetic anisotropy of end-capping prisms (which ensure the absence of an external magnetic field energy), gave a thickness of domains

$$d = \sqrt{\frac{2W_d l}{K_1}}, \qquad (1.8.46)$$

where W_d is the surface energy of the domain wall, l – the thickness of the plate, and K_1 – the magnetic anisotropy constant). Domain wall thickness d is about 100 nm.

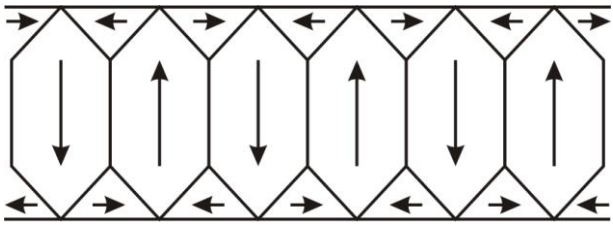


Fig. 1.8.15. The domain structure in a ferromagnetic plate with easy axis anisotropy

Fig.1.8.16. The labyrinth-like domain structure. Light and dark regions are domains with opposite directions of the magnetic moments

The domain structure, shown in Fig. 1.8.15, is realized in the uniaxial (one axis of easy magnetization), as well as cubic crystals in a certain range of values of the domain dimensions d. In thicker samples, more complex surface structures are formed, or the so-called *labyrinth-like* structure (Fig. 1.8.16). In all these cases, the expression (1.8.46) can be used to estimate the thickness domains. For many crystals, both ferromagnetic and ferrimagnetic, is $d \sim 10^{-2}\sqrt{l}$ cm. Experiments show that the character transversal dimensions of the domains in the iron, for example, are of 10-100 microns, and the thickness d of domain walls is about 100 nm.

To image the domain structures, two ways are mainly used – Bitter powder method and magnetooptical method. The powder method is suitable for the

observation of domains that cross the surface. It consists in the following: the surface is covered with slurry comprising a ferromagnetic Fe_3O_4 powder, which is collected at the domain walls, and the domain structure is observed in the microscope. Magneto-optical method based on the Faraday effect, applicable to observations of domain structures in thin, transparent in the visible light region, films. The use of the polarized light and the analyzer allows one to observe domains (in the form of light and dark regions) with different values of the magnetization projection on the normal to the film. The domain structure shown in Fig. 1.8.16, was observed by this method.

As follows from (1.8.46), small samples with sizes $l < d$ in all three dimensions will not be separated on domains. Such samples are called *single-domain*. For those substances, for which the above estimation of d is valid, samples with $l \leq 100$ nm will have single-domain structure.

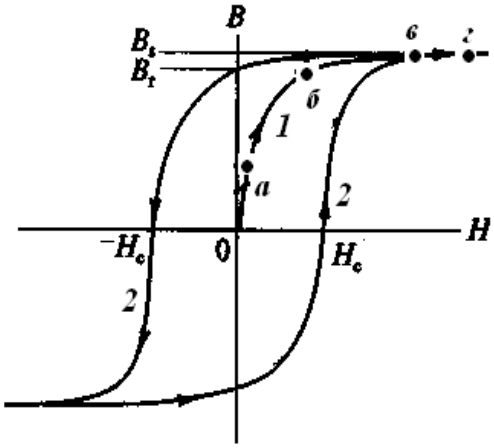


Fig. 1.8.17. The magnetization curves of ferromagnet $B(H) = H + 4\pi M(H)$: *1* – initial magnetization curve, *2* – hysteresis loop; B_s – saturation induction, B_r – remanence, H_c – coercivity

The existence of domains explains the nature of the magnetization reversal of many ferromagnetic materials in a relatively weak magnetic fields. Ferromagnets, having a domain structure, show such characteristic features as non-linear dependence of magnetization (and hence the magnetic induction inside the ferromagnet) on the magnetic field, and *the hysteresis* of the magnetization curve (Fig. 1.8.17). The reason of the hysteresis consists in the effect of various defects that prevent rotation of the magnetic moments in the domain boundaries and the movement of the latter. Therefore, the curve of

magnetization of the ferromagnet is closely linked with the processes occurring in the magnetic domains.

The initial (virgine) magnetization curve 1 has four characteristic segments. Segment 0-*a* is called the region of reversible motion of domain walls, pinned by defects. In such movement, the area and energy of the domain walls increases. If we begin to reduce the modul $|H|$ since the point *a*, domain wall is returned to the original position due to trying to reduce its surface energy. The segment a - b corresponds to the irreversible domain wall motion. In such motion, walls overcome the defects that prevent their movement, and reduction of the field H since point *b* will not bring walls back to their previous position, because they again should overcome the "passed" defects, but in the opposite direction. Segment *b-c* corresponds to the magnetization reversal mechanism owing to the rotation of the magnetization vectors M which are oriented not entirely successfully in targeting domains as a whole. This rotation, observing in a relatively strong fields $|H|$, is called the mechanism of magnetization reversal owing to the rotation of the magnetization vector. Segment *c-d* corresponds to the complete turn of the magnetic moments of ferromagnetic domains along the vector H. The value of the vector B (or M), which corresponds to the point *d* in the initial curve, is called *the saturation induction* B_s (or *saturation magnetic moment* M_s). The modulus $|H|$, where $B = B_s$ is achieved, is called *the saturation field* H_s. The M_s shows the maximum attainable value of the magnetic moment per unit volume of the ferromagnet.

If we now start to reduce H from H_s to zero, the picture of the location of the magnetic moments, corresponding to the saturation, will be saved in the first approximation, because now the atomic magnetic moments in the domains will remain their orientation due to their interaction. Therefore $B(H)$ or $M(H)$ curves in Fig. 1.8.17 will go not by line 1, when the magnetic field decreasing, but by line 2. As a result, when the value of $H = 0$ is approached, vector B (or M) will have a finite value, called *a residual induction* B_r (or *remanence* M_r). When changing H from H_s to 0, the point, that specifies the state of the magnet, is at the point B_r on the ordinate axis in Fig. 1.8.17.

If the H increases in the direction against B_r, the process of reversal is beginning, which already was discussed above. In this case, B (or M) becomes zero at the value $H = H_c$, which is called *the coercivity* (see Fig. 1.8.17). In doing so, there will again a ferromagnetic domains with different orientation of M arise in the sample, but the vector sum of all the magnetic moments of the ferromagnetic material comes to zero.

With further increase of H in the opposite direction, the sample again becomes magnetized, but in the opposite direction. At $H = - H_s$ in Fig. 1.8.17 the saturation $B = - B_s$ will again be reached. If H again to change from $- H_s$ to

H and back, the dependence of *B(H)* or *M(H)* will repeat segment 2. As a result, the curve looks like loop, which is called *the hysteresis loop*.

Thus, the hysteresis loop is characterized by three main parameters that determine the applied properties of ferromagnets (see below): residual induction B_r (or remanence M_r), the coercivity H_c and the saturation induction B_s or the saturation magnetization M_s. These values vary widely for different magnetic materials.

It can be shown that the area of the hysteresis loop, built for a ferromagnet in the coordinates of *B* - *H*, has a definite physical meaning. It is numerically equal to the energy (or done work A) spent on cyclical reversal of the magnetic substance with unit volume per reversal cycle. According to thermodynamics, this work is calculated using the formula

$$\frac{A}{\Delta V_0} = \mu_0 \int (HdJ). \qquad (1.8.47)$$

Thus, to determine the value of *M* for ferromagnet, it is not enough to know the magnitude of the magnetic field *H*, in which it is located: it is necessary also to know the "pre-history" of the ferromagnet. In particular, after the "switch-off" the field *H*, the magnetization of the magnet will depend on the magnitude of the field.

According to (1.8.31), the ratio $B/\mu_o H$ is relative permeability μ, and the slope *dB/dH* of *B*(*H*) curve is called *a differential permeability*. Relative magnetic susceptibility $\chi = (B/H)$ and the relative permeability $\mu = (B/\mu_o H)$ also have a complex dependences on *H* (see Fig. 1.8.18)

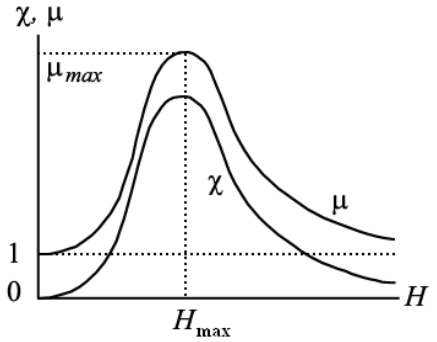


Fig. 1.8.18. Dependences of the relative magnetic susceptibility $\chi = (B/H)$ and relative permeability $\mu = (B/\mu o H)$ on the magnetic field intensity *H*

Application of magnetic materials. Various technical applications in power industry and energy saving require materials with different parameters of the hysteresis loop and, primarily, coercivity - from 10^{-1} to 10^{6} A/m The most

practical application have the materials with particularly low (soft magnetic materials) and very large (hard magnetic materials) values of H_c. Soft magnetic materials have a narrow hysteresis loop (low H_c values) and low losses (including, on heating) for their reversal in alternating electric field. For soft magnetic materials H_c < 800 A/m $\approx 10^{-4}$ T. Materials having a broad hysteresis loop (large coercivity values) are called hard magnetics and have H_c > 800 A/m.

Soft magnetic materials are used in devices that have remagnetizing in small magnetic fields. These include magnetic field sensors, the heads for reading of the magnetically recorded information, the cores of transformers. In most of these cases, it is desirable to have the material with maximal values of μ, and the minimal values of H_c and the hysteresis loop area. Such materials should have the maximally free movement of domain walls during magnetization reversal, and also reduced influence of the magnetic anisotropy and magnetostriction. To do this, it is necessary to reduce the concentration of defects in the alloys, that prevent the free movement of domain walls, and to use the alloy compositions with a weak magnetic anisotropy and magnetostriction. To use in alternating magnetic fields, it is desirable to have large electrical resistivity in soft magnetic materials. The mostly typical soft magnetic materials which qualify thes requirements are listed in Table 1.10.2.

Table 1.8.2. The mostly common soft magnetic materials and their magnetic properties

Groups of materials	H_c, **A/m**	B_S, **T**	$\mu_{max} \cdot 10^{-3}$
Pure iron	6-70	2.15	7-60
Fe-Si	30-50	1.9-2.1	3-7
Fe-Ni	0.2-4	0.5-0.8	100-1000
Amorphous alloys	0.2-0.4	0.9-1.2	400-600
Soft magnetic ferrites	15-180	0.4	0.3-4

Initially refined iron was used as a soft magnetic material, because its purity resulted in the decrease in the defects concentration. Then a set of doping impurities was found, to enhance the grain sizes (which also resulted to a significant decrease in the concentration of defects) and ensure the formation of the crystalline texture with the preferred orientation of lattice in grains along the easy magnetization directions in given product. Thus alloys of Fe-Si (transformer steel) were obtained.

Then the compositions of complex Fe-Ni based alloys were developed with the lowest possible parameters of the magnetic anisotropy and magnetostriction, which allowed to get even greater increase in μ (up to 10^6 or more). The next step was the use of amorphous and nanocrystalline (polycrystalline material

with a crystal grain size of several interatomic distances) materials in which magnetic anisotropy is much less because there is virtually no crystal lattice. Moreover, the electric resistivity of such materials was considerably larger than for crystalline alloys, due to random arrangement of atoms causing additional scattering of electrons (see above).

Some of the listed classes of soft magnetic materials (like, Fe) have a relatively low resistivity, which contributes to the appearance of large parasitic eddy currents during their magnetization reversal. Therefore, products that are often remagnetized, demands to use soft magnetic materials as a system of isolated thin plates or deposited layers. Soft magnetic ferrites have very high electrical resistivity and are widely used as soft magnetic materials for the compacted monolithic powder items. Their disadvantages are the lower values of μ, B_r, B_S and much greater fragility than the metallic soft magnetic materials. The effect of hysteresis in magnetic materials can be used for production of magnetic recording media.

Hard magnetic materials, in addition to large values of H_c, should also have significant quantities of B_r, defining the B values in the magnet, and the maximal product $(BH)_m$, measured in the second quadrant of Fig. 1.8.17, where $H < 0$ and $B > 0$. The $BH)_m$ determines approximately the maximum magnet torque in the field of H. It is also desirable to have a high temporal and thermal stability of these parameters and satisfactory mechanical strength and ductility.

To get the maximal values of H_c, we should performe some obligatory conditions. First, it is necessary to ensure disability of reversal due to the motion of domain walls. To do this, we should create a structure, in which a small single-domain ferromagnetic particles are surrounded by layers of paramagnetic material. In this case, reversal may be realized only owing to rotation of the vector M of domain that is feasible only in relatively large fields H. Such a structure, consisting of single-domain particles, is obtained generally by fine grinding of ferromagnetic substance (with subsequent mixing with the paramagnetic binder and sintering), or using separation of a homogeneous solid solution on two phases (the paramagnetic and ferromagnetic). Secondly, to make the rotation of the domain M more difficult, we can serve the substances with a very strong magnetic anisotropy (alloys Fe - Nd - B, Sm - Co, and some types of ferrites) or provide elongated shape of domains (in the alloys of Fe-Cr-Co, Fe-Co-Ni-Al).

All the parameters (H_c, B_r, $(BH)_m$) are increased when material has the same orientation of the easy magnetization axes (or in some cases, the long axes of domains) along one direction. The latter is achieved by orienting the milled powder particles in a strong magnetic field (about 1 T) or carrying out the initial stages of separation of solid solution on phase in an external magnetic field (of the order of 0.1 T). For this reason, we can use single-crystalline magnets,

formed as two-phase structure, that also improves all the magnetic parameters compared to polycrystalline materials.

The most common permanent magnets (see, Table 1.8.3) are prepared in accordance with the above described principles. All magnets, produced by "powder technology", are generally fragile and not enough durable. The same shortcomings also have magnets based on Fe-Co-Ni-Al alloys. The magnets based on the supersaturated Fe-Cr-Co solid solutions are free form these shortcomings when they are divided into two phases, but their use is constrained by the relatively small values of H_c.

Table 1.8.3. The most common hard magnetic materials and their typical magnetic properties

Groups of materials	H_c, **A/m**	B_r, **T**	$(BH)_m$, **kJ/m^3**
Fe – Nd - B	1000-1200	1.2-1.4	600-800
Sm - Co	1200-1500	1.0-1.1	400-600
Fe-Co-Ni-Al	50-120	1.0-1.2	40-60
Fe-Cr-Co	40-70	1.3-1.6	40-60
Ferrites	30-100	0.3-0.5	10-15

Currently, the best set of "magnetic" parameters have magnets based on Fe-Nd-B alloys, obtained by the "powder technology". They presents single-domain particles of ferromagnetic materials that are prepared from finely milled powder with composition $Fe_{14}Nd_2B$, that is compacted and subjected to heat treatment. The sintered material having a complex a tetragonal lattice (with 68 atoms in the cell) has a very strong magnetic anisotropy along the axis $\vec{c}$. The powders from low-melting substances or neodymium alloy is use as a binder. The powder mixture corresponding to the future of the magnet is pressed in molds in a magnetic field, then sintered at a temperature of 1000°C. It is clear that this method does not allow to achieve the high strength. A significant drawbacks of this material are also relatively low Curie temperature and a strong decrease of the parameters during heating.

Magnets based on alloys of Sm – Co are manufactured by the same principles and technology. Magnetic materials based on Fe-Cr-Co alloys have a good set of magnetic and mechanical (close to stainless steels) properties, see Table 1.5.2. They, in particular, may be manufactured as sheets and wires, and also using die-stamp and cutting processes.

There are large groups of "semi-hard" magnetic materials with desired values H_c, based on Fe-Co-V, Fe-Cr-Co-V, Fe-Co-Mo and others alloys. Their main applications are switchgear, cores of pulse relays, magneto-mechanical

displays and other devices in which the magnetization of the core occurs up to saturation by short pulse current. Similar properties have materials for hysteresis motors, the torque of which is provided by the reversal of rotor composed from a semi-hard material and put into rotating magnetic field of complex configuration.

1.8.5. Thermal properties of crystals

Thermal impacts are one of the main types of external influences on the crystal system, which determine their properties as well as processes occurring in them. *Thermal impacts* on the crystalline solid is provided to the heat energy supply and extraction, which leads respectively to the temperature increase or decrease and a change of internal energy in the crystal.

Since every crystalline solid consists of two subsystems – ions and the electrons, any thermal impact will always alter their energy spectra, the parameters of the dispersion laws, dynamics of movement of particles which form crystal (ions, molecules, electrons), the number and type of defects arising in a crystal lattice and also the type (structure) and the state of the crystal. As a result, these changes and the character of the defects ensemble appeared in the crystal will determine its basic thermal properties.

In this chapter, the basic thermal properties of crystals (heat capacity, heat conductance, thermal expansion) and also generation of lattice defects (which determines migration (diffusion) of atoms and ions) due to thermal impacts will be considered on the basis of quantum representations of the energy spectrum and dynamics of phonons and electrons.

Thermal properties of the crystal lattice. The oscillatory nature of the atoms (ions) motion in crystalline solids with the ordered (lattice) structure under explosion of temperature results in the change in the crystal internal energy (by changing the phonon number, energy, and momentum) and the potential energy of the atoms (ions) through the change of their chemical bonds energy, or even their type. Changing the number and energy of phonons with temperature determines the heat capacity of crystals. The change in the number of phonons and their momenta determines the lattice thermal conductance. Growth of amplitude of atomic vibrations around the equilibrium position, together with the anharmonism of oscillations, leads to the thermal expansion of the crystal lattice, as well as generating of point defects determining mechanisms of ion migration (diffusion) through the crystal.

The classical theory of the heat capacity of the crystal lattice. As is known from statistical physics and thermodinamics, the heat capacity is the amount of heat dQ, absorbed by the body, divided by the temperature change dT while absorbing. For solids and liquids, the delivered heat dQ is almost entirely (the

volume of solid is constant) used on the change of internal energy dU (change the energy of motion and interaction of particles composing the body). In solids, the change of the internal energy dU_{int} is primarily due to the change of energy dE of atomic vibrations.

In the experiment with condensed substances, heat capacity at constant pressure $C_p = \left(\frac{\partial U}{\partial T}\right)_p$ or constant volume $C_V = \left(\frac{\partial U}{\partial T}\right)_V$ are usually measured. These capacities are related by the known relation $C_p - C_V = 9\alpha^2\beta\rho T$, where α is the coefficient of linear expansion, β – bulk modulus. For example, for the Fe, value of $\alpha = 1{,}2\cdot10^{-5}$ K^{-1}, $\beta = 1{,}7\cdot10^{5}$ cm^{-3}; the density $\rho = 7{,}86$ g/cm^3, so that the value $C_p - C_V = 0.48$ J·mol^{-1} at $T = 300$ K.

In 1818 the Dulong-Petit law has been experimentally established, according to which the heat capacity C_V of all solids at a sufficiently high temperatureы is constant, independent of temperature and equaled $3R \approx 25$ J/mol·K. This law was easily explained based on the classical concept of a solids as a set of independently vibrating oscillators (atoms in the crystal lattice).

In the classical theory of heat capacitance, homogeneous solid is regarded as a set of completely independent of each other particles (classical oscillators) to vibrate with the same frequency ω. Every vibrating particle has three degrees of freedom, so that at an average $(1/2)k_BT$ kinetic energy and $(1/2)k_BT$ potential energy falls for each degree. Therefore, the total average energy of vibrating particle is $3k_BT$. In this case, the internal energy of one mole of a substance containing N particles is:

$$U_{\text{int}} = 3Nk_BT = 3RT, \qquad (1.8.48)$$

where R is universal gas constant. According to the definition of the specific heat, after differentiation of (1.8.48) at a temperature T, we obtain the molar heat capacity of the atoms of a solid at constant volume as the Dulong-Petit law:

$$C_V = \frac{dU_{\text{вн}}}{dT} = 3R. \qquad (1.8.49)$$

Table. 1.8.4 shows the atomic heat capacities of some substances at ambient (room) temperatures. As is seen, in most cases, the law of Dulong-Petit is performed sufficiently well. However, for example, for diamond and boron specific heat at room tempera was significantly lower than 25 J/mol·K. However, this discrepancy of the classical model and the experimental values of the heat capacity is not critical. More importantly, the Dulong-Petit law, in accordance with the classical concepts, should be violated for metallic crystals: they have not only an ensemble of atoms but also the conduction electrons, which according to classical statistics, must have a kinetic energy by $k_BT/2$ for each of the 3 degrees of freedom. This means that the heat capacity of

monovalent metal should be equal to the $C_V = (9/2)R$, i.e. 1.5 times greater than the value of (1.8.49) given by the Dulong-Petit law. Therefore, in the framework of the classical approach remains unexplained why the electrons do not contribute to the heat capacity at high temperatures.

Table 1.8.4. The experimental values of the heat capacity of crystalline substances

Elements	Na	Al	Fe	Ni	Cu	Zn	Sn	Pt	Pb	Cd	B	C
C_V, J/mole·K	27	23.5	24.7	24.7	23.5	24	25.5	24.7	24.7	24.7	14.2	5.7

The second fundamental question unexplained by the classical theory of heat capacity is the temperature dependence of the specific heat. In contrast to the predictions of classic model, the experimental data show that C_V is dependent on temperature below room temperature. As can be seen from Fig. 1.10.19, when approaching absolute zero temperature, heat capacity of crystals decreases rapidly to zero, although, according to the classical theory, it should be remained the same – to go along the dotted horizontal line in Fig. 1.10.19.

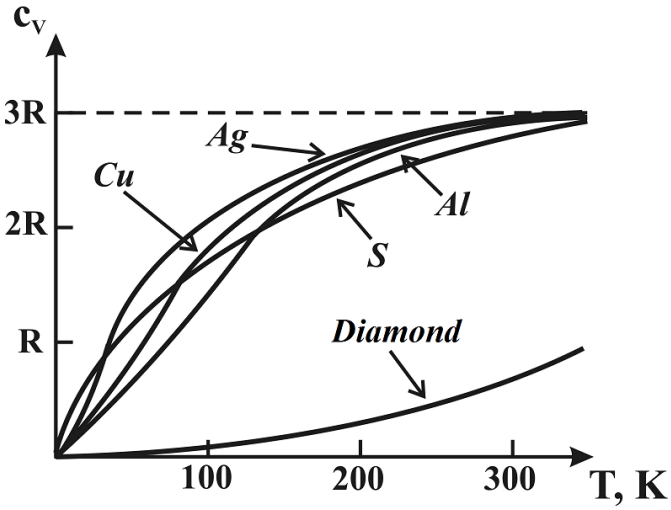


Fig. 1.8.19. The temperature dependence of the specific heat for a number of elements

The discrepancy between specific heats for the classical theory and experiments, especially at low temperatures, are due to two main reasons. First, in the classical theory assumed that the lattice atoms are considered as a lassical, independent oscillators, oscillating with the same frequency. In fact, as shown in Section 1.2, the atoms in crystals form a coherent system, possessing a wide range of vibrational modes. Second, the vibrating atoms are really not classical

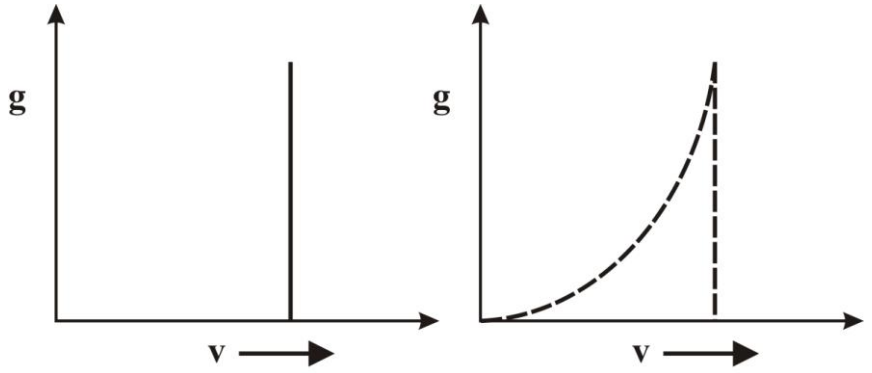
but quantum oscillators with the discrete energy spectrum. Therefore, the further development of the lattice specific heat theory followed by using quantum concepts of crystalline solids.

The quantum theory of heat capacity. As can be seen from Chapter 1.2, each normal mode (each elastic wave of atomic displacements) with frequency ω can be associated with the quantum harmonic oscillator with a mass of vibrating atoms. According to (1.8.59), the energy of such harmonic oscillator (with zero-point energy E_o) is quantized (see Fig. 1.8.17)

$$E_n = 3(N + \frac{1}{2})\hbar\omega = E_0 + N\hbar\omega, \qquad (1.8.50)$$

where N = 1, 2, 3 ... are the numbers of the energy levels for the oscillator.

This means that, due to the transition of a quantum oscillator (phonon) with quasi-momentum $p = \hbar k$ from one state to another, the energy can only be changed by an amount well $E_{ph} = \hbar\omega$. This approach is fully consistent with that described in the Chapter 1.2 as a model of a solid box (volume), filled with the gas of 3N independent harmonic oscillators or phonons (N is number of atoms in the crystal) of various energies.



a) *b*)

Fig. 1.8.20. Einstein (*a*) and Debye (*b*) phonon spectra in crystals

The first quantum model of the heat capacity of the crystals was developed by Einstein. He used the assumption that all quantum oscillators vibrate at the same frequency ω_E, so that the phonon energy is

$$E_n = N\hbar\omega_E. \qquad (1.8.51)$$

This means that phonon spectrum g(ω) of Einshtain crystal has a form as in Fig. 1.8.20a (delta-function). Then, the average energy of the oscillator according to the Bose-Einstein statistics willbe equal

$$\langle E_n \rangle = \frac{\hbar\omega_E}{\exp\left(\frac{\hbar\omega_E}{k_B T}\right) - 1} \tag{1.8.52}$$

It follows that in the low-temperature range (low frequencies), the average energy of the Einstein crystal is equal to

$$U \approx 3Nk_B\omega_E \exp\left(\frac{\hbar\omega_E}{k_B T}\right), \tag{1.8.53}$$

And therefore, by the Einstein model, the heat capacity at low frequencies (low temperatures) is

$$C_V = \left(\frac{dU}{dT}\right)_V \approx 3Nk_B\left(\frac{\hbar\omega_E}{kT}\right)^2 \exp\left(-\frac{\hbar\omega_E}{kT}\right). \tag{1.8.54}$$

At high temperaturah when $\hbar\omega_E << k_B T$ or $T >> \hbar\omega_E/k_B$, this equation is transformed into the Dulong-Petit law (1.8.49). The total energy of the crystal at all temperatures in the Einstein model is given by relation

$$U = \frac{3N\hbar\omega_E}{\exp\left(\frac{\hbar\omega_E}{k_B T}\right) - 1}, \tag{1.8.55}$$

where Einstein heat capacity at any temperature will be of the form

$$C_V = 3Nk_B F_E(\omega_E, T), \tag{1.8.56}$$

where

$$F_E(T) = \frac{\left(\frac{\hbar\omega_E}{k_B T}\right)^2 \exp\left(\frac{\hbar\omega_E}{k_B T}\right)}{\left(\exp\left(\frac{\hbar\omega_E}{k_B T}\right) - 1\right)^2}$$

is called *the Einstein function*. A characteristic feature of the $C_V(T)$ dependence for Einstein's model is the decrease of the heat capacity of the crystal to zero as the temperature is lowered to $T = 0$ K. However, the exponential temperature dependence of the heat capacity of the crystals given by this model is not quantitatively match the experimental curves (see Fig. 1.8.21).

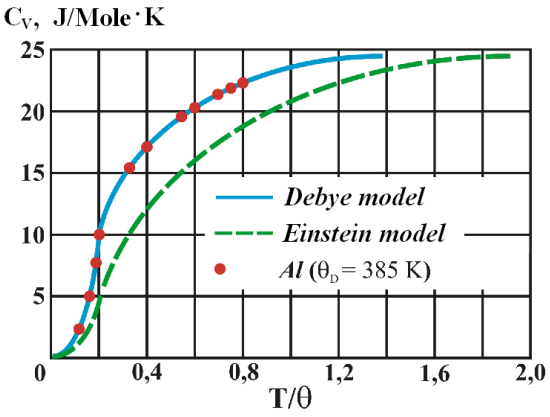


Fig. 1.8.21. Comparison of the temperature dependences of the Einstein and Debye heat capacities with the experimental one for aluminum

Quantum theory of the Debye heat capacity. As noted in Chapter 1.2, the interacting of the atoms in a crystal is so strong that they can not oscillate independently of each other. Therefore, in reality, these N atoms in the lattice are coupled to the system of 3N quantum harmonic oscillators (phonons) vibrating with different (not one as in Einstein's model) frequencies.

Determination of the intrincic oscillations in this system presents significant difficulties. These difficulties were overcome by Debye theory for the first time for the case of an isotropic crystal with a linear dispersion law. According to the Debye model, the total energy E of atomic vibrations in the crystal lattice can be found by summing the energy of all the normal vibrations (phonons) with different frequencies in the bulk of a solid

$$E = \int_{\omega_{min}}^{\omega_{max}} (N + \frac{1}{2})\hbar\omega V_0 f(\omega) g(\omega) d\omega, \qquad (1.8.57)$$

where $f(\omega)$ is the Bose-Einstein function (1.8.35a), $g(\omega)d\omega$ – the number of normal oscillations per unit volume and in the frequency interval from ω up to $\omega + d\omega$, and ω_{min} and ω_{max} – the minimum and maximum frequencies, limiting the spectrum of normal vibrations ((ω_{max} is called the characteristic Debye frequency ω_D).

The expression for the Debye phonon distribution by frequencie (Debye phonon spectrum) $g(\omega)$ has been defined in Section 1.2 as a quadratic dependence on frequency

$$g(\omega)d\omega = \frac{3\omega^2}{2\pi^2 v^3} d\omega \qquad (1.8.58)$$

where v is the velocity of phonons propagation (the speed of sound waves in a crystal). Graphical image of the Debye phonon spectrum is shown in Fig. 1.8.20b. The lowest possible frequency ω_{min} *in* Debye model is believed to be zero, since the body has a large size of the order of *Na*. Then the total number of normal vibrations (phonons) arising in the crystal lattice (potential box) is

$$3N = \int_0^{\omega_D} V_0 g(\omega) d\omega = \frac{3V_0}{2\pi^2 v^3} \int_0^{\omega_D} \omega^2 d\omega. \qquad (1.8.59)$$

Hence, we can find Debye frequency ω_D as

$$\omega_D = v\left(6\pi^2 \frac{N}{V_0}\right)^{1/3}. \qquad (1.8.60)$$

Temperature θ_D, at which phonons with maximal frequency ω_D is generated, is determined by relation

$$\theta_D = \frac{\hbar\omega}{k_B}, \qquad (1.8.61)$$

and is called *the Debye characteristic temperature*. Typical values of θ_D for some substances are shown in Table 1.8.5.

The Debye temperature is a very important physical constant of substances, which characterizes many of the properties of crystalline solids, such as heat capacity, electrical conductivity, thermal conductivity, the broadening of lines in X-ray spectra, elastic properties, etc. The physical meaning of the θ_D temperature is that the full range of normal vibrations in the crystal is generated just when heating to this temperature. Beyond this temperatures, phonons with higher energies are not generated.

The temperature increase above the θ_D leads only to an increase in the degree of excitation of each normal mode (increase in the number of phonons for every value of its energy). If temperatures are $T > \theta_D$, they are called the high temperatures. For many crystalline materials (gold, silver, tin, KCl), room temperature are yet high. However, for other substances, like, for example, silicon, diamond, boron, etc. θ_D values are much higher than room temperature. That is why in Fig. 1.10.19 for these substances the Dulong-Petit law is not valid.

Table.1.8.5. Debye temperatures for some substances

Metals	**θ_D, K**	**Semiconductors**	**θ_D, K**
Hg	60-90	Sn (grey)	212
Pb	94,5	Ge	366
Na	160	Si	658
Ag	225		
W	270	Insulators	
Cu	339	AgBr	150
Fe	467	NaCI	320
Be	1160	Алмаз	1850

The Eq. (1.10.60) allows us to express the propagation velocity v of the vibrations with the Debye characteristic frequency as

$$v = \omega_D \left(6\pi^2 \frac{N}{V_0} \right)^{-1/3} . \qquad (1.8.62)$$

Substituting this relation into (1.2.33), we obtain an expression for the number of normal vibrations (phonons) in terms of the Debye characteristic frequency:

$$g(\omega)d\omega = \frac{9N\omega^2}{V_0\omega_D^3} d\omega. \qquad (1.8.63)$$

The total energy E of the thermal vibrations of the crystal lattice in the Debye model is found by substituting the Bose-Einstein distribution function $f(\omega,T)$ and Eq. (1.8.63) in (1.8.64):

$$E = \int_0^{\omega_D} 9\hbar N \left(\frac{1}{\exp\left(\frac{\hbar\omega}{k_B T} \right) - 1} \right) \frac{\omega^3}{\omega_D^3} d\omega = \qquad (1.10.64)$$

$$= E_0 + \frac{9N}{\omega_D^3} \int_0^{\omega_D} \frac{\hbar\omega^3}{\exp\left(\frac{\hbar\omega}{k_B T} \right) - 1} d\omega$$

where $E_0 = (9/8)\hbar\omega_D N$ is the total zero-point energy (at $T = 0$) of crystalline lattice in the entire frequency range from 0 to ω_D. The resulting integral in

(1.8.64) can be simplified by replacing the variable ω on the $x = \frac{\hbar\omega}{k_B T}$ and substitution θ_D in the form (1.10.61):

$$E = E_0 + 9k_B N\theta_D \left(\frac{T}{\theta_D}\right)^4 \int_0^{\theta_D/T} \frac{x^3}{e^x - 1} dx. \qquad (1.8.65)$$

Consider the dependence of the heat capacity $C_V = \frac{dE}{dT}$ on temperature. The preferred way to do this separately for the low ($T << \theta_D$) and high ($T >> \theta_D$) temperatures, where the integral in (1.8.65) can be calculated exactly. For intermediate temperatures, these integrals are computed numerically. Taking into account that at low temperatures ($T << \theta_D$) the upper limit of integration $\theta_D/T \to \infty$ in expression (1.8.65), and also that

$$\int_0^{\theta_D/T} \frac{x^3}{e^x - 1} dx \approx \frac{\pi^4}{15},$$

we obtain

$$C_V = \frac{dE}{dT} = \frac{12}{5}^4 \pi k_B N \left(\frac{T}{\theta_D}\right)^3 \sim T^3 \qquad (1.8.66)$$

The relation (1.8.72) is called *the Debye law of cubes*. It is closely coinsides with the heat capacity temperature dependence of the crystals shown in Fig. 1.10.14. Note, however, that for a strongly anisotropic crystals and complex vibration spectrum (severe deviation from the linear dispersion law) this cubic relation is violated. Therefore, for example, for layered crystals is $C_V \sim T^2$ and for filamentous substances - $C_V \sim T$.

At low temperatures, both the energy of each normal vibration (phonon energy) and the number of excited normal oscillations with increasingly higher frequencies (increasing the number of phonons with high ω) increases with the temperature increase.

At high temperatures, $T >> \theta_D$, when all the normal fluctuations are excited (because their spectrum is limited by the frequency ω_D) and their energy no longer increases with temperature, the vibrations energy may be increased only by increasing the number of phonons for each value of the energy. Indeed, for high temperatures ($\frac{\hbar\omega}{k_B T} \to 0$, $\exp\left(\frac{\hbar\omega}{k_B T}\right) \approx 1 + \frac{\hbar\omega}{k_B T} + ...$), so that the formula for the total vibrations energy, after integration of (1.8.71), is:

$$E = E_0 + \frac{9N}{\omega_D}\int_0^{\omega_D} k_B T\omega^2 d\omega = E_0 + 3Nk_B T. \qquad (1.8.67)$$

This means that the heat capacity of the crystal at high temperatures does not depend on temperature, as described above Dulong-Petit law. This coincids also with the Einstein's model

In general, as shown by the experimental $C_V(T)$ dependences, they are in good agreement with the theoretical Debye model merely for a simple crystal lattice. How good the Debye model is consistent with the experiments, is seen from Fig. 1.10.21, which shows the dependence of the specific heat of aluminum, copper, and silver on the reduced temperature T/θ_D.

The heat capacity of free electrons. As shown in Section 1.3, according to the classical theory, the specific heat capacity of the electron gas obeys the relation (1.3.15) in the form of $C_e = (3/2k_B)n$ or for one mole of a substance

$$C_V = (3/2)R, \qquad (1.8.68)$$

where n is the concentration of free electrons, and R – the gas constant. This means that free electrons must also contribute to the thermal energy of the crystal, leading to its full heat capacity equal to

$$C_V = 3R_{(\text{ions})} + 3R/2_{(\text{eletrons})} = 9R/2.$$

However, the experiments show that even in metals (with the highest concentration of free electrons), the electron contribution to the heat capacity is almost absent, that is inexplicable within the classical approach.This contradiction between theory and experiment is explained in the framework of the band model of the electron energy spectrum of crystals.

As follows from Fig. 1.8.7, most of the electrons is complitely insensitive to the temperature changes due to step-like character of Fermi-Dirac distribution. As a result, at normal temperatures, electrons does not include contribution to the heat capacity of the crystal, because the thermal excitation affected only a very small their part in the range of ($k_BT \div 2k_BT$) around the Fermi energy ε_F, as was shown in Section 1.3.

According to the mean value theorem, the average energy of the electrons can be written as

$$\langle\varepsilon\rangle = \int_0^{\infty} \varepsilon f(\varepsilon) g(\varepsilon) d\varepsilon, \qquad (1.8.69)$$

which implies

$$\langle\varepsilon\rangle = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} \int_0^{\infty} \frac{\varepsilon^{3/2} d\varepsilon}{\exp\left(\dfrac{\varepsilon - \varepsilon_F}{k_B T}\right) + 1}, \qquad (1.8.70)$$

where the integral is called *the Fermi integral*.

Finding the Fermi integral in (1.8.70) leads to the following expression for the mean energy of a gas of free electrons in the metal

$$\langle \varepsilon \rangle = \frac{3}{5} n \frac{\left(\varepsilon_F(T)\right)^{5/2}}{\left(\varepsilon_F(0)\right)^{3/2}} \left[1 + \frac{5}{8} \left(\frac{\pi k_B T}{\varepsilon_F(T)} \right)^2 \right]. \tag{1.8.71}$$

Hence, using the formula (1.8.36) for the temperature dependence of the Fermi energy, we obtain an expression for the total energy of a gas of free electrons in a crystal:

$$\begin{aligned} U &= \frac{3}{5} n \varepsilon_F(0) \left[1 + \frac{5}{8} \left(\frac{\pi k_B T}{\varepsilon_F(0)} \right)^2 \right] \left[1 - \frac{5}{24} \left(\frac{\pi k_B T}{\varepsilon_F(0)} \right)^2 \right] \approx \\ &\approx \frac{3}{5} n \varepsilon_F(0) \left[1 + \frac{5}{8} \left(\frac{\pi k_B T}{\varepsilon_F(0)} \right)^2 - \frac{5}{24} \left(\frac{\pi k_B T}{\varepsilon_F(0)} \right)^2 \right] = \\ &= \frac{3}{5} n \varepsilon_F(0) \left[1 + \frac{5}{12} \left(\frac{\pi k_B T}{\varepsilon_F(0)} \right)^2 \right], \end{aligned} \tag{1.8.72}$$

Where the total energy at $T = 0$ K is equal

$$U_0 = (3/5) n \varepsilon_F(0) \,. \tag{1.8.73}$$

Differentiating (1.8.72) by temperature, we obtain the following expression for the quantum heat capacity of the electron gas in a model of Sommerfeld

$$C_{el} = \frac{n \pi^2 k_B^2}{3 \varepsilon_F(0)} T = \gamma T, \tag{1.8.74}$$

which shows linearity of C_e with temperature.

This implies communication with the classical electron heat capacity $C_{\text{class}} = (3/2) k_B T$ in form

$$C_{el} = \frac{\pi^2 k_B^2}{3 \varepsilon_F(0)} T C_{class} = \gamma T C_{class} \,.$$

The last formulas implies that only those electrons which have energies close to the Fermi level, can contribute to heat capacity and electrical conductivity of the crystal. The share of these electrons is very low, due to is proportional to the ratio of lattice thermal energy to the Fermi energy $k_B T/\varepsilon_F$. For example, as for the metal Fermi energy is the order of several electron volts, for T = 300 K, the value $k_B T \approx 0{,}003$ eV, so that the ratio $k_B T/\varepsilon_F << 1$. That is why the heat capacity of the electron gas at room temperature is much less the crystal lattice heat capacity which is correspondent to the Dulong-Petit law.

If the phonons and electrons contribute to the heat capacities of crystal, the total heat capacity should be written as follows:

$$C_V(T) = C_{\text{el}} + C_{\text{latt}} = \gamma T + \alpha T^3 . \qquad (1.8.75)$$

As follows from (1.8.75), in the temperature region close to absolute zero, specific heat of the crystal lattice ($C_V \sim T^3$) falls down with the temperature decrease much faster than the heat capacity of the electron gas ($C_{\text{эл}} \sim T$), so that the electronic specific heat gives the main contribution in the capacity of the crystal. However, this occurs only at very low temperatures (order of tens of Kelvin), and so at higher temperatures heat capacity of the electron gas in the crystals is neglected. The temperature T_c, at which these contributions are equal, i.e. $\gamma T_c = \alpha T_c^3$, are usually a tenth of the Debye temperature $T_c \approx (1/10)\theta_D \leq 30$ K.

Anharmonicity of the atomic vibrations. When considering the vibrations of the atoms in the crystal lattice and also their heat capacity, it was assumed that the forces acting between atoms are elastic, so that atoms perform harmonic oscillations with small amplitudes relative their average equilibrium positions. In this case, the potential energy increases as the square of the oscillator deviation from the equilibrium position, and parameters describing the lattice rigidity are not changed with temperature increase. In other words, phonons in potential box are not interacting. This allowed to divide the whole spectrum of vibrations on the independent modes, to calculate thermal energy of the crystal in this approximation, and to obtain relations for the temperature dependence of the specific heat (in the Debye model), well describing its behavior at high and low temperatures.

However harmonic approximation can not explain certain thermal properties of crystals, such, for example, as temperature dependences of elastic moduli, thermal expansion, the lattice thermal conductivity, etc. This is due to the fact that the description of these properties requires accounting of the so-called anharmonic corrections in the interaction forces between the atoms in the lattice is not quite elastic, that is, they depend on the displacement of atoms. In other words, atomic displacements relative to equilibrium position are not linear because contain anharmonic terms of second and higher degree, whose influence is grown especially with the temperature increase. This means also that, in fact, the phonons interact to each other and with other quasi-particles (such as electrons, photons, etc.).

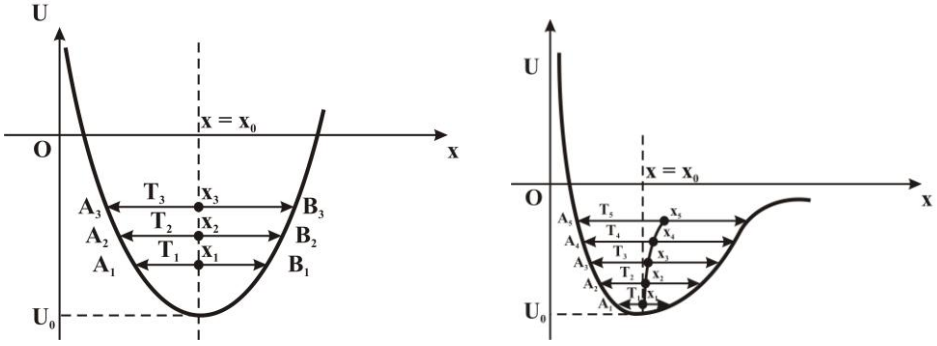
In order to understand the nature of the anharmonicity of atomic vibrations in crystals, we consider the dependence of the interaction energy on interatomic distance. This relationship is generally determined by the balance of the forces of attraction and repulsion, which can be described, for example, using the Lennard-Johnson potential

$$U(r) = U_{att} + U_{\text{rep}} = -\frac{A}{r^m} + \frac{B}{r^n} \qquad (1.8.76)$$

Here, A, B, m, n are constants determined by the type of interatomic interaction. Note that typically $n > m$.

Graphical view of this dependence is shown in Fig. 1.8.22b. In accordance with this image, at absolute zero temperature the atoms are located at a distance r_0, corresponding to minimal interaction energy U_0 (at the bottom of the potential well *abc*). These equilibrium interatomic distances determine the size of the body at absolute zero temperature. When temperature increasing, the atoms begin to oscillate around the equilibrium position at $x = x_0$. The oscillating particle has a kinetic energy that reaches the highest value W_{m} at the time when it passes the equilibrium position at $x = x_0$.

If particle vibrations are purely harmonic, the force F, which occurs due to deviation of atom from its equilibrium position on the distance x, would be elastic, that is strictly proportional to the magnitude of the deviation, and is directed towards the equilibrium position. In this case, changing the potential energy of a particle ΔU would be described with the parabolic equation $\Delta U = (1/2)ax^2$. This parabola is symmetric about the point $x = x_0$. Therefore, the deviation to the left and to the right would be the same in size, so that the center of swing x_{i} would coincide with the equilibrium position $x = x_0$. Heating the body in this case might not arouse its expansion with the temperature increase. Actually, in spite of the increase of the amplitudes of the oscillating particles, the position of the centers of the oscillations (the average interatomic distances) would remain unchanged.



a) *b*)

Fig. 1.8.22. Potential energy on interatomic distance $U(x)$ for harmonic (*a*) and anharmonic (*b*) oscillators

In fact, the plot of the potential energy on the distance, as can be seen from Fig. 1.8.22b, is asymmetrical about the equlibrium position of $x = x_0$: its left branch is raised considerably steeper than the right one. This means that the oscillations of the particles in a solid are unharmonic.

To take into account the asymmetry of the curve $U(r)$, it is necessary to include an additional term - $(1/3)gx^3$ into U(x) dependence

$$U = (1/2)ax^2 - (1/3)bx^3 \tag{1.8.77}$$

(b is a coefficient).

The thermal expansion of solids. Experiments show that almost all crystalline solids are expanded with the temperature increase. The nature of this phenomenon is quite complicated. However, it is believed that the thermal expansion is mainly due to an increase in average distance between the atoms of the crystal. Among the main reasons, which can result in the solid expansion, are the asymmetric dependence of the potential energy of atomic displacement from equilibrium positions, the change in the forces of interaction between the atoms with increasing temperature, the regrouping of different atoms, changing the preferred orientation of the electron density distribution, etc. In this part of Book we will consider mainly the first reason, since it is usually the most important.

Fig. 1.8.22a is easy to demonstrate that the average value of their displacements from equilibrium positions at any temperature T_1, T_2, T_2 , etc. is zero $\langle x_1 \rangle = \langle x_2 \rangle = \langle x_3 \rangle = 0$. This means that the distance between atoms, performing harmonic oscillations when heating should remain equal to x_0.

The same result can be achieved by a formal calculation of the average displacement of the harmonic oscillator from the equilibrium using the mean value theorem:

$$\langle x \rangle = \frac{\int_{-\infty}^{\infty} x \exp\left(-\frac{ax^2}{k_B T}\right) dx}{\int_{-\infty}^{\infty} \exp\left(-\frac{ax^2}{k_B T}\right) dx} = 0 \tag{1.8.78}$$

This means that the coefficient of thermal expansion of solids in the harmonic approximation should be zero, that contradicts the experiment. As is seen from Table. 1.8.6, for the most solids relative expansion when heating to approximately 1 K is about 10^{-6}-10^{-5}.

Table 1.8.6. The coefficients of linear expansion at room temperature for some substances

Substance	**$\alpha \cdot 10^6$, K^{-1}**
Li	56
B	2
Cu	16,6
Ga	18
Ge	5,8
Fe	12
Co	12
Ag	19
Cd	32,5
Au	14

As shown above, the nonsymmetric nature of the $U(r)$ dependence on the interatomic distance leads to the fact that the $\langle x_1 \rangle \neq \langle x_2 \rangle \neq \langle x_3 \rangle \neq 0$, so that, when heating the body, the average distance between the particles should be increased and the body should be expanded (Fig. 1.8.22b).

Calculating the average displacement for the anharmonic oscillator with temperature T on the basis of (1.8.77) gives the relation

$$\langle x \rangle = \frac{3k_B bT}{4a^2} \tag{1.8.79}$$

where a and b are constants in Eq. (1.8.75). Then, the relative linear expansion of solid material can be expressed as the ratio of the average distance between the particles to the normal distance x_0 between them

$$\frac{\langle x \rangle}{x_0} = \frac{3b}{4a^2} k_B T = \alpha T \tag{1.8.80}$$

This result is consistent with the frequently used empirical law of linear thermal expansion for solids near the room temperature, where $l(T) = l_0 \lfloor 1 + \alpha (T - T_0) \rfloor$. Here l_0 is the length of the sample at $T = T_0$, and $l(T)$ – the length of the sample at a given temperature T. The parameter α is called *the coefficient of linear thermal expansion* or symply *coefficient of thermal expansion*, which is defined as $\alpha = \frac{1}{l}\frac{dl}{dt}$. Thus, in the first approximation the thermal expansion coefficient is proportional to absolute

temperature of the solid. Substitution of the numerical values of b, k_B, a and x_0 gives the values of the order of 10^{-4}-10^{-5} for α. This is in good agreement with the data of Table 1.8.6.

Thermal conductivity of crystal lattice. All crystalline solids more or less able to conduct heat. In an isotropic solid, heat transfer in the x-direction obeys the Fourier equation:

$$\Delta W = \frac{dQ}{dt} = -\kappa S \frac{dT}{dx}, \qquad (1.8.81)$$

where Q – *the heat flux density* (heat flow through a unit cross section, perpendicular to the vector Q), T – temperature, $\partial T/\partial x$ – temperature gradient along the x axis, and κ – *coefficient of heat conductance*. The minus sign in the right-hand side of (1.8.81) is due to the fact that the heat flux vector is opposite to the temperature gradient, i.e. is transfered from hot to cold region of the solid sample. In the experiments, basing on formula (1.8.81), the coefficient of thermal conductivity κ is defined as the amount of heat energy ΔW transmitted per unit area of the sample for time Δt and for unit temperature gradient $\partial T/\partial x$.

Experimental data for dielectrics show that κ increases with the increase of temperature from absolute zero, runs at the temperature of 30-50 K by diffuse maximum and decreases at a further increase in temperature. At relatively high temperatures (of the order of the Debye temperature), the thermal conductivity decreases as $1/T$.

Reported above the experimental behavior can be explained on the based of the following concepts. In general, in solids there are two basic mechanisms of heat transfer – the transfer of thermal energy by atomic vibrations (phonons in the model of quasi-particles in the potential well) and free electrons. In metals, both mechanisms act simultaneously, and in dielectrics – only the first.

Consider the phonon mechanism of heat transfer in dielectrics which actually have no free electrons. Since the atoms in the crystal are interlinked, then the amplitude of the atomic vibrations increases on the heated part of solid sample (the energy and the number of phonons is icreased), so that the kinetic energy of atomic vibrations (phonons in the model of quasi-particles) is transferred from the heated part of the sample to a colder part. Macroscopically, this looks like a heat flow. Thus, the heat transfer in the crystal is due to propagation of elastic sound waves.

As noted above, in explaining the phenomenon of thermal conductivity, we can no longer assume that the strictly elastic (harmonic) waves are transfered in the crystal lattice, because such waves will not interact with each other. Such harmonic waves would propagate in the crystal freely without attenuation, consequently phonons in the potential box model would have unlimited mean

free paths. Then the heat flow, even at low temperature gradients, could be present indefinitely long time, before before the heat equilibricity to be established, and thermal conductivity would be infinite.

In real solids, as shown by experiment, heat conductance is finite. The finit value of the thermal conductivity is due to the fact that in a real crystal lattice atomic vibrations are not purely harmonic because the forces of interaction between the atoms are non-linearly dependent on the atomic displacements.

Introducing the anharmonic contribution $b \cdot x^3$ in Eq. (1.8.77) for the potential energy of vibrating atoms, we thus take into account that in the real solids phonons are scattered on each other. Obviously, collisions of the carrying energy phonons should lead to a change in their momenta (Fig 1.2.22 in Section 1.2.), and thereby affect the heat transfer process. If these collisions change the momentum vector by a small angle (see Fig. 1.2.22a), they are less affected by the transfer of heat (the heat flux is almost unchanged). At the same time, as follows from Fig. 1.2.22b, collision processes, leading to a large change of momentum direction (direction of particles motion), cause a decrease in heat flux. Therefore, it is the phonon scattering at large angles determines the thermal resistance of the crystal (its value is inversely proportional to the κ. In this case, according to the quasi-momentum conservation law, we have to deal with the umklapp process for interacting phonons, see Section 1.2.

From the point of view of this approach, we try to analyze the temperature dependence of thermal conductivity. To do so, let us use the expression for the thermal conductivity obtained in kinetic theory of ideal gase, assuming that the heat is transferred not to the gas molecules but phonons:

$$\kappa_{ph}(T) = \frac{1}{3} C_{ph} v_{ph} \lambda_{ph} = \frac{1}{3} C_{ph} v_{ph}^2 \lambda_{ph}^2 \tau, \qquad (1.8.82)$$

where C_{ph} is heat capacity of the crystal, associated with the lattice vibrations, v_{ph} – average velocity of phonons (approximately equal to the sound velocity in the crystal), which can be regarded as weakly temperature dependent; λ_ϕ – phonon mean free path, which is equal to the average distance, which they pass between two successive collisions; $\tau = \lambda_\phi / v_\phi$ – effective relaxation time (mean free path time) of phonons, which the inverse value τ^{-1} corresponds to the frequency of phonon collisions. In Eq. (1.8.80) C_ϕ and λ_ϕ are values which mainly determine the dependence of heat conductivity on temperature.

At high temperatures $T >> \theta_D$, specific heat is approaching a limiting value determined by the Dulong-Petit law ($3Nk_B$), i.e. does not depend on temperature. Therefore, at $T >> \theta_D$ temperature dependence of conductivity is determined mainly by temperature changes in the mean free path of the phonons.

As for $T >> \theta_D$, in accordance with the Debye model, C_{ph} is approximately constant, the number of phonons with $k > \frac{G^*_{hkl}}{4}$ (large-angle scattering, see Fig. 1.2.22 in Section 1.2) is sufficiently high. Then the mean free path will be inversely proportional to the probability of collisions with phonons of this type. Such a probability is proportional to their number, and consequently the temperature T, since at high temperatures, according to the Debye theory, number of phonons is proportional to the temperature. The result is that at high temperatures the phonon mean free path is

$$\lambda_{ph} \propto \frac{1}{T}.$$

It follows that when $T >> \theta_{\text{D}}$

$$\kappa(T) = \frac{const}{T}, \tag{1.8.83}$$

that is observed experimentally.

Let us estimate the temperature dependence of the thermal conductivity at low temperatures ($T << \theta_{\text{D}}$). Since in this case C_{ph} is proportional to T^3, the number of phonons with $k > \frac{G^*_{hkl}}{4}$, scattered by umklapp process is rapidly (proportionally by $\exp\left(\frac{G^*_{hkl}\hbar\omega}{4k_B T}\right)$) decreased with temperature to zero. Hence, the probability of umklapp process also decreases exponentially, and this means that the phonon mean free path is increased exponentially up to infinity with the temperature decrease:

$$\left\langle \lambda_{ph} \right\rangle \propto \exp\left(\frac{\theta_D}{T}\right) \tag{1.8.84}$$

However, this does not happen, due to the effect of the phonons scattering by structural inhomogeneities (defects, grain boundaries, the surface of the crystal, etc.). In this case, λ_{ph} is determined mainly by defect density which is independent on temperature. Therefore, as under the Debye cubes law $C_{\text{ph}}(T) \sim T^3$, the value of $\kappa(T)$ is also proportional to T^3, that is observed experimentally.

If the defect density is low, the increase in thermal conductivity is mainly due to the sharply increasing exponential term for λ_{ph}, so that then

$$\left\langle \kappa_{ph} \right\rangle \propto T^2 \exp\left(\frac{\theta_D}{T}\right) \tag{1.8.85}$$

In this case, when the temperature approaches the T = 0 K, where the probability of umklapp process becomes small, the mean free path λ_{ph} is comparable with the size of the sample and is also independent of temperature. On further lowering the temperature, coefficient of thermal conductance drops sharply down to zero - as well as the specific heat by the T^3 law.

The above-described change in thermal conductivity with temperature is well supported by many experimental data. Fig. 1.8.23 shows a typical dependence of the thermal conductivity on temperature.

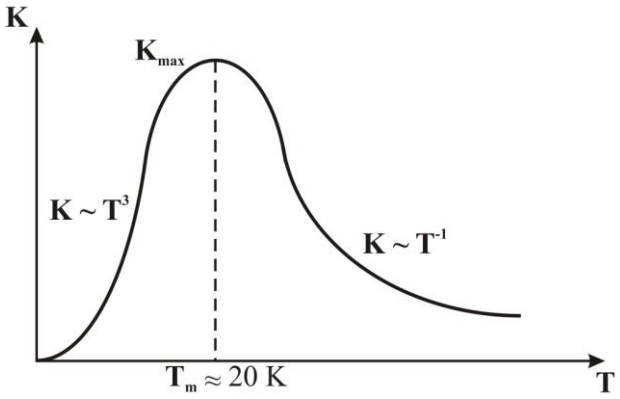


Fig. 1.8.23. Typical temperature dependences of phonon conductivity of dielectrics

The thermal conductivity of free electrons. The coefficient of thermal conductivity of free electrons can be considered in two approximations – classical and quantum. In the classical Drude approach, stated in section 1.3, the electron thermal conductivity was obtained in the form of Eq. (1.3.16)

$$\kappa_{ph} = \frac{1}{2} n k_B T \tau \frac{3k_B^2}{m} = \frac{3nk_B^2}{2m} \tau T ,$$

which takes into account only one mechanism of electron scattering - scattering by stationary atoms. In reality, basing on the ideas of quantum electron gas in the derivation of the thermal conductivity, it is necessary to use quantum statistics and also to take into account that the electron-phonon, electron-electron and electron-impurity scattering give the contributions to heat transfer, leading to the restriction of the heat flux carried by electrons. Since the scattering mechanisms in the quantum theory of the electron gas were discussed in detail in the previous chapter (in connection with the problems of electrical conductivity), here we will focus only on some of the highlights of the heat transfer in the electron gas.

As follows from the quantum theory of the electron gas (see Chapter 1.3), at high temperatures $T >> \theta_D$ major scattering mechanism in metallic crystals are electron-phonon collisions that lead to a temperature-independent relation for the thermal conductivity of electrons

$$\kappa_e \approx (\pi^2/3)\cdot(n\hbar^2/m^* k_B). \quad (1.8.86)$$

At low temperatures ($T << \theta_D$) we shoul take into account that electrons can emit or absorb only phonons with energie of $\hbar\omega \le k_B T$. In this case, the concentration of phonons, involved in these collisions, is $n_\phi \sim T^2$. This implies that the probability of electron-phonon scattering at low temperatures is inversely proportional to the square of the temperature, so that the electron-phonon contribution to the thermal conductivity is equal to

$$\kappa_{ph} \propto \frac{(\hbar\omega_D)^2}{(k_B T)^2} \quad (1.10.87)$$

2. Phase transformations in solid materials

2.1. Thermodynamics of phase transformations in crystalline materials

All the technologies of crystalline materials manufacturing and imparting them the given functional properties are based primarily on the change of such thermodynamic parameters as temperature *T*, pressure p and volume *V*. Changing *T*, *p*, and *V*, we may cause a change in the phase state of the materials. For example, changes in temperature or pressure can transform liquid substance into a solid or gaseous or change the crystal structure of solid material from one type to another (the so-called, *allotropy* or *polymorphism*).

Thermal impact is one of the most important ways to realize phase transitions in crystalline materials, which is widely used in modern materials science. *The thermal impact* on the material is provided to the supply of the heat to the material or its heat removal, which lead, respectively, to increase or decrease the temperature.

The most important source of information about the behavior of pure substances (components), compounds, solid solutions, mechanical mixtures and other compositions under the temperature exposing is the equilibrium phase diagrams. Typically, they provide information about dependence of composition

and stability of phases in the substances on *T*, *p* and *c*. This information makes it possible to study and control many of technological processes, including melting, crystallization, condensation, sublimation, etc. Although these diagrams describe the behavior of substances in the equilibrium state, they also allow one to judge the behavior of substances in a state far from equilibrium, which is essential for the development of technologies for producing of various materials and their further treatments to impart desired properties. This section will outline the basic thermodynamic definitions and relationships that may be useful for further study and analysis of phase transformations during thermal impacts on crystalline materials.

2.1.1. The thermodynamic equilibrium

As applied to crystalline materials, the problem of equilibrium can be considered as one of the important tasks of thermodynamics, which studies the processes running conditions that are accompanying by the energy changes at heat exchange and work doing. The most comprehensive data concerning the material as a thermodynamic system can be obtained when considering the reversible changes in its state, i.e., when at any time, the material state is virtually indistinguishable from the equilibrium. For reversible processes, we can return the system (material) in its original state without any changes in the environment. Under *the material as thermodynamic system*, we shall suppose the sum total of its structural components (phases, grains, defects, etc.) which are interactively with each other and separated from the environment.

The material can be considered as thermodynamic system, if its individual parts can exchange by energy. Since the material is also able to exchange by matter (e.g., by diffusion), the material may be called *physical-chemical system*. Materials are considered as *homogeneous* if there are no within the interphase interfacial areas (mono- and polycrystals of pure elements, solid solutions or chemical compounds) so that the properties of the system are change continuously. *Heterogeneous materials* comprise interfaces between phases, where the abrupt discontinuous changes of properties occur (e.g., water – ice system);

From the thermodynamics viewpoint, all the states of materials are divided into:

- *stationary*, when their parameters are not changed in time (T = const, p = const) and may be characterized by an infinite number of T or p, but their values at each point must remain constant;
- *equilibrium*, when the system is not changing with time, and this is not due to the running of any external (relative to the system) process. (Note that equilibrium state is a special case of steady state, when temperature and

pressure are not only constant, but remain the same in all points of the system);

- *inequilibrium*, when the system state is changed over time and the constancy of the parameters at each point of the system is maintained external (relative to it) processes.

Therefore, if the external factors are changed, the state of the system is also changed. Consider a homogeneous metallic rod consisting of similar atoms of one kind. Let us it is subjected to heating at one end and cooling at the other. For some time, the heating and cooling processes (which are cooperative) will lead to *stationary* (but not equilibrium) *state*, because the constancy of the parameters at each point of the system is maintained external (relative to the system) process. *The system will be in equilibrium state* only if the temperature and pressure for ambient and for each point of the system will be equal. For a rod prepared from alloy, it is necessary to consider the concentration of solutes (in addition temperature and pressure).

As is known, thermodynamics does not study mechanisms which arose transition from one state to another. It studies merely the initial and final states. At the same time, the transition to the final state can run by different ways (by different mechanisms). This is very important because you can artificially "frozen" intermediate state of the system long before reaching its final equilibrium state. The system in this intermediate state may live long time, if the speed of the further running of the process is small. This is *a false equilibrium* in contrast to *the true equilibrium*. System in the true equilibrium is called *stable*, and in the false equilibrium – *metastable*.

This effect of the prolonged living of the system in the metastable state is widely used in the technology of solid materials to give them the desired properties. Thus, at a high temperature we can introduce impurities in the material, for example by diffusion, and obtain their certain distribution profile inside the issue. If the issue is then quenched, the resulting inhomogeneous (and inequilibrium) distribution of atoms is "frozen" and we get a typical example of a metastable state.

If, on the contrary, we want to bring quickly the system to a true equilibrium (in this example – to smooth the impurity profile, getting a uniform distribution of impurities in the material), it is necessary to add extra energy to the system (for example, to heat the issue to a higher temperature to speed up the process of diffusion). The quenching is another example for obtaining of the inequilibrium state of the material (for example, when the austenite phase in the Fe-Fe_3C alloys is quenched, see below).

2.1.2. Changes in energy of materials

When mono- or multi-component material is heated (cooled), this means that its energy changes due to supplying (removal) the heat Q from the outside. Similarly, internal energy of the material is changed on the value of ΔU, if work A is done over it or by it. According to the first law of thermodynamics, this change equals

$$\Delta U = Q - A. \tag{2.1.1}$$

If the heat is absorbed (material is heated), work is done over the material, so that the sign of Q is positive and A – is negative. If the material is cooled (heat is removed), $Q < 0$ and $A > 0$. In other words, the energy change ΔE for heating and cooling processes is not dependent on the way of the material transition from the initial to final state.

In laboratory conditions, the majority of reactions and phase changes are studied at constant external pressure. In this case, the work is performed only due to changing the pressure p, which occurs when changing the volume ΔV:

$$A = p\Delta V \text{ and } \Delta p = 0. \tag{2.1.2}$$

Then

$$Q = \Delta U + p\Delta V. \tag{2.1.3}$$

Since *the heat content* or *enthalpy H* of the system is defined as

$$\mathrm{H} = \mathrm{U} + \mathrm{pV}, \tag{2.1.4}$$

$$\Delta H = \Delta U + p\Delta V, \tag{2.1.5}$$

for $\Delta H > 0$ phase transformations are called *exothermic* because in this case the heat is released. Transformations with $\Delta H < 0$ are called *endothermic* because heat is absorbed. For example, crystal melting is an endothermic process, whereas the melt crystallisation is exothermic process.

2.1.3. Enthalpy-temperature diagram for the materials

Using proper calorimetric system and energy (heat) reservoir, one can easily measure heat which is needed to melt or evaporate the pure substance (material), such, for example, as ice. The result of this experiment is shown schematically in Fig. 2.1.1 as T (temperature) – Q (heat supplied) dependence.

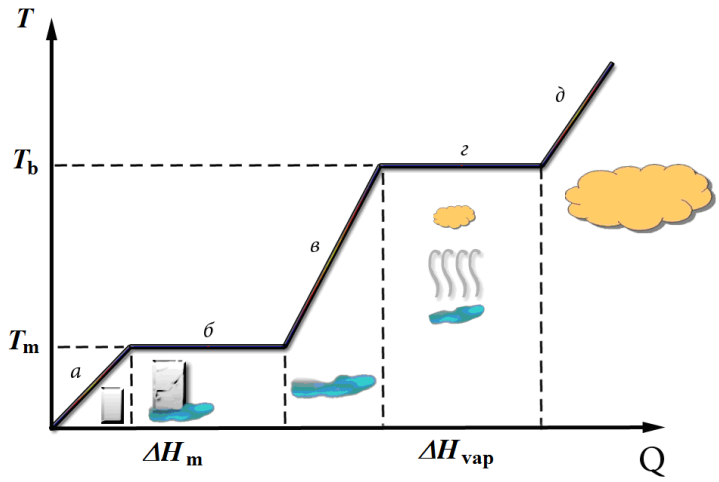


Fig. 2.1.1. Schematic representation of the enthalpy-temperature diagram for ice

As can be seen from the Fig. 2.1.1, while supplying heat to the solid material (ice), its temperature increases along the line and, until it reaches the melting temperature T_m. The heat amount ΔH_m absorbed during the ice-melting process is the length of the horizontal segment *b* of the line. Heating from the T_m to the boiling point corresponds to the straight line *c*. The heat amount ΔH_{vap} absorbed due to vaporization at the water boiling point corresponds to the length of *d* segment of the line. Respectively, vapour heating above the boiling point T_{bp} (vaporization T_{vap}) corresponds to the segment *e* of the line.

Estimating the reverse slopes *dH*/*dT* values of segments *a*, *c* and *e* from Fig. 2.1.1, one can determine heats required to heat one mole of ice, water or steam on 1 K, correspondingly. These values represent the molar heat capacities C_p at constant pressure for ice, water and steam respectively.

From the viewpoint of thermodynamics, Fig. 2.1.1 demonstrates those energy changes that occur in this system (ice) when it is heated (cooled). Actual values of the melting and vaporization enthalpies for ice are 5.98 and 40.5 kJ/mol, respectively. Note that the values of latent heats (enthalpies) of melting or vaporization are the unique characteristics of any pure substances.

2.1.4. Entropy and free energy of the material

When steam (e.g. water) is condensed to form a liquid (water) and the liquid solidifies, forming a crystalline solid (ice), the degree of internal order in the substance increases. In this case, when a perfect crystal is cooled to 0 K, the amplitudes of the atomic vibrations reduce to zero. Since the entropy *S* is

regarded as a measure of disorder in the material, for a perfect crystal the equation $S = 0$ is carried out.

The product of the absolute temperature T and the entropy change ΔS, when occurring the phase transformation, has been called *the entropy factor* $T \cdot \Delta S$. It has the same dimension of the energy density (J/mol), as has the enthalpy change ΔH. At constant pressure p, these two energy changes are related to each other through the relation for the Gibbs free energy:

$$\Delta G = \Delta H - T\Delta S, \tag{2.1.8}$$

where

$$G = H - TS. \tag{2.1.9}$$

According to these relations, when heating, the materials tend to reach the state with a larger disorder so that the entropy factor $T\Delta S$ increases. As a result, the free energy change ΔG due to phase transformations in materials expresses the balance between the two opposing trends – changes in enthalpy ΔH and the change in the entropy factor $T\Delta S$.

If the substance is in an equilibrium state at a constant pressure (e.g., ice and water at a temperature of melting (solidification) $T_m = 0°C$), it can`t achieve a state with the lowest energy at the atmospheric pressure. In equilibrium state of the system melt-crystal, the change of entropy factor $T\Delta S$ and enthalpy ΔH are equal to each other, so that, according to the relation (2.1.1), $\Delta G = 0$. Therefore, the change in entropy at the melting temperature T_m equals

$$\Delta S_{melt} = \frac{\Delta H_m}{T_m}. \tag{2.1.10}$$

Therefore, the melting (or evaporation) can occur only if heat (energy) is applied to the substance from the outside.

Entropy of pure substance at constant pressure increases with the temperature increasing in accordance with the expression $\Delta S = \frac{C_p \Delta T}{T}$, because $\Delta H = C_p \Delta T$.

2.1.5. Free energy – temperature diagrams for the material

The state of any system, including materials, may spontaneously change if it is accompanied by a decrease in free energy. In other words, the negative change of free energy modulo $|-\Delta G|$ is *the driving force of any phase transformation* in the material. If $\Delta G = G$ (after transformation) $- G$ (before transformation) $= 0$, then the system is in equilibrium and the transformation is not running, and if $\Delta G < 0$, the conversion is running.

Graphic representation of the change in free energy with temperature for the water-ice system is schematically represented in Fig. 2.1.2. Reducing the free energy with the temperature increase, shown in this figure, for both phases (water and ice) is the result of the predominance of the entropy contribution $T\Delta S$ at the phase transformation. The growing increase in the negative slope of the curves with the increase of temperature is a consequence of the increasing of entropy for these phases.

2.1.6. Heterogeneous equilibrium. Gibbs phase rule

Multiphase material is in equilibrium state if the values of p and T are the same in all phases. However, under these conditions, a transition of a substance from one phase state to another (due to diffusion or chemical reactions) can occur. An indication that the system is in equilibrium state (i.e. no transition of components from one phase to the other) consists of the equality oof specific thermodynamic or chemical potentials of the given component in the considered phases.

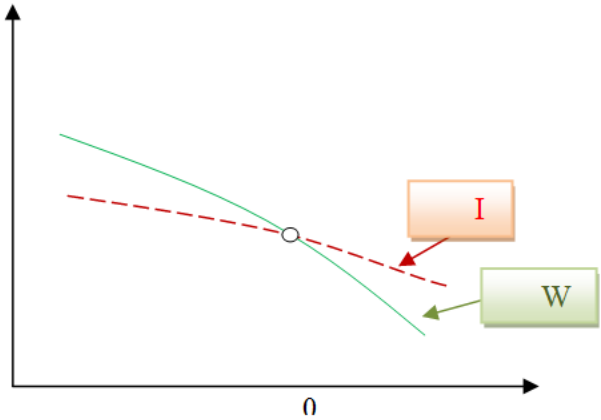


Fig. 2.1.2. Dependence of Gibbs free energy on temperature for water and ice

One of the most important laws of heterogeneous equilibrium in the materials is the *Gibbs phase rule*. It operates with such basic, previously defined, concepts as a *component* and *phase*. In addition to them, each thermodynamic system is also characterized by number of freedom degrees. *The number of freedom degrees* (or *variance of system*) N is determined as the amount of external and internal factors, or coordinates (temperature T, pressure p, a concentration c, etc.) that can be arbitrarily changed in a certain range of values, without changing the number of phases in the system. In other words, the N degrees of freedom of the system are those parameters which play a role

of independent variables. All other parameters will be their functions. Just these independent factors determine the state of equilibrium between the individual phases in the material.

In thermodynamics of materials, *Gibbs phase rule* relates the number of components *C*, the number of phases *F* and the number of freedom degrees (the number of independent variables) *N* in the isolated material by equation

$$N = C - F + 2, \qquad (1.1.11)$$

where the number 2 means two main variables – the pressure *p* and the temperature *T*. The phase rule is a consequence of the fact that the equality of chemical or thermodynamic potentials of the phases is the equilibrium condition. Since the potentials do not depend on the amount (volume occupied) of phases, then the equilibrium in the system does not depend on the amount of a particular phase.

It is found that within a certain limits one can change the independent factors without violating the number and name of coexisting phases. If the number of freedom degrees $N = 0$ (non-variant system), any change in the values of both external and internal factors will cause changes in the number of phases. In the case of mono variant systems ($N = 1$) one can change the value of one of the factors (variables) within a certain range without causing change in the number of phases. Etc.

Thus, the Gibbs phase rule is a mathematical way of expressing the general laws of the stable phases existence in the equilibrium conditions. The phase rule gives the opportunity to predict and check processes occurring in the material during heating and cooling. It shows, for example, if the crystallization process is running at a constant temperature or in temperature range. It also indicates the number of phases which can coexist simultaneously in the material. Therefore, the experimentally built phase diagram should always be checked in their correspondence to the phase rule.

Let us consider the operating of the phase rule for the example of thermal decomposition of crystalline calcium carbonate (chalk) $CaCO_3$. When heated, three phases coexist in $CaCO_3$ – two solid and one gaseous

$$CaCO_3\,(\text{Solid}) = CaO\,(\text{Solid}) + CO_2\,(\text{Gas}). \qquad (2.1.12)$$

As is seen, three different chemical constituents, but only two components are presented in this reaction, because the concentrations of any two constituents fully define the equilibrium of the system. The concentration of any third component can be determined if concentrations of the other two are known. Substituting the appropriate values of $C = 2$ and $F = 3$ in the phase rule (1.1.12), we can see that this system univariant, since $N = C - F + 2 = 2 - 3 + 2 = 1$. Hence, only one variable (either temperature or pressure) can be changed independently. The phase rule can be applied only to

the dynamically reversible processes when the systems are heterogeneous and are in equilibrium state, so that the only temperature *T*, pressure p and concentration c of the components are independent variables. For one-component systems, the maximum number of variables under consideration is equal to two – pressure and temperature. Such a system can be easily represented in the usual Cartesian coordinates.

For two-component (binary) systems the maximum number of variables is equal to three – the *T*, *p* and *c*. In this case, the description of the system requires only a knowledge of the components concentration because the second concentration is equal to (1-*c*). A pictorial representation of the equilibrium diagram of such a system requires a three-dimensional representation. This is, however, difficult to apply, and therefore the double-cross-sections of this diagram in the coordinates of pressure-temperature, pressure-composition, or composition-temperature are most commonly used.

In material science, solid and liquid phases are studied mostly often at *p* = const (0.1 MPa), so that we can neglect the saturated vapor pressure. In this case, the equation (2.1.11) can be accepted with the external factors (variables) equal to 1. Therefore, such systems can be described by using only two variables (temperature and composition), which leads to another phase rule definition:

$$N = C - F + 1. \qquad (2.1.13)$$

In this case, the graphic view of equilibrium diagram of state is given in double temperature-concentration scale. In this case, the invariant equilibrium corresponds to the point on equilibrium diagram (all coordinates are defined as specified constants), univariant – meets the line (one of the coordinates can be chosen arbitrarily), bivariant – pat of the surface, because the two coordinates can be chosen arbitrarily. If the state of the system is determined not only by the temperature and concentration, but still some additional parameter, the degree of freedom will be greater per unit, so that the phase rule can be written as

$$N = C - F + 3. \qquad (1.1.34)$$

The surface area of the grains can be such additional parameter, for example, in nanostructured materials when the value of the surface energy are not able to be neglected. In conclusion, note that the phase rule for single-phase states is not applied.

2.1.7. Phase diagrams for one-component materials

When the phase diagram constructing, one should postpone on the coordinate axes the values of independent parameters (*T*, *p*, *c*, etc.) that characterize conditions of the material existence. To express its composition,

one or two coordinates are used. In such diagrams, the region of the individual phases existence are limited by interphase lines or surfaces.

The example of diagram, which schematically shows the phase equilibrium between solid, liquid and gas phases for such single-component system, as water, is presented in Fig. 2.1.3. The figure shows curves which represent the set of points (coordinates *T*, *p*), in which two phases (in pairs) coexist in equilibrium. Moreover, as seen at the triple point of water (TPW) all three phases – vapor, liquid and solid – coexist. AC line is described by the Clapeyron law.

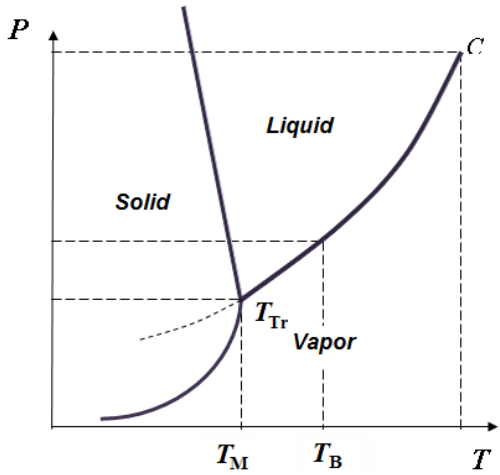


Fig. 1.1.3. Diagram of phase equilibria for water

This diagram shows only the conventional boundaries of phase regions, without specifying the exact values of the parameters that characterize the existing phases. However, it is interesting to analyze the nature of the real interphase lines separating of the coexisting phases. It is easy to see from the analysis of real equilibrium line between liquid and solid phases that, if the ice, for example, is subjected at $T = 2°C$ to high pressure, it will be converted into water. That means that skating sportsman slides not by ice, but by water. This particular sensitivity to pressure in the above example is expressed by negative slope ($dp/dT < 0$) of the equilibrium line liquid-solid. It is an important and unique characteristics of any materials, in which the atomic structure of the liquid phase is characterized by great coordination numbers than in the solid phase. These materials include water, bismuth, germanium and silicon. As to metals, the slope of the solid-liquid phase lines has the opposite sign ($dp/dT > 0$), so that the metal, on the contrary, will be transformed from a liquid to a solid when pressure increasing.

2.1.8. Thermodynamics of phase transitions in crystalline materials

In thermodynamics, all the variety of phase transitions, by Ehrenfest classification, is divided into two main types – the 1st- and 2nd-order transitions. For the 1st-order phase transitions, first-order derivatives of the Gibbs thermodynamic potential $G(p,T)$ are abruptly changed (Fig. 2.1.4a):

$$V = \left(\frac{\partial G}{\partial p}\right)_T, S = \left(\frac{\partial G}{\partial p}\right)_V. \quad (2.1.17)$$

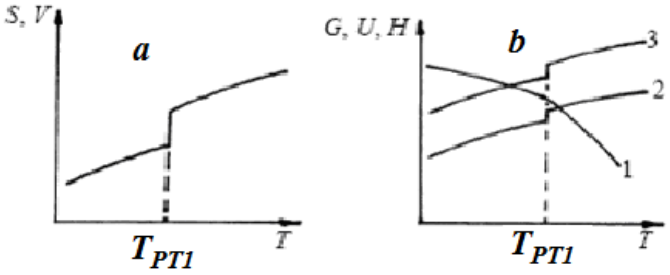


Fig. 2.1.4. Changes in the entropy S and volume V (*a*), the Gibbs thermodynamic potential G (*1*), internal energy U (*2*) and enthalpy H (*3*) (*b*) for the 1st-order phase transition

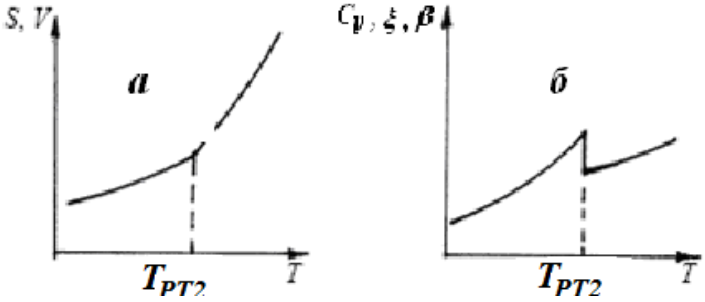


Fig. 2.1.5. Changes in entropy S and volume V (*a*), heat capacity C_p, the coefficient of compressibility ξ and coefficient of thermal expansion β (*b*) for the 2nd-order phase transition

The changes in entropy and volume at the 1st-order phase transition are also accompanied by abrupt change in enthalpy $H = U + pV$, internal energy U (Fig. 1.1.4b) and also many of properties (density, specific heat, viscosity, etc.). Since the potentials G_1 and G_2 of the phases 1 and 2 match ($G_1 = G_2$) at their

equilibrium temperature T_{PT1}, then the change in enthalpy ΔH can be found from the condition

$$H_1 - T_{PT1}S = H_2 - T_{PT1}S ,$$

where H_1 H_2, S_1,S_2 are the enthalpy and entropy of the phases. So

$$\Delta H = T_o(S_2 - S_1) = T\Delta S.$$

Thus, the occurrence of the 1st-order phase transitions is accompanied by the absorption or release of latent heat (heat of phase transition), coinciding with the change of enthalpy ΔH. For the 2nd-order phase transitions the volume and entropy of the system are changed continuously (Fig. 2.1.5a), so that the latent heats are not released and absorbed. The 2nd-order phase transitions are accompanied by an abrupt changes in the values of the second-order derivatives of G by temperature and pressure (Fig 2.1.5b). Heat capacity

$$C_p = T\left(\frac{\partial S}{\partial T}\right)_p = -T\left(\frac{\partial^2 G}{\partial T^2}\right),$$

coefficient of compressibility

$$\chi = -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_T = -\frac{\partial^2 G}{\partial p^2}$$

and thermal expansion coefficient

$$\beta = -\frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_p = -\frac{1}{V}\frac{\partial^2 G}{\partial T^2}$$

are related to such quantities.

2.2. Melting of single-component crystalline materials

Under the melting, we usually understand the transition of the material from the ordered solid (crystalline) state to the non-ordered (liquid) state when heat supplying.

2.2.1. Melting of crystals as a phase transition

Melting of single-component crystalline solids, when their heating, is related to the 1st-order phase transitions. Therefore, for crystalline single-component substances the melting temperature (T_m) and the heat, which is necessary to supply the melting process (*latent heat of melting* or *enthalpy of melting* Q_m), are the main characteristics of the melting process as the 1st-order phase transition. This type of phase transition is studied using the so-called diagrams of heating (melting).

2.2.2. Heating diagram for the melting characterization

Let us consider, how the temperature of single-component material (e.g., metal) is changed during its melting. Experiments show that schematically a plot of dependence of temperature on time when heating the metallic crystal looks like curve in Fig. 1.2.1. Just this dependence is called the diagram of heating or melting.

Consider the reasons for such behavior of heating diagram. Suppose that the heat is supplied to the crystal with a sufficiently slow rate. In this case, the temperature throughout the sample bulk have time to be equalized and it would be considered substantially the same for all points of the sample (i.e. thermodynamic system is in equilibrium). Because of constant heat flow, the time axis t (abscissa) presents also the axis of the heat amount Q, which is absorbed by a substance during heating (Q = const·t). On the segment 1'-1 of the curve in Fig. 2.1.1, heat supply leads to a monotonic growth of the crystal temperature T. The law of the temperature increase in time in this segment depends on the temperature dependence of heat capacity of the crystal and therefore is not necessarily linear. At the stage 1-2 (in the time interval between t_1 and t_2), the temperature is constant, despite the fact that the heat supply is continuously carrying on. Then, when $t > t_2$, the temperature of the sample begins to rise again (segment 2-2' on the heating diagram).

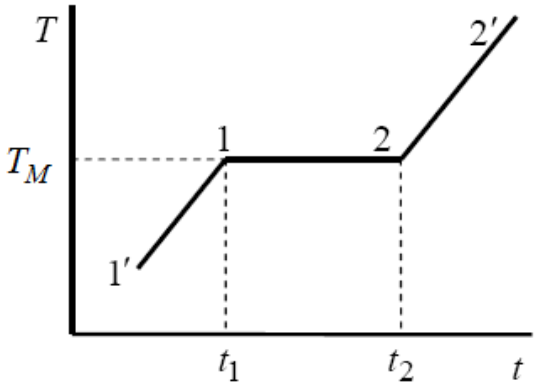


Fig. 2.2.1. Schematic heating diagram during the melting of the crystalline material. On the horizontal axis (abscissa), time t, which is proportional to the heat Q evenly supplied to the body, is postponed

Thus, as follows from the heating diagram in Fig. 2.2.1, melting begins when the temperature of the crystalline material approaches the value, which corresponds to the segment 1-2. As noted above, this temperature is called *the melting temperature* T_m. In doing so, since the beginning the melting and to its

completion (at the segment 1-2 with time duration $\Delta t = t_2 - t_1$) temperature of the single-component substance remains constant and equal to T_m, despite the heat supplying. As mentioned above, the heat absorbed by the crystal at the segment 1-2 is *a latent heat of melting* Q_m. From point 2 in heating diagram, the transition to the liquid state from the crystalline state begins. This is accompanied by the complete destruction of the crystal lattice of the material, so that the internal energy and entropy of the sample increases.

Since the solid and liquid states present two different phases of the material, the melting process is the phase transition. In doing so, the liquid phase (melt) does not occur immediately throughout the crystal. At the beginning, "nuclei" of new phase are formed inside the solid bulk, which then are growing, spreading to the entire bulk of the crystal. In this case, the two phases (liquid and solid) can co-exist for a long time (as, for example, the ice can float in the water without melting). These features of the melting process (energy absorption and the possibility of solid and liquid phases to coexist) are unique features of the 1st-order phase transition.

2.2.3. Dependence of melting point on pressure

Experiments show that the melting temperature is dependent on the environmental pressure *p*. On the phase diagram of the single-component substance, this dependence is represented by melting curve (curve of solid and liquid phases coexistence, AD or AD' in Fig. 2.2.2).

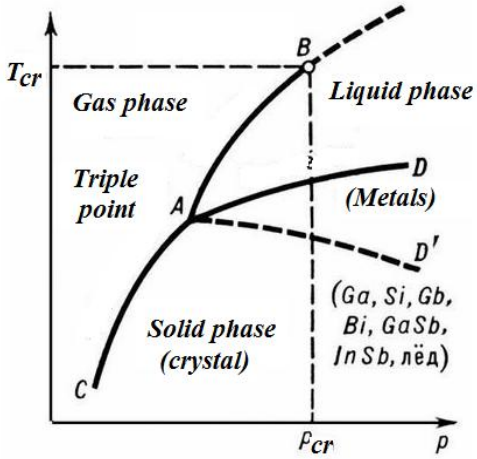


Fig. 2.2.2. The phase diagram of a single-component (pure) substance. Lines AD and AD' are melting curves: material with abnormal volume change on melting is melted by the line AD'. B – critical point, A – triple point, AC – sublimation line, AD – melting line, AB – boiling line

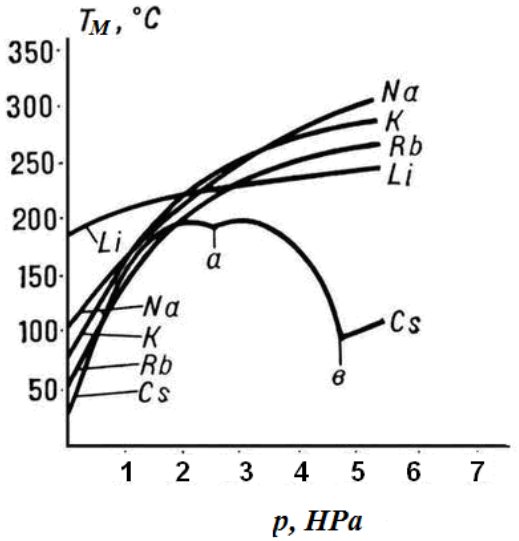


Fig. 2.2.3. Changing the melting temperature T_m (°C) of alkali metals with the pressure p (GPa) increase. The melting curve of cesium (C_s) indicates its existence of two polymorphic transformations (*a* and *b*) at high pressures

The dependence of $T_m(p)$ is determined by the direction of the change of the material volume (ΔV_m) during the melting process (see, Clapeyron-Clausius law). In most cases, the melting of crystals is accompanied by an increase in volume (typically by several per cents). If this occurs, the increase in pressure results in the T_m increases along the line AD in Fig. 2.2.3.

However, certain substances such as water and a number of metals (e.g., Bi, Ga, Sb) shows volume reduction at the melting. Therefore, the melting points of these substances is reduced along the curve *AD'*, when pressure increasing.

For pressures, less than 10^5 Pa, the $T_m(p)$ dependence is approximately expressed by the empirical Simon formula:

$$\frac{p - p_o}{a} = \left(\frac{T_m}{T_o}\right)^c$$

where *a* and *c* are constants for the given substance, and p_0 and T_0 – coordinates of the triple point. When $p > 10^5$ Pa $T_m(p)$ plot can have maxima (Fig. 2.2.4).

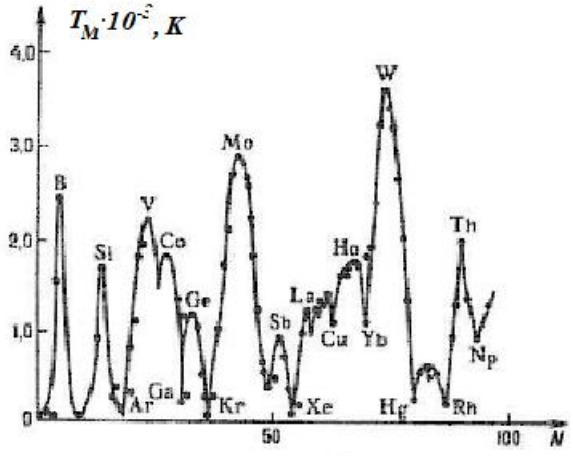


Fig. 2.2.4. The melting points of simple substances at p = 101325 Pa (N is serial number of the chemical element in the periodic table)

The difference between the melting points of polymorphic modifications of the same substance may reach very high values. For example, it can reach 270 K for two cubic modifications of a chemical compound KCl, while two crystals with different isotopic composition (e.g., hydrogen and deuterium) have this difference of about 5 K.

2.2.4. Interconnection between melting points and physical properties of crystals

Melting process is accompanied by the changes in physical properties of crystalline substances. For example, during the melting the increase of entropy occurs that reflects disordering of the crystalline structure of the material. The entropy ΔS increase at the melting is about 3-60 J/(mol·K), while for the most simple substances ΔS does not exceed 6-11 J/(mol·K). Interconnection between changes in enthalpy, entropy and volume of substances and T_m during melting is described by the known Clausius-Clapeyron law.

Melting is accompanied, as a rule, by the increase of heat capacity, atomic volume and also electrical resistance. Substances with the abnormal melting curve AD′ in Fig. 2.2.5 are exceptions. These include some semimetals (Bi, Sb) and semiconductors (Ge), which in the liquid state have a higher conductivity. Resistance to mechanical shear drops to almost zero at melting (because transverse elastic waves will not be able to propagate in the melt), the rate of sound (longitudinal elastic waves) propagation decreases, etc.

Melting process depends on the surface energy of the crystals, that determines the dependence of T_m on grain sizes. The large-grained single-component polycrystalline substances are often melted at a higher temperature than the fine-grained ones. If the size of crystals (grains) $d \geq 10$ nm their melting points can be estimated by the formula

$$T_d = T_M(1 - \frac{4\sigma V}{\Delta H_M d}),$$

where σ is the specific surface energy of the crystallites, and ΔH_m – molar enthalpy of melting. As a result, transition of fine-crystalline unequigranular metal into liquid state occurs not at a constant temperature but in the temperature range (about a few degrees). This interval begins at the melting temperature of the smallest grains, and is completed at the melting temperature of the largest crystallites.

The melting process is closely related to the thermal properties of the crystal such as heat capacity and thermal conductivity. This allows one to describe the melting point of substances with homotypic crystal structure by the well-known empirical Lindemann formula

$$T_M = AM\theta^2 V^{2/3},$$

where A is a constant for a group of substances having similar crystal structure, M and V – the molecular weight and volume, θ_D – Debye temperature of the substance. Also a correlation between the linear coefficient of thermal expansion of metals and their melting temperature was noted (see, Fig. 2.2.5).

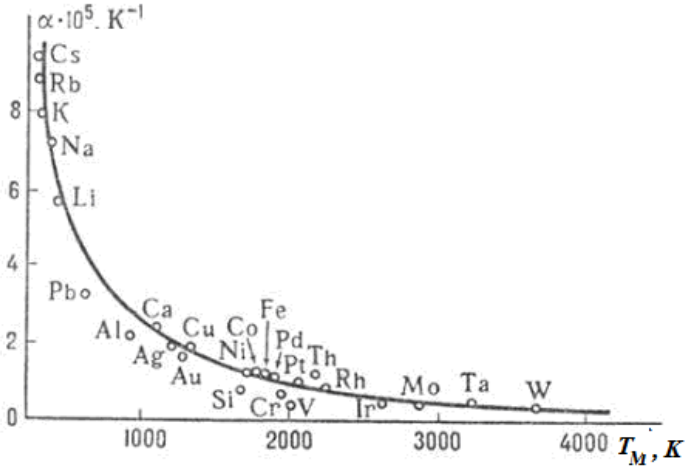


Fig. 2.2.5. The relationship between the melting points and the linear expansion coefficients of metals at room temperature

2.2.5. Dynamics and mechanisms of the melting

Let us ask the question, why matter should melt at all. To do this, first we explain why at a certain temperature (T_m) a part of interatomic bonds in a crystalline solid is distorted or even broken and the crystal is transformed from the ordered (having a crystal lattice) to the disordered state. As follows from the thermodynamics of phase transformations, at the fixed temperature any body tries to minimize its free energy. At low temperatures ($T \to 0$) entropy factor TS is changed insignificantly, and as a result all is reduces to minimizing of the internal energy U. The state of substance with a minimal energy just corresponds to a crystalline solid. When temperature increasing, the contribution of TS to the free energy is becoming increasingly important (because of the growth both T and S), so that at a certain temperature to be preferable to break some of the interatomic bonds. This results in converting the crystal to a disordered state instead of the preserving the crystalline (ordered) structure. In such a case, the internal energy U will rise slightly, but the entropy is greatly increased (due to temperature growth) so that free energy will decrease.

The dynamics of melting. According to the molecular-kinetic concepts, the dynamics of the melting process is as follows. The heat supplying to a crystalline material increases energy (amplitude) of the oscillations of its constituent atoms (ions, molecules). This corresponds to the temperature increase leading to disordering of the crystal. Disordering is accompanied by a distortion of lengths and directions of interatomic bonds (see Fig. 2.2.6), and the formation of various defects in the crystal (see Chapter 1 of this Book).

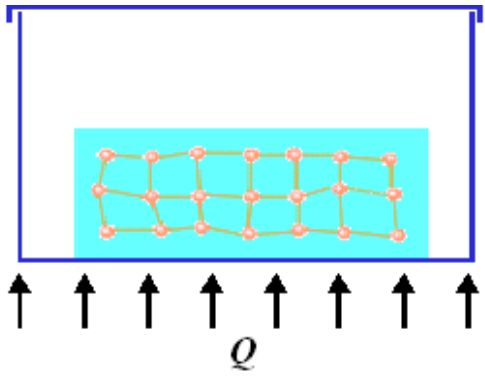


Fig. 2.2.6. Distortion of the crystal lattice (the angles and lengths of chemical bonds) during the melting process in the vessel with the crystal due to the heat Q supplying

The stage of generation and accumulation of defects during the melting process is often called *the pre-melting stage*. On the real heating diagram in Fig. 2.2.7, the presence of this stage of melting results in a smoother bending of the melting curve at heating before it reaches a plateau corresponding to the T_m. According to Fig. 2.2.7, when $t < t_1$, crystalline metal is heated. The time interval $t_1 < t < t_2$ corresponds pre-melting stage, when quasi-liquid layer begins to form at the surface or intergranular boundaries (grain boundaries) of the material. The temperature region $t > t_3$ corresponds to the start of heating of the material melt.

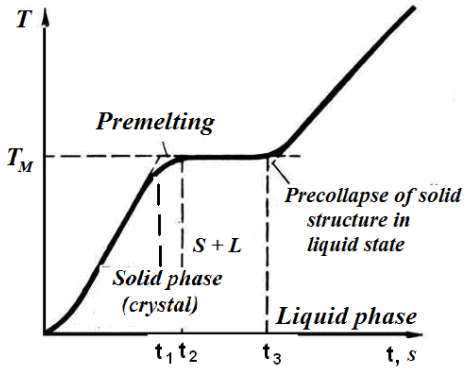


Fig. 2.2.7. The real heating diagram at the melting of single-component crystalline material

The mechanism of melting. It is believed that, when approaching the T_m, concentration of defects in the crystal reaches a certain critical value at which the actual melting begins. All supplied heat at melting (corresponding to the plateau in Fig. 2.2.1 and 2.2.7) goes only on the breaking of interatomic bonds and the destruction of long-range order in the crystal (see Chapter 1), but not on heating of the body. That is why pure single crystals and large-grained polycrystalline materials does not change their temperature during the melting process. As a result, during the melting, liquid regions of submicroscopic dimensions are formed (nucleated) inside the crystal with the ordered atomic lattice. As the crystal is heated, these regions of liquid phase become gradually larger and merge into a one unit, forming a liquid phase, while the crystalline phase, on the contrary, is broken into fragments and reduced in size until they full disappearing.

Note that at a relatively low superheating of the melt, coordination number of atoms in the most cases is the same as in crystalline fragments. This means that the short-range order in the atomic arrangement in these submicroscopic liquid regions is not significantly changed. This, in particular, explains the

lower values Q_m of the melting heats compared with the vaporization heats, and also with the relatively small changes in a number of physical properties of materials during the melting process. The specific melting enthalpy of most elemental substances is in the range of 10-300 kJ/kg, and is significantly lower than the enthalpy of vaporization.

2.3. Crystallization of single-component crystalline materials

By *crystallization* of the material we generally understand its transition to the ordered crystalline state from disordered gas, vapor, liquid, solid amorphous or nanostructured state. Formation of the crystalline material from the melt (just this type of crystallization is very extended in materials production) can be considered as the reverse process to the crystal melting described in the previous section.

2.3.1. Crystallization from the melt as a phase transition

Consider the pure metal at the point of solidification (crystallization), when both liquid and solid phases are at the same temperature. In this case the kinetic energies of the oscillating atoms in these two phases must be the same, while their potential energies are significantly different.

The kinetic energy is determined by only the temperature of the liquid-solid system: the higher the temperature, the greater the kinetic energy of atoms. The potential energy depends on the interatomic distances in both phases. Therefore, since average interatomic distances in the liquid phase (melt) are greater than in solid state, the potential energy of the atoms is higher.

Since during transition of the material from liquid to solid state interatomic distances are reduced, the process of solidification (crystallization), is followed by the release of energy. This released heat is often called *the latent heat* Q_{cryst} or *enthalpy* ΔH_{cryst} of crystallization.

In pure (single-component) homogeneous materials equilibrium temperatures and the heats of crystallization and melting must coincide. As noted in the previous chapter, the release and absorption of latent heat is an indispensable feature of the so-called 1^{st}-order phase transitions. Therefore, when the molten liquid is transformed into the crystal when cooling (crystal lattice is formed), the atoms, ions or molecules tend to be arraged so that their potential interaction energy was minimal. The released excess potential energy is converted into kinetic energy of thermal atomic motion in the crystal. This process of the atomic system ordering would cause a rise in temperature. However, simultaneous removal of thermal energy in the environment occurs,

which is compensated by the energy released during crystallization, so that the temperature during solidification is constant.

Note that many properties of the crystals (density, specific heat, viscosity, etc.) are changed abruptly at the crystallization point in the process of crystallization, similarly to as the melting process. The values of entropy changes during crystallization are 5-12 J/(mol·K) for simple substances, 20-30 J/(mol·K) in inorganic compounds, and for organic compounds they are 40-60 J/(mol·K).

In real conditions, melting process typically occurs at a temperature above the equilibrium, i.e. at overheating, whereas crystallization, on the contrary, begins only at overcooling. The difference between the real melting and crystallization points is called *temperature hysteresis*.

2.3.2. Cooling diagram at crystallization from melt

Consider schematically changes in temperature of the molten material during cooling and subsequent crystallisation (Fig. 2.3.1). In this process, the heat must be removed from the melt so that the temperature had time to equalize over the entire melt volume and can be considered the same for all points of the liquid phase.

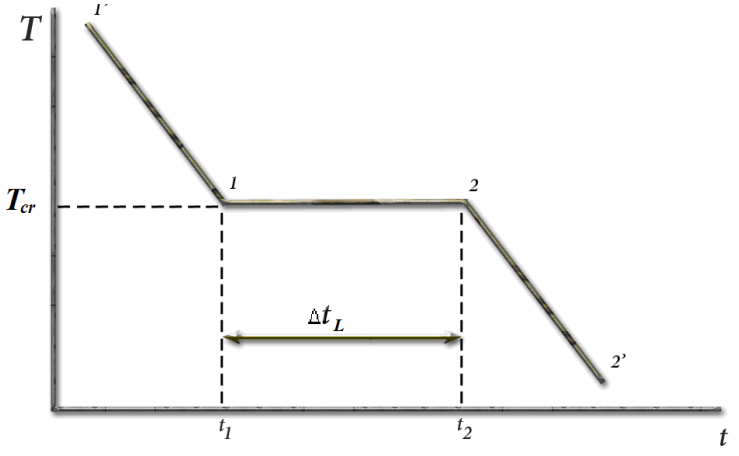


Fig. 2.3.1. Cooling diagram of the melt during solidification

The curve in Fig. 2.3.1, which shows the temperature variation with the melt cooling, is called *the cooling diagram*. Because of the heat sink (heat flow is negative), at the segment 1'-1 temperature of the liquid phase decreases monotonically with the law, which is determined by the temperature dependence of the specific heat of the melt. At slow cooling, see section 1-2 in

Fig. 2.3.1, the temperature of the melt remains constant. Just this temperature is called *the crystallization* T_{crest} or *solidification temperature*. The plateau with the duration $\Delta t = t_2 - t_1$ in cooling diagram occurs because, as noted above, the heat released from the crystal is compensated by the heat of crystallization. After completion of the crystallization process, temperature of the formed crystalline body in the segment 2-2' begins to fall down by law, which is determined by the temperature dependence of the heat capacity of the crystal.

2.3.3. Conditions for the melt crystallization

Consider the conditions which are necessary for solidification of the crystal, in general. If the crystal is not fused, not dissolved and not evaporated, this means that it is in equilibrium with the melt or other medium (solution, steam), in which it is growing. Equilibrium of the crystal with the melt of the same material is possible only at the melting temperature T_m, and the equilibrium, for example, with a solution or vapor occurs only when saturating of the latters. This means that, to create conditions for growth of the crystal, we should transfer the growth medium into inequilibrium state by its supercooling (melt) or supersaturating (solution, vapor). In this case, the crystal growth rate will be higher, the more the deviation from thermodynamic equilibrium.

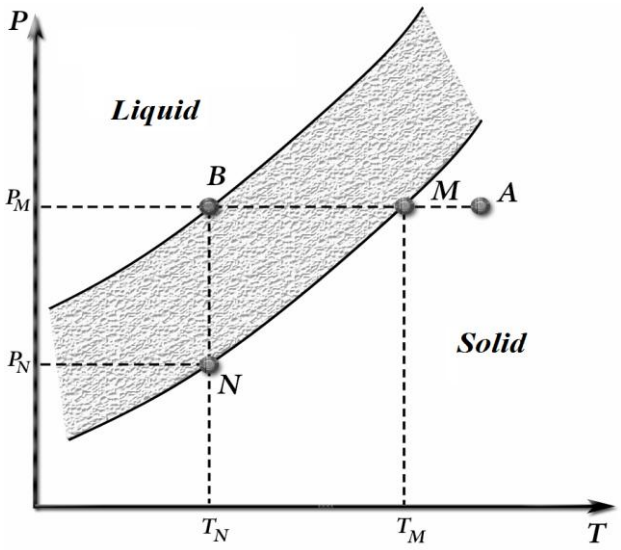


Fig. 2.3.2. To the explanation of the solid phase nuclei formation during cooling of the melt

Thus, crystallization from the melt can be considered as a phase transition of a substance from the super-cooled state of the melt in the crystalline (solid) state, which has a lower free energy. The crystallization process is completed with the exhaustion of the melt or the achievement of equilibrium values of temperature and pressure.

Consider the process of crystallization from the melt studying the *p*-*T* diagram at constant volume in Fig. 2.3.2. This diagram presents connection between the pressure *p* and temperature *T* for the case of the existence of disordered (liquid) and ordered (solid) phases. The line *MN* is the line of phase equilibrium that delimits the regions of the melt and the crystal existence. For all points on this line, both phases exist in equilibrium.

Suppose that the temperature of the melt at constant pressure falls so that the melt passes from the point *A* to the point *M* where two phases should be in equilibrium. Formally, if the peace of grown crystal (it is called a *seed*) was embedded into the melt, it would begin to grow when the system reaches the point *M*. However, the seed is absent, spontaneous occurrence (nucleation) of the crystal at the point *M* does not happen (the reasons for this will be discussed below). Experiments show that the nucleation and growth of the crystal in this case can only be started at overcooling of the melt, i.e. when the system melt-crystal approaches the point *B*, which corresponds to the temperature T_B and pressure P_M. The overcooled degree at the point *B* is measured by the temperature difference $T_M - T_N = \Delta T$.

2.3.4. Homogeneous and heterogeneous crystallization from the melt

To understand the mechanisms of the crystallization process in pure single-component materials, we should primarily consider the structure of the melt, in which a crystalline phase can be formed under certain conditions. Consider the mechanism of homogeneous crystallization on the example of the metallic crystals. According to modern concepts, the structure and properties of the metal in the liquid state is significantly closer to solids than to gases. In particular, the heat capacity of the molten metal is only on 10% higher than for solid metal, whereas the difference in value of the same property between liquid and gaseous states is more than 25%.

Structural studies have shown that, if liquid metal is cooled to a temperature close to the of crystallization point, small groups (aggregates) of atoms are formed in the melt, which are arranged with virtually the same coordination as in the metallic crystal (see Fig. 2.3.3a). The number of atoms in such units (aggregates) is very small, and their state is dynamically unstable. This means that in some places of the melt such aggregates arise continuously, then

disappear, then re-appear again at the same or in other regions of the melt and disappear again. These groups of atoms are characterized by the so-called *short-range order* (in contrast to the crystals possessing long-range order due to a lattice structure, see Part 1 of this Book.) They can be stable only under appropriate conditions determined by the thermodynamic laws. Consider these conditions.

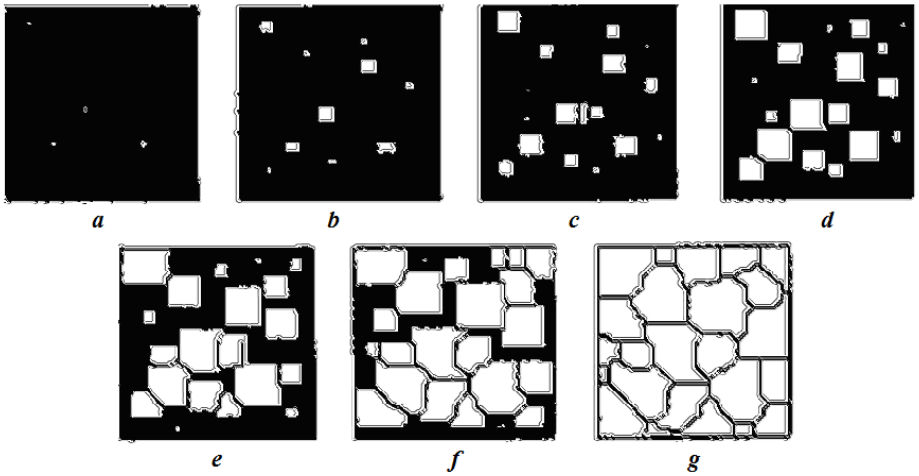


Fig. 2.3.3. Schematic model of the metallic polycrystals solidification (bright areas correspond to the crystalline phase, and the dark – the melt)

If the melt is overcooled below some critical temperature T_c, it will not be able to conserve metastable state, without crystallizing, at temperatures which are arbitrarily low in relation to the temperature of crystallization T_{cryst}. For example, a small liquid drop of pure metal with a diameter of about 100 microns can be overcooled to a temperature of the order of 0.8 T_m. Experiments show that when a certain limit (critical value) of overcooled is approached, a lot of very small crystals are aroused in the melt almost instantly in shape the above-mentioned groups of atoms with short-range order. They are called *nuclei* or *centers of crystallization* (Fig. 2.3.3b). Centers of crystallization are formed independently in random parts of the melt. With increasing the supercooling degree, number of such nuclei, produced per unit time, increases (Fig. 2.3.3c,d). Just around these centers, crystallites start to grow during crystallization process (Fig. 2.3.3d). First, the crystallites have the regular shape, but after touching of each other and their following coalescence their regular form is violated (Fig. 2.3.3e).

The increase in the total mass of solidified metal occurs both due to the appearance of new crystallization centers, and the growth of existing crystallites. However, the growth continues only in those areas where there is free access to the melt. The described scheme of sequential stages of solidification process is completed the formation of a fully crystallized metal, which is called *polycrystalline material*, consisting of *grains* or *crystallites* (Fig. 2.3.3f,g).

The described model of melt crystallization is called *homogeneous* or *spontaneous crystallization*. As follows from this model, homogeneous crystallization process actually consists of two basic processes-nucleation the crystallization centers and their growth as crystallites (grains). Critical overcooling, needed for arising of stable crystallization centers, depends on the type of material, its melting temperature, the volume of the melt, the presence of extraneous particles (e.g., foreign particles and crystals of other substances, which serve as seeds to form crystallization nuclei, and the like), the state of the vessel (crucible) walls surface with the melt, the intensity of the heat release from the melt, melt mixing, and ultrasonic impact, as well as many other factors.

The main reason for the melt overcooling consists in need to spend energy to supply nucleation of crystallization center. This work is called *the nucleus activation energy*, which depends on the number n of particles (atoms, molecules, ions) included in the center.

The activation energy of the nucleus is calculated as the difference of two energies E_1 and E_2. Energy E_1 is necessary for the formation of interface separating the two phases – a metastable (liquid), and a stable (crystalline). Energy E_2 is the decrease in free energy of the melt-crystal system, which comes to crystalline state possessing a lower internal energy. The first of these energy increases proportionally to the nuclei surface, and therefore to square of the particles number in nucleus ($E_1 \sim n^2$). The second component of the energy is reduced proportionally to the volume of the nucleus, i.e. the cube of the particles number ($E_2 \sim n^3$). As a result, for some value of n, which means achieving the critical size of the center, both energies become equal ($E_1 = E_2$). Thereafter crystallization is running spontaneously, as it is accompanied by only a decrease in free energy of the system (due to release of the latent crystallization heat).

In contrast to the above-described process of homogeneous nucleation of crystallites, when crystallization centers are nucleated spontaneously within the melt, heterogeneous nucleation of crystal growth begins on the particles of the other phases-dust particles, particles of oxides, on the walls of the crucibles and on the seeds specially introduced into the melt. In the most of the below described industrial crystallization techniques heterogeneous nucleation is more

often used. For this goal the pre-grown crystalline seeds are produced either of the desired substance or of isomorphic to it substances.

2.3.5. Thermodynamics of homogeneous crystallization from the melt

As is clear from the preceding chapter, spontaneous (or homogeneous) crystallization is due to tending of the material to come in more (equilibrium) stable state, which is characterized by reduction of Gibbs potential *G* (free energy) (see (2.1.4) in the S. 2.1).

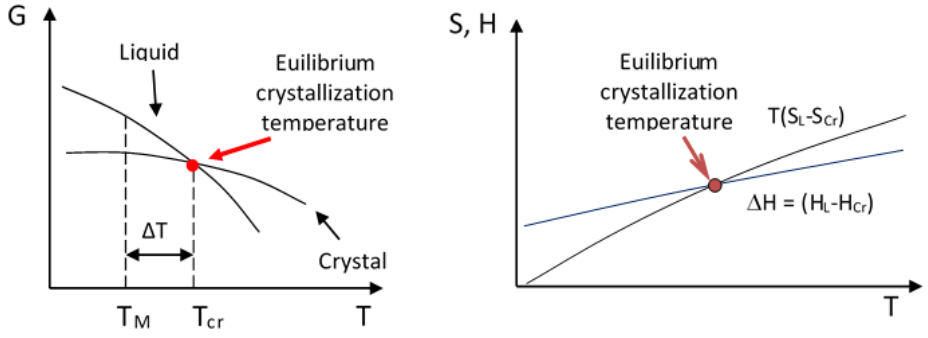


Fig. 2.3.6. Schematic representation of the temperature dependence of thermodynamic potential *G*, enthalpy *H* and the entropy factor *TS* for crystalline phase and the melt

As can be seen from Fig. 2.3.6a, thermodynamic potential *G* in a liquid phase (melt) and a crystal decreases with the temperature growth, although the reduction in *G* (*T*) at heating of the melt is faster than for the crystal. Therefore, the intersection of curves $G_l(T)$ and $G_s(T)$ determines the temperature of phase transition from liquid to solid state during cooling of the system. The temperature of this intersection, at which the thermodynamic potentials of the solid and liquid states are equal (see (2.3.7) below), is called *the equilibrium temperature of crystallization* T_{cryst} (marked by arrows in Fig. 2.3.6).

On the other hand, the temperature dependence of the entropy contribution $T(S_l - S_s)$ the difference in free energies (thermodynamic potential) between the liquid and solid states behaves differently. As can be seen from Fig. 2.3.6b, the entropy factor increases with increasing temperature, and faster than the enthalpy contribution $\Delta H = H_l - H_s$. This means that the melting is controlled to a greater extent by increasing the entropy in the liquid state as compared with that in the solid state. Accordingly, the process of crystallization is determined by a rapid decrease in entropy during cooling in the solid than in liquid state.

Thus, the crystallization begins in the case if the thermodynamic potential of the substance in the solid state is less than its thermodynamic potential in a liquid state, that is usually achieved by overcooling of the liquid metal to a temperature T_{oc} below.

Since the liquid metal, characterized by only a short-range order in arrangement of atoms, has more internal energy than the solid state with the long-range ordered atomic structure, the heat release should be observed by crystallization. The latent heat (ΔH_{cryst}) and the equilibrium temperature (T_{cryst}) of crystallization are interconnected. The nature of this connection can be understood from the following considerations. Since thermodynamic potentials in liquid and solid states are equal at the equilibrium crystallization temperature, it follows from (2.1.9) in section 2.1 for the thermodynamic potential

$$G = H - TS \tag{2.3.6}$$

that

$$H_l - T_{cryst}S_s = H_s - T_{cryst}S_s. \tag{2.3.7}$$

Then a crystallization heat, which is determined as difference of enthalpies in liquid and solid states, is equal

$$Q_{cryst} = \Delta H_{cryst} = H_l - H_s = T_{cryst}(S_l - S_s) = T_{cryst}\Delta S. \tag{2.3.8}$$

This relation, re-written as

$$\Delta S = \Delta H_{Cr}/T_{Cr}, \tag{2.3.9}$$

characterizes the order degree in the arrangement of atoms at crystallization, and establishes the desired relation between ΔH_{cryst} and T_{cryst}. The crystallization het for various metals, depending on the magnitudes of interatomic bond energies, varies from 2500 J/mole (Na, K, etc.) to 20000 J/mol (W, etc.). The above thermodynamic model of the crystallization process allows to understand both the reasons for the differences between ideal (Fig. 2.3.1) and real (Fig. 2.3.7) cooling curve and the reasons for the discrepancy in the melting and crystallization temperatures observed in the experiments.

In theory, for small overcooling $\Delta T = T_{cryst} - T_{oc}$ crystallization process is depicted by the cooling curve of type 1 in Fig. 2.3.7, which is close to the crystallization curve in Fig. 2.3.1. Cooling curves of the form 2 in Fig. 2.3.7 demonstrate more real progress in crystallization processes when cooling. As seen from the curve 2, really to begin the crystallization we need to overcool the melt to the temperature T_{oc} lying substantially below the equilibrium crystallization temperature T_{cryst}. Only in this case, thermodynamic (energetic) conditions, needed for the occurrence the crystallization process, are created.

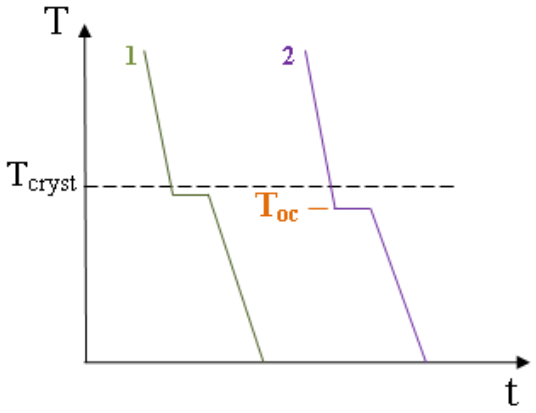


Fig. 2.3.7. Types of cooling curves for crystallization with small (*1*) and large (*2*) overcoolings

Experiments have shown that in some cases, after a large supercooling and the beginning of crystallization, temperature of the crystal melt is increased stepwise, making closer to its equilibrium value T_{cryst} (curve b, Fig. 2.3.8). Consider the reasons for such behavior.

During the crystallization of a relatively small volume of pure element, the heat elimination from the crystal-melt system is compensated by the released crystallization heat. This heat is not enough to superheat the melt, whereby crystallization will occur at lower than the equilibrium temperature T_{cryst} (Fig. 2.3.8a). However, if the volume of liquid metal is large, the released crystallization heat rises the temperature of the melt to nearly equilibrium crystallization temperature T_{cryst} (curve b in Fig. 2.3.8).

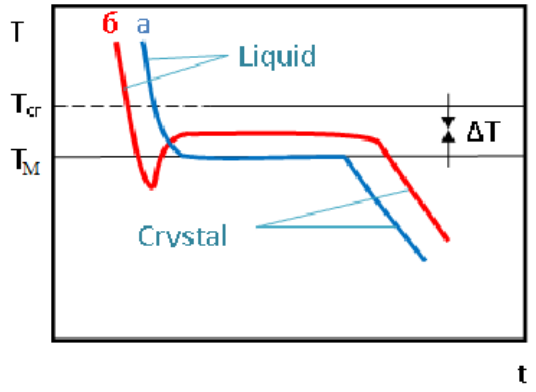


Fig. 2.3.8. Cooling curves of the melted metal with large (*a*) and small (*b*) masses of the melt in crucible

The difference between the equilibrium (T_{cryst}) and real (T_{oc}) temperature of crystallization is called *the degree of overcooling* ($\Delta T = T_{cryst} - T_{oc}$). The degree of overcooling depends on the nature of metal. It increases with increasing of the metal purity, and also with increasing of cooling rate.

Influence of cooling speed on the character of cooling curves is shown in Fig. 2.3.9. At slow cooling, appropriating the speed of cooling V_1, the degree of supercooling ΔT_1 is small and crystalzation occurs at a temperature close to the equilibrium T_{cryst}. With increasing of cooling rate (curves V_2 and V_3) the degree of supercooling increases, and the process of crystallization occurs at increasingly lowered temperatures.

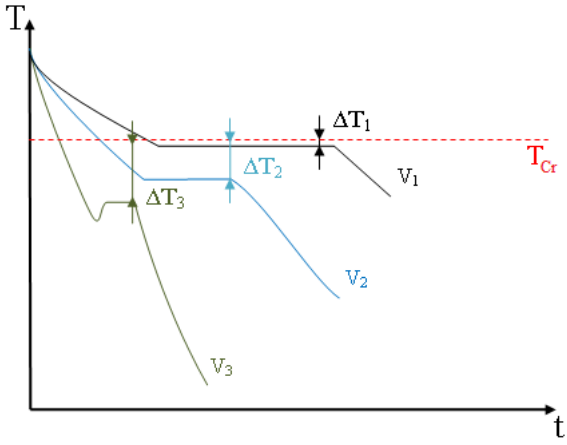


Fig. 2.3.9. Cooling curves of the molten metal at different cooling speeds $V_1 < V_2 < V_3$

Usually the degree of supercooling in the crystallization of metals in the industrial conditions is ranged from 10 to 30°C, while the degree of overheating in the melting of metals usually does not exceed some °C. At high speeds of cooling the degree of supercooling can reach hundreds °C.

2.3.6. Thermodynamic conditions for homogeneous nucleation of crystallization centers in the melt

As was noted above, the process of homogeneous crystallization begins from the stage of crystallites nucleation, which changes thermodynamic potential G of the melt-crystal system. As follows from Fig. 2.3.6a, the $G(T)$ dependence is determined by two contributions G_V and G_s. On one hand, crystallization reduces $G(T)$ on the value $G_V = V\Delta G_v$ due to a decrease in free energy of the system with volume V. On the other hand, due to the appearance on the

interface between the melt and the crystal nucleus G is increased by an amount $G_s = s\sigma$. Therefore, total $G(T)$ can be reperesent as

$$G = G_V + G_s = V\Delta G_v + s\sigma, \qquad (2.3.10)$$

where V is volume of the crystallization center, s – its surface, σ – specific surface tension at the liquid-crystal interface. It follows from (1.3.6) and (1.3.9) that the difference of the specific thermodynamic potentials at the transition of liquid to the crystalline state is equal

$$\Delta G_v = \Delta H_{Cr}\Delta T/T_{Cr} \qquad (2.3.11)$$

If we assume that the nucleus has the spherical form with radius r, then the total change in the thermodynamic potential, taking into account (2.3.11), is equal to

$$G = (4/3)\pi r^3\Delta G_v + G_s = (4/3)\pi r^3\Delta G_v + 2\pi r^2\sigma. \qquad (2.3.12)$$

As follows from equation (2.3.12), the $G(T)$ curve has a maximum (see Fig. 2.3.10) at some *critical value of* r_c. Nuclei with sizes above the critical cause G to decrease, and therefore they become stable, i.e. are capable to grow. Nuclei having sizes less than the critical r_c are unstable and dissolved in the melt because they cause an increase G.

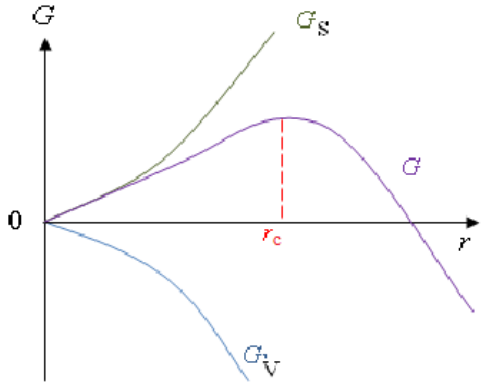


Fig. 2.3.10. Dependences of the total thermodynamic potential and contributions to it on nuclei sizes by nucleation of crystalline phase

To determine the critical size of crystalline nucleus r_c, we make differentiation Eq. (2.3.12) by r and equate it to zero:

$$-\frac{4\pi r_c\Delta H_{Cr}\Delta T}{T_M} + 8\pi r_c\sigma = 0.$$

As a result, one can get critical radius of nuclei

$$r_c = \frac{2\sigma T_{Cr}}{\Delta H_{Cr}\Delta T} \qquad (2.3.13)$$

When $T \to T_{cryst}$,critical radius of the nucleus $r_c \to \infty$. This just indicates the need for supercooling of the melt to begin crystallization process.

As the surface tension is changed weakly with the supercooling degree ΔT increase, Eq. (2.3.11) for the ΔG_v increases linearly. Consequently, critical size of the nucleus should be reduced according to (2.3.13) with ΔT increase. This means the possibility of a greater number of nuclei that are capable to grow when the supercooling degree increase. We can easily ascertain, if estimate the critical size of the nucleus at different degrees of supercooling. We make this calculation for iron.

If we take specific latent crystallization heat for iron as $\Delta H_{cryt} = 1{,}5\cdot10^3$ J/cm^3, $\sigma \approx 2{,}0\cdot10^{-5}$ J/cm2 and the crystallization temperature $T_{cryst} \approx 1800$ K, we find from Eq. (2.3.11) that for $\Delta T_1 = 10$ K, the value $\Delta G_v \approx 8{,}3$ J/cm^3. Substituting this value and the σ value in Eq. (2.3.13), we get $r_c \approx 100$ nm. Similar calculations carried out for supercooling $\Delta T_2 = 100$ K show that in this case $r_c \approx 10$ nm.

Keeping in mind that the lattice period for high-temperature modification of iron equals about 0.3 nm, it is possible to estimate the number of unit cells per nucleus of critical size. When the degree of supercooling is 10 K, critical size of nucleus in iron melt will hold about $3.55\cdot10^7$ unit cells, and for the $\Delta T_2 = 100$ K – $3.55\cdot10^4$, i.e. on three orders of magnitude lower.

Let us obtain another important relationship that characterizes the crystallization process, substituting Eq. (2.3.12) into Eq. (2.3.13),

$$E_{act} = G_V = \frac{4\pi\sigma^3 T_{Cr}^2}{3(\Delta H_{Cr})^2(\Delta T)^2}. \qquad (1.3.14a)$$

This energy is nothing else than the activation energy of crystallites nucleation or the total change in value of the thermodynamic potential ΔG in (2.3.12).

The area of the solid-liquid interface for the critical nucleus of the cubic form is $s = 4\pi r_c^2 s$, and its surface energy equals

$$G_s = \frac{4\pi\sigma^2 T_{Cr}^2}{(\Delta H_{Cr})^2(\Delta T)^2}. \qquad (2.3.14a)$$

A comparison of Eqs. (2.3.14a) and (2.3.14b) gives that $G_V = (1/3)\Delta G_s$, that is right for a critical nucleus of any shape.

Thus, as follows from Eqs. (2.3.13) and (2.3.14), the critical size of the nucleus $r_c \sim 1/\Delta T$, and its activation energy is $E_{act} \sim 1/(\Delta T)^2$. The E_{act} value can

be regarded as the height of the energy barrier, to create interface between solid and liquid phases to begin nucleation of the crystallization centers. This barrier is overcome due to thermal fluctuations, which are caused by randomness of the thermal motion of the atoms in the melt.

2.3.7. The speed of crystal phase nucleation in the melt

In the kinetic theory of nucleation of a new phase, the appearance of nucleus with the size greater than the critical, which is capable to further growth, is regarded as the fluctuation process. If the deviation ΔG of thermodynamic potential from the equilibrium state of the system is small, the number of stable nuclei in unit time per unit volume of the melt appears. The probability to appear the I_V nuclei is determined by (*1*) the probability f_{fl} to create the critical-size nucleus during homogeneous crystallization *due to the fluctuations of the thermal atomic motion* and (*2*) the probability f_{j} further joining of new atoms from the melt to the nucleus of critical size.

Probability of the critical-size nuclei formation can be described by the equilibrium distribution function:

$$F_{\text{fl}} = N_{\text{o}}\exp(-\Delta G_{\text{V}}/k_{\text{B}}T), \tag{2.3.21}$$

where k_{B} – Boltzmann constant and N_o – the number of places that can hold atoms in the crystallization center (nucleus). The last approximately equal to the number of atoms per unit volume of the material (sometimes it is necessarily to consider additionally some entropic effects).

The critical nucleus of the solid phase becomes the center crystallization only when it gets the opportunity to join new atoms from the melt. This probability is given by

$$F_{\text{join}} \approx \omega\exp\left(-\frac{\Delta E_{act}}{k_B T}\right), \tag{2.3.22}$$

where ω is the frequency of atomic vibrations (number of attempts to atomic jump from one equilibrium position to another), ΔE_{act} – the activation energy of the atomic jumps across the crystal-melt interface.

In a first approximation, the expression for the *intensity* (or *rate*) *of nuclei formation* per unit volume of the melt can be written as

$$\text{I}_{\text{centers}} = N_o\omega\exp\left(-\frac{\Delta E_{act}}{k_B T}\right)\exp\left(-\frac{\Delta G_V}{k_B T}\right). \tag{2.3.23}$$

The rate of crystalline phase nucleation in the melt plays an important role in the formation of the microstructure of polycrystalline ingots. In this regard, consider the results of experiments, which studied the effect of rates of nuclei formation (RNF) and growth (RNG) at homogeneous crystallization. RNF has

the dimensionality of $mm^{-3}s^{-1}$, while the dimensionality of RNG – mm/s. In fact, by the RNG is determined by withdrawal of the face of crystalline nucleus from its center. Both processes are connected with the migration of atoms and dependent on temperature.

Graphical behavior of the RNF and RNG, depending on the degree of supercooling, are shown in Fig. 2.3.11.

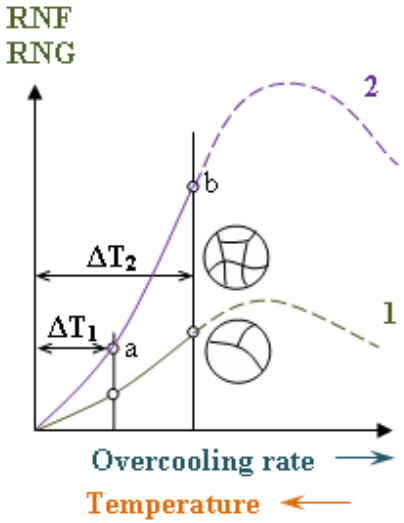


Fig. 2.3.11. Dependence of the centers nucleation rate RNF (*1*) and the crystal growth rate RNG (*2*) on the degree of overcooling ΔT. The insets schematically show large-grained (bottom) and fine-grained (top) polycrystalline structures formed upon supercooling ΔT_1 and ΔT_2, respectively

Consider the reasons for such dependencies of RNF and RNG on the degree of supercooling. When the equilibrium crystallization temperature, i.e. at $\Delta T = 0$, the values of RNF and RNG are equal to zero, so that crystallization does not occur. With the supercooling degree ΔT increase, difference of thermodynamic potentials $\Delta G = G_l - G_s$ is growing, so that RNF and RNG are also growing due to high mobility of atoms, reaching a maximum. The subsequent reduction of RNF and RNG in Fig. 2.3.11 can be explained by the decreased mobility of atoms (diffusion coefficient) with decreasing temperature as a result of strong supercooling. Small values of the diffusion coefficient (high melt viscosity) make it difficult to incorporate atoms from the melt into the crystal lattice of nuclei. As a result, due to very strong supercooling, RNF and RNG must again become equal to zero and the melt is not crystallized, transforming into amorphous state.

For metals, in which large supercoolings are not realized under normal conditions, the rising branches of the RNF and RNG curves are characteristic, so that the both rates are increased when ΔT increasing. For small supercooling degrees when the nuclei of critical size are large, and the rate of nuclei formation is small, a coarse structure in polycrystal is formed (see the bottom inset in Fig. 2.3.11 corresponding supercooling ΔT_1). Small supercooling degrees are achieved by casting molten metal in a crucible with a low thermal conductivity or in the pre-heated crucible.

The growth of ΔT occurs when the liquid metal casts in the cooled crucibles (form), as well as due to reducing the crucible wall thickness, which leads to the rapid heat removal from the melt. Since RNF increases more rapidly than RNG in this case, a fine-grained polycrystals are frowing from the melt (see the top inset in Fig. 2.3.11 corresponding supercooling ΔT_2).

2.4. Mechanisms and models of crystal growth

2.4.1. Layered-like mechanism of the crystal growth

As shown above, a homogeneous crystallization from requires nucleation of stable crystallization center in the parent phase (the melt or solution), which serves as a seed for the further growth. If the nucleus has been formed, further spontaneous joining of atoms to it occur, resulting in growth of crystal and development of crystal faces in equilibrium conditions (without supercooling of oversaturation, to form a crystalline polyhedron.

To describe this process, Kossel and Stransky created molecular-kinetic model of growth for alkali-halide crystals. The Kossel-Stransky model believed that the growing crystal surface (face) is flat, and its growth occurs with a slight oversaturation or supercooling.

In this model, elementary process of sequential ions joining to different positions at the crystal surface is used as a growth mechanism (see Fig. 2.4.1a). In other words, it is assumed that the growth process is carried out as a series of sequential steps with performing of “repeated step” work.

Added to the nucleus or seed particles (ions) will be considered for convenience in the form of cubes of the same size. In this case, the work done by joining the particle to the crystal is considered to be equal to the sum of energies of its interaction with neighboring particles. This total work (energy) depends on the number of chemical bonds which ion establishes with the close

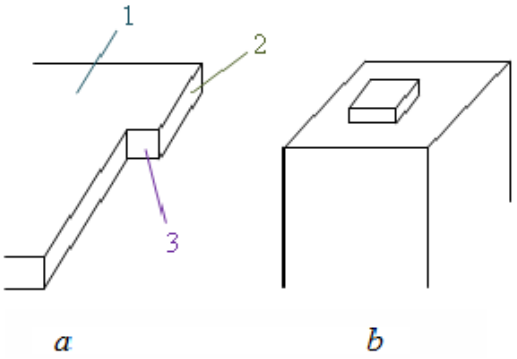


Fig. 2.4.1. Joining the particles to the surface of the growing perfect crystal: *a* – different positions at the face of the crystal, *b* – two-dimensional nucleus at the surface of the crystal

ion-neighbors in the nucleus. In this case, the "repeated step" energy (work), released due to joining the particle, will depend on the joining positions, which are numbered 1, 2 and 3 in Fig. 2.4.1a:

1) if the particle is attached to an arbitrary place of the planar area of the seed (position 1 in Fig. 2.4.1a), this leads to the formation of only one its chemical bond with the cubic-like crystal;
2) if the particle starts a new particles chain (row) at the growth step in the upper layer of the growing crystal face (position 2 in Fig. 2.4.1a), this corresponds to the formation of the two chemical bond;
3) if the joined particle continue the above ions chain on the growth step (position 3 in Fig. 2.4.1a), this corresponds to the formation of three bonds with a crystal seed.

According to calculations by the Kossel-Stransky, for such a model of NaCl crystal growth energy ε_i, released by incorporating of an ion to the above positions 1, 2 and 3 in Fig. 2.4.1a, respectively, equal

$$\varepsilon_1 = 0{,}066e^2/r, \quad (2.4.1)$$

$$\varepsilon_2 = 0{,}181e^2/r, \quad (2.4.2)$$

$$\varepsilon_3 = 0{,}874e^2/r, \quad (2.4.3)$$

where e is ion charge, r – the shortest distance between the oppositely charged ions in NaCl lattice. These relations show that the position 3 is mostly energetically favorable, since in this case the energy of the crystal is lowered above all.

The presented Kossel-Stransky model indicates that at crystal growth, firstly, the growth steps formation must be completed and before completion of the flat

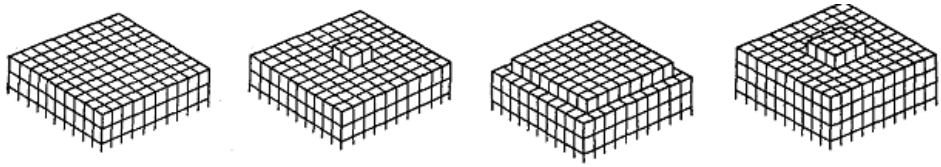
face formation in the crystal seed. And only then the formation of a new plane becomes energetically favorable. Thus, in order to force the face to grow, it is necessary to form on it two-dimensional (flat) growth nucleus (Fig. 2.4.1b or 2.4.2), to which then new ions will be joined by repeated steps mechanism, illustrated in Fig. 2.4.1a.

According to the above stated thermodynamics of crystal growth, the two-dimensional nucleus, consisting of the new flat part of crystal lattice, should reach a critical value. If the nucleus size will less than critical, it will be dissolved. If the size of the flat nucleus approaches critical value, it will join new particles (because it is energetically favorable) and is growing in all directions, forming new flat atomic layer (grid) at the top the crystal face (Fig. 2.4.2). Such a crystal growth from two-dimensional nucleus is called *the model layer-by-layer growth or layered-step growth.*

2.4.2. Layered-spiral mechanism of crystal growth

The above-described molecular-kinetic theory of crystal growth predicts that the probability to nucleate the two-dimensional nucleus at the growing crystal face is very sensitive to the degree of supercooling (or supersaturation). Thermodynamic calculations by Kossel-Stranski model show that the crystal growth can only begin with a relative supercooling not less than tens percents relative to equilibrium crystallization temperature. However, the experiments show, that real crystal growth start to go at very small supercooling (supersaturation), of the order of some percents. This contradiction has forced to develop a new theory of real crystal growth.

Kossel-Stransky theory, which has led to the relations (2.4.1)-(2.4.3), have been carried out for defect-free alkali-halide crystals with atomically smooth faces, which, in the best case, had steps (terraces) of atomic scale (type 3 in Fig. 2.4.2). This mechanism of layer-by-layer growth due to lateral displacement of their steps, when joining the atoms of the melt (or solution), has got an additional name as *layered-step growth mechanism*.



1 2 3 4

Fig. 2.4.2. The growth of the perfect crystal face by the expanding of two-dimensional nucleus: *1* - *4* – sequential stages of growth

Experiments show that yet one possible mechanism of crystal growth can be realized in real condition. It consists in the joining og the particles (atoms, ions) to the screw dislocations appearing on the surface of growing due to temperature (concentration) gradients in the melt (solution). As can be seen from Fig. 2.4.4, the screw dislocations output at the surface of the crystal looks like a step which is similar to position 3 in Fig. 2.4.1 and can easily join a new particle.

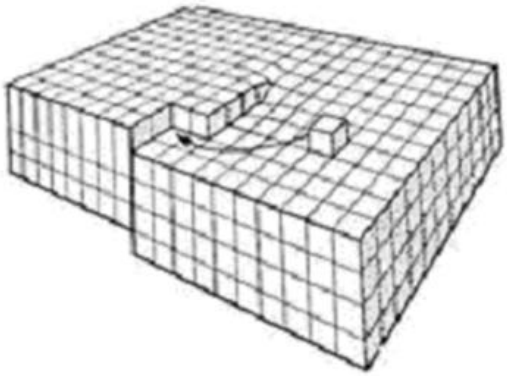
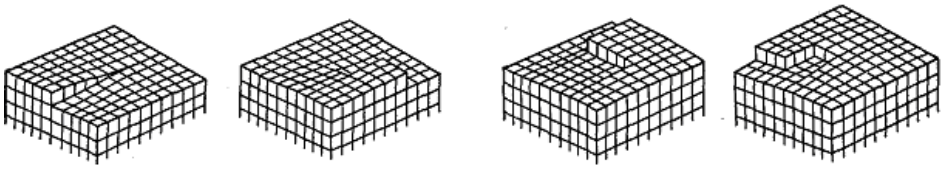


Fig. 2.4.3. Schematic view of a screw dislocation output on the surface of the crystal face and joining to it the new particle



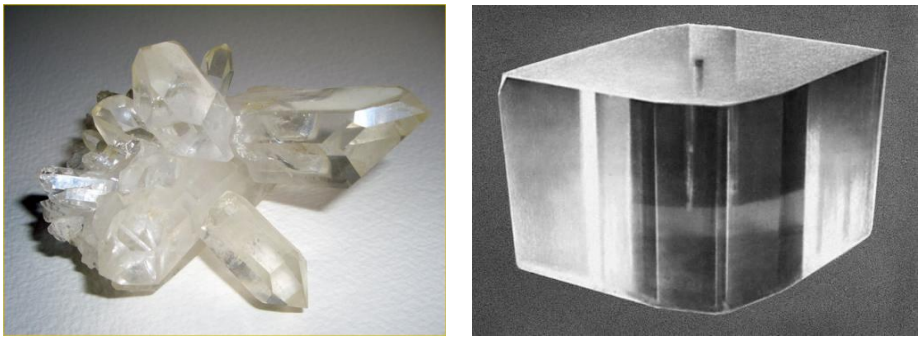
1 2 3 4

Fig. 2.4.4. The successive stages of the crystal face growth by a layered-spiral mechanism

The described character of crystal growth is called *a spiral-layered mechanism* and presented in Fig. 2.4.4. Since the growth of the crystal in this case is carried out by attaching the individual particles to gradually advancing spiral steps, then such growth does not require the stage of two-dimensional nucleation. The calculation shows that spiral-layered growth mechanism, carried out by screw dislocations, is energetically more favorable and can occur at arbitrarily low supercoolings or supersaturations.

2.4.3. Shape and structure of crystals and crystalline aggregates

Morphology of crystalline solids and their properties depend on their chemical composition (electronic structure of atoms), the nature of the interatomic interactions (chemical bonds) and atomic arrangements (type of crystal lattice), as well as growing conditions. When growing in a thermodynamic equilibrium conditions, the crystals may have a regular form with flat faces (Fig. 2.4.5). Such crystals are called *single crystals*.



a) *b*)

Fig. 2.4.5. Images of crystals grown in less equilibrium (*a*) and more equilibrium (*b*) conditions: *a* – group of intergrown single crystals of quartz, *b* – *a* single crystal of Rochelle salt

Crystalline solids remain solid, i.e. conserve their shape, at heating up to the melting point. When heating of amorphous substances, they are softened in the large temperature interval, becoming firstly viscous, and only then liquid. When cooling, process goes in the opposite direction. The crystalline solid state is more stable than amorphous. The amorphous state can be realized in many materials, using such methods as rapid cooling of the vapor or liquid state (see below).

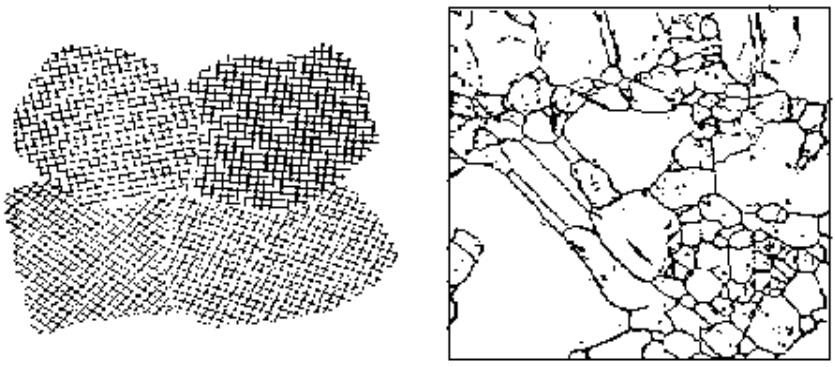
2.4.4. Atomic, nano, micro and macro structure of crystalline aggregates

The nature of the atomic structure of solid materials (ordered or amorphous), their external form (regular or random), the size and shape of the individual constituents (grains, phases) are determined by conditions of their production (growth). All these details of the structure of solid crystalline aggregations are described by the name of "structure".

Atomic structure shows the arrangement of ions, atoms or molecules, forming the crystal, in space (see Fig. 2.4.6a). It is usually studied by difaction techniques using X-rays, electrons and neutrons beams. By analysis of the diffraction patterns, obtained due to interaction of X-rays (or beams of electrons and neutrons) with atoms in a crystal lattice, you can get extensive information about the structure of crystals. In addition, the arrangement of atoms in space can be examined by transmission and tunneling microscopes. Transmitting electron microscope allows to observe the arrangement of atoms in space. Using tunneling microscope, we can "see" the distribution of atoms and molecules at the surface of the crystals, as well as the growth nanostep heights and other details of the atomic scale.

Exterior shape and sizes of the crystals formed in the crystallization process depend on the conditions of their growth: the direction and rate of heat removal and the melting point of substance (at crystallization from the melt), the speed of heat release and substance supply to the substrate (at crystallization from steam or gas phase), on the content of impurities, and many other factors. As can be seen from Fig. 2.4.5, with more or less the equilibrium conditions of the crystal growth, can be with plane faces, and even in the form of a regular polyhedron when perfect growing conditions.

In the case of non-equilibrium crystallization, for example in the case of nucleation of many crystallization centers in the melt, the shape of growing crystals will be irregular, because they will be touched each other preventing their growth. In this case the formed crystal is called a *polycrystal* (Fig. 2.4.6). Polycrystals contain *grains* or *crystallites* every of which present single crystals with the same with atomic structure (lattice).



a) *b*)

Fig. 2.4.6. Schematic image of polycrystalline structure (*a*) and image of a flat etched surface of the metallographic section in optic microscope (*b*) of polycrystalline material

To indicate the nature of the structural components of polycrystalline aggregates, a classification (identification) is frequently used, which is based on the size of their grains (crystallites). If the grain size of the polycrystalline material is in the range of 1-100 nm, we speak about *nanostructure* (Fig. 2.4.7).

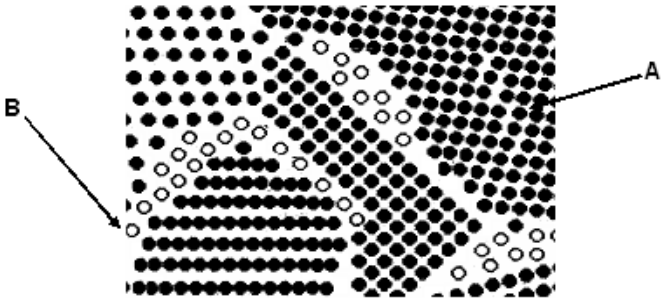
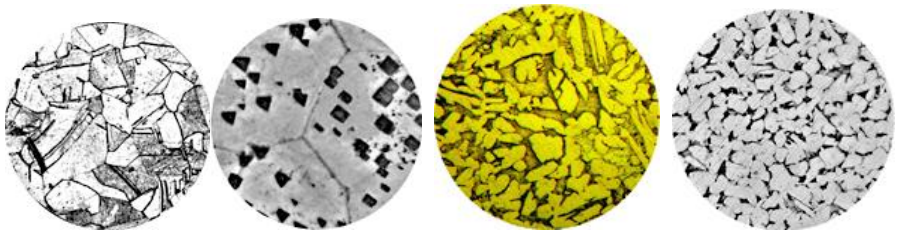


Fig. 2.4.7. Schematic image of nanocrystalline structure of polycrystalline material. Black and light circles present atoms belonging to nanograins (A) and grain boundaries (B) correspondingly

If polycrystalline material has grain sizes of about 1 to 1000 μm, we call this structure as microstructure (Fig. 2.4.8). If the grains are millimeters and more in size we say about macrostructure (Fig. 2.4.9).



a) *b*) *c*) *d*)

Fig. 2.4.8. Microstructure of silicon iron (*a*), copper (*b*), bronze (*c*) and brass (*d*)

Example of the schematic arrangement of identical atoms in a two-dimensional section of grains in nanocrystalline material is shown in Fig. 2.4.7. As is seen, such nanostructured materials can be divided into two basic structural constituents: the ordered which includes atoms localized within the grains, and intergranular, localized at the grain boundaries. As is seen, nanograins are characterized by the ordered atomic structure. The location of the atoms in grain boundaries may be different from the arrangement of the

atoms within the nanocrystallites. Atomic structure of the boundaries is not simple and depends on many factors, primarily on the relative orientation of the atomic lattices in neighboring grains.

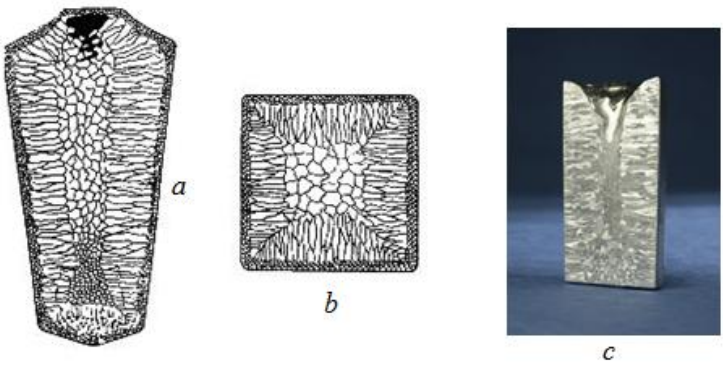


Fig. 2.4.9. Schematic image of macrostructure of steel ingot obtained by casting in a mold for a longitudinal section (*a*) and cross-section (*b*). Macroscopic real cast ingot in a longitudinal section (*c*)

To monitore such small structural constituents of polycrystals with nano-sized grains (nanocrystals), we should use electronic microscopes and also atomic force and tunneling microscopy with atomic resolution. Electron and atomic force microscopy also allow to identify defects such as dislocations and their boundles, the growth steps, grain boundaries, etc.

The microstructure, we can observe by optical or electron microscope of relatively low resolution on specially prepared samples. They are prepared by grinding and polishing of the pieces, cut of polycrystalline ingot, and their subsequent etching by any chemical reagents (usually in mixtures of acids). After etching of the plane metallographic section, we can see on its surface grain structure under a microscope. This method is called *metallographic analysis*. Examples of the microstructures in some polycrystalline materials are shown in Fig. 2.4.8.

Microscopic methods provide opportunity to determine the sizes and shape of the grains, the presence of different crystallites and their relative volume amounts, a form of foreign inclusions and microvoids, the orientation of the crystallites, the availability of special crystallographic features (twinning, slip lines, grain boundaries, outputs of growth dislocations decorated by impurities, etc.).

Macrostructure of the crystals is usually seen with the naked eye or with small magnifications, for example, given by a magnifying glass. Fig. 2.4.9 shows the scheme and a real image of the grain macro structure in the cross-section of steel ingot produced by custing into the mold. The shape, orientation and size distribution of macro crystallites, formed with such nonequilibrium solidification, depend primarily on the character and rate of heat removal from the melt and mold wall.

When the macro structure studying, we can reveal the nature of the fracture in a polycrystal and the so-called shrinkage voids and pores in ingots. It is possible to determine the size and shape of the large grains. Using specially prepared samples (polished, etched), we can find cracks, chemical heterogeneities, texture, etc. The study of macrostructure, in spite of its simplicity, is a very valuable method to study materials. Consider below in more detail some examples of the formation of micro and macro structure of real crystals obtained in different ways.

2.4.5. Casting of polycrystalline materials

High equilibrium crystallization of the material from a single crystallization center leads to the formation of a single crystal. In technical (industrial) material science, such single crystals are grown only in strictly defined and carefully controlled conditions, that dramatically increases their cost. Therefore, production costs of such crystals are justified only for special applications such as semiconductor devices, micro- and nanoelectronics, solar cells, functional materials for various kinds of sensors and detectors, etc.

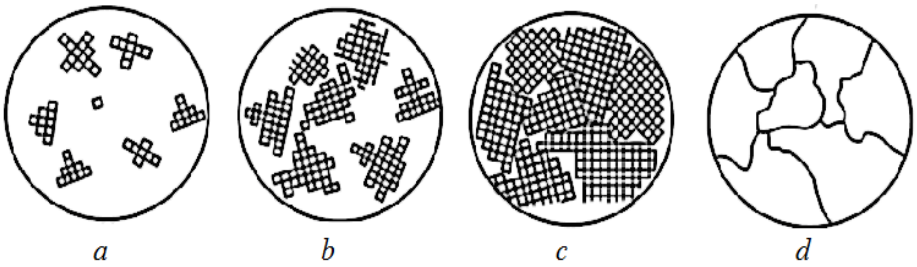


Fig. 2.4.10. The scheme of polycrystalline growth (squared-like lattice structure approximate the grain structure): *a* – crystallites nucleation, *b* – crystallites growth; *c* – the formation of a polycrystal with multitude of grains with randomly oriented crystal lattices; *d* – view of the grain boundaries in a polycrystal under microscope

Under normal, inequilibrium growth of crystalline materials from the melt, the solidification starts from nucleation of many crystallization centers

(see, Fig. 2.4.10a) in the melt, which will lead to the formation of polycrystalline structures. As described above and follows from the figure, the nucleation of these centers startbegins when the temperature of the melt becomes lower than the crystallization point. Crystallites, when cooling, increase in size due to the continuous addition of atoms to them (Fig. 2.4.10b) and grow until touching adjacent growing grains (Fig. 2.4.10c), which corresponds to the completion of the formation of a polycrystal (Fig. 2.4.10d). As can be seen from Fig. 2.4.10d, grains in a polycrystalline sample are typically single crystals of small sizes without expressed smooth faces.

2.4.6. Dendritic growth of crystals

Metallic crystals, grown from the melt or supersaturated solution at the rapid cooling, display very often the dendritic structure similar to the branches of trees (Fig. 2.4.11).

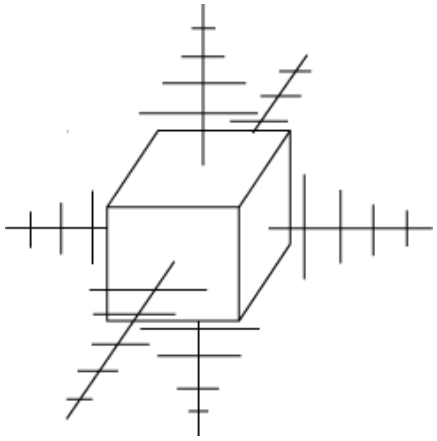


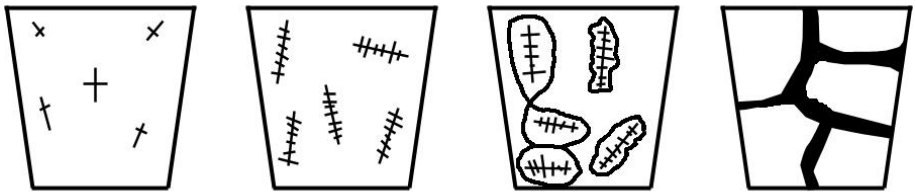
Fig. 2.4.11. The scheme for the growth of dendritic crystal with a simple cubic structure

Typically, the formation of dendrites occurs during rapid solidification of the melt or due tp inhomogeneous supplying of substance to different parts of the growing crystal. The necessary conditions for the formation of dendrites are strong supercooling (for example, at custing) and inadequate intermixing.

Dendrites appear quite often in the casting of molten metal into the mold or crystallisation form in the metallurgical industry. Fig. 2.4.11 shows a schematic diagram of a three-dimensional growth of a dendritic crystal on the cubic form nucleus with six faces. For simplicity, the addition of atomic crystallographic cells under continuous solidification is not shown.

The possibility to realize the scheme in Fig. 2.4.11 is due to the fact that the maximum rate of crystal growth is observed by such crystallographic planes and direction that have the highest density of atomic packaging. These directions are normal to the faces of the growing crystal, as shown in the diagram in the form of lines. As a result, in such directions long "branches" are growin that are called the axes of the first order. With the growth, on the dendrite axes of the first order the second order "branches" begin to grow, then the third order "branches"appear, etc.

During the solidification of the melt and the formation of a polycrystalline material a plurality of crystallization centers for future dendrites are nucleated (Fig. 2.4.11a). If heat removal from the melt with nuclei increases dendrites grow (Fig. 2.4.11b) until they become large enough and will not touch each other (Fig. 2.4.11c). Since the crystallization centers are formed arbitrary, the crystal growth axis of the primary dendrite branches will also oriented randomly. As a result, grains will grow of these dendrites in different directions. Because of mutual touching of crystallites, the growth slows down or even stops and grains become an irregular shape. Ultimately, the melt filling the interdendritic space will also crystallize, giving polycrystal composed of several spliced grains with a regular structure, but generally with non- regular shape of each crystallite (Fig. 2.4.11d) .



a) *b*) *c*) *d*)

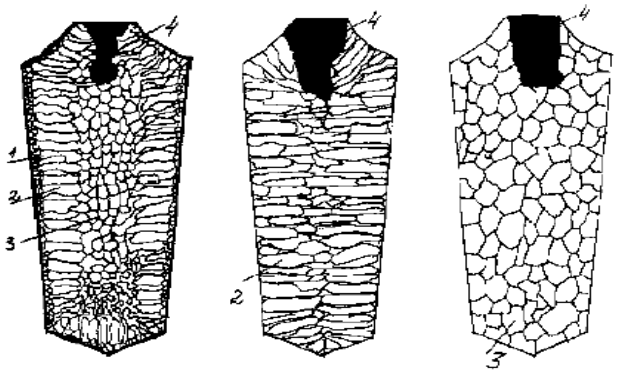
Fig. 2.4.12. Scheme of the polycrystalline sample at the nucleation and growth of dendrites.The bright areas correspond to the liquid phase (melt), and dark – a crystalline phase. Straight lines within the grains indicate the directions of growth of primary and secondary branches (axes) of dendrites

Note that the dendritic structure often can not be detected in the study of the polycrystalline material microstructure. However, as will be shown below, the dendrites may typically occur around the so-called shrink pores or cavities formed in the process of rapid crystallization, such as casting or welding.

2.4.7. The influence of the nature of heat removal on the features of the polycrystalline ingots macrostructure

Consider the particular structure of the polycrystalline material formed at a sufficiently rapid cooling of the melt, e.g., at the casting to the metallurgical mold. In this case, the crystallized polycrystal is called *an ingot*. Macro and micro structure of an ingot depends on many factors. The amount and properties of impurities in the pure metal, casting temperature, cooling rate during crystallization, as well as configuration, temperature, thermal conductivity, surface roughness of the inner surface of the mold, etc. etc. are the mostly important conditions.

Conditions of heat removal play the most significant role in the shape and size of grains in ingots. This is because the crystallization process is running advantageously in the opposite direction to heat removal (i.e., primarily in the direction perpendicular to the wall of the mold). For such directed heat removal, grains will generally have elongated columnar-like shape. When a heat removal from the growing crystal is given in all three directions with approximately the same speeds, equiaxial crystallites (approximately with the same sizes in all three dimensions) will form when the melt solidifies.



a) *b*) *c*)

Fig. 2.4.12. Schematic representation of different types of macrostructure in cast ingots: *a* – a small degree of supercooling ΔT (normal cooling), *b* – a strong overheating and rapid cooling (high ΔT), *c* – very slow cooling (small ΔT). *1* – zone of fine equiaxial crystals, *2* – zone of columnar crystals, *3* – zone of equiaxial crystals, *4* – shrinkage hole

Fig. 2.4.12 shows the scheme of the formation of ingot structure produced by casting of molten metal into a simple vertical metallic mold. As will be shown, the structure of the ingot (the size, shape and orientation of the crystallites) after completion of crystallization depends on the degree of

supercooling ΔT, directions of heat removal and the presence of ready centers of crystallizatin (roughness on the inner surface of the mold). Let us analyze them.

As can be seen from the diagram in Fig. 2.4.12a, with a small degree of supercooling ΔT macrostructure of cast ingot consists of three zones. In such ΔT firstly liquid metal is supercooled at points of contact with the cold walls of the mold, where supercooling is reached fastest. Slight cooling ΔT and faster heat removal normally to the mold wall surface leads to nucleation of multitude crystallization centers on its inner surface, which promotes the formation of equiaxial fine crystals (grains) in zone 1 near ingot surface.

The orientation of the crystallites in the zone 1, to a large extent, depends on the state of the internal surface of the mold (its roughness, the presence of adsorbed gases and moisture) as well as the presence in the metal oxides and other nonmetallic inclusions in the melt. Their touching of each other, due to random orientation, stops their growth very soon, so that this zone is usually very thin and is not always visible with the naked eye. Then, as far as the cooling of the melt is running, the predominant grain growth goes normally to the mold walls due to maximal heat removal in this direction. This gives columnar grains in Zone 2. Finally, towards the middle of the ingot, where fewer crystallization centers are created due to the lower ΔT and lack of directed heat removal (it becomes the same in all directions), Zone 3 is formed with equiaxial crystallites, but larger sizes than in zone 1.

By applying different technological methods, it is possible to change the thickness of zones 1-3 or even exclude them from the structure of the ingot at al. For example, a strong overheating of the melt prior to casting (large value of ΔT) and its rapid cooling during crystallization will lead to the formation virtually alone columnar grains in the ingot (Fig. 2.4.12b). A very pure metal ingots with high thermal conductivity have typically such a structure. Zone with the columnar crystalline structure has the highest density, but in the center of ingot, where columnar grains are in contact, insoluble impurities are collected and the so-called shrinkage pores are formed (see below). Such a structure can also be formed in the welds due to very rapid cooling of the melt (due to its small volume and direct contact with a large mass of the base metal), which increases the brittleness and reduces strength.

Lowering the temperature of casting alloys (small ΔT), flushing of liquid metal by inert gases, the application of ultrasonic vibrations, the addition of the modifying agents, as well as the increase the walls mold temperature (heating) result in a reduction or even disappearance of the zone with columnar crystals. Such ingots have structure, consisting of equiaxial crystallites in in the bulk (Fig. 2.4.12c).

One of the features of the cast ingot macrostructure is the formation of large shrinkage pore 4 in the upper part of the ingot. It is formed due to lack of liquid

metal for filling interdendritic space (due to the lower density of the melt in comparison with the density of the crystalline phase). Metal under shrink comes out loose and contains a lot of cavities.

Ingots with shrinkage pores are often cracks during subsequent machining. They also lead to failure of welds. Therefore, the ingot parts with a shrinkage pores and loose metals are not usually used in the manufacture of critical issues and other products.

2.5. Quenching of single-component materials

The quenching is one of important ways of thermal impacts on the structure and properties of crystalline materials and products on their base that are widely used in materials science and metallurgy.

Quenching is a form of heat treatment, which results in the change of the structural-phase state when the object of quenching (material or product) is first heated to a high temperature (called the *quenching temperature* T_Q), maintained at this temperature to bring the object to an equilibrium state, and then abruptly cooled to $T << T_Q$ (room or lower temperature). Quenching is used to fix the high-temperature state, or prevent (suppress) the diffusion processes taking place during slow cooling and causing the formation of undesirable, equilibrium micro- and macrostructure of the crystal. The quenching is very often consists of the transfer material in a molten state and then rapid cooling of the melt to transform it into disordered solid state (in the limit, the *amorphous* or *glassy*).

Quenching in the solid state is possible only for crystalline substances, if their equilibrium state at a high temperature is different from the equilibrium state at a low temperature (e.g., by the concentration and type of defects in the crystal lattice, by crystal structure of single-component substances (polymorphism), by phase composition in multi-phase materials, etc.).

Quenching is effective only if actually achievable cooling rate is sufficient to suppress undesirable processes whose suppression is just the purpose of quenching. Primarily one try to suppress atomic diffusion and the formation of long-range order).

The state of crystalline material, resulting from quenching, is inequilibrium and only relatively stable below T_Q. Subsequent heating of the quenched materials they are transferred surely to a more stable crystalline state (*recrystallization*).

Quenching, as a method for fixing of the high-temperature state of pure (single-component) material, is used for many purposes, the most important of which are: the study of lattice defects (e.g., vacancies), which remain only in

small amount when slow cooling to low (room) temperatures; fixing the high temperature modification of the crystalline structure in polymorphic substances; giving other than the equilibrium properties to the material by changing its micro- and macrostructure (high hardness and mechanical strength, magnetic hardness, low conductivity, etc.) and also the obtaining an amorphous (glassy) or nanocrystalline material state by quenching from the melt.

2.5.1. Quenching the solid state

Quenching of solid materials or products heated to a high temperatures (e.g., tubes, gears, cutting tools, etc.), as well as materials slugs (plates, foils, slabs, wires, etc.) is often carried out by their immersion in a *queqnching medium* (air, water, oil or liquid polymer or metals) to quench. Usually quenching to quenching liquids is carried out to approach high cooling rate to avoid creating undesirable internal microstructure and provide uniform mechanical properties with the minimized residual stresses, in order to avoid distortions of the working parts. Quenching of crystalline solids is a complicated process and usually involves three stages: formation around the quenched parts of vapor film of the quenching liquid; boiling of quenching liquid, including the nucleation of the boiling process, and the convective stage.

The main challenge in quenching technology is to ensure conservation of the homogeneous (and very high) rate of heat removal from the quenched object. Nonuniform cooling can lead to residual stresses (as a consequence, deformation of the quenching object) that causes its bukling. In this sense, if the quenching medium is a liquid, the most critical is the stage of the formation of the steam shell of the quenching fluid around the objects quenched. This shell separates the item and quench liquid, that decreases the rate of heat removal from the quenched object and, thereby, reduces the efficiency of quenching. Typically, it is necessary to destroy this vapor shell by vigorous mixing of the medium (liquid) near the surface of the quenched body.

It is evident that the quenching rate depends not only on the type and mixing degree of the quenching medium, but also on the thermal conductivity and heat capacity of the quenched object. Simulation of fluid flows within the quench medium vessel (tank) detects strong influence of their heterogeneity around the quenching object (particularly, in the bodies of complex forms) on the quality of quenching.

2.5.2. Quenching from the melt

As stated in the previous section, the higher cooling rate of the melt, the lower the temperature to which liquid metal may be supercooled prior to crystallization. This creates the possibility to make *quenching from the melt*,

which has gained a great practical significance in recent years. It turned out that at very high cooling speeds of the melt (above 10^6 K/s) to room temperature, diffusion of atoms can be slow enough that allows to suppress the formation of crystallization centers and growth of crystallites. In this case, as mentioned above, solidification can result in the formation of amorphous structure. The materials obtained (metals or semiconductors) are called *metallic* or *semiconducting glasses*.

The most difficult to obtain an amorphous structure for pure metals (due to their high thermal conductivity and the rapid heat removal). The stabilization of the amorphous structure in pure metal requires extremely high cooling rates of the melt ($l0^7$-10^9 K/s). These speeds are usually reached such ways as: ejection of melt drops and jets on the cold rotated plates; spraying of jets of molten liquid by liquids or gases with a high cooling capability, laser ablation or melting of the surface of the body, ultra-fast cooling of vapor, electrolytic deposition, plsma or ion-beam sputtering with high-speed deposition on cold substrates, etc.

Ejection of melt drops and jets on the cold rotated plates is considered as the most effective way to produce amorphous materials, suitable for practical applications. Amorphous metals in this case can be obtained as very thin ribbons, wires, fine granules, flakes, etc. It is essential that such amorphous metals have important technical applications because of their unusual properties that are unattainable in the ordered crystalline materials. Solidifying to form an amorphous structure is fundamentally possible for virtually all metals. At the present time, amorphous structure was produced in more than 20 pure metals and semiconductors, as well as in more than 100 alloys. To stimulate formation of the amorphous structure in refractory transition metals by quenching from the melt, special doping elements (*C, P, R, N, S*, etc.) are added to melts.

Naturally, the metal with the structure of the supercooled liquid, i.e. in the amorphous state is thermodynamically unstable. Therefore, upon subsequent heating, when the atoms become defined diffusion mobility, amorphous or glassy metal will be crystallized in the solid state.

2.5.3. Nanocrystalline and nanostructured materials

At condensation of vapor on the cooled substrate or the cooling of the melt at a moderate rate (when the migration of atoms is not completely suppressed), solidification process may not lead to the formation of glassy solid without long-range order (as at the ultra-fast quenching). In this case, the formation of some intermediate state – *nanocrystalline* – occurs. This state can also be obtained under certain modes of heating the amorphous material (through the mechanism of solid recrystallization) .

Nanocrystalline materials are condensed in form of nanosized structural elements (clusters, crystallites, grains, granules), which contain the number of atoms not more than 10^2-10^4. The atomic structure in such nanoscale materials may possess not only short-range but also long-range order within the grains. However, this order is not always coincide the same in the bulk crystal lattice in single crystals or polycrystalline sample of the same composition. Therefore materials composed of nanosize structural elements are often more properly to call as *nanostructured*, but not nanocrystaline.

Nevertheless, nanostructured state of materials can be realized in the form of a nanocrystalline structure (with long-range order), wherein the crystallite sizes are less than 10-15 nm. It is possible both in nanogranular materials and multilayered structures with alternating layers of different materials with thicknesses of about 5-10 nm or more.

Already from the time when chemists began to investigate the so-called ultra-fine materials (in the form of colloidal solutions), it has become clear that the nanostructuring is an extremely powerful and unusual impact on many of physical and chemical properties of solids. That is why the nanostructuring of materials now is considered as an effective (and sometimes even the only) method to change (improve) the properties of solids. For example, in metals such changes detected by reducing mean sizes of structural elements (crystallites, grains, granules, clusters) down to 100 nm, but they are most visible at the crystallite sizes of less than 10 nm.

The increase of the relative proportion of atoms belonging to the interface (or surface) as compared to their number in grains, clusters or granules. Often just this is what plays a crucial role in changing the properties of materials at their nanostructuring.

Let us estimate the fraction of atoms belonging to the surface of the spherical nanoranules with a decrease in their size *d*. We assign the interface (interface) between the contacting grains certain thickness h. Then the fraction of atoms belonging to the interfaces between nanogranules in their total amount in nanogranule can be described by the following equation

$$\frac{\Delta V}{V} = \left[\frac{\pi}{6}d^3 - \frac{\pi}{6}(d-2h)^3\right]\left(\frac{\pi}{6}d^3\right)^{-1} \cong \frac{6h}{d} \qquad (1.4.1)$$

The calculation by this formula shows that for the interface thickness $h \approx 0.5$-1.5 nm (corresponding to three or four atomic layers) number of their atoms can account for up to 50 % of all atoms in the grain.

Note that such defects, as vacancies , dislocations, stacking faults, which are typical for coarse-grained polycrystalline materials are absent in nanostructured materials. Such their defect-free structure is due to coming of defects to the surface of the grains or granules. On the other hand, as grain boundaries are

themselves defects, this leads to a high diffusion mobility of atoms and impurities at the interface, which can be 5-6 orders of magnitude higher than in the bulk materials.

Consider how the atoms are arranged in a nanocrystalline material. As already noted, it is assumed that one of the important reasons for the change in the physical properties of nanomaterials is an increase in the proportion of intergranular phase (light atoms in Fig. 2.5.1). It is assumed that the arrangement of atoms forming the interface differs both from crystalline and amorphous structure: they have no even short-range order and are characterized by a lower density than the crystalline phase. We can use the model of hard spheres to characterize the arrangement of atoms in a nanocrystalline material, as schematically shown in Fig. 2.5.1.

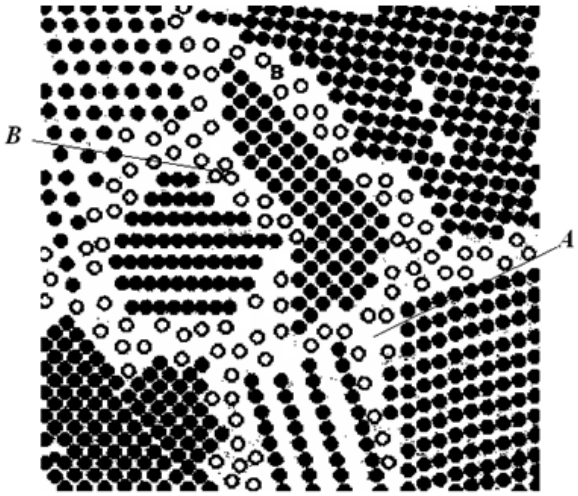


Fig. 2.5.1. Schematic representation of the crystalline structure (black circles) and grain boundaries (white circles) in the nanocrystalline material. *A* – a special boundary, *B* – random boundary

As is seen from the experiments and sketch in Fig. 2.5.1, nanocrystalline material can be roughly divided into two constituents: the ordered atoms which are localized within the crystallites (granules) and disordered atoms in intergranular interface. In the nanomaterial every grain has always a long-range order in the arrangement of atoms. The atomic structure of the boundaries (interface) is significantly affected by many parameters, including the angle of misorientation of the crystal lattices in the neighboring crystallites and the presence of foreign atoms. Fig. 2.5.1 shows difference in the structure of

intercrystallites boundaries: the arrangement of atoms in the boundary B (random) differs from that of the boundaries of type A (special).

2.5.4. Effect of quenching by high-rate cooling of the melt on the properties of single-component materials

Quenching of pure metals and single-phase alloys is used for research purposes as well as for giving new and unusual properties not possessed by the materials obtained by slow crystallization from vapor, liquid solution or melt. For example, quenching is applied to study the influence of vacancies on the mechanical, electrical and other physical properties of substances. The purpose of quenching in this case consists in fixing at a low temperature vacancy concentration which is stable at high temperature. Subsequent heating up to a temperature, which stimulates the vacancies mobility, causes an increase in resistance to plastic deformation (this is called a quench hardening), a reduction of the internal friction, etc. When studying the dependence of the excess concentration of vacancies on the quenching temperature, one can find the energies of their formation and migration, and also determine the activation energy for self-diffusion in pure elements.

Quenching in the solid state, fixing the high-temperature (usually more damaged) state, leads to an increase in electric resistivity ρ of crystalline materials. For example, the resistivity of amorphous metal is 2-3 times higher than that of similar crystalline metal. Temperature coefficient of resistivity $d\rho/dT$ of such materials becomes often negative, indicating that the amorphized metal is similar to the behavior of semiconductors in electrical sense.

Amorphous metals are often successfully combine high values of strength, hardness and wear resistance with good ductility and corrosion resistance. Great practical importance is also the possibility to obtain amorphous metals directly in the casting process in the form of tapes or wires with diameters a few micrometers, escaping such operations as forging, rolling, drawing, intermediate annealings, etching, etc.

Macroscopic bulk materials transferred into nanostructured state acquires also such properties that can not be obtained in materials with conventional polycrystalline or monocrystalline structure.Examples include materials which are used in the manufacture of blades for gas turbines. Using heat-resistant nanostructured alloys, we can improve their wear resistance and toughness.

Recently discovered that the nanocrystalline materials possess the low coercivity and high magnetic permeability. The magnetic properties of this material are insensitive to the agreed mechanical influences. Found that nanostructured materials differ significantly from conventional polycrystals by other fundamental properties such as elastic constants, specific heat, thermal

expansion coefficient, etc. X-ray or ultraviolet optics uses special multilayer mirror coatings which consist of alternating nanolayered elements with high and low density, such as tungsten and carbon.

2.6. The change of structure in single-component crystals under thermal impacts

The presence of polymorphic transformations in single-component substances will cause the modification of the structure at the atomic level as a change in the type of crystal lattice under thermal impacts (cooling or heating). Therefore, single-component compositions, prepared in non-equilibrium conditions (e.g., by quenching, deformation, and also at high concentrations of structural defects due to violation of the growth conditions, etc.), will tend to pass in more equilibrium state. This transition will cause a variety of changes in the atomic , nano-, micro- and macro-structure of the material. For example, when quenching from the melt, the created amorphous (glassy) solids will change their atomic structure from disordered (without long-range order) to the ordered (with long-range order and lattice structure) state.

Heat treatment at a certain temperature of nanostructured, nanocrystalline or severely deformed materials will change the character of their grain structure (average size, number and relative orientation of crystallites) and defect subsystem (concentration and type of defects). Indeed, in the obtained by casting fine-grained polycrystalline ingots of pure substances, annealing will change mean sizes of grains, their number and mutual orientation, and also the concentration and type of lattice defects.

All the above mentioned potential changes in atomic, grain and defect structure of single-component substances under heating or cooling are widely used in the technology of construction and special materials, and also products on their base in order to impart them the needed functional properties. For example, manipulating the size of grains in polycrystalline materials, we can tune their mechanical properties. As is known from experiments, the tensile strength σ of polycrystals increases with grain size d decrease, obeying the Hall-Petch law

$$\sigma = \sigma_0 + kd^{-1/2} \qquad (2.6.1)$$

where σ_0 and k are parameters that are functions of the atomic and microcrystalline structure of the material. In particular, σ_0 is the stress necessary to ensure the free movement of dislocations and k – coefficient characterizing the strength needed for blocking of dislocations movement (their fixing) caused by mechanical stresses.

Note that reducing the average grain size is practically the only mechanism for simultaneously increase both tensile strength and toughness (the ability to absorb energy during dynamic mechanical impacts). Increase in toughness due to fine-grained structure of the material is explained by two reasons. They are reducing of cracks` nuclei sizes and better blocking of their extension. Actually, in a polycrystalline sample with randomly oriented relative to each other grains, crack growth is blocked because the generated crack is forced to change the direction of its motion when moving from one grain to another (crossing grain boundaries). As a result, the trajectory of its extension makes longer and, consequently, the resistance to its spread is increased.

On the other hand, grain boundaries in polycrystals are defects contributing to the formation of routes with the enhanced diffusion of atoms. For example, at high temperatures, grain boundaries lead to loss of the material strength due to enhanced migration of atoms along them at high temperatures. This enhanced diffusion results in permanent deformation of the polycrystal under hot deformation. This phenomenon is called *creep* of polycrystalline material. Hence, to use structural material at high temperatures, it is necessary to increase the grain sizes to lower density of grain boundaries and, thereby, to minimize the area of grain boundaries per unit volume. Let us consider in more detail the changes in atomic-crystalline, grain and defect structure of materials under different types of thermal impacts.

The materials, produced by one or another way, rarely have a full range of properties needed for certain applications. Therefore, after manufacturing, material or product must be processed in some way to give it the needed shape and the desired functional properties. For this purpose, together with the deformation of crystalline materials, various kinds of thermal impacts, including the so-called heat treatments of materials are widely used.

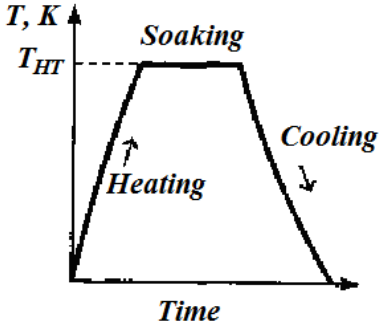


Fig. 2.6.1. General scheme of heat treatment procedure

Heat treatment is technological processes, consisting of 3 stages – (*1*) heating up to a certain predetermined temperature T_{HT}, (*2*) exposure (soaking) at this temperature for a certain time and (*3*) subsequent cooling at the determined rate to change the structure and properties of the material or product (Fig. 2.6.1).

Heat treatment is used for processing of grown single crystals (copper, silicon, germanium, etc.), ingots and casts (steel, cast iron, etc.), semi-finished products, welded joints, parts of machines and mechanisms, tool ware, etc. The heat treatment is often used as an intermediate operation in the chain of production processes (to manufacture half-finished products), and also as its final stage. For example, after casting, forging, etc. components or half-finished products are subjected to heat treatments to reduce their hardness and strength to improve their machinability on machine tools and stamping presses. On the other side, the ready-made items (products) are subjected to heat treatments on the contrary to increase their hardness, strength, elasticity and durability of the work.

2.6.1. Annealing of single-component materials

Single-component substances (pure elements and chemical compounds), are usually subjected to two kinds of heat treatments – quenching and annealing (see Section 2.4).

Annealing is a type of heat treatment, when atomic, nano-, micro- and macro-structure of pure, single-component materials become more equilibrium. As a rule, materials having non-equilibrium structure due to violation of the conditions of growth, as well as plastically deformed metallic ingots and products are subjected to annealing.

Annealing gives mostly rise softening and increase plasticity of materials (for example, metals) due to removal of residual stresses (reducing the defect density in the crystal lattice). The annealing temperature depends on type of metal and the specific type annealing. Speed of cooling from the annealing temperature after annealing is usually small – within the 30-200°C/h.

Annealing of metals has a few varieties, which are defined their prehistory. The presence of crystal defects and the occurrence of internal stresses (e.g., due to plastic deformation or disturbance of growth conditions) lead to the storage of free energy in the metal. This means that metal is in a non-equilibrium, and therefore unstable, state. So long-time exposure at room or low temperatures of low-melting metals and high-temperatures of the refractory metals have to facilitate the transition of non-equilibrium metal into more steady structural state. This transition in the equilibrium state with temperature increase is due to growing of defects mobility and enhancing of diffusion processes. Since the

plastic deformation is an important intermediate stage of manufacturing for almost all metallic items (for example, to impart a specified shape), we first consider annealing of metals subjected to moderate plastic deformation.

In physical metallurgy, all the processes occurring during annealing of deformed metal can be divided into two main stages – recovery and recrystallization (see the diagram in Fig. 2.6.2).

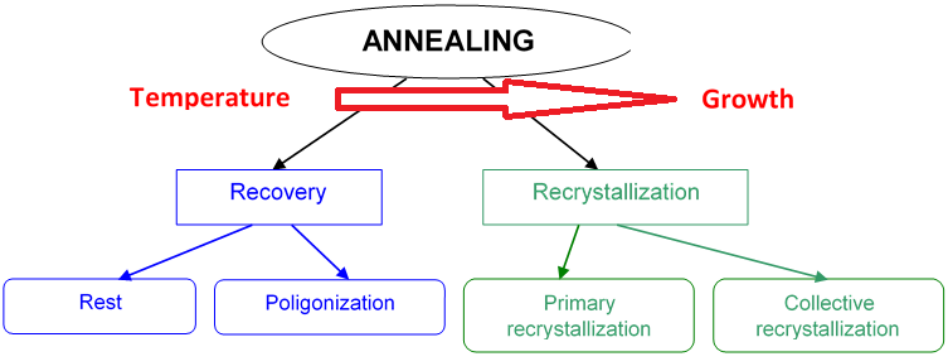


Fig. 2.6.2. Scheme of annealing modes for the plastically deformed metals. The horizontal arrow shows the direction of temperature increasing for different annealing stages

All the changes in the defect ensemble and properties, that are not accompanied by a modification of the micro- and macrostructure of the deformed metal, are called a *recovery*. In other words, the size and shape of grains in the deformed polycrystalline material at the recovery stage are not changed during annealing.

Recrystallization is the process when nucleation and growth of grains occurs during annealing in solid state. The new grown grains replace the original ones in polycrystalline or deformed material.

Like any 1st-order phase transition of order-disorder type, both stages of annealing are accompanied by the release of latent heat transition and decrease in free energy (thermodynamic potential). As is seen from Fig. 2.6.2, recovery occurs at relatively low temperatures (for pure metals it is usually below (0.1-0.2) $\mathbf{T}_m$), while recrystallization runs at higher temperatures (for commercially pure metals – 0.4 $\mathbf{T}_m$).

2.6.2. Changes in atomic structure of deformed metals on the stage of recovery

Consider the changes that occur in the system of lattice defects and the atomic structure of the deformed pure metals at different recovery stages. According to scheme of annealing steps, shown in Fig. 2.6.2, recovery is usually subdivided into two stages – *rest* and *polygonization*. Note that rest is always developed at heating of the deformed metals, while polygonization demands some additional conditions.

Rest is the stage of recovery of the cold-deformed (at room temperatures) metals, when annealing results in decrease of the amount of intrinsic point defects, due to their recombination. During the rest the cold-deformed metal stores the internal energy due to reducing the concentration of point defects (mainly vacancies and interstitial atoms). There are no obvious changes in the material microstructure after the rest stage.

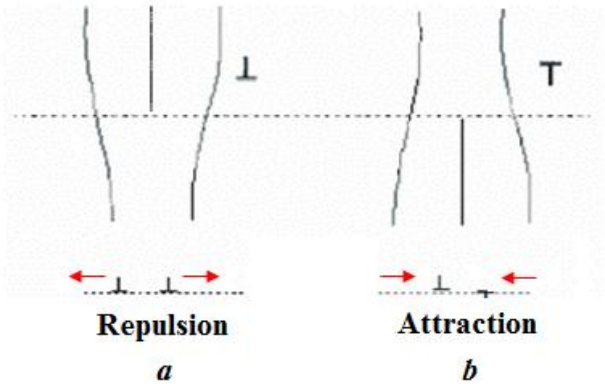


Fig. 2.6.3. Repulsion of positive (*a*) and attraction of positive and negative (*b*) dislocations during the rest of the crystal

For example, rest of pure iron occurs yet at heating up to 400°C. Reducing the density of point defects in step is rest can lead to reduction in the strength by 20-30 % compared with the initial state, while the ductility, inversely, will not increase much.

In a number of metals (aluminum, iron, etc.) dislocation climb can also occur during the recovery stage. This results in the reduction of their density due to annihilation of dislocations with opposite signs, and as a consequence an additional reduction of internal (residual) stresses. Fig. 2.6.3 shows a diagram of dislocations interaction during their climb along slip planes. As is seen, dislocations of different signs are attracted (Fig. 2.6.3a), while with opposite

signs are repel (Fig. 2.6.3b). The annihilation of dislocations leads to an noticeably increase the density of the metal.

When dislocations with one sign climb along parallel crystallographic planes, they tend to line up under each other (in neighboring planes) creating the wall normal to slip plane. If dislocations of opposite signs slide in parallel planes, they line up in the wall under the angle of 45° to the slip planes. Such rebuilding of dislocations within each crystallite (grain) is equivalent to the formation of new low-angle boundaries insude of grains. This phenomenon is called *polygonization* and new subgrains are called *polygons*. Polygonization usually (but not always) accomplishes recovery stage. Note that in some cases the above restructuring of the dislocations results in the formation of cellular dislocation structure.

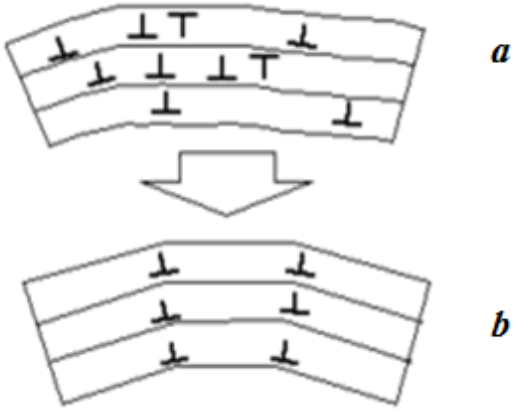


Fig. 2.6.4. The scheme of transformation of the dislocation structure due to polygonization: crystallite of deformed metal before (*a*) and after (*b*) the completion of the polygons formation

Polygonisation mechanism, shown schematically in Fig. 2.6.4b usually is realized in 3 stages. Dislocations of different signs, caused by plastic deformation (strain hardening), are usually randomly distributed in the slip planes within the grains (Fig. 2.6.4a). When heated to the temperature sufficient to realize the process of atomic self-diffusion, dislocation of opposite signs annihilate (Fig. 2.6.4b), forming the dislocation walls (Fig. 2.6.4b). As a result of polygonisation, low-angle boundaries separating subgrains (polygons) are formed, which have the form of polyhedra and practically free of dislocations.

The above described polygonization with the formation of sub-grains separated by a planar dislocation walls, is called *stabilizing polygonization*.

Since the walls between subgrains are flat, they have low mobility and are stable, so that the polygonal structure of the metal is conserved almost to the melting point. After stabilizing polygonization recrystallization is usually stopped.

In some cases, the second type of polygonisation – *prerecrystallyzed polygonization* – is selected. It occurs usually in metals with cellular dislocation structure formed during deformation at room temperature (such deformation is called *cold deformation*). In so doing, dislocation walls are tightened at the annealing (due to climbing of dislocations from the cells to the boundaries), transforming these cells into subgrains.

The curvature of the walls of these cells-subgrains can be so great, that they become highly mobile. As a result, after the formation of sub-grains (polygons) and continuation of annealing, subgrains-cells are coarsed, so that their number decreases. This will reduce the total area of the subgrain boundaries and reduce the total energy stored in the polycrystal.

Note that the process of subgrains coarsening has much in common with the process of grain growth during recrystallization (see below). For example, if structure of grains can be approximated as a system of spherical sub-grains with radii R and boundaries with the surface energy σ_s due to polygonization, the stored energy and the force acting on the boundary are uniformly distributed. In this case, we can assume that the boundary is subjected to a pressure, which contributes to the displacement of polygon boundaries. This pressure is given by

$$P = -\alpha R \frac{d}{dR}\left(\frac{\sigma_s}{R}\right), \qquad (2.6.2)$$

where α is constant. Since σ_s depends on the angle of adjacent polygons misorientation, the pressure acting on their boundary remains constant during coarsening of subgrains.

Note that sub-grains, grown on this stage, can subsequently become the centers of the primary recrystallization (see below) during the annealing at elevated temperatures (see the diagram of Fig. 2.6.2).

Polygonisation process may be impeded by limiting the dislocations mobility due to their pinning by impurities (the formation of the so-called Cottrell clouds around the dislocations), stacking faults, and also by reducing the concentration of vacancies, which provide the dislocation climb. Note that in the most of technical-pure metals polygonization is observed only after some deformation. Thus, this process is seldomly observed in copper (which has many stacking faults due to low energy of their nucleation), but it is developed well in aluminum, iron, and molybdenum (where the energy of stacking faults formation is very large).

Restructuring of the defects of the cold-deformed metal during recovery stage leads usually to the changes in the properties of polycrystalline material as a whole. In so doing, the magnitude and nature of these changes is different for the stages of rest and polygonization (see Fig. 2.6.5). So at the first stage of recovery – the rest stage – usually the excess vacancies and interstitials are annealed (annihilate). This is manifested in the existence of a low-temperature maximum in the lower curve c in Fig. 2.6.5. As is seen, this results in a significant decrease of the electrical resistivity (middle curve b in Fig. 2.6.5), but has little effect on the hardness (upper curve a in Fig. 2.6.5), tensile strength σ_B and ductilityδ in polycrystal.

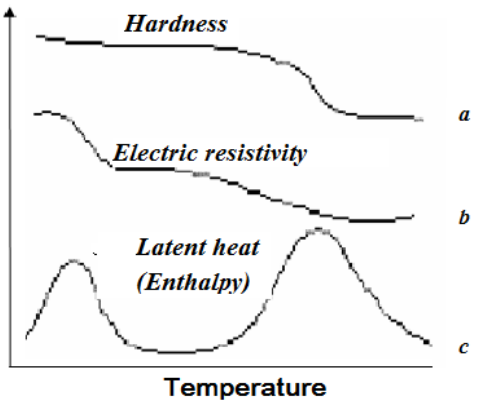


Fig. 2.6.5. Schematic view of the changes in the properties of the cold-deformed metal with increasing temperature at the recovery stage

Increasing the annealing temperature during the stage of recovery results in the development of polygonization (high-temperature maximum in curve *c* in Fig. 2.6.5), which give rise of the dislocations ordering. It seems, at this stage there is a greater decrease in hardness (curve a in Fig. 2.6.5) and also sharp drop in σ_B, compared with a weak decrease in resistivity (curve b in Fig. 2.6.5). Plasticity δ is dramatically increased at this stage.

Increasing the annealing temperature and transition to the polygonization stage in polycrystalline metals leads to an increase in the mobility of dislocations, which causes their additional annihilation and redistribution inside the crystallites. The last results in the appearance of sub-grains and their possible coarsening that determines the further reduction of the strength characteristics. Polygonization in metals enters into competition with recrystallization, so that, as noted above, after the formation of polygons, recrystallization, as a rule, does not occur.

2.6.3. Mechanisms of recrystallization in the deformed metals

To obtain the desired shape of the metal issue, or to optimize the microstructure of ingot after crystallization, plastic deformation (rolling, extrusion, forging, etc.) and subsequent thermal heat treatments are typically used. In accordance with the scheme, shown in Fig. 2.6.2, relatively low-temperature annealing (at recovery stage) changes only defect subsystem in grains of polycrystals. With a relatively high-temperature annealing of the deformed metals, another process – recrystallization – can occur in polycrystalline materials.

Let us study mechanism of nucleation and development of this process which is not only reconstruct defect subsystem (mainly dislocations, vacancies and interstitials), but also results in a significant transformation of grain structure in polycrystalline materials.

2.6.4. The impact of deformation on the stored energy and the grain structure in polycrystalline materials

To study the nature and mechanisms of changes in the dislocation and grain structure in the deformed metal at different stages of recrystallization, first remember to what energetic impacts plastic deformation of the metal is resulting.

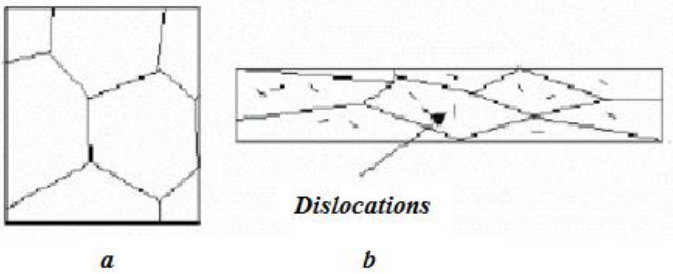


Fig. 2.6.6. Grain structure and the dislocation density in metallic ingot before (*a*) and after (*b*) cold plastic deformation

As can be seen from Fig. 2.6.6, plastic deformation of polycrystalline ingot causes changes not only structure of grains along the deformation axis, but also increase the concentration of point defects and dislocations in the grains. Increasing the concentration of defects in polycrystalline material means the increase of internal energy stored by polycrystal. The stored energy in the deformed crystal due to relative plastic deformation ε, arising under the influence of external stress σ, can be expressed by the formula

$$U_{st} \approx 0{,}05 \cdot \sigma \cdot \varepsilon. \qquad (2.6.3)$$

The value of U_{st} does not exceed 5% of the total work by deformation of the crystal. The rest part of this work is converted into heat.

Let us try to estimate the energy stored in the crystal by increasing the concentration N_D of dislocations during deformation. The dislocation density in the annealed metal equals $N_D \approx 10^{10}$ m^{-2}, whereas in the cold-deformed metal it may reach values of the order of 10^{15} m^2. As a result, the amount of the dislocation density change due to deformation can be estimated as $\Delta N_D \approx 10^5$ m^{-2}. If we know ΔN_D, we can estimate the dislocation component of the stored energy from the relation

$$U_{st} \approx \mu b^2 \Delta N_D, \qquad (2.6.4)$$

where μ is shear modulus and b – Burgers vector. Calculation by this formula gives a value of about $3 \cdot 10^6$ J/m^{-3}, which corresponds to the pressure, acting per unit area of grain boundary of the order of 3 MPa. Just this pressure activates movement of grain boundaries (increasing sizes of more perfect grains) at the deformed polycrystal annealing (see Figs. 2.6.7 and 2.6.8 below). This leads to an increase in dislocation-free grains, which is identified as *recrystallization*.

2.6.5. Nucleation of new grains during recrystallization of the deformed polycrystalline materials

Recrystallization of the cold-deformed polycrystalline material begins from the formation of centers of new grains nucleation. The nucleation of these centers is traditionally associated with inhomogeneous distribution of dislocations in the deformed polycrystals. The most of dislocations, generated due to deformation, are generally situated near or at grain boundaries and their concentration on the recovery stage remains practically unchanged. However, just these regions may stimulate recrystallization. It is generally believed that the nucleus of new grain arises on the dislocations bundles of (see Fig. 2.6.7).

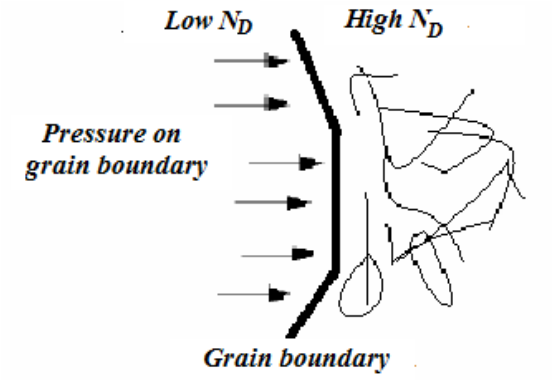
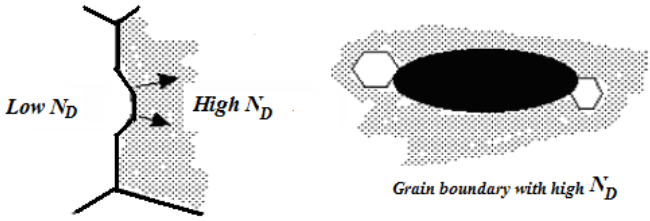


Fig. 2.6.7. The origin scheme of driving force for grain boundaries motion



a) *b*)

Fig. 2.6.8. Scheme of new grains nucleation during the recrystallization in the region with high dislocation density (*a*) and due to high internal stresses at the interface of precipitate and grain (*b*)

Nucleus of new crystallite may occur either due to bending of grain boundaries or owing to the presence of differences in the density of dislocations at the boundary and in the grain bulk (Fig. 2.6.8a). Precipitates of foreign phases in grains also effect on the nucleation of recrystallization process in polycrystalline materials (Fig. 2.6.8b).

Large particles of irregular shape are usually more effective in stimulating the recrystallization. This is attributed to the great differences in the strains around particles: the largest deformation occurs at more sharp (with smaller radii of curvature) segments of these particles. Therefore, the recrystallized grain will be substantially free of dislocations.

2.6.6. The role of grain boundaries during recrystallization of the deformed polycrystalline material

In order to understand better the conditions of new grains nucleation and growth during recrystallization, we remember some of grain boundaries characteristics, and at first, their structure and energy on the example of the simplest case of symmetric tilt boundaries. Tilt boundaries at small misorientation angles θ between neighboring grains can be present in the form of dislocation wall (Fig. 2.6.9a). Dislocation density (inter-dislocation distance *D*) in this wall is related with this the angle θ by relation

$$\theta \approx \frac{b}{D} \tag{2.6.5}$$

where *b* is Burgers vector, and *D* – the distance between neighboring dislocations. This value allows to estimate the grain boundary energy, in terms of dislocation density in it, as magnitude which is inversely proportional to the *D* value

$$U_{GB} \sim 1/D. \tag{2.6.6}$$

Dependence of grain boundary (GB) energy on the angle of neighboring grains misorientation, following from Eq. (2.6.5), is shown in Fig. 2.6.9b by solid line. However, this model of grain boundaries as dislocation walls is valid only for low-angle boundaries with $\theta < \theta_{cr} \approx 15 - 20^{o}$, which correspond to $D < b$. When larger θ angles dislocation cores becomes overlapping and dislocation model of grain boundaries stops to work.

Note also that in addition to low-angle GBs of general type with the random θ values, the so-called special grain boundaries are present in polycrystalline materials. They have the well-defined values of θ (which are determined mainly by lattice type and orientation of the GB plane relative to the crystallographic planes in neighboring grains) and are characterized by the ordered arrangement of atoms. It is at these, special, misorientation angles of neighboring grains a few maxima are observed on a real curve $U_{GB}(\theta)$, see dashed curve in Fig. 2.6.9b.

Now return to the question of driving force leading to the nucleation and growth of new grains during recrystallization of the cold-deformed polycrystalline material. Assume that new grain is nucleated when the misorientation angle between the neighboring grains with bended GB (where the highest number of dislocations is present) reaches a critical value $\theta = 10^{o} \approx 0.2$ radians (i.e. for a boundary with $D < b$), Fig. 2.6.10. Then the estimated from Eq. (2.6.5) critical density of dislocations N_{Dcr}, at which new grain is able to be nucleated (i.e., GB will be able to move), gives

$$\theta \approx \frac{b}{D} = 0,2rad \approx \frac{3 \cdot 10^{-10} m}{D}. \qquad (1.6.7)$$

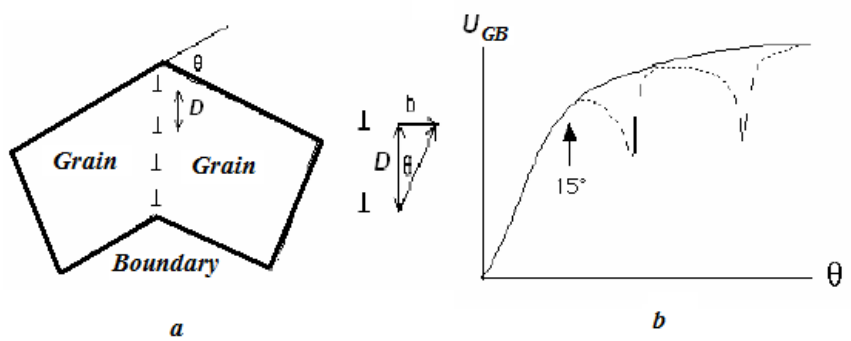


Fig. 2.6.9. The structure of symmetric tilt GB (*a*) and the GB energy UGB (*b*) on misorientation angle θ between crystal lattices of neighboring grains for the model of dislocation wall (solid curve) and taking into account the presence of special GBs (dashed curve)

Eq. (2.6.7) gives that the distance between dislocations in the wall is $D_{cr} \approx 1{,}5 \cdot 10^{-9}$ m. Choosing a typical grain size $d \approx 0.1\text{-}1$ μm for such misorientation angle, we get the following value for critical dislocation density for nucleation of a new grain

$$N_{Dcr} = \frac{1}{dD} \approx 10^{15} \div 10^{16} \text{ m}^{-2}. \qquad (2.6.8)$$

In fact, the required N_{Dcr} value will be slightly higher, as some of dislocations will be annihilated at the recovery stage. Such large density of dislocations can occur only in small, localized regions of the crystal (at the boundary region with the smallest radius of curvature). More N_{Dcr} values and thus higher rates of new grains nucleation, the smaller grains will occur after recrystallization (Fig. 2.6.10).

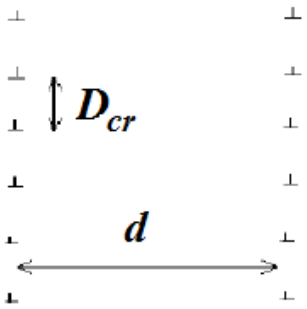
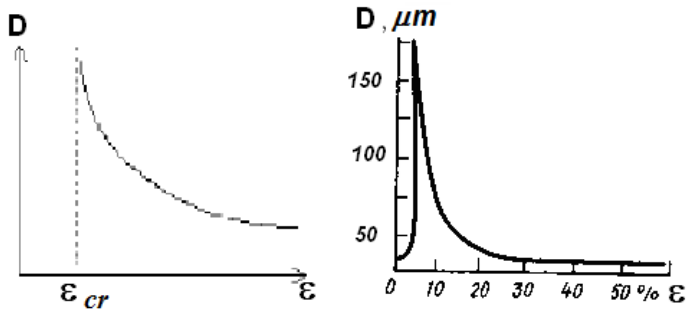


Fig. 2.6.10. The critical density of dislocations inside of grain boundary

2.6.7. Dependence of the recrystallized grain size on the strain ratio

As follows from the above mechanism of grains nucleation in the cold-deformed crystal, the need to create a critical dislocation density for the nucleation of new grains should automatically lead to the existence of a critical value of the strain ratio ε_{cr}. In Fig. 2.6.11a, this critical value of ε_{cr} is marked by vertical dashed line. According to the model, if strain ratio is below this value, recrystallization should not occur at all, because the whole effect of deformation vanishes at the stage of recovery (due to polygonization).



a) *b*)

Fig. 2.6.11. Modeling (*a*) and real (*b*) dependences of recrystallized grain sizes on degree of preliminary deformation

The real dependence of recrystallized grain sizes on the degree of deformation of the metal has a curve with a sharp peak (Fig. 2.6.11b), which sets the critical value of the strain ratio ε_{cr}. Real values of ε_{cr} for pure metals usually do not exceed ~ 2-10%. For example, for aluminum ε_{cr} is close to 2%, while for iron and copper - 5%. In the refractory metals it is even higher. Note that the critical strain can cause recrystallization even in a single crystal.

2.6.8. The kinetics of grains nucleation during recrystallization stage

Most of the theories describing the kinetics (rate) of new phase nucleation during recrystallization (as one of the first-order phase transitions), are based on the assumption that the thermal-fluctuation model for new grains nucleation is validity. According to these theories, the vibrational atomic motion (which is enhanced with increasing temperature) in domaged grains leads to the spontaneous nucleation of small nuclei of the new phase. The formation of these nuclei is associated with the performance of well-known thermodynamic (energy) condition – the balance between the work for formation of the interface between the nucleus and the existing grain and latent heat released when more perfect grain with low energy is formed. If nucleus becomes greater than some critical dimension, it is thermodynamically stable and may start to grow.

With regard to the process of recrystallization, the described model of new grains nucleation in places which are enriched by dislocations (for example, at the grain boundaries and interfaces) deal with the main problem: small contribution of the dislocations into the stored energy (U_D ~ 0.1 J/m^3) as compare to large energy density of the grain boundaries (U_{GB} ~ 0.5 J/m^3). Such low values of U_D give unreasonably low values of the rate of new grains growth as compared with the experimentally observed.

Therefore, in the present, the model proposed by Kahn is considered as more adequate. This model assumes that nuclei of new grains occur at the stage of the recovery in form of sub-grains or dislocation cells, formed during the stages of rest and polygonization. In this case, the time of recovery stage can be regarded as some "incubation period" for nucleation sub-grains or cells with low-angle boundaries, which begin to accumulate dislocations and to result to their more and more misorientation with respect their neighbors. Growth of lattice misorientation of growing grains with previously existing results in the increase of boundaries mobility so that the rate of growth of new grains is elevated. If any growing grain in the local region of the material accidentally acquires an advantage over its neighbors (e.g., due to a locally higher density of dislocations, larger size or more favorable orientation of crystal llattice), this subgrain will grow faster than its competitors. Since it is growing, its crystal lattice has become increasingly misoriented with respect to the lattice of the surrounding material (the initial grain), so that it can be identified as a new, completely free from deformation (dislocation-free) grain.

2.6.9. The kinetics of grain growth during recrystallization stage

If new grain in polycrystalline material already occurred, the boundary between the new growing grain and the adjacent material (old grains) can be considered as two-dimensional defect with a specific surface energy. As a result, as noted above, some thermodynamic driving force arises (as a negative pressure acting on the boundary from more damaged grain), which should lead to reduction of the boundaries total area in polycrystalline materials. In other words, if the size of the growing new grains increases due to decrease in the total number of (old, more damaged) grains in the polycrystalline sample, the total area of grain boundaries will decrease.

So, in the simplest case, we assume that the boundaries motion is due to tending to reduce the surface energy of all new grains in polycrystalline material. Other contributions to the driving force of the boundaries motion (associated with the elastic deformation and temperature gradients) this model will not take into account. Let us believe that the growth rate is proportional to the driving force of GB motion, which is proportional to the surface energy of all grains (i.e., the GBs area, that is proportional to the square of the grains diameter d_o). Then, it is easy to show that the time required to reach a new set of grains with diameters d is given by the same parabolic law for the diffusion length

$$d^2 - d_o^2 = kt, \tag{2.6.9}$$

where d_o and d are the initial and final dimensions (diameters) of grain, respectively. The temperature coefficient k is given by the exponential (activation) law

$$k = k_o \exp\left(-\frac{Q}{RT}\right), \tag{2.6.10}$$

where k_o – constant, T – absolute temperature, and Q – activeation energy of mobility for growing GB. Ideally, the Q value must coincide with the activation energy of atomic self-diffusion in the crystal, although this is not always true.

In general, it is known that these equations are valid only for high-purity metals. Dissolution of even very low impurities concentrations in metals leads to the disruption of these equations. This is because the energy Q, needed to activate the growth of grains of the same phase, is small. Therefore, the grain growth is very slow even in the pure metal, slowing down more and more in the presence of foreign phase particles and impurities.

Let us compare the kinetic model of grain growth during recrystallization, based on the above model of nucleation of new grains from sub-grains or dislocation cells that are nucleated at the recovery stage. For such a model, the kinetics of recrystallization, which is observed in the experiments, is shown in Fig. 2.6.12. Note, that this curve looks very similar to the kinetic curve in Fig. 2.3.5 for crystallization from the melt.

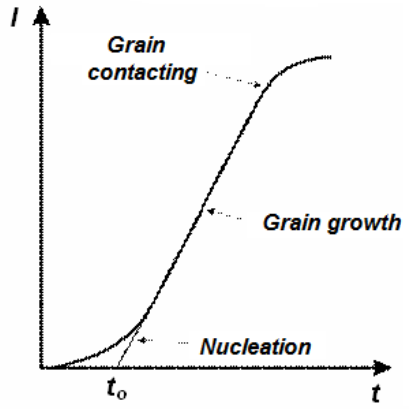


Fig. 2.6.12. Change in share I of the recrystallized polycrystalline material in time t.

As follows from Fig. 2.6.12, recrystallization process is triggered after some "nucleation" period t_o, when new grains start to nucleate, after which they will grow at a constant rate killing (absorbing) the old highly-damaged grains.

Although, as noted above, a constant rate of new grown grains (linear part of the curve in Fig. 2.6.12) does not exactly follow the classical theory of nucleation of a new phase, however, represented below mathematical description of this process provides a sufficiently good approximation.

Under the above assumptions, for $t > t_o$ grain size d will grow linearly in time

$$d = 2G(t - t_o), \tag{2.6.11}$$

where t_o is incubation period to nucleate grain, and $G = \mathrm{d}d/\mathrm{d}t$ – the rate of grain growth. If N nuclei create N grains with spherical shape during time dt, the portion of the volume occupied by new grains is presented by equation

$$I = \frac{4}{3}\pi N G^3 \int_0^t (t - t_o)dt = \frac{2\pi}{3} N G^3 t^2. \tag{2.6.12a}$$

This equation is valid only at the early stages of recrystallization, where $I << 1$ and the growing grains do not touch yet each other. When the grains come into contact and prevent to grow to each other, their growth rate will be reduced and described by Johnson-Mel equation

$$I = 1 - \exp\left[-\frac{2\pi}{3} N G^3 t^2 \right], \tag{1.6.12b}$$

which takes into account the portion (1-I) of the non-recrystallized material. The Johnson-Mel equations (2.6.12) provide a better description of the recrystallization process, although some of the assumptions (spherical grains, constant grain rates of nucleation and growth and small nucleation times t_o) are not well-grounded. Therefore, a more accurate description of the recrystallization process requires more complex models.

2.6.10. Factors affecting the grain nucleation and growth rates during recrystallization

In accordance with the above described model, the annealing temperature is one of the mostly important factors, which determine the grain growth rate during recrystallization process. However, in addition to the annealing temperature, there are several other factors that affect this rate. In particular, recrystallization rate is highly dependent on the degree of plastic deformation and, to a lesser extent, on the method and conditions of deformation. Recrystallization of heavily deformed materials will flow faster than the weakly deformed. As noted above, metallic materials deformed with strain ratios below their critical values can not be recrystallized principally.

In some cases, deformation can occur strongly homogeneously, while in some of them it is running merely in certain crystallographic planes. The lack of orientation gradients and other irregularities prevents the nucleation of stable nuclei to form new grains.

The change of grains orientation (texturing) during deformation affects the amount of stored energy, and therefore the speed subsequent recrystallization. Therefore, the mobility of grain boundaries depends on the grains orientation, so that, for example, the crystallographic texturing of the polycrystalline material will accelerate recrystallization process.

If plastic deformation is carried out at elevated temperatures (the so called hot deformation), it will lead to competition of the stored energy (due to lattice damage) and the energy loss due to the crystal defect annealing at the recovery stage. Therefore, the hot-deformed materials are recrystallized slower than cold-deformed ones (at room temperature). Recrystallization occurring during hot deformation is called *dynamic recrystallization*.

The kinetics of recrystallization strongly depends on the purity of the deformed polycrystalline material. The impurities present in metal segregate easily at the grain boundaries (in particular, on random GBs) because their structure is more friable to allow the impurities penetrating and fixing there. Segregation of impurities reduces the grain boundaries energy, i.e. increase critical pressure required to start GB motion: the formed Cottrell clouds increases the stress needed to shift GB. In other words, contamination of GBs in polycrystalline material leads to anomalous decrease in their mobility, that was called *the impurity retardation of boundaries*. For example, pure aluminum may be recrystallized even at room temperature, whereas technical aluminum is recrystallized only when heating. Incorporation of 0.004% of impurities increases recrystallization temperature of iron approximately on 100°C. However, what contribution is retarded by impurities to a greater extent – the nucleation rate or boundary mobility (i.e., the rate of grain growth).

Note that the impurity retardation of GBs is not the only mechanism of their mobility reduction. Foreign particles also retard GBs (similarly to the dislocation pinning mechanism), that is called by *Zener retardation*.

The above mentioned allows us to calculate certain parameters of grain growth during recrystallization. For spherical growing grains with diameters D have volume $(4/3)\pi(D/2)^3$ and surface area of $4\pi(D/2)^2$. Then the energy of such boundary, separating this grain from the neighboring, will be equal to $U_{GB} = 2\pi(D/2)^2\sigma_s$, where σ_s is grain boundary energy per unit area. Consequently, taking into account that the boundary belongs to two neighboring grains, the boundary energy per unit volume between the growing and old grain is equal to

$$u = \frac{3\sigma_s}{D}. \qquad (2.6.13)$$

Exactly this energy causes the forces acting on the GB with the equivalent pressure of about 0.1 MPa for typical values of σ_s = 0,3 J/m^2, and D = 10 μm. Hence it is clear that this value is sufficient to move the boundary in a pure crystal, but not too large in the presence of foreign particles, which can easily pin such boundaries (see Zener retardation).

2.6.11. Changes in the microstructure and properties of the deformed metal at the recrystallization stage

Consider what changes in the dislocation and grain structure occuring at different stages of the deformed polycrystalline materials recrystallization due to the above described mechanisms of the atomic rearrangements. Thus, plastically deformed crystals can be recrystallized only after some critical strain ratio ε_{cr}, and after approaching a certain annealing temperature called by *the recrystallization temperature* T_{recr}, which is the least heat temperature supplying nucleation of new grains at the given (critical) dislocation density. Temperature of recrystallization is usually characterized by a relative share a of the melting point of the metal (in absolute Kelvin scale)

$$T_{recr} = aT_m, \qquad (2.6.14)$$

where the a value depends on the degree of metal purity and plastic deformation ε. For technical purity metals $a \approx$ 0.3-0.4 usually decreasing with ε increase. Reducing the impurities content may reduce it to 0.1-0.2. The temperatures of recrystallization for some metals of technical grade are shown in Table 2.6.1. These values of T_{recr} correspond to the temperatures at which the strongly deformed metal is fully recrystallized for 1 hour.

Table 2.6.1. Recrystallization temperatures for some metals

Metal	T_{recr}, K
Copper (99.99 %)	120-125
Aluminium (99.99 %)	80-100
Iron (99.99)	450
Zink	20
Tin	20

The sequence of the grains transformations during the annealing of the deformed polycrystalline sample at $T \geq T_{recr}$ is shown in Fig. 2.6.14. According

to this scheme, 3 stages of recrystallization are distinguished depending on temperature and time of annealing – primary, collective (coalescence) and secondary recrystallization.

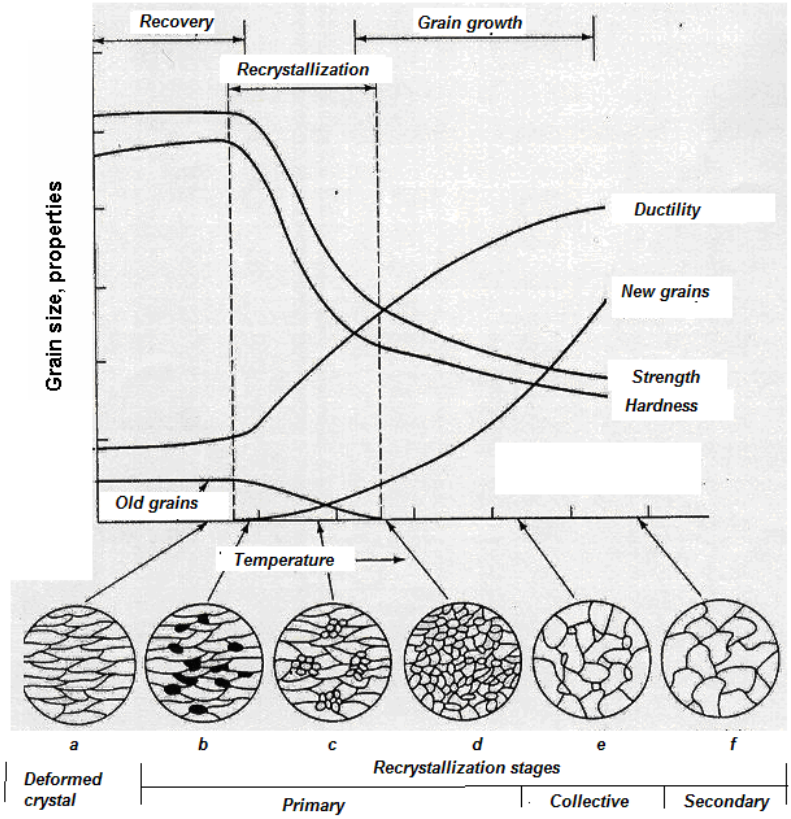


Fig. 2.6.14. Diagram showing the change of properties (hardness, strength, ductility, and size of the old and new grains) and the microstructure of the deformed metal during recrystallization annealing: *a* – the initial structure of cold-deformed metal (after completion of recovery stage), *b* – the beginning of primary recrystallization stage, *c* – the development of primary recrystallization stage, *d* – end of primary recrystallization stage, *e* – grain growth at the stage of collective recrystallization, *f* – the formation of equilibrium structure after secondary recrystallization stage

As noted above, nucleation of new grains during recrystallization occurs in the regions with the highest dislocations density, usually at the grain boundaries of the deformed elongated grains (Fig. 2.6.14a) after completion of recovery stage (Fig. 2.6.14b). The higher degree of plastic deformation (higher

dislocation density), the higher density of nucleated primary recrystallization centers. These centers last arise due to redistribution and partial annihilation of grain boundary dislocations, representing a submicron regions with minimal number of defects in the crystal lattice. In this case, between the growing recrystallization center and the deformed grain of polycrystalline metal high angle boundaries appear surely.

In the course of recrystallization annealing time, the formed centers for nucleation of new grains are increased in size due to migration of atoms over boundary between the deformed environment and perfect lattice of new growing grain. As a result the boundary of new grain is moved into the depth of the deformed part of metal (Fig. 2.6.14c). Considered stage of *the primary recrystallization* is completed by a total replacement of the deformed initial grains by more or less equiaxial grains throughout the bulk of the deformed polycrystalline metal (Fig. 2.6.14d).

Secondary recrystallization (Fig. 2.6.14e), developing at higher temperature after completion of the collective recrystallization stage, results in inhomogeneous grain growth (grain sizes can differ many times).

Recrystallization is usually completely removes the hardening created by plastic deformation, and polycrystals acquires more equilibrium structure with minimal density of defects in crystal. This is easily seen from the analysis of changes in hardness, strength and ductility during annealing of the deformed polycrystalline metal schematically shown in Fig. 2.6.14. As can be seen, redistribution of dislocations due to polygonisation at the recovery stage leads to a slight increase of hardness, strength remains almost unchanged and flexibility increases. Nucleation (Fig. 2.6.14b) and the growth in the size (Fig. 2.3.14c) of new grains at the stage of primary recrystallization leads to a drastic change in the mechanical properties – reduction of hardness and strength with an increase in ductility. The increase in grain size also leads to decrease in the electrical resistivity ρ due to decrease in the square of grain boundaries, which scatter electrons.

Stage of *collective recrystallization*, which corresponds the increase in the size of new grains in Fig. 2.6.14d, causes further reduction of hardness and strength, and retardation (and sometimes even decrease) of the ductility growth. The latter results in the deterioration of sliding (due to mismatch of slip systems in neighboring grains) owing to severe misorientation of the enlarged grains. At the stage of secondary recrystallization strength and hardness of polycrystalline metal is somewhat reduced, while the ductility reaches saturation. In general, the properties of metal after recrystallization becomes close to the properties of the metal, manufactured in equilibrium conditions.

2.6.12. Polymorphic transitions in crystalline materials

Some pure elements of the periodic table can be in different crystal forms or modifications depending on temperature. Such elements are called polymorphic.

Polymorphism is a property of certain pure elements to have more than one type of crystalline lattice in solid state when conditions change. Therefore, the heating or cooling of the material will change the type of its structure if crystal possess by polymorphism. If this structure change is reversible, the polymorphic transformation is known as *allotropy*. Polymorphic transformation occurs when at lower temperature crystal lattice of metal has a lower free energy.

Among the elements of the periodic table polymorphism attends in carbon, iron, zirconium, titanium, hafnium, tin and others. For example, the carbon may have a crystalline structure of graphite or diamond.

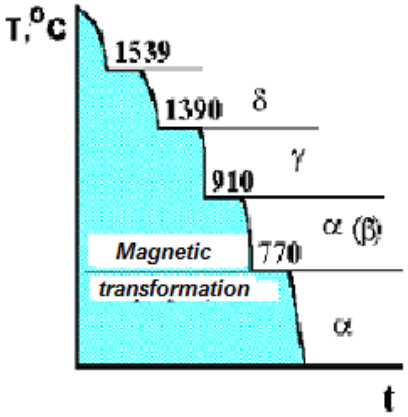


Fig. 2.6.17. The cooling curve of pure iron

The most famous example of allotropic element is iron. As can be seen from the cooling curve of iron (Fig. 2.6.17), when passing through the crystallization temperature of 1539°C during the solidification, iron acquires the *bcc* structure, which is called *δ-iron* or *high-temperature ferrite*. When the temperature decrease to 1390°C δ-iron is transformed to γ-iron with the *fcc* structure, which is called *austenite*. With further cooling γ-iron and the transition of the temperature 910°C it turns again into *bcc* structure called *α-iron* or *ferrite*. As can be seen from Fig. 2.6.17, iron cooling curve has yet one step at 770°C, indicating the existence of phase transition, which is associated with the transformation of a nonmagnetic γ-iron into magnetic α-iron.

Polymorphic transformation, as any the 1st-order structural phase transition, is carried out by the nucleation of a new phase and its subsequent growth.

Therefore, mechanism of polymorphic transformation is very close to recrystallization.

The formation of nuclei at polymorphic transformation is going in compliance with the principle of structural and dimensional conformity. As with conventional recrystallization, new phase nuclei occurs at the grain boundaries of the old phase or in regions with a high level of stored free energy (on dislocation bundles). Growth of new grains of polymorphic phase occurs by random, mutually non-bound transitions of individual atoms (or atomic groups) across the interface between new and old phases. The newly formed crystallites of a new phase are regularly oriented with respect to the initial crystalline modification. As a result of polymorphic transformation, new grains with different size and shape are formed, which leads to an abrupt change in material properties.

2.7. Equilibrium diagrams and phase transformations in two-component crystalline materials

Pure elements (including metals) have sufficiently limited application in technics. Among crystalline materials, alloys are widely used as engineering structural materials in machines and mechanisms, in instruments and for other purposes. Alloys are also used in power technologies, as well as for energy savings (boilers, turbines, pipes, walling, building envelopes, etc.). This is due to the fact that very often only alloys may give the desired combination of functional characteristics, which allow to operate one or other issues in a wide variety of conditions. Alloys play a special role, when material must be characterized by mutually exclusive functional characteristics, for example, high strength and ductility simultaneously; low electrical conductivity, but a high thermal conductivity; high strength at high temperatures and in corrosive environments, etc.

Very often crystalline structures in alloys, that have important beneficial properties, are thermodynamically non-equilibrium, and it is important to be clear about the conditions of their phase stability. Therefore, this Section will describe the main types of alloys and the mostly important regularities of the formation of their structure, its stability and changes under thermal and other impacts. General similarities and differences in the behavior of alloys and their constituents at various phase transitions will be also discussed. In particular, considerable attention will be paid to the formation of the alloys structure during crystallization under different conditions, including their chemical composition, cooling rate, the interaction between components and phase stability. Phase transformations in the solid state, including changes in the structure and properties during plastic deformation and mechanisms

strengthening of metals and alloys are considered in detail, especially depending on the regimes of thermal treatments. The processes that occur during the heating of the deformed metals and alloys causing significant reconstruction of its structure and properties modification are also expounded.

2.7.1. Types of crystalline alloys

As is known, crystalline materials may consist of one kind or of several types of atoms. In the latter case we have complex compositions in form of alloys and chemical compounds, composites, ceramics, etc. Modern power industry and energy saving use mostly multi-component materials, like alloys or more complicated compositions. So their atomic, nano, micro and macro structures can gain strong differences from those that are implemented in simple single-component materials.

Let us consider the main features of major types of alloys. *Alloy* is a substance which is obtained by fusing of several chemical elements, which are called *the alloy components*. Components can join together to interact, forming chemical compounds and solid solutions or to remain as pure components to form their mechanical mixture.

The alloy parts, which have the same chemical composition, crystal structure and properties and are separated by internal boundaries (interfaces) between different parts of the alloy with different composition, structure and properties, are called *phases*. The microstructure of alloys is revealed by *metallographic analysis* using optical or electron microscopy, X-ray difraction and microspectral analyses, and other methods. The main elements of the microstructure in metals and metallic alloys are crystallites (grains) and phases (with their specific distribution by shape and sizes), intergarain boundaries and interfaces, dislocations and their boundles, stacking faults, nano and micro cracks and pores, concentration inhomogeneities at the boundaries and in the bulk of grains (micro and nanoregions with segregated elements), etc.

If polycrystalline alloy contains crystalline grains with the same phase only, it is called *single-phase alloy*. As shown in Chapter 1, pure monocrystalline or polycrystalline metal or semiconductor are single-phase materials. As will be shown below, the so-called solid solutions or chemical compounds are also single-phase alloys.

The alloy, comprising two or more phases, is called *heterogeneous* or *multiphase alloy*. The most of steels, brasses, bronzes, duralumin and other structural alloys are related to multiphase alloys. For example, high carbon steel after heating, crystallization and slow cooling consists of two phases: almost pure iron and compound Fe_3C (cementite).

The term "alloy" now has a broader meaning than at the time of its occurrence, when they were obtained predominantly by alloying. Now much of materials are obtained by other technological methods, like powder metallurgy method (compaction of powders with subsequent sintering at high temperatures), as well as the diffusion method (penetration of one substance into another in solid state at high temperatures). Alloys can be also prepared by ion-beam, laser and plasma sputtering of compound targets, crystallization from vapor in vacuum or gas mixtures, chemical or electrochemical deposition, etc. Components of alloys may interact differently with each other, forming crystalline phases with different chemical composition, type of interatomic bonds, atomic structure and nano, micro, macro structure.

In the liquid state, the majority of metal alloys, used in technology and engineering, are homogeneous liquid solutions whose structure near the crystallization temperature tends to establishment of short-range order (perfect arrangement of atoms within a certain distances, Fig. 2.7.1a). If the alloy melt is transformed into solid state, conserving its uniformity on atomic level and forming long-range order (crystal lattice), such solid phase is called *the solid solution*.

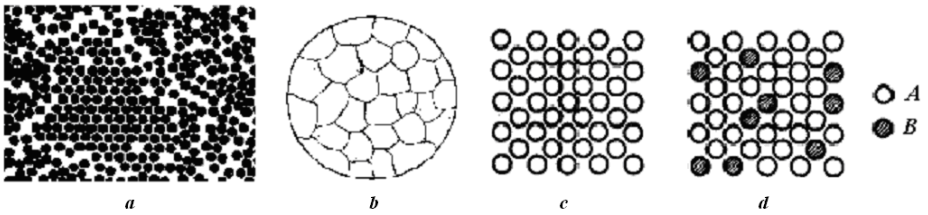


Fig. 2.7.1. Schematic representation of the atomic structure of liquid melt (*a*), microstructure of solid solution of alloy A-B (*b*), schemes of atomic arrangements in the initial lattice of the component-solvent A (*c*) and randomly arranged atoms of the solved B atoms (*d*) in a binary substitutional solid solution *A-B*

Chemical and X-ray analysis of solid solutions indicate the presence of two or more elements in them. At the same time, metallographic analysis display homogeneous grains of the same composition. In other words, the microstructure of the solid-solution alloy in the equilibrium or quenched state looks like pure metals in microscope (Fig. 2.7.1b).

In *chemical compounds*, a new type of crystalline lattice is formed after alloying, which differs from lattices of every component. Chemical compounds are essentially the same solid solutions, but with the fixed composition and different structure. Therefore, alloys and chemical compounds appear under the microscope as solid solutions and pure metals.

Chemical compounds are formed between the metals (they are called *the intermetallic compounds*), and also between metals and nonmetals. They have the following main features:

1) constant composition, which is reflected by stoichiometric formula A_mB_n;
2) different type of crystal lattice than components have. For example, steel and cast iron have one common phases – compound Fe_3C (cementite). In so doing, iron has either *bcc* or *fcc* cubic lattices, carbon – hexagonal, while cementite – a complex rhombic lattice.

Alloys can form not only solid solutions or chemical compounds (as well as intermediate phases), but also the so-called *alloy-mixtures*. Two-component (binary) alloy-mixtures present mechanical mixture of the two phases. If the A-B alloy components are completely insoluble in each other, the alloy comprises a mixture of the pure components A+B (phase A and phase B). If the components are partially soluble in each other and form α – and β-solid solutions (based on components A and B, respectively), the alloy is a mixture of two phases $\alpha+\beta$ (see below in more detail). Let us consider briefly the main types of alloys in terms of their atomic and phase structure.

Solid solutions. Solid solutions are the basis of most industrial constructional and special alloys, which are used in power engineering, power industry and for energy saving purposes.

It should be emphasized that solid-solution crystals are the mostly similar to the element-solvent because retain its crystal lattice and type of interatomic bonds. Depending on the arrangement of atoms in the crystal lattice of element-solvent, the following types of solid solutions are differed: *substitutional, interstitial and subtractional solid solutions.*

Binary substitutional solid solutions of AB type are formed as a result of the substitution of solvent atoms (for example, of class A) in the crystal lattice on the atoms of the other substance without changing the crystal lattice (Fig. 2.7.1d). of If any number of solvent atoms (type A) may be replaced by atoms of type B, we say about *unlimited solubility*. Therefore, if the increase in the concentration of atoms, then more and more of sites A will be replaced by B atoms in the lattice, so that at last we make transition from metal A to metal B.

Unlimited solubility of the components is possible provided that both metals have the same crystalline structure. Consequently, an important condition for the formation of a continuous series of solid solutions is isomorphism of components A and B.

When two chemical elements with the same crystalline lattice have very different atomic radii, the formation of solid solutions between these metals greatly distorts the lattice, resulting in the accumulation of elastic strain energy in the crystal. When this energy reaches a certain value, crystal lattice becomes

unstable, so there comes the so-called *solubility limit* of the second component in the lattice-solvent. As will be seen later, when concentration of solute element exceeds this limit, a two-phase alloy arises. Thus, the second condition for the formation of unlimited solid solutions is a relatively small difference in atomic size of allying atoms. Finally, as was observed experimentally, the unlimited solubility arises mainly for elements that are very close in the periodic table, i.e. with the close structure of the valence electrons shells.

If the crystal lattices of alloying components are different (are located in different groups of the periodic table), they are often disposed to the formation of chemical compounds, not solid solutions. Thus, to form continuous (unlimited) solid solutions, the alloyed components must meet the following conditions:

1. Elements or their polymorph modifications must have the same crystalline lattices (isomorphic to each other).
2. The atomic radii of the elements should differ by no more than 10-15 %. This condition is called *the size factor* or *the Hume-Rothery rule*.
3. Elements must belong to the same group (or sub-group), or adjacent (related) groups of the periodic system, that determines their close *electronegativity* (*electron affinity*).

Fulfillment of the above conditions for the formation of continuous solid solutions is a necessary, but not sufficient, as evidenced by data concerning the formation of solid solutions in real alloys. Therefore, these conditions are often supplemented by the need to take into account the proximity of melting temperatures, elastic constants, etc. of the components

An example of alloys with unlimited solubility in the solid state are alloys of copper with gold, silver or nickel, germanium-silicon alloys, etc. In the polymorphic metals unlimited solubility occurs within one modification of the space lattice. For example, α-Fe has continuous series of solid solutions with chromium (as both elements have *bcc* lattice), and γ-Fe – with nickel (*fcc* lattice).

If both elements do not meet the above condition 2, then they can only be partially soluble in each other. As noted above, the greater the difference in size of the atoms in the other properties of the alloying components, the lower the solubility. The limit of solubility decreases dramatically when condition 1 is disturbed, and fused components have different types of crystalline structure. Solubility limit also decreases with decreasing temperature in most cases. If the elements are of different groups of the periodic system (so that condition 3 is not fulfilled), the interaction between atoms can lead to the appearance of chemical compounds or intermediate phases.

As mentioned above, dilute of atoms in substitutional solid solutions cause local distortions of the crystal lattice, which lead to the change in the average period of the crystal lattice. The lattice periods vary depending on the difference between the atomic diameters of the dissolved element d_A and solvent d_B. If d_A-$d_B = \Delta d > 0$ (dissolved element larger than solvent atom), the lattice unit cell (and its period a) increases. If Δd d < 0, the lattice parameter a decreases with the increasing of atoms B concentration. In some solid solutions this change is proportional to the concentration x_B of the dissolved atoms

$$a_{AB} = a_A + (a_B - a_A) \cdot x_B/100, \qquad (2.7.1)$$

expressed in atomic percentages. Linear growth of the lattice period according to (2.1.1), observed by the addition to the solvent of the second element, is called *Vegard's rule*.

The crystal lattice distortion during formation of solid solutions cause also the changes in their properties with x_B growth. Typically, the increase of hardness, strength, electrical resistivity, and conversely, decrease of ductility, temperature coefficient of resistivity is observed when x_B increasing. However, examples are known when addition of impurities results in ductility increase, such as in α-brass in Cu-Zn alloys. In general, irrespective of the alloying elements A and B, the relative strengthening of solid solution is proportion to the relative change of the lattice parameters $(a_B - a_A)/100$, and lattice compression (when $(a_B - a_A) < 0$, i.e., the interatomic distance decreases) leads to a greater hardening than its extension (when $(a_B - a_A) > 0$).

Thus, ideally, substitutional solid solutions are macroscopically homogeneous phases, in which the atoms-components in any local volume are distributed randomly (Fig. 2.7.1d). However, in real alloys at the atomic scale, this uniform distribution of the components may be disrupted resulting in the presence of separate *crowds* (*clusters*) of the dissolved atoms.

Subtractional solid solutions are formed as a result of incomplete (unlike substitutional solid solutions) substitution of the solvent atoms in the lattice by the dissolved atoms, when part of the lattice sites in solvent is unoccupied by atoms (i.e. vacancies are formed). Subtractional solid solutions are usually formed when chemical compounds act as components. Iron oxides present an example of such solid solution alloy: FeO (it is called *wustite*) has a fcc lattice with Fe^{2+} vacancies due to deficiency of iron atoms (see Fig. 2.7.2).

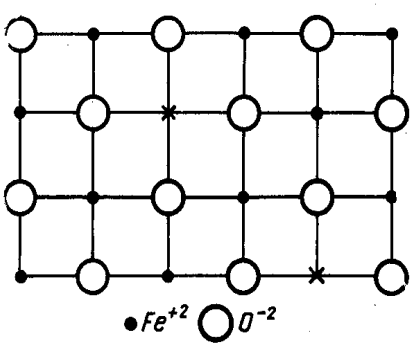
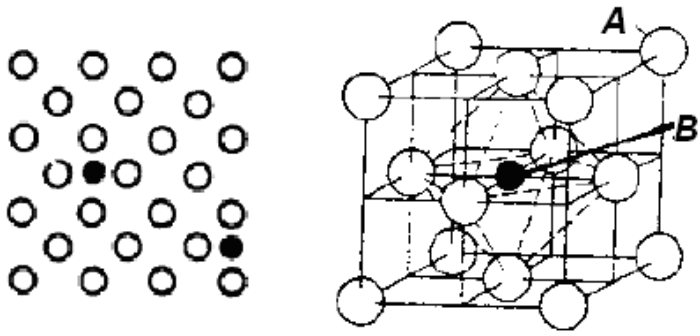


Fig. 2.7.2. Arrangement of atoms in subtractional solid solutions on the example of wustite FeO with iron deficiency: vacancies are indicated by crosses

As you can see, all the sites of oxygen sublattice are filled, while some of sites in the metallic sublattice (Fe atoms) are free. Thus, FeO oxide has a structure with a large deficit of metallic ions, which defines, for example, the presence of semiconductor properties in FeO oxide.

Interstitial solid solutions typically occur when alloying of large ionic radii transition or rare-earth metals (Fe, Mn, Co, Ti, Mo, W, V, Zr, etc.) and nonmetallic elements with much smaller atomic radii (N, N, S, V). In this case, the non-metal atoms are introduced into the interstices in the crystal lattice of solvent (Fig. 2.7.3).



a) *b*)

Fig. 2.7.3. Two-dimensional (*a*) and three-dimensional (*b*) schematic images of interstitial solid solutions with a disordered arrangement of *B* atoms in the interstices of the crystal lattice of solvent *A*

Interstitial solid solutions are always limited. The main condition to form solid solution by interstitial mechanism is the size factor (R_B/R_A). According to the closest ball package model (see Part 1 of this e-Book), the size R_B of interstitial atom *B* must be less, equal to or only slightly larger than the interstices (pores) in the lattice of atoms *A*, answering the relationship

$$\frac{R_B}{R_A} < 0{,}59.$$

Interstitial solid solutions occur mainly when the solvent has the closest lattice, like *hcp* or *fcc*, which contain octahedral pores which size is about 0.41*R*, where *R* is radius of solvent atoms. In the *bcc* lattice solubility of atoms dissolved in interstitials is smaller, since the pore size is not greater than 0.29*R*.

Interstitial solid solutions with the great technical application are presented, for example, by solid solutions of carbon in the α-Fe (ferrite) and γ-Fe (austenite): *fcc* austenite solves up to 2.14 wt.% of carbon, whereas the *bcc* ferrite hardly dissolves carbon because its maximum solubility does not exceed 0.02 wt.%.

Distortions of the crystal lattice, which accompany the formation of interstitial solid solutions, are much higher than during formation of substitutional solid solutions. Therefore, the lattice parameter of the solvent crystal lattice is always increases at alloying, and therefore the properties of interstitial solid solutions are changed more rapidly. For example, increasing the concentration of the dissolved element significantly increases electrical resistance, the coercive force, hardness and strength but significantly decreases the ductility and toughness in the interstitial solid solutions. In alloys, containing more than two elements, may dissolve the second and third elements in the same solvent by substitutional and interstitial mechanisms. So, being alloyed with manganese and carbon, iron originally forms substitutional solid solution with manganese and interstitial solid solution with carbon.

Chemical compounds and intermediate phases. As stated above, the crystalline material formed by the elements from two different groups in the periodic table with different types of crystalline lattices can form *chemical compound* with another type of crystal lattice. Any binary chemical compound is characterized by well-defined ratio of its constituent atoms (components), which is written down by stoichiometric formula A_nB_m, wherein *n* and *m* are expressed as integers. As noted above, chemical compound may act as the alloy component. In the latter case, when it dissolves one or both of its constituent components, the stoichiometry (exact atomic ratio) of the compound is disturbed and the resultant solid solution is called *an intermediate phase*. Thus, there is evidence of an intermediate phase has both features of solid solutions and the chemical compound ($A_mB_n + A$ or $A_mB_n + B$).

Depending on the nature of the elements, the chemical compound or intermediate phase may have any chemical bonds, which primarily determine the crystals properties. Type of crystal lattice in the compound always differs from the lattices of the components, its melting point is always constant and the properties are very different from the properties of the components.

Alloys-mixtures. If the alloyed components (e.g., *A* and *B*) are insoluble in each other, or do not meet the Hume-Rothery regulations and preparation conditions to form chemical compounds or their intermediate phases, they may form the so-called *alloys-mixtures* (of *A*+*B* type in the case of insoluble components). As the number of elements in the periodic table forming a continuous series of solid solutions is small, alloys-mixtures are the largest group of engineering alloys. Most of the technically important alloys such as steel, cast iron, brass, bronze and others belong to this group. For example, according to the phase composition, alloys of Pb-Sb system consisting of absolutely insoluble in each other components refer to the alloy of (*A*+*B*) type, and brasses (copper-zinc alloys) containing more than 39 at. % Zn belong to the alloys of (α+β) type, wherein α- and β-phases are limited solid solutions of zinc in copper and copper in zinc, respectively. Steels and cast irons are alloys-mixtures of $\alpha + A_mB_n$ type, where α – is solid solution of carbon in iron and A_mB_n – is a chemical compound Fe_3C (cementite).

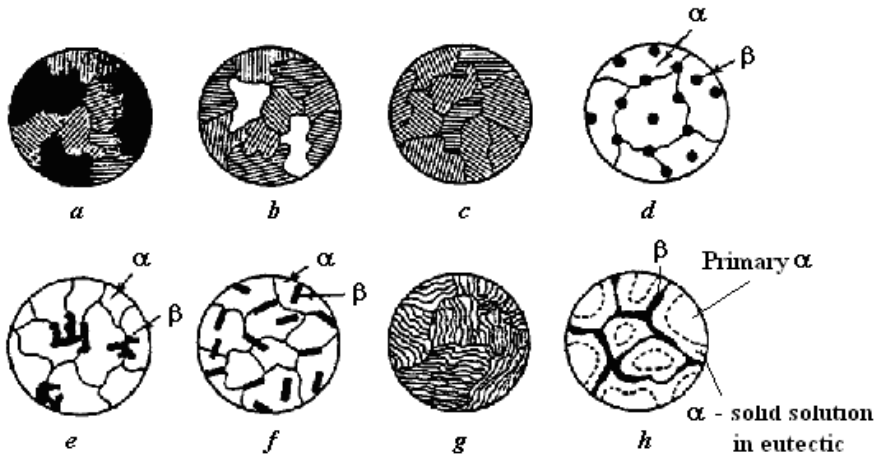


Fig. 2.7.4. Schematic representation of possible microstructures in eutectic α+β alloys: *a* – small content of β-phase, *b* – roughly equal contents of α- and β-phases, *c* – small concentration of α-phase, *d* – spheroidal structure, *e* – type structure of Chinese characters, *f* – needle-like structure, *g* – lamellar (platelet) structure, *h* – discontinuous structure

Alloys-mixtures are obtained in different ways. If the alloy is a mixture obtained by crystallization from the liquid phase (melt), it is called *eutectic*. If the mixture of different phase is obtained due to phase transformation in the solid state (below the crystallization temperature), the resulting alloy is called *eutectoid*.

The third way to form alloys-mixtures is possible. For example, during formation of alloys-mixtures in the liquid-phase, one phase (with higher crystallization temperature) may initially crystallize, and then the formed crystalline phase can interact with the rest melt, crystallizing into another solid double-phase system. In this case, the solid mixture formed after complete crystallization is called *peritectics*. Schematic representation of microstructures in various types of alloys-mixtures under microscope (after processing of the metallographic plane section by corresponding chemical etchants) is shown in Fig. 2.7.4.

Alloy in Fig. 2.7.4a consists of grains of α-solid solution (dark fields) surrounded by eutectic (α + β), often having a lamellar microstructure. Alloy in Fig. 2.7.4c includes primary crystallites of β-phase as solid solution of *A* component in *B* component (bright fields) surrounded by eutectic. Structure of eutectic alloy itself in Fig. 2.7.4b (as in any of the above alloys in Fig. 2.7.4a and 2.7.4c) presents a mechanical mixture of lamellar particles of α and β phases.

Most often, microstructure of eutectic alloys has lamellar (platelet) or needle form (Figs. 2.7.4 f, g). Sometimes eutectic microstructure of so-called discontinuous type is detected (Fig. 2.7.4h) in which there is no clear boundary between the primary α- or β-phase and part of the same phase in eutectic mixture. In other words, there is a tendency to disintegration (separation) of these two phases in a polycrystalline ingots.

2.7.2. Phase equilibrium in alloys

As noted above, system of alloys can consist of one or several phases, i.e. to be homogeneous or heterogeneous. Depending on temperature, pressure, and concentration of components in different phases, the system can be in different states, either liquid or solid. Phase transitions take place in alloys during their transition from one state to another. The results of such transitions are new phases formation and (or) old phases disappearance.

The equilibrium state of the alloy as a thermodynamic system is characterized by time-constant values of temperature, pressure and concentration of components in phases which are the same at all points. Properties of alloy depend on many factors but, first of all, they determined by composition of phases and their proportions. Therefore, study of equilibrium

phase diagrams is important for analysis of structure, properties and transformations in alloys. As will be seen below, these data on composition of phases and their proportions can be obtained from an analysis of these diagrams.

Phase diagrams of alloys and their construction. *Equilibrium phase diagrams* (or *phase state diagrams* or *phase diagrams*) of alloy is a graphical representation of all possible phase states for it as a function of main state parameters (temperature *T*, pressure *p* and chemical composition *x*). Each real state of system is presented on phase diagrams by so-called *figurative point* the position of which is determined by the values of coordinates (*T*, *p*, *x*). The areas of three-dimensional space (in three-dimensional state diagrams) or planes (in two-dimensional diagram) correspond to the regions of one phase existence. Phase coexistence conditions are determined by surfaces or lines of phase separation. Change of phase state of the system with changing the state parameters of alloy is considered as a movement of figurative point on phase diagram.

For solid materials (pressure is commonly close to atmospheric), only temperature *T* and ratio of incoming to the system components *x* (composition) are considered as state parameters (coordinates of phase diagram). In this case, on the phase diagrams of two-component (binary) systems, there are lines of liquid-solid phase equilibrium (so-called liquidus and solidus lines), equilibrium lines between different solid phases as well as special points where three phases can coexist on *T*-*x* plane. As a result, state diagram graphically illustrates the regions of stable states (component concentration and temperature) for which equilibrium phases exist under these conditions. For more complex systems consisting of more than two components and three phases, construction of phase diagrams is the only method that can help to determine, in practice, how many and which specific phases form a system for given values of the state parameters.

Analysis of the relative position of bulk parts, surfaces, lines and points on state diagram allows to unambiguously and clearly define the conditions of phase equilibrium, the appearance of new phases and chemical compounds in the system, formation and decomposition of liquid and solid solutions, etc. State diagrams are used in materials science, metallurgy, chemical technologies (in particular, for development of methods of substances separation), electronics and microelectronics, etc. They can help to determine the direction of the processes associated with phase transformations, to choose the technology of alloys processing (e.g., modes of thermal and mechanical impact), to find the optimal alloy compositions, etc. If state diagram is known, complete information about any alloy formation can be provided, optimal temperature of

its preparation can be determined and, as will be seen below, it is possible to make a conclusion about its properties.

For phase diagrams plotting, it is necessary to know from calculation the dependence of thermodynamic (chemical) potentials of all components of system on T, p and x. Approximate methods of calculation with usage of computer modeling techniques are intensively developed, in particular, for multi-component alloys. However, nowadays, phase diagrams are plotted on the basis of experimental data which allow determining the dependence of melting or crystallization temperatures on the composition, as well as by study of liquid – vapor and liquid – liquid equilibrium. X-ray phase analysis, data on the microstructure of solidified melts, measurement of physical properties and plotting the composition-property diagrams are widely used.

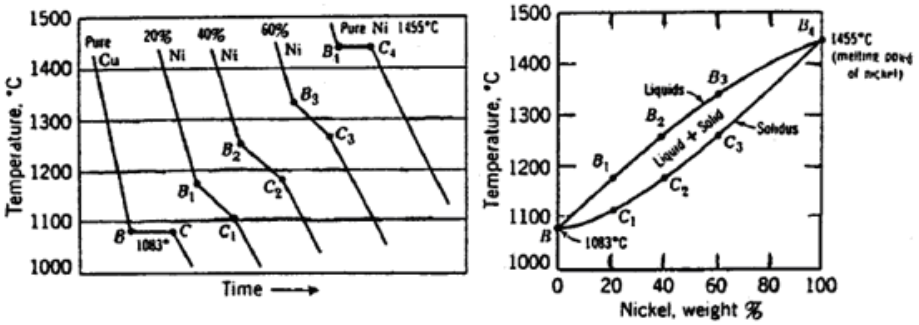
Experimental construction of state diagram is based on the developed method of thermal analysis, which was discussed in Chapter 1. This method is based on the fact that any phase transformation (not only for clean components, but also for alloys) is accompanied by thermal effect (release of the latent heat of phase transformation), as well as changes in physical properties (microhardness, electrical resistivity, specific volumes, thermopower, coefficient of expansion, etc.). In this case, peculiarities of the heating or cooling curves in the form of kinks or plateau (see below) are used to construct phase equilibrium lines that allow analyzing and classifying system of alloys in a wide range of their components ratios.For example, temperature change during heating or cooling of a set of alloys with various compositions as a function of time is usually registered for the determination of critical temperatures at which crystallization or melting of alloys with various compositions occurs. In this case, the line of liquid – crystal phase equilibrium can be constructed basing on kinks or plateau on the cooling curves (see Chapter 1).

To detect phase transformations in solid state (e.g., due to polymorphism of components or formation of chemical compounds), any physical value which significantly changes because of phase transitions (relating to electric, magnetic, mechanical or other property) is commonly observed. This is due to the fact that these types of phase transformations are accompanied by very small thermal effects, in contrast to the liquid-solid transition.

As already mentioned above, state diagram of binary alloys such as *A*-*B* are plotted in a temperature-composition plane coordinate system. Temperature is plotted on the Y-axis whereas chemical composition, such as concentration x_B of *B* component in *A*, is plotted on the X-axis. Then, any figurative point (T, x_B) of state diagram indicates temperature and composition of alloy, vertical line x_B = const shows temperature change of alloy with a certain composition, and horizontal line T = const shows composition changes at a certain temperature.

For thermal analysis and experimental construction of state diagram of alloys formed by two *A* and *B* components, series of alloys with different component concentration should be prepared. This set of alloys is called *alloys system*. Then, after melting of each alloy, its cooling or heating curve is recorded (the principle is the same as for definition of melting or crystallization temperature of pure elements, see Chapter 1), for which thermal effects or properties changes in the process of crystallization or melting are registered.

Heating or cooling curves allow determining experimentally the critical point (temperature) at which phase transformations for each alloy composition occur. Obtained values of critical temperatures are plotted on the vertical lines corresponding to chemical composition of the alloy. If we connect critical points for alloys with studied compositions with lines, the line of phase equilibrium can be obtained. Combination of these lines gives state diagram. An example of such a construction of state diagrams for solid solutions (for example, Cu-Ni alloys) is shown in Fig. 2.7.5.



a) *b*)

Fig. 2.7.5. Example of binary equilibrium diagram of Cu-Ni: *a* – cooling curves of alloys with different compositions, *b* – line of phase equilibrium

Since the position of phase equilibrium lines on the diagram depends on the cooling rate of alloys, temperatures of the critical points during phase diagrams construction are determined at a sufficiently slow cooling or heating processes. These diagrams are close to equilibrium.

Let us consider in more detail the construction of phase diagram similar to that shown in Fig. 2.7.5. This type of diagram is called a *binary phase diagram of alloys with unlimited solubility of the components in liquid and solid states* (see Fig. 2.7.6).

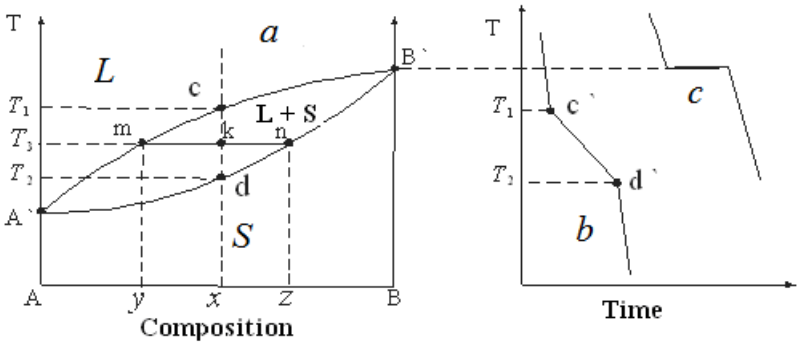


Fig. 2.7.6. Scheme of construction of binary equilibrium diagram for alloys with unlimited solubility of the component in liquid and solid states: *a* – state diagram, *b* – cooling curve of the alloy, *c* – cooling curve of *B* component

Constant crystallization (melting) temperatures of components *A* and *B* detected by the position of plateau on the cooling or heating curves determine the position of the extreme points *A`* and *B`* on the phase diagram (Fig. 2.7.6c). As can be seen from Fig. 2.7.6b, the cooling curves of alloys (in contrast to those for pure components, Fig. 2.7.6c) have no plateau at melting or crystallization. This means that the alloy crystallization (melting) occurs in the range of temperatures $T_1 \div T_2$. Corresponding critical points (*c*' and *d*' inflection points) on the cooling curve of the alloy with composition *x* (Fig. 2.7.6b) arising as a result of thermal effects during phase transformations determine this temperature range in which crystallization of the alloy with such composition take place. This makes it possible to plot *c* and *d* point on the state diagram in Fig. 2.7.6a which indicate the beginning and the end of crystallization of the alloy with composition *x*. Constructing a set of critical points of the beginning and the end of crystallization of alloys with different compositions it is possible to obtain lines of diagrams of a binary system forming a continuous series of liquid and solid solutions between *A* and *B* components (Fig. 2.7.6a). Let us analyze in detail state diagram shown in Fig. 2.7.6a.

On this diagram, the locus of points which define the temperature of crystallization beginning for alloys (*A'mcB'* curve) is called the *liquidus line*. Single phase region above the liquidus line on the state diagram corresponds to the *A*-*B* solutions in liquid state (indicated by letter L). The locus of all points that determine the temperature of crystallization end for alloys (*A'dnB'* curve) is called the *solidus line*. The region below the solidus line on the phase diagram corresponds to *A*-*B* solutions in solid state (indicated as S). As can be seen, the region between liquidus and solidus lines corresponds to two-phase mixture of

the areas occupied by liquid and solid solutions with different compositions (L+S).

So, any binary phase diagram is a graphic image of phase composition of alloys of this system as dependence of the critical temperatures of phase transformations on chemical composition of alloy plotted on the horizontal axis. The leftmost point on the X-axis in Fig. 2.7.6a corresponds to the 100% content of *A* component, and the rightmost point corresponds to 100% content of the second (*B*) component. As noted in Chapter 1, the solid solutions enriched by *A* component are often called α-solid solutions, and solid solutions enriched by *B* component are called β-solid solutions. Alloy compositions usually expressed in atomic or weight (mass) percent. Recalculation of atomic percent into mass and vice versa can be carried out by the formulas

$$a = \frac{100 M_A \alpha}{M_A \alpha + M_B \beta}, \tag{2.7.1}$$

$$\alpha = \frac{100 a}{a + b \dfrac{M_A}{M_B}}, \tag{2.7.2}$$

where M_A and M_B are the atomic weight of *A* and *B* components, respectively; α and a – atomic and mass concentrations of *A* component in at.% and weight %, respectively; β and b – atomic and mass concentrations of *B* component in at.% and weight %, respectively.

Phase equilibrium conditions in two-component systems. As noted above, any point on state diagrams which shows composition and temperature of alloy is called *figurative point*. Figurative point corresponding to the composition of one of the equilibrium phases at any constant temperature is called a *node*. For example node *m* in Fig. 2.7.6a corresponds to temperature T_3 and composition *y*, while node *n* – to the same temperature T_3 and composition *z*.

Segment *mn* of straight line connecting two nodes of the same temperature on the state diagram is called *conode*. Thus, the conode – is the locus of figurative points of two-phase states with a constant chemical composition of each phase (in the case of *mn* conode it concerns liquid and solid phases). Indeed, if analyze chemical composition of liquid phase and solid solutions of alloys with all compositions from *y* to *z* along the *mn* conode at T_3 temperature in Fig. 2.7.6a, the composition will always correspond to *m* point in liquid phase and to *n* point in solid solution. So-called *rule of concentrations* follows from this. It is used to determine chemical composition of phases in the two-phase regions of state diagram of any alloys system.

The concept of conode allows also to formulate a *rule of segments* (also known as *lever rule*) which allows to determine relative concentration of two

phases in alloy. Indeed, the figurative point *k* lying between liquidus and solidus lines of the alloy with composition *x* at T_3 temperature divides *mn* conode into two sections proportional to the amount of two phases of this alloy at T_3 temperature. Let us prove this statement.

Let us denote relative amount of liquid phase in the alloy with composition *y* at temperature T_3 as *p*, and relative amount of α-solid solution in the alloy with composition *z* at the same temperature T_3 as *g*. Since

$$p + g = 1, \tag{2.7.3}$$

the average concentration of *B* component in the alloy can be determined by the relation

$$x = yp + zg, \tag{2.7.4}$$

where *y* and *z* – are *B* component concentrations in liquid and solid phases, respectively.

Two unknown parameters can be easy found from two equations (2.7.3) and (2.7.4):

$$p = (x\text{-}y)/(z - y) = mk/mn \tag{2.7.5a}$$

$$g = (z - x)/(z - y) = kn/mn. \tag{2.7.5b}$$

Hence we obtain the desired ratio for relative content of two phases in the alloy corresponding to figurative point *k*:

$$\frac{p}{g} = \frac{mk}{kn}. \tag{2.7.6}$$

This is the mathematical expression of the *lever rule*: relative amounts of equilibrium phases of alloy with given composition at a given temperature are proportional to conode segments for this alloy. Since many conodes corresponding to different temperatures can be constructed for the range of crystallization of each alloy, it is a universal rule.

Application of the Gibbs phase rule for analysis of alloys. As any thermodynamic system, each system of alloys in equilibrium obeys the Gibbs phase rule formulated in Section 2.1 of Chapter 2 in the following wording

$$N = C - F + 2 \tag{2.7.7}$$

or

$$N = C - F + 1, \tag{2.7.8}$$

which relate the number of components *C*, the number of phases *F* and the number of degrees of freedom *N* in the isolated equilibrium system (alloy).

Therefore, the experimentally constructed phase diagram are always checked for correspondence to the Gibbs phase rule, which gives the opportunity to theoretically prove the direction of transformation processes for identifying equilibrium state of the system. The phase rule gives the possibility to predict

and verify processes in alloys during heating and cooling, for example, it shows if process of crystallization occurs at a constant temperature or temperature range, and indicates how many phases can coexist in the system.

Let us analyze cooling curves and phase diagram shown in Fig. 2.7.6 for solid solutions using the phase rule (1.1.33). The pure *A* and *B* components are in two-phase state (*L*+*S*) at a constant temperature (plateau on the cooling curves) during solidification. They do not have degrees of freedom for changing temperature, because $N = 1 - 2 + 1 = 0$ in accordance with calculation by the formula (1.1.31). Alloys *x*, *y* and *z* are in two-phase state (*L*+*S*) in the range of crystallization, but their $N = 2 - 2 + 1 = 1$, i.e. within the range of crystallization change of temperature is possible without changing the number and names of coexisting phases. However, the concentration of components in these phases during cooling process (temperature changing) is not constant, but varies in accordance with the *rule of concentrations*.

According to the phase rule, in a binary system with a constant pressure more than three phases cannot coexist at the same time. This is possible when $N = 0$, i.e., when $F = 2 + 1$. In other words, three phases can coexist only at certain composition and temperature of alloy, for example, in the alloy with eutectic composition for binary mechanical mixtures (see below). If analysis of the binary system shows that there is more than three phases, this means that the alloy is not in the equilibrium state or the number of phases or components in studied alloys system is determined incorrectly.

2.7.3. State diagrams of two-component alloys

Phase diagram of the alloys with unlimited solubility of components in solid state (solid solutions with unlimited solubility). Let us summarize the results of the analysis of phase diagram shown in Fig. 2.7.6 for solid solutions with unlimited solubility of components:

1. Number of components in the alloys is C = 2 (A and B components).
2. Number of phases is F = 2 (liquid phase L and crystals of solid solution S).
3. The main lines of phase equilibria on this diagram:

- *A'mcB'*– the liquidus line, above which the alloys are in liquid state (L);
- *A'dnB'*– the solidus line, below which the alloys are in solid state (S).

4. Pure A and B components crystallize at constant temperature (see cooling curve of B component in Fig. 2.7.6c).
5. All alloys crystallize similarly to alloy with composition x whose cooling curve is shown in Fig. 2.7.6b. Crystallization process of the alloy with composition x is following. The alloy is cooled in a liquid state up to the point c. Schematic image of this under microscope is shown in Fig. 2.7.1b. Crystallization process takes place at c-d segment. It occurs at decreasing

temperature, because, according to the Gibbs phase rule in two-component system in the presence of two phases (liquid and crystals of α-solid solution), the number of degrees of freedom is equal to one (N = 2 - 2 + 1 = 1). At temperature corresponding to the point c, centers of nucleation of α-solid solution on the basis of A component begin to form, and inflection point (critical point) is detected on the cooling curve associated with a decrease of cooling rate due to the latent heat of crystallization release. At point k alloy is a two-phase mixture of α-phase crystallites in the liquid melt (Fig. 2.7.1a). When temperature corresponding to the point d is reached, the alloy solidifies at constant temperature. After solidification of the whole liquid melt the alloy is cooled in the solid state under further temperature decrease. This alloy consists of homogeneous crystallites (grains) of α-solid solution with a microstructure similar to the microstructure of pure component.

6. Quantitative analysis of alloy structure and phase composition.

Using state diagram it is possible to determine not only number of phases, but also chemical composition of each phase and proportion of the volumes occupied by these phases for each alloy at any temperature. For this purpose let us use the *rule of segments*.

To determine the composition of phases corresponding to the figurative point *k*, let us draw a horizontal line (conode) through this point to its intersection with the nearest lines of phase equilibrium diagrams – the liquidus and solidus lines. Composition of liquid phase is defined as the projection of the point *m*, which lies at the intersection of conode with the liquidus line, on the concentration axis – it corresponds to the point *x*. Composition of the solid phase is determined by the point *y* which is the projection of point *n* (intersection of conode with the solidus line) on the axis of concentrations.

From Fig. 2.7.6 can be concluded that composition of the liquid phase changes when moving along the liquidus line, and the composition of the solid phase changes when moving along the solidus line. With decreasing temperature, composition of the phases changes in the direction of reducing *B* component concentration.

Let us determine mass (weight) ratio of liquid and solid phases at a certain temperature (e.g., at point *k*). According to the lever rule, mass of phases at point *k* is inversely proportional to the segments which are formed in the result of dividing conode *mn* by this point. Let us assume that the amount of alloy (*p* + *g*), according to Formula (2.7.3), corresponds to the length of the entire conode *mn*. Then, according to (2.7.5), the segment *mk* adjacent to the liquidus line determines the amount of solid phase *g*, and the segment adjacent to the solidus line *kn* (or to axis of component) determines the amount of liquid phase *p*.

Phase diagram of alloys with the lack of solubility of components in the solid state (mechanical eutectic mixture). Let us consider the alloys system both components of which are indefinitely soluble in a liquid state, but in the solid state are generally insoluble in each other and do not form chemical compounds. The state diagram and the cooling curves of such alloys system are shown in Fig. 2.7.7.

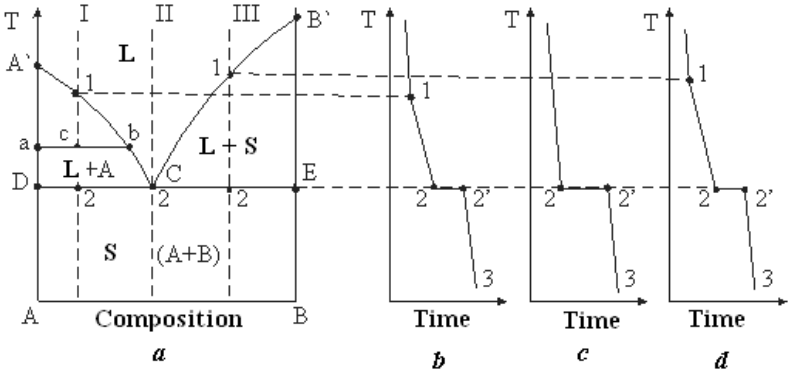


Fig. 2.7.7. State diagram (*a*) of binary eutectic system for which formation of a mechanical mixture of the pure components during crystallization is characteristic. Cooling curves are plotted for alloys of I (*b*), II (*c*) and III (*d*) types

In this case, on the state diagram shown in Fig. 2.7.7a, the components – are *A* and *B* chemical elements and the phases – are liquid (*L*) and crystals of *A* and *B* elements (*S*). When the alloys are cooled below A`C line, crystallites of pure *A* component begin to plate out, but below the CB` line, crystallites of *B* component begin to precipitate. During the crystallization of the alloy whose composition is determined by the point C, the crystallites of both *A* and *B* type precipitate simultaneously. Mechanical mixture of two kinds of crystals which simultaneously crystallize from liquid is called *eutectic*, and transformation as well as crystallization temperature at point C are called eutectic transformation and eutectic crystallization, respectively.

Let us carry out a more detailed analysis of the phase diagram:

1. Number of components is $C = 2$ (*A* and *B* components);
2. Number of phase $F = 3$ (crystals of *A* component, crystals of *B* component, liquid phase *L*);
3. The main lines of diagram:

- liquidus line A`CB` consists of two branches converging to the point C;
- solidus line ECF is parallel to the axis of concentrations and tends to the axes of the components, but does not reach them;

4. Pure *A* and *B* components crystallize at constant temperature (point A' and C').
5. *Eutectic alloy* corresponds to concentration of components at point C (alloy II) and, as seen from Fig. 2.7.7c, has a cooling curve similar to cooling curves of pure components, i.e. has a plateau (as in Fig. 2.7.6c).

Microstructure of *eutectic* is crystallites with fine-dispersed (generally plate-like) mechanical mixture of pure *A* and *B* components (Fig. 2.7.7b) crystallizing simultaneously at constant and the lowest (for considered system) temperature.

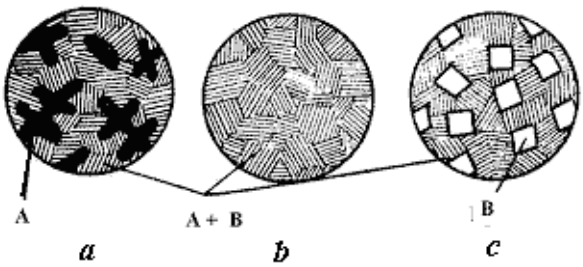


Fig. 2.7.8. Scheme of alloy microstructure for hypoeutectic I (*a*), eutectic II (*b*) and hypereutectic III (*c*) compositions

Crystallization process of the alloys with eutectic composition occurs as follows (Fig. 2.7.7c). Section 0-2 on the cooling curve of the eutectic alloy II describes cooling of liquid alloy, section 2-2` – eutectic crystallization, 2`-3 – cooling of crystallized alloy with eutectic composition. Crystallization of the eutectic is actually takes place at a constant temperature, because, according to the rule of phase, the number of degrees of freedom will be equal to zero ($N = 2 - 3 + 1 = 0$) in a two-component system in the presence of three phases (liquid and crystals of *A* and *B* components). Crystallization is completed at point 2'.

6. Crystallization process of hypoeutectic alloy I occurs as follows (Fig. 2.7.7b). Section 0-1 on the cooling curve of the alloy I corresponds to cooling of the liquid alloy. Crystallization centers of excess *A* component start to form at temperature corresponding to point 1 (black crystallites in Fig. 2.3.3a). Inflection is detected on the cooling curve (critical point) associated with a decrease of cooling rate due to the latent heat of crystallization release. Crystallization process takes place in section 1-2. It occurs at a decreasing temperature, because in accordance with the rule of phases a number of degrees of freedom will be equal to unity ($N = 2 - 1\ 2 = 1$) in a two-component system in the presence of two phases (liquid and

crystals of *A* component). Liquid phase composition during cooling varies along the liquidus line up to the eutectic. In section 2-2 ', alloy crystallizes in the form of grains consisting of eutectic plates (*A*+*B*) similarly to crystallization of eutectic alloy II. Below the point 2', alloy consisting of crystals of initially crystallized excess of *A* component and eutectic (*A*+*B*) is cooled (Fig. 2.7.8a).

7. *Hypereutectic alloy* III crystallization process occurs as follows. Cooling curve of alloy III in Fig. 2.7.7d consists of three sections: 0-1 – cooling of liquid alloy, 1-2 – precipitation of crystal of *B* components (white crystallites in Fig. 2.7.8c), 2-2' – eutectic crystallization and 2'-3 – cooling of crystallized alloy (*A*+*B*). As can be seen, this type of alloy crystallization has all the features of alloy I crystallization, described above in Section 6.
8. For quantitative analysis of structure and phase composition of the alloy I let us plot the conode *acb* drawn through a given point *c*. Conode crosses the liquidus line and the axis of *A* component, so the solid phase composition of the alloy I is 100% of *A* component and the liquid phase composition corresponds to point *b*. Using the rule of lever for the alloy I shows that the amount of liquid phase in it is determined by the ratio of *ac/ab* at temperature T_a, and amount of solid phase as *cb/ab*, where *ab* – is the length of conode.

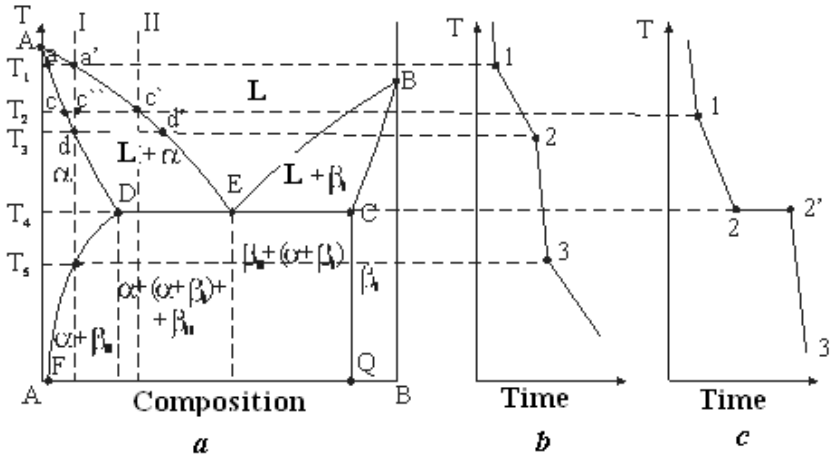


Fig. 2.7.9. Phase diagram (*a*) and cooling curves of the alloy I (*b*) and alloy II (*c*) of eutectic system, when solubility of one component in other depends on temperature

The alloy of type III can be analyzed by the same way. Any conode leans on the axis of the pure components for eutectic alloy II below the temperature

corresponding to point *C*, and this alloy always consists of a mixture of components (*A*+*B*).

Phase diagram of alloys with a temperature-dependent solubility limit of components in solid state. Let us consider the case of the eutectic alloys, when both components have limited solubility in solid state, but solubility limit of at least one of the components depends on temperature. An example of such a system, when solubility of *B* component in *A* component changes (decreases) during cooling of the alloy is shown in Fig. 2.7.9. Let us analyze this system.

1. Number of components is C = 2 (A and B components);
2. Number of phases is F = 3 (liquid phase L as well as crystallites of α-solid solution (solution of *B* component in *A* component) and β-solid solution (solution of *A* component in *B* component)).
3. The main line of the diagram:
   - liquidus line Aa`c`d`EB consisting of two branches that meet at point E;
   - AcdDESB solidus line, which consists of three parts;
   - DF – line of maximal solubility of B component in A component (temperature dependent);
   - CQ – line of maximal solubility of A component in B component (independent on temperature).
4. Pure A and B components crystallize at a constant temperature.
5. Regions of existence of α- and β-solutions adjoin to the axes of temperatures for A and B components. The maximal solubility of B component in A component and A in B are determined by solubility lines DF and CQ, respectively. It should be mentioned, that DF line is not parallel to the axis of temperature (temperature-dependent solubility).

Cooling curve of the alloy I is shown in Fig. 2.7.9b. The alloy is in a liquid state if it locates higher than point 1 ($T > T_1$). At point 1 the process of crystallization begins at which crystals of α-solid solution precipitate. At the same time, concentration of *B* component in α-phase and in liquid varies along the curves *acd* and *a'c'd'*, respectively. At point 2 of cooling curve, crystallization is completed, and obtained at equilibrium crystallization crystals of solid solution have composition of the initial liquid. These crystals do not change during cooling down to the point 3. Below this point, α-solid solution becomes supersaturated, and crystals of β-solid solution separate out from it. Concentration of *B* component in β-solid solution is determined by the point Q. Composition of the α-solid solution changes along the DF curve due to the precipitation of crystals of β-solid solution. Crystals of β-solid solution, separated out from the liquid phase L, are called primary crystals and are denoted as β_I. Crystals separated out from the solid solution are called secondary crystals and are labeled as β_{II}.

6. Alloys which correspond to the compositions lying between DF and CQ lines of solubility (are located outside solubility) are two-phase alloys because they are composed of mixtures of α- and β-solid solutions. Cooling curve of the alloy II for this area is shown in Fig. 2.7.9c. When the horizontal line DEC (eutectic line) is achieved, nonvariant transformation occurs (C = 2 - 3 + 1 = 0). In this case, there are three equilibrium phases – fluid L, crystals of α-solid solution, whose composition is determined by point D, and crystals of β-solid solution with composition corresponding to point C. When the alloy is cooled below temperature T4 (below the horizontal line DEC), eutectic transformation of following type occurs:

$$\alpha_E \rightarrow \alpha_D + \beta_C,$$

when crystals of both solid solutions separate out from liquid crystals. Here, α_E and α_D – are α-solid solutions whose composition corresponds to E and D points, respectively, and β_C – is solid solution of A component in B component with composition corresponding to C point. Thus, eutectic of (α_D+β_C) type forms during crystallization of the alloy II besides the primary crystals of α-solid solution (separated out during cooling from point 1 to point 2). Further cooling is accompanied by a decay of α-solid solution with precipitation of secondary crystals β_{II} at temperature T_5 in Fig. 2.7.9a (at point 3 on the cooling diagram in Fig. 2.7.9b) due to supersaturation of α-solid solution with B component.

Using the rule of segments for the alloy I at temperature T_1, proportions of liquid and solid phases can be determined which are equal to $c\mathrm{T}_1/cc'$ and t_1c'/cc', respectively.

7. Behavior of the alloy II in Fig. 2.7.9a is similar to behavior of the alloy II in Fig. 2.7.9.

Phase diagram of the alloy with components forming a chemical compound. Phase diagram of the binary systems whose components form a chemical compound is complicated and consists of a few simple diagrams. Number of components and number of diagrams depends on number and types of chemical compounds formed by basic components of the system. Diagram of binary system which produces a single A_mB_n type of chemical compound, which is insoluble in components formed it, is shown in Fig. 2.7.10 is a. A and B components are insoluble in this compound, too, and form eutectics with it.

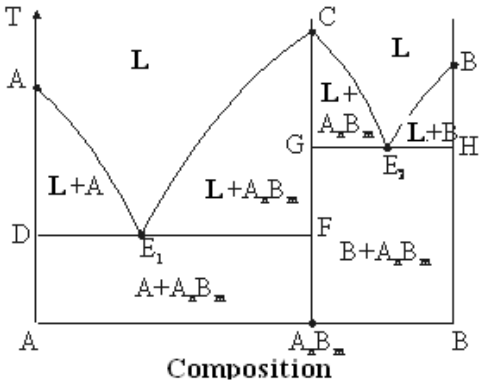


Fig. 2.7.10. State diagram in which there is a stable chemical compound

Since chemical compound is characterized by a certain ratio of the components (m/n), it is reflected in the phase diagram as a vertical line intersecting the axis of compositions at point corresponding to this ratio of components. Chemical compound is stable if it can be heated up to melting without decomposition, and is unstable if it decomposes during heating. Let us analyze this system.

1. Number of components is C = 3 (A, B and AmBn components);
2. Pure components and chemical compound crystallize at constant temperature.
3. Phase number and kind of simple diagrams are determined by the nature of interaction between the components. In the case presented in Fig. 2.7.10, diagram consists of two diagrams of eutectic type in each of which AmBn compound is considered as one of the components. Thus, the components of this system are A and B elements, and solid phases of this system are A, B and AnBm. Number of phases (of four possible phases in this system) in equilibrium is F = 3 (L+A+AnBm or L+B+AnBm).

Melting of chemical compound A_nB_m occurs at a constant temperature that corresponds to the phase rule (N = 1 - 2 + 1 = 0) and reflected in presence of a plateau on the cooling curve.

4. The main line of the diagram:
   - liquidus lines AE1C and CE2B for the left and right simple diagrams, respectively, each of which consists of two branches meeting at the eutectic points E1 and E2;
   - solidus lines DF and GH for the left and right simple diagrams, respectively;

5. On this state diagram there are two eutectic points – E1 and E2. Eutectic E1 consists of crystallites of A component and AnBm. Eutectic E2 consists of B and AnBm crystallites. Primary crystals of chemical compound separate out during melt cooling when crossing E1C and E2C lines.
6. Processes of alloy crystallization for this phase diagram realize similarly to the crystallization of alloys which form mechanical mixture of pure components crystals during crystallization. Therefore, cooling curves of alloys of this system are similar to those shown in Fig. 2.7.7. The only difference is that in addition to precipitation of crystallites of pure A and B components formation of crystallites of AnBm compound takes place.

2.7.4. The relationship between properties of alloys and type of phase diagram

Physical and chemical properties of alloys are largely determined by their phase composition, i.e. depend on type of chemical compounds or phases formed by components of this alloy. On the other hand, phase composition of alloys reflects on their equilibrium state diagrams. Consequently, correlation between properties and the type of equilibrium phase diagram of any alloy system should exist. For the first time, N. Kurnakov pointed out the relation between state diagrams and properties of alloys. To establish this relation between phase composition of alloys and their physical and chemical properties so-called structure-property diagrams are usually constructed which are called Kurnakov diagrams.

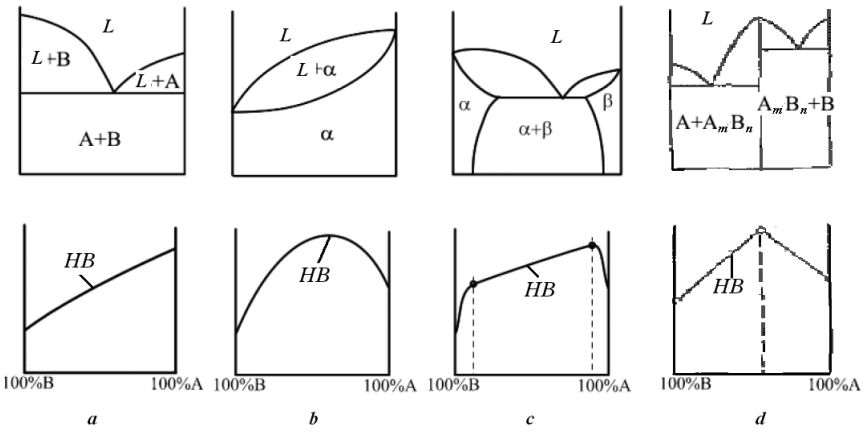


Fig. 2.7.11. Schematic illustration of state diagrams and concentration dependences of Brinell hardness for binary A-B alloys

Schemes of the simplest types of phase diagrams of binary alloys and corresponding Kurnakov diagrams are presented in Fig. 2.7.11. The comparison of these diagrams allow to derive a number of useful rules that help to predict properties of alloys depending on the type of equilibrium phase diagrams, and vice versa – on the basis of equilibrium phase diagram it is possible to predict changes in properties of corresponding alloys. In this case it comes to such properties as micro-hardness, hardness, electrical resistivity, thermal electromotive force coefficient, the coefficient of expansion, Young's modulus, etc., which are widely used in metallography.

Let us formulate these rules.

1. In formation of binary mixtures, many of their mechanical properties (e.g., electrical resistance, hardness) vary linearly (additively) with concentration of components, so that the values characterizing alloys properties are in the range between their respective values for pure components (Fig. 2.7.11a).
2. While binary solid solutions with unlimited solubility forming, alloys properties vary according to monotonic (without features such as inflection points and breaks) curvilinear dependences (Fig. 2.7.11b). However, some properties (firs of all, electrical resistivity, micro-hardness, thermal electromotive force coefficient) may differ significantly from those of the pure components. Therefore, whereas, for example, resistivity changes slightly in formation of mechanical mixture (always demonstrating only the values that lie between the magnitudes of resistivity of pure components), it however can increase (significantly) in formation of solid solution and achieve values higher than for any of the components. Therefore, decomposition of a solid solution into two (or more) phase always leads to the increase of electrical conductivity. The latter rule is called the Kurnakov law.
3. As follows from Fig. 2.7.11c, while solid solutions with limited solubility forming, the properties change according to curvilinear dependences (as in the case of solid solutions in Fig. 2.7.11b) in the concentration range corresponding to a single-phase solid solutions. At the same time, they change linearly (as for mechanical mixtures in Fig. 2.7.11a) in two-phase region. In this case, the endpoints on the additive line corresponds to the properties of extremely rich α- and β-solid solutions which form this mixture.
4. In the case of the formation of chemical compounds, singularity in form of maximum or minimum is observed on the property-composition curve (Fig 2.7.11d). Thus, it is possible to identify stoichiometric ratio of components in a given chemical compound by determining concentration that corresponds to a singular point on the structure-property diagram.

The method of analysis of properties changes as a function of composition changes as well as construction of composition-property diagrams forms the basis of *physics and chemical analysis of alloys* developed by N. Kurnakov. This is one of the main methods of alloys study and it is widely used in scientific research. Regularities mentioned by N. Kurnakov are the basis for the development of alloys with desired properties because they allow predicting character of the alloys properties changes as the result of changing chemical composition. However, these regularities concern alloys in equilibrium state. This means that their application is limited. Study of Kurnakov diagrams is only a supplement to study and construction of equilibrium phase diagrams.

2.7.5. Structural transformations in alloys under thermal impacts

As mentioned earlier, any initial crystalline material obtained by casting or other methods must be converted into a product having a predetermined shape and desired functional properties. For this purpose, any crystalline alloy, both after manufacture and after transformation into a product, is subjected to different types of impacts – mechanical, thermal (heat treatments), thermomechanical, thermochemical, etc. which can lead to phase transformations in it. As for pure metals, quenching and different types of annealing are the basic types of heat treatment.

Alloys used in various power designs and energy engineering are usually quenched for they hardening. Alloys undergoing transformations in solid state under equilibrium conditions (e.g. eutectoid transformation in iron-carbon alloys (steels)) can be strongly hardened by quenching. In this case, the strength increases either because of reduction of grains size forming eutectoid mixture due to the martensitic phase transition, or because of decrease of eutectoid reaction temperature. In alloys for special applications (electrical, magnetic, heat-resistant, etc.), such quenching is made that allows to change their structure-sensitive physical and chemical properties – electrical resistivity, coercive force, saturation magnetization, corrosion resistance, etc.

In most engineering alloys (steels, bronzes, brasses, etc.), high-temperature solid solution state is usually detected as a result of alloy quenching, so that supersaturated solid solutions and their mixtures are formed after quenching. However, significant hardening of solid solutions alloys do not usually occurs just after quenching. For hardening of these alloys after quenching special kind of annealing is used. As a result of such annealing, strengthening phases separate out in the alloy during decomposition of supersaturated solid solution. This decay process is carried out either by repeated low temperature heating (below the quenching temperature) or during ageing at room temperature. The first type of annealing is called *tempering*, and the second type is called *aging*. The term "tempering" is usually used for steels and other alloys

undergoing polymorphic transformation during annealing after quenching (ferrite steels, two-phase aluminum bronzes, some titanium-based alloys). The term "aging" usually refers to the alloys which do not undergo polymorphic transformation after quenching (aluminum-based alloys, austenitic steels, nickel alloys, etc.). In the result of tempering or aging of preliminary quenched alloys phase transformations occur, that make their phase structure more equilibrium (corresponding to state diagram), but fully equilibrium state does not reach during alloy processing, as it takes place after usual annealing. Combination of quenching and tempering or aging processes almost always assumes obtaining increased values of functional properties of alloy (hardness, strength characteristics, coercive force, electrical resistivity, etc.).

Temperature and exposure in tempering or aging is chosen in such a way to improve the needed structure and properties. Cooling rate from the annealing or aging temperature to room temperature, with rare exceptions, does not influence on structure and properties of alloys.

Listed types of thermal treatments and resultant mechanisms physical and chemical properties changes will be discussed in more detail below with respect to concrete alloys.

Heat treatment of deformed and cast alloys. Heat treatment of many alloys is carried out regardless of the fact if there are phase transformations in alloys in solid state or not. This treatment is used primarily for reduction the residual stresses in products, recrystallization of plastically deformed semi-manufactured articles, and decrease of inhomogeneity of components distribution in ingots or castings. The task of these types of annealing is transformation of alloy to more equilibrium state. Basing on heat treatments purposes, there are different types of annealing: stress reduction annealing (heating), recrystallization annealing (for changing grain structure) and the diffusion annealing or homogenization (to homogenize components in alloy volume).

Let us consider in more detail these types of annealing used in material science of alloys.

Annealing for residual stresses relaxation. Many technological impacts on processed alloys and their products accompanied by the appearance of residual stresses (local deformations) which forms because of defects in crystal lattice, as in pure metals. Significant residual stresses occur in castings and semi-manufactured articles nonuniformly cooled in the preparation process (casting, rolling, forging), in cold deformed semi-manufactured articles or intermediates, in roads during straightening, in welded connections, during quenching, etc.

Residual stresses can reach yield limit. Residual stresses arising in listed above cases are often undesirable because they can cause deformation of parts during processing (cutting, milling, stamping etc.). When summing up during

operation under external loads, residual stresses can cause warpage of parts of complex constructions and their premature failure. The appearance of residual stresses in seams of welded structures increases probability of brittle fracture during subsequent deformation of the product. In many alloys employed in the presence of corrosive ambient, the residual stresses can cause cracking of material.

To reduce the residual stress of products made of alloys (or alloys themselves at the intermediate stages of products manufacturing) are subjected to special heat treatment (annealing). Existing internal residual stresses cause local plastic deformation of material during heat treatment which reduce this deformations to the yield strength level at heating temperature (i.e., stresses become smaller than at room temperature).

In steels and cast irons, significant reducing residual stresses occur during annealing at temperature of 450 °C. Stresses drop to very low values after they tempering at 600°C. In copper- and aluminum-based alloys, significant reduction in residual stresses occurs at substantially lower heating temperatures. For example, in cold-deformed brass semi-manufactured articles, residual stresses can be almost completely removed in the process of annealing at 250-300°C.

After ageing at certain temperature, products are cooled slowly to prevent appearance of new stresses associated with quenching. Acceptable cooling rate depends on the weight of the product, its shape, material thermal conductivity, and it is usually in the range of 20-200°C/hour.

If compare this kind of alloys annealing with pure metals annealing, the processes of residual stresses removal that occur during annealing of deformed or cast alloy is the closest to the stages of rest and return (they are described in detail in Chapter 1). As any first order phase transition (order-disorder), these two stages are accompanied by the release of latent heat and a decrease in free energy. Rest of alloys takes place at relatively low temperatures, and the return – at higher ones.

As in the case of pure or commercially pure metals, size and shape of grains in alloys at such annealing does not change, but only the number of point defects reduces as well as partial annihilation and spatial redistribution of dislocations takes place.

Recrystallization annealing. Recrystallization annealing of deformed alloys or parts is carried out by almost the same mechanisms as for metals, but at higher temperatures. The cooling rate of this type of annealing is not critical. Recrystallization of alloys is accompanied by nucleation and growth of new grains in them with fewer structural defects. For alloys in form of solid

solutions almost all features and recrystallization laws are characteristic as for commercially pure metals.

Plastically deformed alloys, as pure metals (see above), can recrystallize only after a certain critical degree of deformation. Nucleation of new grains during recrystallization occurs in areas with the highest density of dislocations: this usually takes place at the boundaries of deformed grains. Thus, the higher the degree of plastic deformation, the more points of recrystallization appear. Over time, centers of new formed grains grow due to the diffusion of atoms from deformed area to area with a perfect crystal lattice (in new nucleated grains) that results in motion of new grains boundaries deeper into the deformed alloy.

Recrystallization annealing is often used as a pretreatment of cast alloy (ingot) prior to cold deformation (cold rolling, drawing, extrusion, etc.). The purpose of such annealing is to change the microstructure of the alloy leading to deformed alloy strength decrease and ductility recovery; to impart given grain size and etc. Temperature of recrystallization beginning is related to the temperature of melting. Recrystallization annealing temperature is usually chosen 100-200°C above the recrystallization temperature determined by the known relation (1.6.14) $T_{recr} = aT_{melt}$. Note that if the coefficient a of pure metals is typically equal to 0.1-0.2 and for technically pure metals is 0.3-0.4, it is significantly higher for solid solutions and is equal to 0.5-0.8. Practically recrystallization annealing is performed for low-carbon steels at 600-700°C, for brass and bronze at 560-700°C, for aluminum alloys at 350-450°C, for titanium alloys at 550-750°C.

Recrystallization annealing can also be used as final processing of semi-finished products. As in metals, recrystallization in some solid solution is accompanied by texture formation (preferred orientation of the crystallites in parts volume) which creates anisotropy of properties. This allows to improve various properties along certain directions in parts (e.g., magnetic properties of the magnetic circuits made of transformer steel and permalloys, coefficient of elasticity in some spring alloys, etc.).

In mechanical and instrument engineering solid solutions that do not have phase transformations in the solid state (ferritic and austenitic steel, single-phase brasses and bronzes) are widely used. In these materials, combination of cold plastic deformation with subsequent recrystallization annealing is the only possibility of adjusting grain size and, consequently, characteristics of alloys.

Diffusion annealing (homogenization). In real conditions of melt cooling, diffusion processes which are necessary for aligning components concentrations in volume of growing crystals of solid solutions are typically slower than crystallization process. This may result in composition heterogeneity of alloy in volume of ingot as a whole or in volume of individual crystallites. The radial

non-uniformity of components distribution in grain is called *intracrystalline segregation (liquation)*. It appears in the grains central part of polycrystalline alloy (crystallites) corresponding to the place of crystallite nucleus formation from melt being often enriched with more high-melting component of alloy (it crystallizes first), and the peripheral parts of the crystallites being enriched with low-melting component. If crystallization is fast enough (e.g., during casting), the diffusion of components (and, as a consequence, their homogenization in grain) has no time to finish.

In Fig. 2.6.1 change in the average concentration of solid solution of alloy with composition *I* in the range of crystallization temperatures under deviation of crystallization process from equilibrium is schematically shown by an example of system with eutectic transition. The alloy with such composition is in the region of bounded solid solutions. For example, while rapid cooling, solidus line in the region of α-solid solutions is shifted down (dashed line in Fig. 2.6.1) and, therefore, α-crystallites have less doping component *B* in comparison with equilibrium concentration reached during slow cooling of the melt. If the temperature T_2 of the end of alloy I crystallization is lower than the equilibrium eutectic T_3 temperature (due to the decrease of *B* component content in solid solution), as shown in Fig. 2.7.12, the remaining by the time of eutectic temperature reaching liquid solution solidifies by the eutectic reaction, and the eutectic component appears in the structure of alloy (which should not be in the alloy with this composition with equilibrium structure).

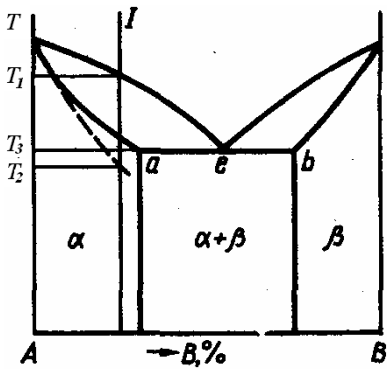


Fig. 2.7.12. The state diagram of components with limited solubility in solid state and eutectic transformation (dashed line shows the change of solidus line under non-equilibrium crystallization)

As noted above, one of the consequences of non-equilibrium growth of crystalline alloys (such as casting) is the segregation in the form of inhomogeneous distribution of components and impurities in ingot and (or) in grains of polycrystals. For example, in steel ingots obtained by casting, concentration of carbon and such contaminants as sulfur and phosphorus increases in the direction from surface towards center and from bottom to top along the ingot (see Fig. 2.4.12). This chemical heterogeneity of different zones of ingot is called *zonal segregation* and adversely influences on the mechanical properties of steels.

In real ingots, other types of phase segregation take place in addition to zonal segregation. For example, solid phase "floats" on the surface during crystallization, if it is lighter than liquid phase, or sinks to the bottom of melt in crucible, if it is heavier. This type of phase segregation resulting from the difference in densities of liquid and solid phases, as well as crystallization of immiscible liquid phases, is called *gravitational segregation*. Ingot layering with respect to density may cause its embrittlement which is unacceptable for many structural alloys because it deteriorates theirs plastic deformation during machining. To reduce gravitational segregation large cooling rates are use for melt, but this leads to increase in the content of defects and residual strains in ingots.

Intracrystalline segregation in a form of heterogeneous radial distribution of impurities in grains, especially in case of eutectic phase appearance in grain structure, reduces ductility of alloys complicating subsequent alloys processing by pressure. For elimination of micro- and macrosegregation in alloys diffusion or homogenizing annealing is used.

Diffusion (*homogenization*) *annealing* is bulb exposure of alloys at high temperatures resulting in reduced segregation inhomogeneity of solid solution due to intensification of diffusion processes which did not manage to complete during ingot crystallization from the melt.

Ingots of steel alloy, as well as many alloys made of aluminum, copper and other, are undergone diffusion annealing. Diffusion annealing of steel ingots is carried out at temperatures of 1100-1300°C with 20-50 hours exposure. In steel ingot, as a result of this type of annealing, more uniform distribution of phosphorus, carbon and doping elements in the grains of the solid solution is reached.

Diffusion annealing of aluminum alloys ingots for segregation eliminating in them is carried out at temperatures of 420-520°C with 20-30 hours exposure. Segregation is especially undesirable in aluminum alloy ingots. Since dendrite axes in the result of segregation contain less doping elements than spaces between axes and grain boundaries, secondary crystallites are mainly separate

out during cooling between dendrite axes and along grain boundaries, often in very unfavorable form (e.g., as continuous brittle shells at grain boundaries).

Heat treatment of quenched alloys. Principal possibility of using one or another type of heat treatment of the quenched alloy depends on type of its equilibrium phase diagram. From the point of view of differences in heat treatment, the following main groups of alloys are usually distinguished:

1. Alloys that do not have phase transitions in solid state;
2. Alloys with variable solubility of components in solid state;
3. Alloys having phase transitions in solid state.

Heat treatment of alloys with variable solubility of components in solid state. The dependence of solubility of components in solid state on temperature allows significantly hardening the alloys by their heat treatment. This has led to the extensive use of this type of alloys which are called aging alloys. Aging alloys based on aluminum, copper, iron, nickel, cobalt, titanium and other elements are used as structural materials with heightened and high strength.

Let us consider the principle of alloy hardening during aging by the example of system in which there is a chemical compound (Fig. 2.7.13a).

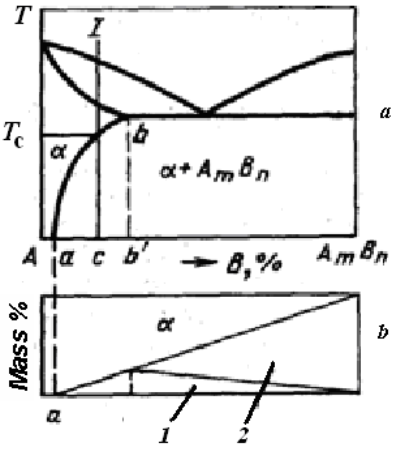


Fig. 2.7.13. The equilibrium diagram of components with variable solubility in solid state:*a* – eutectic equilibrium diagram; *b* – number of primary (*1*) and secondary (*2*) crystallites of AmBn compound in the alloys with different composition at room temperature (height of the rectangle is proportional to the mass of alloy)

Alloys whose compositions lie between point *a* and A_mB_n chemical compound (Fig. 2.7.13a) can be considered as thermally hardenable alloys. In these alloys, secondary crystallites of A_mB_n compound separate out during

cooling from α-solid solution. The degree of hardening will be higher, the greater mass of the secondary crystallites in equilibrium alloy (Fig. 2.7.13b).

Let us consider, for example, the alloy *I* with composition corresponding to point *c* which in equilibrium has a two-phase structure consisting of α-solid solution crystallites with concentration of *a* point and relatively large A_mB_n secondary crystallites. Resistance to dislocation motion leading to the strength enlargement increases as the distance between particles of the hardening phase A_mB_n decreases. This means that the alloy *I* will be harder when large number of smaller inclusions of A_mB_n phase form instead of few large ones. Inclusions separated from each other by 25-50 interatomic distances create the greatest obstruction to dislocation motion.

In most aging alloys, desired finely dispersed structure of inclusions is formed by heat treatment consisting of two procedures – quenching and subsequent aging. Before quenching, alloy is heated to a temperature ensuring the decay of secondary crystals. For considered alloy I such temperature will be slightly above the temperature T_c (see Fig. 2.7.13a). When rapid cooling from temperature $T > T_1$, single-phase alloy in the form of α-solid solution oversaturated with *B* component forms as a result of quenching because the process of precipitation of secondary crystallites of A_mB_n compound is completely suppressed (has no time to occur). The distortion of crystal lattice of α-solid solution due to its oversaturation has relatively low influence on increase of hardness and strength, as well as reducing the ductility of the alloys.

Oversaturated α-solid solution is a non-equilibrium structure with an increased level of free energy. Therefore, as soon as the mobility of atoms is quite large (for example, when heated to a temperature slightly below T_c), α-solid solution begins to decompose with the separating out of A_mB_n compound, i.e. the process of aging begins.

Aging occurring at increased temperatures is called *artificial*. In alloys made of low-melting metal (e.g., aluminum) aging after quenching can occur even during exposure at room temperature. Therefore, this aging is called *natural*.

While aging, the concentration of the supersaturating component in solid solution decreases (for example, carbon in iron for steels) because this component is consumed for the formation of the second phase (cementite). As the alloy aging leads to its hardening, the aging process is often called *dispersion hardening*.

Type of precipitates (their crystal structure, size and character of conjugacy with the lattice of solid solution) depends on the type of alloy and conditions of aging, i.e., temperature and time of exposure. In most alloys several types of precipitates appear during aging.

It is necessary to overcome a certain energy barrier for any rearrangement of the atoms in solid solution, including nucleation of precipitates. The magnitude of this barrier is called activation energy. The activation energy of nucleation and growth of precipitates depends on difference between lattices of precipitates and base solid solution. Energy of activation is small for low difference between the lattices, but it increases with lattice difference. Kinetic energy of the atoms increases with increasing temperature, and, therefore, the probability of overcoming a high energy barrier increases.

During decomposition of oversaturated solid solution of the type shown in Fig. 2.7.14 the reduction of free energy is the largest for equilibrium A_mB_n crystallites precipitation. It should be mentioned, that nucleation of such crystallites having, as a rule, a complex crystal lattice is possible only at sufficiently high temperatures. At low temperatures of aging, non-equilibrium precipitates with more simple crystal structure can be generated.

In general, phase precipitations of following types may appear as the result of oversaturated solid solution decomposition (these are listed in increasing order of nucleation activation energy): 1) Guinier-Preston zone; 2) crystallites of metastable phase; 3) stable phase crystallites.

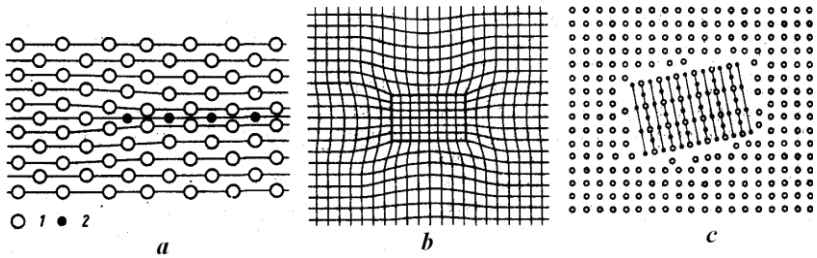


Fig. 2.7.14. Types of precipitates from oversaturated solid solution: *a* – Guinier - Preston zone, *1* – solvent atoms, *2* – solute atoms, *b* – crystallite of metastable phase (coherent precipitation),*c* – crystallite of stable phase (incoherent precipitation)

Guinier-Preston zones (GP zones) are precipitations of solid solution with a much higher concentration of solute which retain solvent lattice. The accumulation of solute atoms in GP zone causes a local change of the lattice parameter of solid solution. When a significant difference in the size of *A* and *B* atoms is observed, as, for example, in Al-Cu alloys (atomic radii of Cu and Al are 0.143 and 0.128 nm, respectively), GP zones have the shape of disks the thickness of which (taking into account lattice distortions) is a few interatomic distances (Fig. 2.7.14a), and a diameter is 10-50 nm. When difference in atomic size is small, such as in Al-Zn alloys (Zn atomic radius is 0.138 nm), GP zones have the shape of spheres.

The movement of dislocations is strongly hampered at high concentrations of GP zones in alloy because application of higher external voltage is required for overcoming of more stable (than matrix) precipitations and its surrounding area with the distorted lattice by dislocation.

Metastable phases (precipitations) have a different lattice than solid solution (matrix). However, correlation in atoms arrangement in certain atomic planes of matrix and precipitates lattices can occur on assumption of certain structural and dimensional correspondence of both lattices. This will lead to the formation of so-called coherent (or semicoherent) interface between the matrix and precipitation. In a certain difference in crystal structures of both phases, coherent boundary results in appearance of transition area with a distorted lattice (Fig. 2.7.14b). High dispersion is characteristic for these metastable phases. It greatly increases the resistance to dislocations movement.

Stable A_mB_n phase has different, usually more complicated, lattice, if compared with solid solution. It has lower symmetry and more atoms per unit cell. Therefore, crystallites of A_mB_n chemical compound with a stable structure separate out in most alloys in the form of sufficiently large particles. The significant difference between crystal structure of solid solution and stable crystallites leads to the formation of incoherent phases interface (Fig. 2.7.14c) and, consequently, smaller distortions of solid solution lattice near the interface. Therefore, alloy hardening in the result of A_mB_n stable crystallites precipitation is less than in the cases of GP zones and metastable coherent crystallites formation.

The process of phase precipitation in decomposition of the solid solution of type I in Fig. 2.6.2 is studied using so-called aging curves. Aging curves (Fig 2.7.15) are usually plotted in coordinates of property – duration of aging (at certain and strictly constant temperature). Hardness or strength is typically used as measured property.

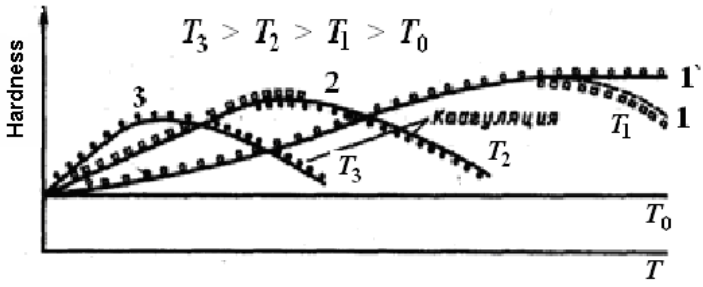


Fig. 2.7.15. Change in hardness of the alloy *I* (in Fig. 2.6.2) during aging: *1* – formation of Guinier-Preston zones; *2* – formation of metastable phase; *3* – formation of stable A_mB_n phase

If the aging temperature T_0 is selected as much lower than T_c temperature in Fig. 2.7.15 (decomposition of oversaturated solid solution does not occur), then, consequently, there will be no change in hardness (strength) of the quenched alloy, see horizontal line T_0.

Aging at higher temperature T_1 will cause increase in strength due to formation of GP zones. If this temperature is not sufficient for activation of metastable crystallites nucleation, the hardness (strength) reaches a maximal value and will not be change for arbitrarily long time (solid line with saturation in Fig. 2.7.15 corresponding to 1` curve). If the temperature T_1 is sufficient for metastable crystallites nucleation after a certain period of aging, the hardness will decrease after reaching a maximal value (line 1 in Fig. 2.7.15). Such a behavior of aging curve (hardness decrease after maximum reaching) is due to the fact that hardness will firstly decrease as a result of the GP zones replacement by metastable crystallites and then it will decrease with aging time increasing due to these crystallites transformation into stable A_mB_n crystallites. If the exposure time is sufficiently large, *coagulation of stable crystallites* occurs. Coagulation is growth of crystallites of the phase which is distributed in form of inclusions in base alloy. The growth of crystallites of the second phase occurs by the decay of the smallest and, as the result, unstable particles and subsequent diffusion of solute component to more stable (large) particles. Coagulation makes alloy structure closer to the equilibrium one.

Aging at temperature T_3 does not cause a significant hardening of alloy because stable A_mB_n crystallites separated out at this temperature have relatively large size and coagulate faster than at temperature T_2. The degree of age hardening can be very high. Thus, the hardness and temporary tensile of aluminum alloys, called duralumin, under optimal conditions of aging are increased in 2 times, and for beryllium bronzes – in 3 times. Heat treatment leading to a stable structure (stable A_mB_n crystallites after coagulation) is called *stabilization*.

Chemical heat treatment of metals and alloys. Chemical heat treatment (CHT) is technological processes that lead to the diffusion saturation of surface layer of finished article (for example, part) by various chemical elements in order to impart other surface properties. CHT is used for increase of hardness, wear resistance, fatigue resistance and contact durability as well as to protect against various types of corrosion.

There are three stages of chemical thermal treatment process.

At the first stage, active elements are formed at the surface of treated workpiece which must diffuse in its surface layer. These atoms of saturating element arise due to dissociation (chemical reactions) of saturating medium (solid, liquid, gas). An example is the dissociation of ammonia:

$2NH_3 \rightarrow 3H_2 + 2N$.

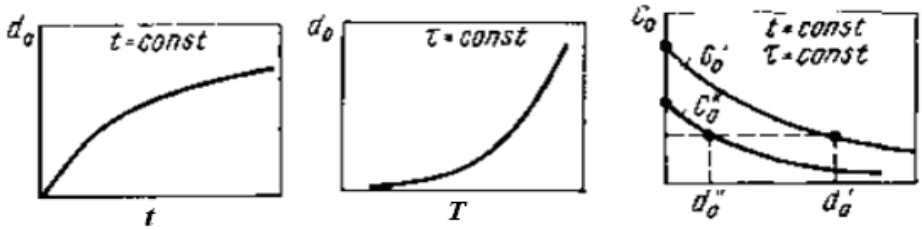
At the second process step, these active elements are adsorbed by working surface of product with the formation of chemical bonds between ions of saturating element and main metal (chemisorption). This results in a concentration gradient of these elements.

The third step is the diffusion of adsorbed active elements deep into saturable product which accompanied by the formation of diffusion layer of solid solutions, chemical compounds and other phases.

The first and the second stages of CHT process realize much faster than the third one, the diffusion step, at which structure and properties of the diffusion zone is formed. Therefore, just the third stage determines the rate of CHT process.

There are many ways of CHT, but the processes of diffusion saturation from active liquid or gas environments is the most widely used in industry. Activated gas is more rational initial environment, i.e. environment without neutral (non-reactive) impurities where the active diffusing element is formed as a result of dissociation or reduction reactions. In some cases, the initial gas environment is activated by ionization in a glow discharge. CHT is typically used to improve hardness, wear resistance, cavitation and corrosion resistance of products, for increasing their reliability and durability.

Thickness of diffusion layer d_0 formed after CHT depends on solubility of saturating element in material, heating temperature T, exposure time t (Fig. 2.7.16a,b) at this temperature and concentration of saturating atoms on the surface (Fig. 2.7.16c). Total thickness of diffusion layer is the shortest distance from the saturation surface to the layer with structure of the product core. Effective layer thickness is the distance from the surface to the region of layer with given parameters (concentration, hardness or other characteristics).



a) *b*) *c*)

Fig. 2.7.16. Dependences of thickness of diffusion layer d_0 formed after CHT on exposure time τ (*a*), heating temperature T (*b*) at this temperature and concentration of saturating atoms C on the surface (*c*)

Phase and structural changes taking place on the diffusion stage of CHT process can be predicted on the basis of binary phase diagrams, if only two elements are involved in diffusion interaction. It is assumed that the diffusion process is not intensified and the resulting diffusion zone is in equilibrium. The structure of the diffusion layer is formed in the same sequence as single phase regions on two-phase diagram A-B with the saturable element B at given temperature T_1 (Fig 2.7.17). Sudden changes of diffusing element concentration are detected during transition from one phase to other (Fig. 2.7.7b). An example of the final microstructure of diffusion layer after CHT corresponding to the system shown in Fig.2.7.17a is presented schematically in Fig. 2.7.18.

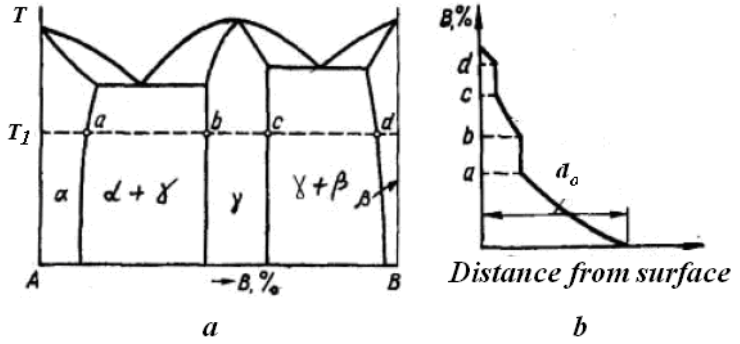


Fig. 2.7.17. State diagram (*a*) and phase distribution (*b*) in diffusion layer at CHT

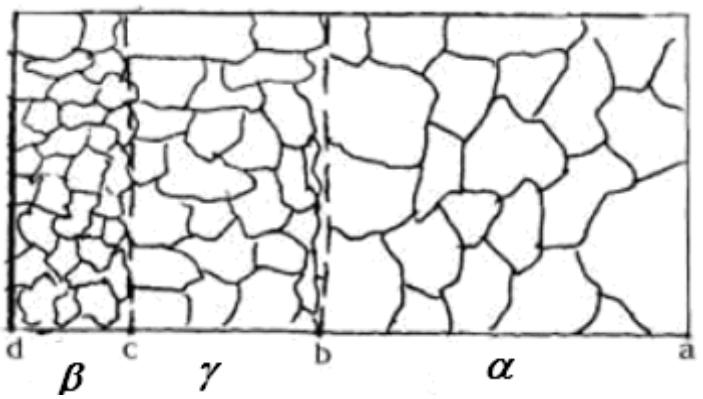


Fig. 2.7.18. Scheme of diffusion layer microstructure formed after CHT at a temperature T_1 corresponding to the system shown in Fig. 2.7.7a

Let us give some examples of CHT for iron-cementite alloys. Diffusion saturation of steel with carbon, nitrogen or these elements together is one of the most common processes of CHT. Carbon and nitrogen relatively quickly diffuse

into subsurface region of steel products, forming rather thick layers of interstitial solid solutions with iron. The technological process of diffusion saturation of steel with carbon is called *cementation*. Such machine parts which must have a wear-resistant working surface and viscous core (gear wheels, shafts, etc.) are typically subjected to cementation.

The process of diffusion saturation of surface area of steel parts with nitrogen is called *nitriding*. Nitriding is used to improve wear resistance and fatigue strength of such machine parts as crankshafts, cylinder liners, worms, rollers, etc.

For many parts of heat power engineering heat-resistant coatings is required because their surface must resist oxidative action of working ambient or environment. Conventional methods of producing such coatings are aluminizing (saturation with aluminum), chroming (Cr saturation) and siliconizing (saturation with silicon) from powder mixture containing diffusing element, activator (NH_4C1, NH_4J, etc.) and neutral powder (chamotte, alumina, etc.) in order to prevent mixture sintering. Saturable parts together with powder is placed in a sealable metal container, heated in furnaces up to 1000-1200°C and held for several hours to obtain the diffusion layers with predetermined thickness and structure.

3. Properties of materials for power industry and energy saving

As noted in the introduction, all aspects of production and use of energy (mining and fuel processing, the conversion of one form of energy into another, the distribution and efficient use of energy) are directly dependent on the availability of materials with special properties. Much of these materials are used as structural materials. Working in various systems, they can be subjected to high static and dynamic mechanical loads (envelope and other load-bearing units, turbines and electric engines; reactors` walls, boilers, piping and other components and parts of power generating systems being under high-pressures, etc.). Furthermore, the materials may serve electrical conductors and have special magnetic properties. They also can be used in heat exchangers and other devices of electric and heat engineering. In many cases, they have to withstand radiation, corrosion or mechanical loads at high temperatures, to reflect or absorb well heat radiation, etc

All the variety of materials used in the power industry and for energy saving can be divided into two major groups – metallic and non-metallic. In this chapter we review the structure and properties of metals and metallic alloys for different functional purposes, as well as their applications.

3.1. Metallic materials

3.1.1. Structural metallic materials

Structural metallic materials are used for production of power machinery parts and components, various enevelopes, appliances and other items, which are subjected to the static and dynamic (including, cyclic and shock) mechanical loads. The most important criterion, upon which we define the main *operational*, *technological* and *economic* requirements for structural materials (not only metallic!), is their high structural strength. By *structural strength* is meant the complex of mechanical properties, which ensures reliable and continuous operation of the material in use of the product thereof. It provides long-term performance of specific units, machinery, tools, components of equipment and devices for testing in the power industry.

Construction metallic materials to be used in power industry are classified according to different criteria, while many of them (e.g., steels, cast irons, some nickel-, cobalt-, copper-, aluminum-, titanium-based alloys) are universal. Nevertheless, along with the generic materials, metal structural materials with strictly defined (narrow) functionality are used. So most often metallic materials are separated by their application in quite concrete groups of issues, constructions and devices:

- Materials providing rigidity, as well as static and cyclic strength (steels, special high-strength alloys);
- Materials resistant to abrasion (wear-resistant);
- Materials with high elastic properties (springs, diaphragms, the other damper devices);
- Materials with low density (aluminum and titanium alloys);
- Strong materials with a high specific strength (aluminum and titanium alloys);
- Materials that retain strength at high temperature impact (heat-resistant);
- Materials resistant to working environment impact, in particular at high temperatures (corrosion-resistant, heat-resistant);
- Materials for special purposes (frictionproof, cold-resistant, radiation-resistant, high- and low electric conducting, high- and low heat conducting, superconducting, soft and hard magnetic metals and alloys, etc.).

Structural strength of metallic materials is a complex operating characteristic, which, along with the *strength criterion*, also includes the *criteria of reliability* and *durability*. The criteria for the material strength under static loads are tensile strength σ_B or the yield stress σ_T, which characterize the material's resistance to plastic deformation. Since plastic deformation is unacceptable for the most metallic working parts, their load-carrying capability is generally limited by yield. For an approximate estimation of the static strength Brinell hardness HB is used sometimes, for example, empirical relationship $\sigma_B = HB/3$ is true for steels).

The magnitudes of the selected strength criteria allow to estimte permissible operating strains, to which issue or device can be subjected in operation. However, improving the strength of the material and, consequently, the permissible operating strains σ is accompanied by increasing of elastic deformations

$$\varepsilon_{el} = \sigma/E, \qquad (3.1.1)$$

where E is modulus of elongation. Hence, to limit the magnitude of possible (and unwanted) elastic stiffness ε_{el} at a predetermined mechanical loading, material should have a high modulus of elasticity (or shear modulus). This requirement is called the criterion of *material stiffness*. Just this criterion (but not strength criterion) is more important when preservation of the exact dimensions and shape of the conjugated components is required during operation.

The opposite claim is also possible. For example, for springs, membranes, dampers, as well as a number of sensitive elastic elements in test instruments, on the contrary, it is important to provide their large elastic displacements. The material in this case must have a high value both σ_{el} and a low E values because $\varepsilon_{el}^{max} = \sigma_{el}/E$.

Reliability criterion reflects the properties of the materials to withstand *brittle fracture*. To prevent brittlness, metallic construction materials must have sufficient ductility, i.e. high values of the relative strain (reduced elongation δ and contraction ψ) before failure, as well as high impact strength.

In accordance with the strength theory, for high-strength metallic materials small cracks either operating or technology origins (for example, generated during welding, heat treatments or intense dynamic loads) are usually sources of fracture. These cracks are usually nucleated on the large-scale defects in the crystalline structure (like dislocation bundles, nonmetallic precipitates, etc.).

For ductile metallic materials danger the existence of cracks is small, since the movement of dislocations at the local plastic deformations around the crack causes relaxation (reduction) of local stresses and their equalization. In brittle

materials, the dislocation motion is blocked so that plastic deformation is impossible. Therefore, with increasing of average (at all issue) strain, the local stresses may be increased so much that will cause breaking of atomic bonds, resulting in the formation and development of cracks. As a result, unlike ductile metals, after reaching a certain critical size spontaneous avalanche growth of crack comes, which is accompanied by the destruction of issue. This destruction is just called by brittle destruction.

Thus, the ability of structural metallic materials to work under operational conditions can be characterized by the following criteria:

- the criteria for strength (tensile strength $\sigma_в$, yield strength $\sigma 0.2$, etc) define the permissible operating stresses;
- the elastic modulus E determines the allowable elastic deformations (stiffness) at a given geometry of detail;
- flexibility (δ, ψ), impact strength KC, the temperature threshold of cold brittleness t50 characterize the reliability of the material in use;
- number of cycles to failure, rates of wear, creep, corrosion determine the durability of materials.

High-strength alloys for power units (turbines, machinery, boilers, pipelines and other constructions) should have the stifness and strength sufficient to limit elastic and plastic deformations, with guaranteed reliability and durability. From the diversity of metallic materials, such iron based alloys as cust iron and especially steels satisfy these requirements to the greatest degree.

Steels have a high inheritability of iron elasticity modulus ($E = 2{,}1 \cdot 10^6$ MPa) and thus its high stiffness. In this, they are only inferior to tungsten, molybdenum and berillium, which are used only in special applications, due to their high cost. High rigidity and affordability cause widespread use of steels for the manufacture of power units, envelopes and metallic constructions in power industry, nuclear power plants and many other systems.

The high stiffness of steels is combined with their relatively high static and dynamic (including cyclic) strength. Their values can be varied over a wide range of values by changing the carbon concentration, alloying elements and microstructure. Last is regulated by methods of heat and chemical-thermal treatments.

Metallic alloys, which are based on copper, aluminum, magnesium, nickel, cobalt and titanium, as well as non-metallic materials (natural rock material, plastics, ceramics, concrete, wood) accede in the strength characteristics of steels. The exceptions are some composite and ceramic materials.

An important advantage of steels are their manufacturability, so that their products are comparable inexpensive. It is because of these advantages, steel is the main construction material in power industry.

Wear-resistant materials. Many movable parts in the power units and control devices are in contact with each other. This leads to their mutual friction resulting in heat release and deformation, as consequence. The last alter the properties of the contacting (conjugated) surfaces of movable parts and components causing premature wear.

The operational capability of materials due to friction of the mating parts is dependent on several factors. *Internal factors* are determined by the type of chemical bonds and also the atomic, nano-, micro- and macrostructure of the material. *External factors* are characterized by the type of friction (sliding, rolling) and the mode of its implementation (the speed of the relative movement of the rubbing surfaces, the magnitude and nature of the load applied to them, temperature). An important factor is also the aggregate state (gaseous, liquid, solid) and the type of *working environment* that separate (lubricates) rubbing surfaces.

Parts subjected to wearing, are usually divided into two categories: (*1*) the details forming a *friction pair* (sliding and rolling bearings, train of gears, etc.) and (*2*) the items, whose friction wearing is caused by the influence of the working medium (liquid, gas, dust, etc.).

The primary reason of wear of the conjugated parts is the work of the friction forces. These forces cause the release of heat and deformation of contacting portions of the detalil surface. This, in turn, will cause oxidation and change of the structure in the surface layers of moving parts, and also their hardening or softening, the development of cracks, etc.

According to the adhesion-deformation model friction of surfaces of the real mating parts is explained by the roughnesses of micron or sub-micron size (Fig 3.1.2). As a result, the actual contact of rubbing surfaces arise only in local areas of the surface of friction pairs (touching spots), where there is a mutual incorporation of micro roughnesses in each other. Therefore, the process of friction has an adhesive and deformation nature, at the same time, so that the full coefficient of friction is determined as the sum of the deformational and adhesion components

$$f = f_{\mathrm{D}} + f_{\mathrm{AD}}. \qquad (3.1.2)$$

Deformational component of friction coefficient f_D is due to embedding of roughness peaks between irregularities resulting in multiple deformation of the surface layers in touching spots of contacting details. *Adhesive component* of friction factor f_{AD} is associated with the formation of direct interatomic interaction, which create the *adhesive welding bridges*. According to this model, the component f_D of friction is defined by three types of deformation interaction of the material in the touching spots: *elastic contacting* (Fig 3.1.1a), *plastic deformation* (Fig. 3.1.1b) and *microcutting* (Fig. 3.1.1c).

Plastic deformation and especially microcutting of the incorporated roughness spikes in the contact zone lead to *wear* of the frictional surface layers in friction pairs. Note that microcutting of surface spikes is also possible due to the presence of hard particles which come from working environment (*abrasion wear*).

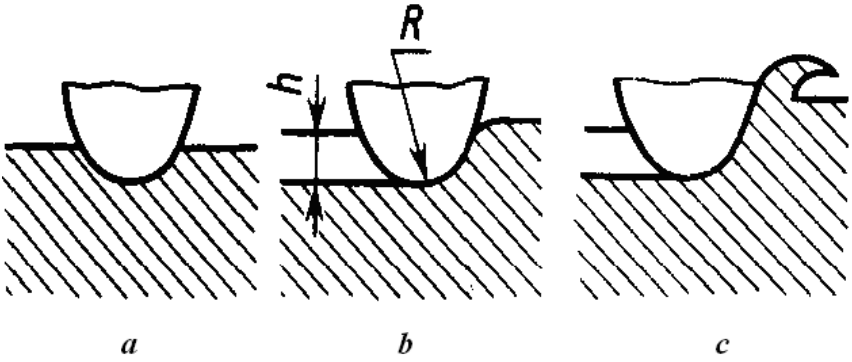


Fig. 3.1.1. Types of deformational interaction between the frictional surfaces: *a* – elastic contacting, *b* – plastic deformation, *c* – microcutting

According to the adhesion-deformation model of friction, there can be two kinds of adhesive interaction: either by smearing and destruction of surface films (Fig 3.1.2a), or by smearing, accompanied with tearing-out of material from the depth of rubbing details (Fig. 3.1.2b).

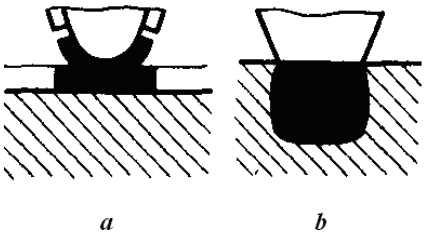


Fig. 3.1.2. Types of adhesive interaction of frictional surfaces: *a* – smearing and destruction of surface films; *b* – smearing, leading to deep tearing-out of material

According to the adhesion- deformation friction model, to improve the wear resistance of the material, it is necessary to increase the hardness of the rubbing surfaces (Fig 3.1.3) and/or the decrease in the strength of the adhesive bonds. *Increased hardness* (and therefore *higher strength*) of the rubbing surfaces hampers their plastic deformation in touching spots, providing an increase in the proportion of elastic deformation of the roughness spikes, and suppresses

microcutting mechanism in the formation of the friction coefficient. *Decrease in the strength of adhesive bonds* allows to avoid roughness smearing mechanism. The most effective way to achieve this goal is the use of lubricants (liquid, solid or gaseous) that share the roughness spikes.

Materials with a low coefficient of friction are called *antifriction materials*. They are widely used for the manufacture of sliding or rolling bearings, which provide reducing of friction loss costs in power industry, quiet operation of machinery and equipment, reduction of heat release in friction pairs, as well as the stability of the moving parts to vibrations.

The main functional characteristics of the materials for production of bearings are antifrictionality (low friction factor), fatigue resistance (resistance to crack formation), high thermal conductivity (for the rapid removal of heat from the friction surfaces), good wettability of lubricants (no ruptures in their hydrodynamic layer), etc.

Among metallic materials, two-phase alloys exhibit high antifrictionality. More often they consist of a soft matrix and solid precipitates of the second phase. Such materials include tin-based alloys (like *babbits*) and copper-base dones (*bronze, brasses*). Among the iron-carbon alloys, gray cast-irons are used in bearings production. They contain a solid iron carbide-based matrix, where soft graphite inclusions (which act as a lubricant) are randomly distributed.

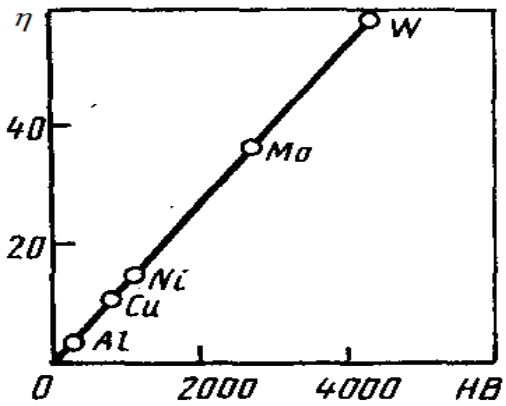


Fig. 3.1.3. Relative wear resistance η of metals with different Brinell hardness

Babbitt metals, and especially bronze have the best antifriction properties, but less resistance to fatigue. Gray cast-irons have a high fatigue strength, but the worst antifriction properties.

Frequently, combined antifriction materials like metal-metal or metal-nonmetal composites (in fact, they include gray cast-iron) are used. For example, in the so-called self-lubricating bearings, iron-graphite, iron-copper, bronze-graphite, as well as of metal-polymer (nylon, teflon, etc.) composites are used.

3.1.2. Alloys with high elastic properties

Metallic materials with high elastic properties are widely used in power engineering, as well as in the manufacture of certain types of test equipment. In power industry, steel is commonly used to manufacture shock absorbers, dampers, power springs for various purposes. In measuring and testing devices special metal alloys are required to produce various kinds of elastic elements (tension bracings, suspensions, membranes, springs, relay plates, bellows, flexible corrugated pipes, etc.).

Main features of metallic elastic elements is that the residual deformation under high static or cyclical shock loads is fully excluded. Therefore, all the spring alloys should have (besides the usual mechanical properties like strength, ductility, toughness, fatigue endurance) a high resistance even to small plastic deformations.

When short static loading, resistance to small plastic deformations is characterized by the *elastic limit* (see Chapter 1). Its value is characterized by the slope of the linear part Ob in the strain diagram $\sigma(\varepsilon)$ in Fig. 3.1.4, which should not depend on the duration of the load. However, during prolonged compression or tension of spring or bending of relay contact plate, residual deformation may retain (ε_{res}) when load is fullly removed. This means that part of the elastic deformation during the load becomes plastic and linear strain plot diagram in Fig . 3.1.4 shifts (from the line Ob to the line $C\varepsilon_{res}$) .

Therefore, during prolonged static or cyclic loading, the material of elastic element (e.g., a spring) should be characterized by *relaxation resistance*, which is defined as *resistance to stress relaxation*. The last is estimated by reducing of operating strains after repeated or long-term loading of elastic element from the value σ_1 to the value σ_2, which is necessary for achieving a given elastic deformation ε_1 (Fig 3.1.4).

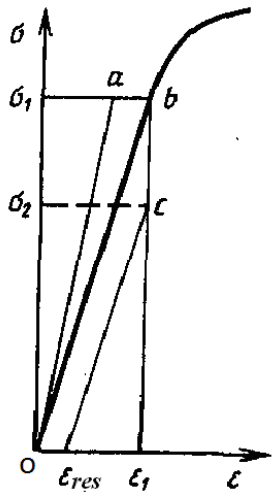


Fig. 3.1.4 . Strain diagram σ(ε) explaining the relaxation and elastic aftereffect

Relaxation of stresses is dangerous because elastic elements change size and shape due to the partial transition of elastic deformation into plastic one after unloading. This, in particular, increases the measurement error of the corresponding quantities in the testing equipment.

The physical mechanism of the elastic properties relaxation in the materials is due to the fact that the elastic elements are manufactured of polycrystalline materials. In this case, the stress relaxation during prolonged load occurs by microplastic deformation in the individual grains, which are accumulated with time. Even if the applied load create stress lower than the elasticity limit, microplastic deformation in the grain may occur even at low stresses due to either bending of dislocations or breakdown of some of them from the precipitates of strengthening phases. At the enhanced stresses, a lot of dislocation, which was blocked at low stresses, can break away and start moving at the same time.

In this regard, in order to achieve high values of elastic limit and relaxation resistance in the alloy, we should block movement of practically all the dislocations. Furthermore, this structure should not have zones with a high level of initial local stresses (e.g., due to cracking). Otherwise, the latter, being summarized with external stresses during operation of the elastic elements, can lead to local displacement of dislocations in the mentioned zones (i.e. relaxation effect).

One of the most important operational requirements for materials, which are used in the manufacture of elastic elements, is stable linearity and

reproducibility of the deformation characteristics $\sigma(\varepsilon)$, Fig. 3.1.5. If this characteristic is not linear, it is impossible to provide the necessary and constant accuracy of the instrument over the entire range of measured values.

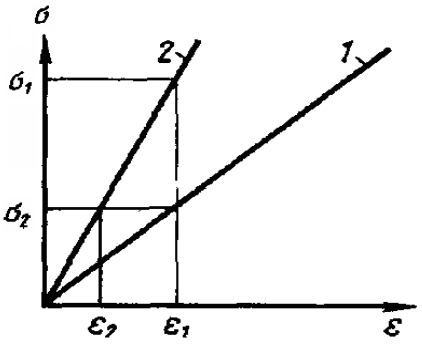


Fig. 3.1.5. Deformation characteristics of the bronze (*1*) and steel (*2*) elastic elements

The quality of the elastic element is also defined by the load needed to create elastic deformation of a certain magnitude. Therefore, it is desirable that $\sigma(\varepsilon)$ dependence allowed as much as possible reversible movement of the elastic element parts (for example, the the spring compression), because the bigger it is at the determined effort, the greater the sensitivity of the elastic element. The maximum achievable elastic deformation of element is equal to $\varepsilon_{max} = \sigma_{0.002}/E$ ($\sigma_{0.002}$ is apparent elasticity limit), and hence its sensitivity is determined by the elasticity modulus (a slope of the linear part of the $\sigma(\varepsilon)$ in Fig. 3.1.5). As can be seen, for the same stress σ_2 value elastic displacement of a first spring element is greater than for the second one (because $\varepsilon_1 > \varepsilon_2$). This figure shows that bronze spring (line 1 in Fig. 3.1.5) provides higher sensitivity and relatively lower measurement error than steel spring.

3.1.3. Alloys with low density

Light structural metallic materials are widely used in various power units (e.g., in aerospace engines) that requires both low density ρ and high strength. In this case, the main criteria for selecting the construction materials are the values of *specific strength* $\sigma_B/(\rho_g)$ and specific stiffness $E/(\rho_g)$, where g is gravity acceleration. These materials include primarily non-ferrous metals (magnesium, beryllium, aluminum, and titanium) and a number of alloys on their base (see Table. 3.1.1). Among non-metallic materials various composites are related to the materials with high strength and low density (see below).

Table 3.1.1. Strength, specific strength and specific stiffness of light metallic materials

Material	σ_B, MPa	(σ_B/ρ_g), km	$E/(\rho_g)\cdot 10^{-3}$, km
Magnesium-based	430	21	2.3
Aluminium-based	700	23	2.4
Titanium-based	1500	30	2.6
Stainless steel	2000	27	2.6
Berillium	680	37	6,1

The most lightweight metal – aluminum – has a low strength. Virtually all of the properties (including mechanical) are determined by its purity and structural state. When the concentration of impurities and structural defects arise during the production of aluminum (due to plastic deformation), its strength, hardness, elasticity increases. Due to the low strength aluminum is used for details and elements of designs which do not subjected to high loads, but requires lightness and high technological effectiveness (weldability, ductility). Therefore, only piping, heat exchangers, tanks for transportation of light liquids, some of the building envelope (windows, frames, doors, light reflectors, etc.) are produced from commercially pure aluminum. Due to the high thermal conductivity, it is used in industrial refrigerating systems. The high electrical conductivity of aluminum makes it indispensable in electrical applications (electrical wires, cables, tires, capacitors, etc.). High reflectivity in the visible region of the spectrum can be used in the manufacture of aluminum reflectors (powerful lamps, headlights), reflecting screens, etc. Since aluminum has a low effective neutron capture cross-section, it is used in the manufacture of walls of nuclear reactors. With appropriate doping, aluminum alloys gain a high specific strength, including high resistance to dynamic and inertial loads. In aluminum alloys tensile strength reaches 500-700 MPa at a density of not more than 2850 kg/m^3.

As can be seen from Table. 3.1.1, the specific strength of some aluminum alloys reaches value (σ_B/ρ_g) = 23 km, which is close enough to the properties of high-strength steels (27 km). Many aluminum alloys have high corrosion resistance (with the exception of copper alloys), and also high thermal and electric conductivities. From Table. 3.1.1 also follows that aluminum alloys have a significantly higher specific strength, ductility, and are more resistant to corrosion than magnesium alloys.

Aluminium alloys are classified according to various criteria: by the technology of production (wrought, casting, sintered), the ability to heat

treatment (heat-treatable, non-heat-treatable), industrial applications (electrical engineering, thermal engineering, low-density), and the like. The most common heat-treatable aluminum alloys include copper and magnesium as dopants. These ternary alloys, called *duralumin*, are characterized by acceptable combination of strength, ductility and corrosion resistance. Fig. 3.1.6 shows that an increase of magnesium in the aluminium-copper alloys causes an increase in their strength. Hence duralumin is used in structural elements subjecting to relatively high mechanical loads, but working mostly under compressive stresses.

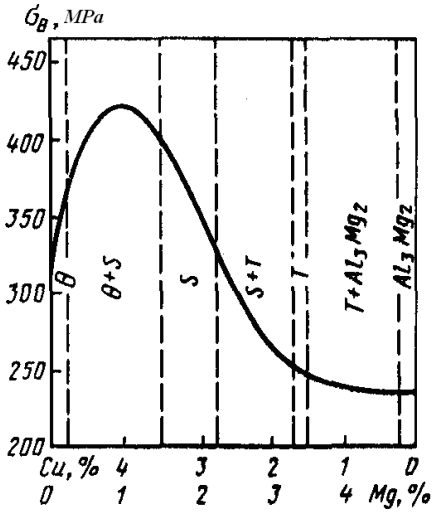


Fig. 3.1.6. Dependence of duralumin strength on the Cu/Mg ratio as their constant content 5 wt.%

Light eutectic alloys of the Al-Si, called *silumins*, have good mechanical properties. Density of the most silumin alloys is close to 2650 kg/m^3 that is slightly less than for pure aluminum (2700 kg/m^3). Doped silumin alloys are used for production of cast parts of critical details of medium and large sizes (compressor housing, crankcase, and valve caps in internal combustion engines, etc.).

Magnesium and its alloys have not only a low density, but also a good cutability (unlike aluminum), ability to load impacts and also suppress vibrations. However, the thermal and electrical conductivities of magnesium is lower than that of aluminum (or 1.5-2.0 times lower). Modulus of elongation for magnesium is about 1.5 times smaller than that of aluminum, although the specific stiffness is close to aluminum. The main advantage of magnesium alloys is their high specific strength. Tensile strength of some magnesium alloys

reaches values of 250-400 MPa at a density of less than 2000 kg/m^3, which gives the values of $(\sigma_B/(\rho_g)$ ~ 20 km (see Table. 3.1.1). Along with a low corrosion resistance, magnesium alloys have also other disadvantages: low modulus of elasticity, low casting properties, the tendency to saturate gas, oxidation and inflammation.

However, high specific strength of individual magnesium alloys contribute to their wide application in the construction and power units of aerospace vehicles (aircrafts, missiles), as well as cars. As a result of high shock-absorbing capacity, magnesium alloys are useful for the manufacture of testing equipment packages. Low capacity to absorb heat neutrons makes these alloys useful for nuclear technologies.

3.1.4. High-resistant metallic materials

Constructional metal materials undergo surface oxidation during heat treatments and exploitation in air at elevated temperatures that can cause their subsequent destruction. The ability of metals and alloys to resist by the corrosive effect of oxygen at high temperatures is called to *heat resistance*.

The oxidation of metals and metal alloys includes several successively proceeding stages. The first step is adsorption of oxygen from air by atoms of surface metal layer. In the The heat, generated during adsorption, causes the dissociation of oxygen molecules into atom constituents. This process is accompanied by the redistribution of electrons and ionization of both metal and oxygen atoms in accordance with the reactions

$$Me \rightarrow Me^{2+} + 2e; \quad (3.1.3)$$

$$O + 2e \rightarrow O^{2-}; \quad (3.1.4)$$

$$Me^{2+} + O^{2-} \rightarrow MeO. \quad (3.1.5)$$

As can be seen, the surface metal atoms lose electrons in accordance with the oxidation reaction (3.1.3), whereas oxygen atoms are capturing electrons at the reduction (3.1.4). Chemical oxidation process is completed by Coulomb interaction of ions (3.1.5) to form oxides at the surface.

The possibility of spontaneous oxidation is determined by the sign of the change of a standard thermodynamic potential ΔG_T^0 in the reaction (3.1.3) at a given temperature (affinity for oxygen): if $\Delta G_T^0 < 0$, the oxidation is possible.

The values of ΔG_T^0, listed in Table. 3.1.2, show that gold and silver at room temperature are not oxidized. In alloys containing chromium and nickel, the formation of chromium oxides is more likely.

Table 3.1.2. Standard thermodynamic potentials ΔG_T^0 for the oxidation reactions

Oxide	Au_2O	Ag_2O	NiO	Cr_2O_3
ΔG_T^0, kJ/Mole	+3	+1.2	- 24.1	- 40.5

The rate of oxidation depends on the density of the forming surface oxides. If oxides are tight, they hinder further oxidation, thus exerting a protective effect. The protective properties of oxides are determined by the coefficient of φ, which is the ratio of the volume occupied by the mole of oxide and mole of metal. Protective properties have only oxides having the values $1.0 < \varphi < 2.5$. When $\varphi < 1$, oxide is friable, so that access of oxygen to the metal surface remains free. When $\varphi > 2.5$, oxide can cracking and peeling from the metal surface under the influence of large internal stresses that arise due to the large difference in volumes of oxide and metal. As a result, the surface is re-oxidized.

The oxide films formed at the surface of metal at room temperatures are called *natural*. The crystal lattice of these oxides is close to the lattice of metal and therefore, despite their low thickness (3-10 nm), they have good *protective properties*. Oxides of aluminum, nickel and chromium are examples of thin dense oxides. When metal heating, oxide thickness increases, and its crystallographic structure is changed. In this case, although the oxide lattice, directly neighboring with the surface of metal, is different from its lattice, they have dimensional and structural matching. As far as distance from the metal surface increases, the oxide crystallographic structure becomes increasingly different from the structure of metal approaching the crystalline lattice of bulk oxide.

Metal oxides are generally intermediate phase with ionic bonds. The stoichiometric composition of oxides, that is reflected by the chemical formula, can be broken, leading to an excess of either metal or oxygen ions. This results either in the excess of metal ions, located in interstitials of the crystalline lattice, or vacancies due to the lack of oxygen ions.

The more defective crystal lattice of oxide with variable composition due to the presence of vacancies or interstitial ions, results to lower protective properties of the oxide. Oxides CuO, Al_2O_3 and Cr_2O_3, which are practically undamaged, have the best protective (high-resistant) properties.

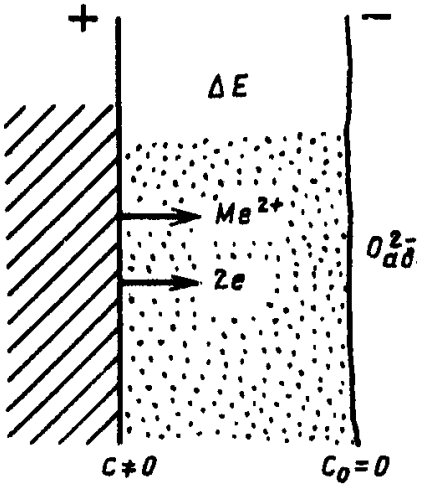


Fig. 3.1.7. Scheme of growth of dense oxides on the metallic surface

Oxidation rate is higher for the friable metal oxides, since the movement of oxygen ions to the metal surface is facilitated. In dense oxide ion transport (typically metal ions) is more difficult. It is provided by a variety of physical processes depending on the oxidation temperature (Fig 3.1.7).

Heat resistance of metals. Usually two types of oxidation are recognized: low temperature and high temperature, which are determined by different mechanisms of oxide growth.

At low temperatures ions move by migration in electric field, which occurs near the surface. This field is due to the transition of free electrons from a metal atoms to oxygen atoms in the oxide at the outer surface which is in contact with the external environment.

At elevated temperatures, when the thickness of the oxide increases, the distance between the inside (at metal surface) and the outer surfaces of the oxide becomes much larger, so that the electric field strength decreases rapidly. As a result, the field migration of ions is stopped, and their movement is ensured by diffusion due to the difference in the concentrations of metal ions on the interior and exterior surfaces of oxide layer. The oxidation rate in this case is somewhat higher than the migration in an electric field, but also small compared to the rate of oxidation in the presence of loose oxides.

The dependence of the oxidation rate *dh/dt* on temperature is described by the equation

$$dh/dt = (dh/dt)_o \, exp[- E/(RT)], \qquad (3.1.6)$$

where h – the thickness of the growing oxide, R – universal gas constant, and E – activeion energy of oxidation. In the Arrhenius coordinates $ln(dh/dt)$-1/T (see Fig. 3.1.8), this dependence looks like a straight line whose slope determines the rate of oxidation of E. The oxidation rate is estimated either by the change of metal mass [g/m^2·hour], or the increase of oxide film thickness h [μm/hour).

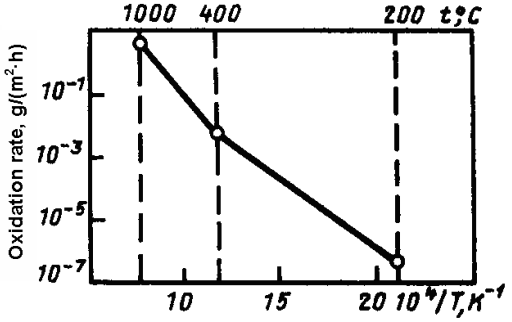


Fig. 3.1.8. Effect of temperature on the rate of oxidation of copper on air

For the metals, in which chemical composition and structure of the oxides is changed when heating, the temperature dependence of the oxidation rate has several linear parts with different slopes. Fig. 3.1.8 presents the temperature dependence of the rate of copper oxidation in air. It follows that in the temperature range 200-400°C, the rate of copper oxidation is within the limits 10^{-7}-10^{-3} g/m^2·h, so that a sufficiently dense (non-defect) oxide CuO is formed. At T > 400°C, more intense oxidation of copper begins to form a loose oxide Cu_2O, which replaces flawless oxide CuO. The Cu_2O has a higher concentration of vacancies in the cation sublattice oxide due to the deficiency of metal ions as compared with the incoming oxygen.

The temperature dependences of the rate oxidation allow to estimate the heat resistances of metallic materials and a maximum operating temperature at their exploitation. Both characteristics determine the possibility of metal to work at a given operating temperature and duration of the exploitation.

Heat resistance of metal depends on the external factors (temperature, composition air environment, the speed of its motion, the oxygen partial pressure) as well as the internal factors (chemical composition, surface state and structure of metal, like chemical bonding, grain size, etc.).

Comparative evaluation of the rate of oxidation in air for pure metals in the accessible working temperature range are given in Table 3.1.3.

Table 3.1.3. Heat resistance of pure metals

Group	Heat resistance	Metals	Governing factor
1	Very poor	Mg	Loose oxides
2	Poor	Nb, Ta, Mo, W, Ti, Zr	Dense oxides with poor protective properties
3	Satisfactory	Cu, Fe, Ni, Co	Dense oxides with great defectness
4	Good	Al, Zn, Sn, Pb, Cr, Mn, Be	Dense oxides with high protective properties
5	Excellent	Ag, Au, Pt	Low affinity to oxygen

Heat resistance of the alloys. The basic requirement to all doping elements to improve the heat resistance of metals is their higher affinity for oxygen than the parent metal. Only in this case, the doping element can increase the heat resistance of the metal. Therefore, metals such as Ag, Au and Pt (Table 3.1.3), which themselves have high heat resistance, can not be used as alloying elements.

The enhanced heat resistance of the low-doped steels and alloys is usually caused by the fact that the alloying element *B* is incorporated into the crystal lattice of oxide of basic metal A (iron). This results in the formation the doped oxide $(AB)_mO_n$ with the lowered defectness of its crystal lattice.

Metals, whose oxide has an excess of ions in interstitials (Fig. 3.1.9a) should be doped by elements with higher valence. In this case, to maintain electrical neutrality, ions $3A^{2+}$ will be replaced by $2B^{3+}$, resulting in decrease in the amount of metal ions in the interstitials. Metals having oxides with the deficiency of metal ions Me in the lattice (Fig. 3.1.9b), on the contrary should be doped with elements of lower valence. In this case, change A^{2+} ions on $2B^{+}$ will reduce the number of cation vacancies.

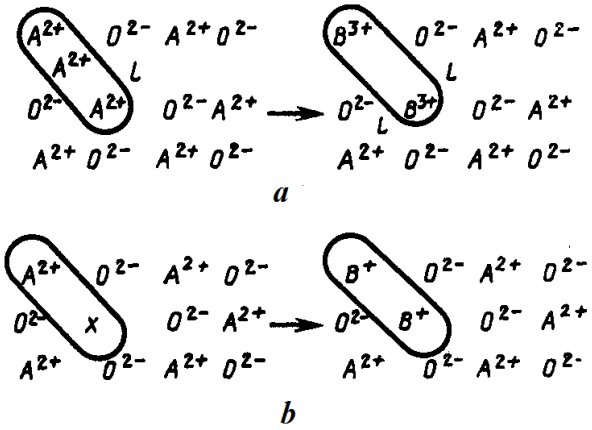


Fig. 3.1.9. Scheme of changes in the concentration of defects in the doped oxide with an excess of metal ions (*a*) and their deficiency (*b*)

Good heat resistance of heavily alloyed steels and many doped alloys is attributed to the fact that the alloying element (*B*) forms its *own oxide* B_mO_n which has the best protective properties than the main oxides A_mO_n (*A* – main alloy component). For the formation of native oxide, alloying element *B* has to have not only greater affinity for oxygen, but its concentration and diffusion mobility must provide the formation of oxide which is uniform by thickness across all oxidized surface. The doping element oxide itself should be tight, do not crack, has a low concentration of electrons (low conductivity), high temperature of sublimation and melting point, and does not form a low-melting eutectic phases.

High heat resistance of heavily alloyed steels and doped alloys on the basis of transition (refractory) metals is often caused by the formation of double oxides like $AO \cdot A_2O_3$ with spinel structure. In this structure, the oxygen ions are arranged in lattice sites, while the ions of bivalent and trivalent metals having a smaller ionic radii occupy interstitial sites. Variable composition of these oxides is due to the presence of empty interstices, allowing migration of metal ions.

High protective properties of double oxides are usually attributed to high-density packaging (as the interstitial pores are occupied by metal ions) and small crystal lattice parameter. Dopant ions B, partially or completely replacing the base metal ions *A*, lead to a decrease in the diffusivity and electric conductivity and thus improve the protective properties of the oxide.

Heat resistance of industrial *aluminum alloys*, except Al-Mg alloys, is practically as high as that of pure aluminum, because the aluminum affinity to

oxygen is higher than that occurring in these alloys based on Cu, Zn, Si and Mn. Some deterioration of heat resistance in alloys with a multiphase structure is caused, in this case, by failure of nonstoichiometric composition of alumina (Al_2O_3) and its heterogeneity by thickness. Aluminum alloys, containing magnesium, are inferior in heat resistance to pure aluminum due to the formation of loose MgO oxide on its external surface.

Magnesium alloys also exhibit increased heat resistance due to doping. Doping with beryllium (up to 0.02-0.05 wt.%) improves heat resistance of magnesium and eliminates spontaneous combustion when processing. Magnesium alloys with Mn, Zn and A1 also have a higher heat resistance than tat in pure magnesium.

Heat resistance of brass and bronze is always higher than in pure copper. This is because the elements of the fourth group in Table. 3.1.3 have a greater affinity for oxygen than copper. Therefore, at sufficient concentration, they form own oxides when heating, which have a better protective properties than Cu_2O. As a result, alloys of copper with Be, A1 and Mn have a higher heat resistance than copper alloys with Zn, Sn and Si.

Titanium alloys absorb oxygen more actively than pure titanium, that does not promote the formation of tight surface oxides with protective properties. Therefore, heat resistance of doped titanium is not improved neither for α-solid solutions, nor for eutectic ($\alpha+\beta$)-alloy phase. In this case, they can improve the heat resistance only by applying special heat-resistant coatings.

Heat resistance of iron and steel is generally increased by doping with chromium, aluminum and silicon. For example, increasing the chromium concentration in the steel to 30 wt.%, results in the formation doped iron oxides, which leads to an increase in heat resistance with the increasing the temperature and duration of heat treatment, as well as decrease of partial oxygen pressure.

Low carbon steel at high chromium content acquires single-phase structure of ferrite. In the process of long-time work at high temperatures, crystallite sizes of ferrite are increased, that is accompanied by embrittlement due to intergranular carbides formation (a decrease of impact strength). To prevent the growth of grain size of ferritic steel, they are additionally doped with titanium as carbide-stabilizing element.

Heat resistance of refractory (transition) metals (such as Mo, W, Ta and Nb) is very low due to the friability of oxides formed. The use of vacuum and protective environments during processing and operating of refractory metals, in some cases, causes some technical difficulties. Bulk doping of refractory metals does not lead to the enhancement of heat resistance, while it may be effective for the enhancement of high-temperature strength. Higher heat resistance of these metals is usually achieved through the use of refractory coatings.

3.1.5. High-temperature strength of metallic materials

Many of parts in modern power units and systems (boilers, steam-powered machines, rocket and cars engines, turbines, etc.) are not only heated in the process of high temperature operation, but also are subjected to high mechanical loads. This requires a materials with high strength at elevated temperatures. High temperature, the magnitude of the mechanical stresses applied and the duration of the load have cruicial importance for the choice of the structural material. For example, steam-power systems work hundreds of thousands of hours, while rocket engines – only a few minutes. In some cases, the temperature of the detail parts is very high and close to the melting point of the material, while for the other – they are only significantly above the recrystallization threshold.

The main mechanisms of the effect of temperature on strength properties of metallic materials are the change of interatomic forces and the intensity of diffusion processes. Raising the temperature reduces the interatomic bonds in the crystals, thus decreasing the elastic modulus, tensile strength, yield strength and hardness. Intensification of diffusion processes at temperatures of the order of $0.3T_m$ leads to a significant change in the number and type of defects in the crystalline structure and promotes the recrystallization. The last results in the increase of grain size, and consequently leads to a further degradation of the strength characteristics of the material. Therefore, as follows from Fig. 3.1.10, lower melting point T_m of the metal material, the lower its permissible operating temperature, at which it can be used in load-bearing units.

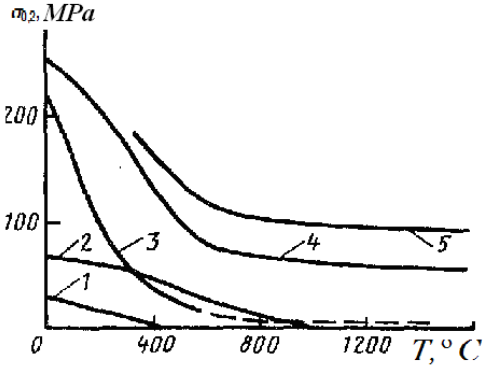


Fig. 3.1.10. The dependence of the yield strength of metals on temperature:
1 – aluminium, *2* – copper, *3* – titanium, *4* – tantalum, *5* – tungsten

Thus, when operating at high temperature, the material should satisfy the requirement of constancy of the mechanical characteristics (or only their small deterioration) as compare with those at normal temperatures.

High-temperature strength is a capability of the material to withstand deformation and destruction for a long time at elevated temperatures. The high-temperature strength is important in choosing the material, when the part, made of this material, works at temperatures above the recrystallization threshold (about $0.3T_m$ for pure metals). Inasmuch as the material behavior at $T > 0.3T_m$ is determined by the diffusion processes, processes of creep and stresses relaxation are characteric for these conditions.

Creep presents slow increase in time t of relative plastic deformation ε when the metallic sample is subjected to constant load $\sigma < \sigma_Y$ (yield strength) at a constant temperature T (Fig. 3.1.11). It is seen that $\varepsilon(t)$ curve, in this case, consists of three parts, each of which corresponds to a specific stage of creep. Stage I (part ab of the curve) is called a *non-steady-state creep*, when the deformation rate ($d\varepsilon/dt$) is reduced to some constant value. Stage II (part bc) is characterized by the *steady-state creep* with a constant deformation rate (linear part of $\varepsilon(t)$ curve). At stage III (part cd) deformation rate starts to increase progressively up to the moment of the sample destruction at the point *d*. Time of occurrence and duration of each stage depends on T and σ. If σ = const, temperature rise reduces the duration of the linear stage II and accelerates the destruction. The increasing of stress affects similarly on the $\varepsilon(t)$ curve during the test a at constant temperature.

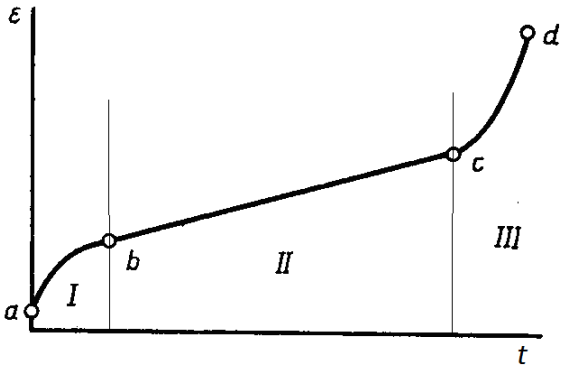


Fig. 3.1.11. Creep curve at a constant external stress (load) σ and temperature T:
I – non-steady-state stage; *II* – steady-state stage; *III* – stage of destruction

As a criterion of high-temperature strength of materials, the creep limit and long-term strength are used. *Creep limit* is the stress, under the influence of which the material is deformed by a certain amount for a certain time at a given temperature. The symbol of creep limit includes the temperature of measurement (in °C), strain (in %) and time of measurement (in hours). For

example, the creep limit $\sigma^{550}_{1/10000}$= 100 MPa means that the application of the stress of 100 MPa for 10,000 h at 550 °C will create plastic deformation ε = 1% in the marial tested.

Under *long-term strength* we understand the stress, which causes destruction of the material at a given temperature (in °C) for a certain time (in hours). For example, the limit of long-term strength σ^{600}_{10000}= 130 MPa means that at a temperature of 600°C, the material is destroyed at the stress of 130 MPa after 10,000 hours of work. The long-term strength is always less than the tensile strength σ_B, which is defined at short-time tests at constant temperature because it describes the material in the long-time of loading process at high temperatures.

Creep of metals causes *relaxation of stresses* in pre-stressed parts. This is manifested in the fact that at high temperatures creep causes the deformation of parts. This, for example, may cause a gradual reduction of stresses on the fasteners (nuts, bolts), loss of tension of wires and cables, etc. The criterion for relaxation resistance is the reduction of stress $\Delta\sigma = \sigma_o - \sigma_t$ for time t at a given temperatour T.

Motion of dislocations at a temperatures above the recrystallization threshold (~ $0.3T_m$ for metals) occurs due to sliding them as a whole and (or) their climb ofof their parts. When heating, the process of dislocation slip is enhanced because of the disappearance of Cottrell atmospheres (due to the solubility of impurities) and accelerating the diffusion inflow of vacancies.

Grain boundary sliding represents a shift of grains with their rotation relative to each other along the joint boundaries (see Part 2) under the action of tangential stresses. In so doing, slip deformation is greater the grains smaller (due to the higher density of grain boundaries).

Diffusive transport is associated both with the motion of vacancies along grain boundaries (it is enhance, see Part 1), and by grains bulk. Under the influence of external tensile stresses the energy of vacancy formation decreases. Therefore, concentration of vacancies between two boundaries is increased (Fig. 2.17) and they begin to move to the areas with their lower concentration.

The above basic mechanisms of high-temperature creep indicate that to improve the heat resistance of metallic materials we need to stop dislocation motion and to slow migration of atoms at higher temperatures. This can be achieved by increasing the chemical bonds strength and barriers hindering the motion of dislocations. The main way to increase heat resistance is to block the movement of dislocations in the material by creating large-grained structure with a uniform distribution fine reinforcing phases (precipitates) within the grains and at their boundaries.

In steels and alloys used in power units and systems, such multiphase structure with fine particles of hardening phases is obtained by quenching and subsequent tempering (aging). To obtaine optimal structure (large grain size, and the needed amount small sizes of hardening particles) in heat-resistant steels and alloys, a complex doping is used. These materials are more complicated by chemical composition than ordinary alloyed steels and alloys.

Strength of the interatomic bonds in the most metals falls with the temperature and therefore is insufficient to provide heat resistance. Although doping of metals and creation of solid solutions increases heat resistance, but the effectiveness of such hardening is small and is conserved only to the temperatures of the order of $(0.6\text{-}0.65)T_m$. Therefore, the high temperature strength of materials with the structure of the solid solutionsis less as compared with the ones reinforced with fine particles of solid phases.

Hardening phases in heat-resistant steels are generally iron carbides and (or) the alloying elements. Efficiency of the material hardening is determined by strength, mechanism of the formation, size, nature and distribution of the hardening phase precipitates in the material: the smaller their sizes and higher density, the stronger their ability to block the movement of dislocations and hence higher heat resistance effect. The larger grains, the lower the density of grain boundaries and weaker grain boundary sliding and grain-boundary diffusion.

To improve the heat resistance of polycrystalline steel and alloys, special elements (boron, cerium and other rare earth metals) are introduces in a small amount (about 0.1-0.01 wt.%). They are concentrated at grain boundaries and slow grain boundary sliding. Grain boundaries in nickel alloys are reinforced by carbides which are introduce by adding of about 0.1 wt.% C.

Basic heat-resistant superalloys. Among the most common heat-resistant iron-based alloys (mainly steel), and nickel are apreaded. *Heat-resistant steel* include perlite, martensite and austenite in their structure, that allows their use at temperatures up to 700°C. Below 450°C ferritic steels are used.

In power engineering *pearlitic steels* are mostly spreaded. They are intended for long-time operation at temperatures of 450-580°C. Heat resistance of this type of steels is secured by rational doping ferrite, as well as homogeneous distribution of carbides in it. In pearlitic steels high temperature strength criterion is creep strength with allowable deformation not higher 1% for times not less than 10^4 - 10^5 hours.

Steel containing 0.12-0.15 wt.%C, are used mainly in steam power units for the production of boilers, pipesof steam superheaters, steam pipelines and other items for operating temperatures not exceeding 570-580°C.

Nickel superalloys are more heat-resistant and corrosion-resistant than steel, but much more expensive. These alloys can be operated at temperatures of 700-1000°C and therefore used in the manufacture of turbine blades and critical parts of modern gas-turbine engines. The most of nickel-based heat-resistant alloys containe about 10 wt.%Cr. To further improve the heat resistance of products based on these alloys additional thermochemical treatment of surface is used.

Light superalloys are made on the base of aluminium, magnesium and titanium. They are lighter than steel but less heat-resistant. They can therefore be used at substantially lower temperatures: operating temperatures of aluminum and magnesium alloys do not exceed 300-350°C, and titanium – 500-600°C.

For operating temperatures above 1000 °C, materials on the base of refractory transition metals (vanadium, niobium, tantalum, chromium, molybdenum, tungsten) and their alloys, which melting point exceeds 1700°C, are used. However, at high temperatures all the essential refractory metals (except chromium) have low corrosion resistance due to rapid oxidation.

3.1.6. Corrosion resistant metals and alloys

When operation in hostile environments structural materials must have not only appropriate mechanical properties, but also high resistance to corrosion.

Metal corrosion is spontaneous destruction of metallic materials due to their chemical or electrochemical interaction with the environment. *Corrosion-resistant* metals and alloys are capable to withstand corrosive effect of the medium, i.e. when the corrosion process is developed with low speed.

Classification of corrosion processes. Depending on the properties of the environment in which an issue works, and the character of physical-chemical impacts on the material, it is accepted to distinguish *chemical* and *electrochemical corrosion*. The most common type of corrosion is a chemical gas corrosion at high temperatures due to interaction of issues with oxygen or active gas media (halides, sulfur dioxide, hydrogen sulfide, sulfur vapor, CO_2, etc.). Gas corrosion destroys such important components and power units as gas turbine blades, nozzle jet engines, boilers, walls of pipelines, etc. The formation of heat scale from iron oxide is an example of this type of chemical corrosion in iron-based alloys.

Electrochemical corrosion is developed due to contact with metals and alloys of fluids possessing properties of the electrolyte. Galvanic corrosion occurs when the electrical contact of two metals forming the so-called galvanic couple. *Galvanic couple* (see Fig. 3.1.12) is formed from metals in the electrolyte which have different electrode (electrochemical) potentials E_0

(see Table 3.1.4). Metal having more negative electrode potential is called *the anode*. It is dissolved because is giving positive ions in solution. The metal having the higher electrode potential, is called *the cathode*, because accepts the excess electrons from an external circuit. In doing so, this cathode is not destroyed. The lower metal electrode potential E_0 relative to standard hydrogen potential, taken as the zero level, the easier metal gives ions to the solution and the lower the corrosion resistance.

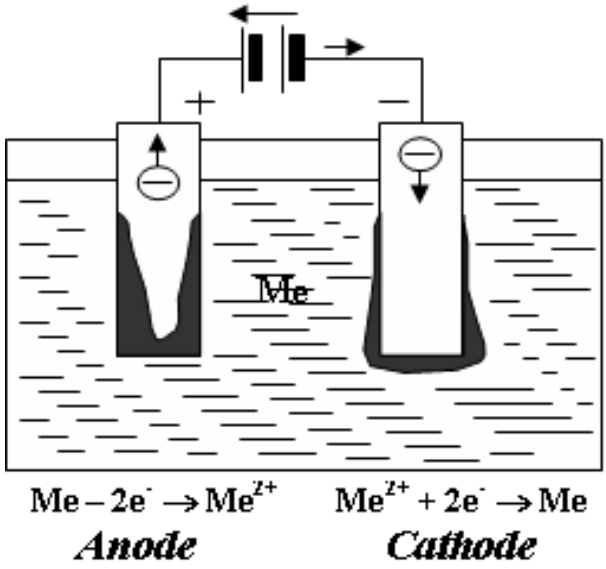


Fig. 3.1.12. Schematic representation of the electrochemical corrosion process

Table 3.1.4. The values of the electrode potential E_o for different elements

Element	Mg	Al	Zn	Cr	Fe	Co	Ni	Sn
E_0, V	-2.37	-1.66	-0.76	-0.74	-0.44	-0.28	-0.25	-0.14
Element	Pb	H	Cu	Hg	Ag	Br	Au	
E_0, V	-0.13	0	+0.34	+0.79	+0.80	+1.06	+1.50	

Progress of the electrochemical process is determined by the potential difference between the cathode and the anode. For example, for a galvanic pair of Cu-Zn the potential difference is 1.1 V. In the case of closed circuit, copper in this pair is a cathode and zinc and is an anode (it dissolves). When an

external voltage above 1.1 V is applied to this galvanic couple, the anode and cathode are interchanged, resulting in the dissolution of copper and recovery of metallic zinc.

The foregoing scheme of galvanic couple underlies most of electrochemical corrosion in multiphase alloys. In this case the role of the anode and cathode play different phase components presenting in the alloy structure. Being in a liquid medium, which has properties of the electrolyte, these phase components gain different electrode potentials both by magnitude and sign. The greater the difference in the electrode potentials of the individual phase components, the more will carry corrosion of alloy.

Thus, the main cause of corrosion is an electrochemical heterogeneity of metallic surface. Alloys with a uniform solid solution structure are more corrosion resistant. However, even in such alloys galvanic corrosion can occur due to the formation of anodic areas, due to the presence of large-scale structural defects (grain boundaries, dislocation pileups, Cottrell atmospheres around dislocations, intragranular liquation, etc.).

Schematic representation of the main varieties of electrochemical corrosion on example of duralumin is shown in Fig. 3.1.13 . As can be seen, in the case of single-phase uniform solid solution corrosion is expanded homogeneously (by depth) over the entire surface of the workpiece (Fig. 3.1.13a). This type of corrosion is called *general* or *uniform corrosion*. This means that the thermomechanical treatment, which leads to homogenization of the alloy structure, makes it more stable against electrochemical corrosion.

In the heterogeneous (multiphase) metallic materials corrosion *is localized*, covering only those portions of the anode surface, which have a low values of the electrode potential (Fig. 3.1.136). «Spots» of such corrosion create stress concentrators that promote nucleation of cracks (see Part 2).

Grain and interphase boundaries in heterogeneous alloys are the anodes, while grains (phases) themselves are cathodes. In this case *intergranular corrosion* is developed, which is the most dangerous. It is almost invisible from the surface and extends into the metal depth along the grain boundaries (Fig. 3.1.13c). As a result, intergranular corrosion results in arising of pores, leading to cracks nucleation, so that material is easily disrupted after application of the load.

Simultaneous exposure of corrosion media and mechanical load (including in the form of residual stresses) in the material occurs stress corrosion, which is manifested in the form of stress corrosion cracking (occurrence of cracking net). The parts of alloy, where a concentration of stresses are presented, are characterized by lower values of the electrode potential, which imparts them the

anode-like properties. High residual stresses cause enhanced corrosion of welds, as well as all sorts of fastening elements (bolts, rivets, nuts, etc.).

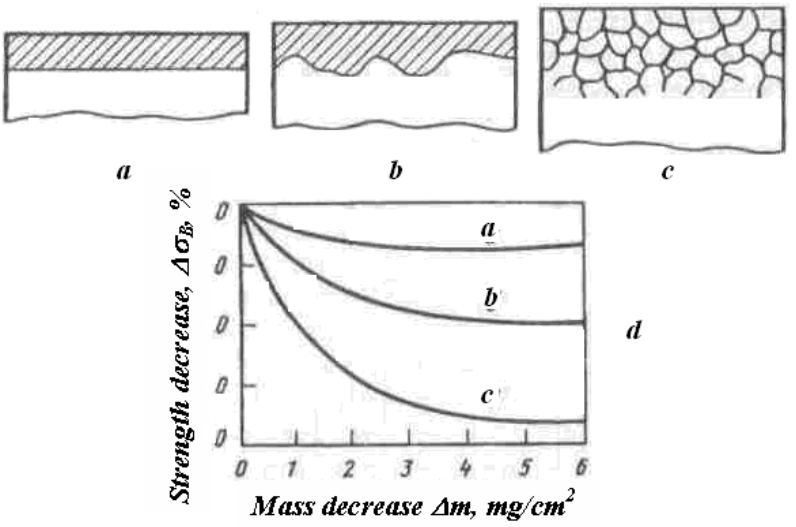


Fig. 3.1.13. Schematic representation of the main types of electrochemical corrosion (*a*, *b*, *c*) and their relative impact on strength in duralumin (*d*): *a* – uniform corrosion, *b* – localized corrosion,*c* – intergranular corrosion

3.1.7. Metallic materials with special thermal properties

Many power technologies demands metallic materials that allow to maintain stable sizes of units (or strictly regulated their changes during operation) in certain temperature ranges, as well as the ability to pass through itself strictly controlled (small or large) thermal flows. Materials, which qualify these requirements, are applied to the class of materials with *special thermal properties*. The main characteristics of this kind of metallic materials are the temperature coefficient of expansion (TCE), as well as thermal conductivity and heat capacity.

Materials with a predetermined coefficient of thermal expansion. Depending on the physical nature of the TCR changes with temperature, all the metallic materials are divided into nonmagnetic and ferromagnetic. Most ferromagnetic materials for these purposes are based on iron-nickel alloys. The non-magnetic alloys are mainly manufactured on the basis of chromium, nickel, titanium, zirconium, copper and manganese doped with various elements.

Fe-Ni alloys with expansion coefficient $\alpha \leq 3{,}5 \cdot 10^{-6}$ $^{o}C^{-1}$ relate to ferromagnetic materials with a minimum value of TCE. They are used for the manufacture of measuring devices and instrumentation, cryogenic systems, for coupling junctions, bimetallic strips in mechanical switches (relays), etc.

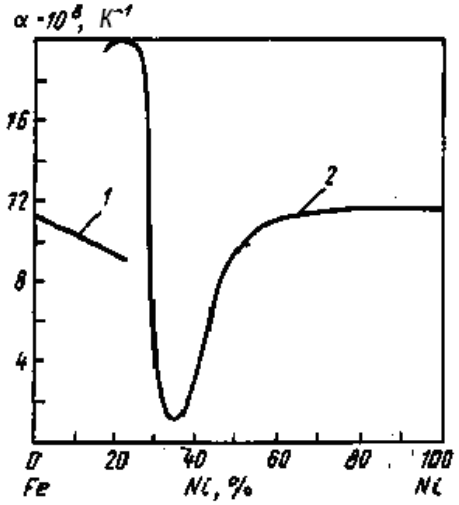


Fig. 3.1.14. The coefficient of thermal expansion in Fe-Ni alloys: *1 – bcc* – phase, *2 – fcc* – phase

As can be seen from Fig. 3.1.14, in Fe-Ni alloys TCE dependence on the concentration of nickel has comlex shape that allows one to create alloys with low TCE values. Experiments show that the mostly low value of TCE in the temperature range from -100 to +100 °C corresponds to an alloy with a nickel content of about 36 %. Such alloys are called *invar*. Anomalous decrease of TCR values in invar alloys has a ferromagnetic nature. It is due to the high saturation magnetostriction in the ferromagnet.

Thermal conductivity of metallic materials. In metallic materials the thermal conductivity can be represented as a sum of electronic κ_e and phonon κ_{ph} components:

$$\kappa = \kappa_e + \kappa_{ph}. \qquad (3.1.1)$$

In metals at not too low temperatures (of the order of the Debye temperature θ_D) $\kappa_e > \kappa_{ph}$ and that is why in the case of such substances Wiedemann-Franz law is performed: the ratio of the coefficient of heat conductivity κ and electric conductivity σ is approximately the same $L = \kappa/\sigma T$, where L is Lorentz constant (see Part 1). If, as in the case of certain semiconductors (e.g., in Germanium), the concentration of free electrons is very low, κ_e decreases accordingly and

Wiedemann-Franz law is no longer satisfied. Because of the reasons mentioned, generally, those metals, which are better conductors of electricity, are also better heat conductors (aluminum, silver, copper).

Magnitudes of κ_e andκ_{ph} can be compared with the corresponding thermal resistances $W_e \sim 1/\kappa_e$ и $W_{ph} \sim 1/\kappa_{ph}$. As in the case of the Matthiesen rule for the electric resistance (see Part 1), the electronic contribution to the thermal resistance of the electron gas can also be presented in the form of two contributions

$$W = W_o + W_{e\text{-}ph}, \qquad (3.1.2)$$

where $W_{e\text{-}ph}$ dis ue to scattering of electrons by thermal vibrations of atoms (electron-phonon scattering), a W_o is due to scattering of electrons by impurity atoms and lattice defects

At low temperatures, $T < 0.1\ \theta_D$, where, according to the Gruneisen model, ideal resistivity is proportional T^5, $W_{e\text{-}ph}$ contribution can be expressed as the ratio

$W_{e\text{-}ph} \sim \alpha T^2$,

where α – constant characteristic for a given metal. W_o magnitude in this range can be expressed by the relation

$W_o \sim \beta/T$,

where β is dependent on the type of metal.

Thus, at the low temperature limit, electronic contribution to the thermal resistance has the form

$$W_e \sim \alpha T^2 + \beta/T \qquad (3.1.3)$$

and must pass through a minimum, which will be shifted to lower temperatures with increasing of perfection of the metal (i.e., the less of impurities and defects).

Indeed, for many pure metals, their total thermal resistance is changed at low temperatures in such a way that the coefficient of thermal conductivity κ passes through a maximum (Fig 3.1.15), and $W = 1/\kappa$ is expressed by the equation which is similar to equation (3.1.3). This shows that the contribution of electrons in metals has the highest thermal conductivity values of θ_D.

Almost complete independence of the temperature coefficient of thermal conductivity at high temperatures ($T > \theta_D$) can be understood (see Part 1), if we consider that, according to the classical theory of heat conduction , κ can be written as an expression of

$$\kappa_e = A \cdot C_e \cdot l_{e\text{-}ph} \cdot v_F.$$

Here C_e is specific heat of electrons, which is proproportional to the absolute temperature T; $l_{e\text{-}ph}$ – the mean free path of electrons between two "collisions" with phonons, which is inversely proportional to the probability of electron

scattering; v_F – the average velocity of the electrons with the Fermi energy in the conductance band.

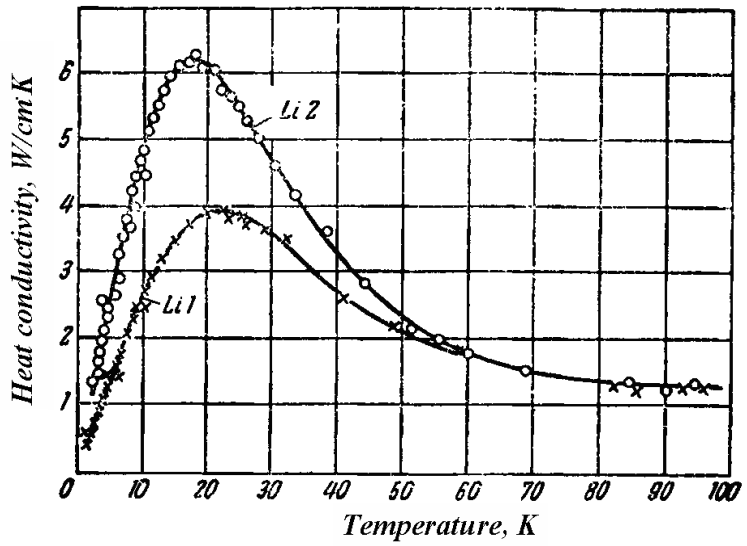


Fig. 3.1.15. Thermal conductivity of two samples of lithium with different purity: Li1 – pure, Li2 – less pure

As shown at the study of the electrical conductivity in Part 1, the probability of electron scattering by phonons at high temperatures is directly proportional to T, so that the product $C_e \cdot l_{e\text{-}ph}$ is independent of temperature. Therefore, κ_e is practically unchanged with temperature.

As follows from (3.1.3), the presence of various kinds of impurities leads to a sharp decrease in the coefficient of thermal conductivity of metals due to additional scattering of electrons on them (Table 3.1.4). For example, if copper has κ = 396 W/(mK), for copper doped with arsenic atoms κ = 142 W/(m·K). Note that, unlike pure metals, thermal conductivity of alloys increases with increasing temperature (see, bronze, constantan, brass in Table. 3.1.14).

Table 3.1.4. Thermal conductivity of metals and alloys at different temperatures

Thermal conductivity of metallic materials, W/(m·K)					
Temperatures	*200 K*	*300 K*	*400 K*	*600 K*	*800 K*
Aluminium	227	237	240	230	220
Aluminium bronze	-	105	-	130	145
Vanadium	31	30.7	31.3	33.3	36
Wolfram	185	174	159	137	125
Iron	94	80	70	55	43
Gold	323	317	311	298	284
Cobalt	122	100	85	67	58
Constantan	-	22	24	32	-
Brass	-	110	-	140	150
Copper	413	401	393	379	366
Molibden	143	138	134	126	118
Nikel	107	91	80	66	68
Nichrom	-	12	-	-	23
Tin	73	67	62	L	L
Platinum	73	72	72	73	76
Lead	37	35	34	31	L
Silver	430	429	425	412	396
Titanium	25	22	20	19	20
Crom	111	94	91	81	71
Zink	118	116	111	103	L

The mechanism of heat transfer, determining the thermal conductivity of crystalline materials, primarily depends on type of chemical bonds: in metallic materials heat is transferred by electrons and in covalent dielectric materials or ionic bonds – by phonons. Among the most thermally conductive crystalline dielectrics is diamond. In semiconductors with very low charge carriers concentration thermal conductivity is mainly supplied by phonons. More perfect crystals, the higher their thermal conductivity. Single crystals conduct heat better than polycrystalline as grain boundaries and other crystal defects scatter phonons and electrons.

When a metal contains more impurities, grains are finer and crystal lattice is more distorted, the thermal conductivity is lower. The larger grain size, the higher the conductivity. The doping causes a distortion of the crystal lattice, therefore, solid solutions have lower thermal conductivity as compared with

pure metals. Constituents, representing dispersed mixture of several phases (eutectics, eutectoids, peritectics), reduce the thermal conductivity. Structures with uniform distribution of precipitates have a lower thermal conductivity than the basic metal (matrix) of the alloy. Ultimate type of such a structure are porous materials.

3.2. Nonmetallic inorganic materials

As noted in Section 3.1, in power industry and energy saving, besides metallic and alloys also non-metallic materials are widely used. They provide significant energy, technology and economic efficiency. *Nonmetallic materials* include a wide range of inorganic and organic substances of natural and artificial origin: ceramics, composites, polymers, silicate and specific glasses, glass ceramics, stone (rock) materials, wood, mineral and synthetic resins (e.g. bitumen), etc.

The advantage of many of these materials, at the relative cheapness and less power-intensity, consist in combination of often non-combined properties (in metallic materials): a high strength and toughness, resistance to heat and fire resistance, chemical resistance, nonsusceptibility to aging (degradation), elasticity and low density, transparency to electromagnetic radiation in various ranges of spectrum, specific electric and magnetic properties, etc.

3.2.1. General information about the non-metallic inorganic materials

Two large groups of natural and synthetic materials can be numbered among non-metallic inorganic materials (NMIM): mineral (ceramics, composites) and the noncrystalline polymeric type materials (glasses, ceramics, certain plastics), as well as their composition. Natural and synthetic mineral materials basically include oxides and oxygen-free compounds of metals and non-metals. Among them, silicon oxide (silicates) and also pure oxides and oxygen-free compounds of aluminum, magnesium, zirconium, iron, etc., as well as their compositions with the silica are widely used. These pure oxides and oxygen-free compounds are used for preparation of ceramics, bricks, concrete, mineral wool, glass, etc.) and often have a more valuable application properties than conventional silicate materials. *Inorganic polymeric materials* include mineral and synthetic glasses and glass ceramics having the atomic structure of the carcass type, and some other non-crystalline material.

All of these NMIM species possesses by flammability, high heat stability, chemical resistance, lack of inclination to aging, as well as high hardness with considerable resistance to compressive loads. However, they are highly brittle, non-elastic, do not tolerate rapid temperature changes (thermal shocks), low

resistant to tensile and bending loads, and significantly heavier than organic polymer materials.

By the structure NMIM are divided into single crystalline, polycrystalline (cust, ceramic, sintered), amorphous, including glassy (vitriform), glass ceramic and polymeric (with the atomic structure of the carcass type). By a set of properties, that define the application in power industry and energy saving technologies, NMIM include construction materials, nuclear materials (nuclear fuel, units of thermonuclear facilities, containers for radioactive substances, etc.), heat power industry and renewable energy sources (high-temperature strong, heat-resistant, fire-resistant, corrosion-resistant), materials with special electrical (semiconducting, conductive, dielectric, superconducting, piezoelectric, ferroelectric) and magnetic (soft and hard ferromagnets, low- and high-frequency ferromagnets) characteristics; materials with special optical (reflective coatings for screens and energy-efficient lighting systems) and thermal properties (heat-insulating, heat-conducting). Inorganic materials with sensory properties that are used in sensors of temperature, pressure, flow rate, concentration, humidity, pH , etc. are included very often in a separate group.

3.2.2. Ceramic materials

General information about ceramic materials. Ceramic material represent a complex heterogeneous inequilibrium system consisting of inorganic polycrystalline (granular) substances with ionic, covalent or metallic chemical bonds (mineral and synthetic oxides, nitrides, carbides, refractory metals and alloys, etc.), obtained by sintering of powder mixtures.

To produce such traditional (known since antiquity) ceramics as porcelain, earthenware, tiles and bricks, different varieties of clays are used as raw materials. *Clay* consists of crystallites hydrated aluminosilicates, i.e. compounds containing different proportions of kaolin Al_2O_3, earth silicon SiO_2 and crystallization water H_2O, as well as a number of special additives. After forming a viscous soft mass produced from a mixture of clay and water, raw ceramic ware are fired at high temperature to remove water and mainly to provide conditions for the various chemical reactions. The microstructure of the final product, typically consists of crystallites of refractory component, primarily oxides, which are enclosed in a glassy (non-crystalline) matrix.

Currently, the term "ceramics" has become more widespread than ever, meaning. Now it is used for all materials obtained by pressing and heat treatment of pure oxides, carbides, nitrides and mixtures thereof. Besides clay ceramics, it also includes cements and concretes, and also cermets (carbide solid particles in a metallic matrix) or hardmetals obtained from metal powders). In some classifications they include also glass-ceramics, because many of them are also producted from mineral oxides .

To save power, such ceramic materials as brick and tile, porcelain and earthenware, cement and concrete are normally used. For other purposes (instrumentation, electrical and radio engineering, electronics) special ceramics are used: magnetic, electric, superconducting, etc. As structural ceramics, tool ceramics, including cermets (metal or metal-carbide systems), are used.

Cement consists of complex compounds of lime CaO, silica SiO_2 and alumina Al_2O_3, which react with water during setting and hardening of the cement mixture and converting it into *concrete*. Cement is prepared by roasting of clinker – a mixture of finely ground limestone with some kinds of clay. The particle size defines both a sort of cement and type of the future concrete. For conversion of cement powder into the concrete, it is mixed either with fine-grained sand (earthsilicon) and water to yield the slush or with sand, crushed rock (as a filler) and water, which after hardening become concrete.

Concrete is used for a variety of purposes – from building enclosures (residential buildings) to a high-pressure container production for nuclear reactors and defensive systems in nuclear power plants. Annual consumption of concrete per capita is much higher than the consumption of steel.

The new, first of all, instrumental types of ceramics (cermets or powder alloys) include primarily materials produced by powder metallurgy – by pressing and sintering of different powder mixtures. Previously, such agents have been used only as abrasives, which are for example aluminum oxide (argil Al_2O_3) or silicon carbide (carborundum SiC). By *cermets* are usually related to the two types of materials :

(1) ceramic powder, comprising particles of a crystalline refractory oxides, carbides and nitrides (with the amount not less than 5-15 wt.%), which is enclosed in a metallic matrix, or (2) ceramics, consisting of compacted and heat treated (sintered) refractory metal powders.

In the cermets of the type (1) metal is destined for formation of the plastic matrix to couple solid particles to a single particulate product for overcoming their brittleness. Such cermets are similar to composite materials (see, below). Cermets of type (2) replace refractory alloys that are impossible or very difficult to obtain by metallurgical methods (melting or casting). In some cases, different additives like carbon materials (graphite nanotubes, graphene, etc.) are introduced into cermets.

Cermets with high metal content can be useful as high-temperature structural materials, for example for the manufacture of gas turbine blades and other parts of power devices that require high heat resistance. The combination of extremely high hardness and impact toughness (albeit limited), which is stored at high temperatures has led to the widespread use of cermets with a high content of solid phases for the manufacture of cutting tools (cutters, mills,

drills). Along with the structural and instrumental ceramics, special ceramics for electronics, radio engineering and optoelectronics, having special electrical (including superconducting), dielectric, magnetic, and optical properties are intensively developed.

Ceramic materials based on oxides. Classic ceramics (bricks, porcelain, tiles, etc.), which are used in construction and for domestic use since ancient times, are made of three main components – clay, silica and feldspar. Clay mixture consists essentially of mixtures of compounds of hydrated alumina Al_2O_3 and silica (quartz) SiO_2. Clay may also contain small quantities of other oxides such as CaO, Fe_2O_3, MgO, etc. For production of brick and other classical Al_2O_3- and SiO_2-based ceramics, feldspar is added. The last is alkali metal aluminosilicate, often consisting of a mixture of potassium $K_2O{\cdot}Al_2O_3{\cdot}6SiO_2$, sodium $Na_2O{\cdot}Al_2O_3{\cdot}6SiO_2$ and calcium $CaO{\cdot}Al_2O_3{\cdot}2SiO_2$ feldspars. Feldspar has a lower melting point than the alumina and silica. During the firing, spar forms a glassy phase (as in glass-ceramics) that binds crystalline refractory components of alumina and silica in a single unit.

The increase in the proportion of feldspar leads to greater density of low-temperature ceramics and their translucency. Increase of clay content significantly facilitates molding and improves mechanical and electrical properties, but requires higher firing temperatures.

When water is added, the mineral particles are covered with water film forming a colloidal suspension, which results in plasticity of clay moldable mixtures. In the formation of water films, a crucial role belongs to electric charges and thin plates shape of the clay particles. At certain water concentrations, clay is elastic at low stresses. However, if the stress exceeds the yield stress value, the clay is plastically deformed with a rate proportional to the excess stress.

Properties of oxide ceramics. Physical properties of oxide ceramics (both classical and others) are determined by their composition and production technology. The most valuable features of oxide ceramics as structural materials are their strength, chemical and thermal resistance. Furthermore, as most ceramics are composed of metal oxides, further oxidation, as well as other chemical reactions are almost impossible. It is therefore they are not susceptible to combustion. All these properties are due to the presence of covalent chemical bonds in oxide ceramics. The strength of the chemical bonds determines not only a high melting temperature but also relatively high compressive strength, toughness and hardness of ceramics (Table 3.2.1).

Lack of plasticity in most ceramics leads to low values of their tensile strength, which is less than the theoretical strength for fragile crystals. The latter, according to the theory of the strength of brittle crystals is $E/10$ (E –

modulus of elasticity). Experimentally well established that ceramic materials are much more resistant to compressive stresses than the shear:

$$\sigma_{comp} > \sigma_{str} > \sigma_{sh}.$$

This relationship, as well as the brittleness of ceramics is caused by point defects (impurities, single vacancies), vacancy pores and micro- and nanocracks. The presence of pores, which are stress concentrators, contributes to the development of cracks. Quantitatively brittle ceramic materials have a viscosity which is equal to approximately 1-2 $MPa/m^{1/2}$. Recall that the viscosity of metals is about 40 $MPa/m^{1/2}$.

Understanding the role of defects in brittle fracture of ceramics has stimulated the search for ways to increase their strength. The first method is to remove the pores and agglomeration of the particles forming the ceramics, which are sources of crack. In this case, raw material powder is thoroughly purified and is very finely milled (to a particle size < 1 μm). Next, a special suspension of the particles in the organic liquids (like methanol with addition of a polymer) is prepared. Polymer is adsorbed by the particles and prevents their agglomeration. After compacting, powder was subjected to "controlled packaging", followed by sintering. The resultant material should have a density equal to the theoretical, i.e. must be substantially continuous (non-porous).

The second way to increase the strength of ceramics is to improve the failure stress to values close to the theoretical limit $E/10$. This is necessary for this goal, either to increase the ductility (e.g., by doping) to a value such that cracks do not lead to premature failure, or to eliminate the cause of microcracks nucleation and inhibition of their growth under load. At present, there has been a number of ways to solve this problem. One way is to use carefully prepared thin glass and ceramic fibers, each of which has strength close to the theoretical value. To implement such a strength, fibers are enclosed in a ductile metallic or plastic matrix.

Since many ceramics are operated at high temperatures, their ability to withstand thermal shocks (rapid change of temperature), under which the thermal stresses are nucleated, gains the vital importance. The nature of these stresses can be related, for example, to the thermal expansion, that can be due to the temperature gradients, in particular its anisotropy. This anisotropy is usually associated with low thermal conductivity, as well as non-cubic crystalline structure of ceramics. An additional reason for the destruction of ceramic under thermal shock may be heterogeneity of the ceramics by their bulk.

Table 3.2.1. Mechanic and some other properties of ceramics in comparison with certain non-oxide materials (cermets, glasses, metals and graphite)

Material	Tensile strength σ_B, MN/m^2	Modulus of elesticity E, GN/m^2	Coefficient of thermal expansion α, $10^6\ K^{-1}$	Heat conducti-vity factor κ for 100°C, Wt/(m·K)	$\sigma_B/E\alpha$	$\sigma_B k/E\alpha \cdot 10^{-3}$
Al_2O_3	145	350	8.8	30	47	1.4
MgO	96	210	13.5	36	34	1.2
BeO	145	300	9.0	220	53	11.5
$3Al_2O_3 \cdot 2SiO_2$ clay	82	145	5.3	6.3	107	0,67
WC - Co cermet	1600	600	5.2	60	510	30
TiC – Co cermet	1100	410	9.0	35	300	10.5
Reinforced tool steel	2500	200	10	46	1250	57
Fused silica (SiO_2)	105	725	0.55	1,5	2640	4.0
Sodium-calcium glass	70	69	9.0	1.0	113	0.1
Cold-rolled copper wire	410	117	14,0	380	250	95
Engineering carbon	20	10	3,0	150	670	100

Electrical, magnetic and optic properties of common ceramics are very diverse. Large band gap in many of classical oxide ceramics leads to the fact that they have insulating properties because they do not have free electrons. This is the reason that many of ceramics are widely used as electric insulators. Some of the ceramics with a narrow band gap have semiconductor properties.

As follows from the nature of chemical bonding, many of crystals with ionic chemical bonds comprise electric dipoles, which interact with each other and may also vary due to external factors (temperature, mechanical deformations, electric field, etc.). In some of ceramics, such as barium titanate $BaTiO_3$, the dipole-dipole interaction can lead to the spontaneous polarization (orientation) of the dipoles and the phenomenon of ferroelectricity.

The presence of the magnetic moments in the atoms of some ceramics, due to the existence of uncompensated electron spins, the interaction between them leads to ferromagnetism of the whole class of ceramics, known as ferrites. The

latter are non-conductive or poorly conductive magnetic materials which can be used in many high-frequency radio devices.

Impurities in some type semiconductor ceramic lead to the fact that they can absorb the incident electromagnetic radiation, and then again to emit it as a coherent beam, wherein all the waves are in phase and are distributed in the same direction. Such materials can be used as active elements in solid-state lasers and masers and in optoelectronics (light emitting diodes, etc.).

The crystalline structure of the ceramics are numerous and varied. They vary from the cubic structure of MgO to the layered structure of mica and asbestos fiber structure (hydrous magnesium silicate). Such structures often have low symmetry, which may cause the piezoelectric properties – the presence of static charge due to mechanical deformation. As an example we call quartz (single crystalline forms of SiO_2), which is used in electronic oscillators.

The lack of free electrons in many ceramic materials causes the fact that they are poor conductors of both electricity and heat. Therefore, they serve for production of heat insulators.

The above mentined structural features are also responsible for the transparency of high quality and densed ceramics. At the usual methods of production ceramics still have considerable residual porosity which makes them semi-transparent, or often opaque.

3.2.3. Cement and Concrete

One of the important types of materials that can be related to the ceramics (along with porcelain, brick and tiles) and are widely used in power industry and energy efficient technologies (to create different types of enclosures) is *concrete*. For its preparation the *cement* is used, which, as noted above in this chapter, is composed of a mixture of CaO, Al_2O_3 and SiO_2. To prepare concrete, the cement powder is mixed with fillers (sand, coarse pieces of crushed rock, etc.) and water. This mixture of cement with water forms a doughy soft mass which is molded and, after the passing of complex chemical reactions, becomes concrete as a hard brittle material with high compressive strength.

The process of cement mixture hardening and its transformation into the concrete is accompanied by several chemical reactions, induced, above all, by the presence of mortar. To get strong ceramic material from mortar, you must first obtain *quick lime* CaO. It is produced by heating of limestone or chalk ($CaCO_3$) to remove carbon dioxide. Quicklime is then mixed with water, to produce a calcium hydroxide $Ca(OH)_2$, which is called *drowned lime*. Since this reaction is exothermic, it is accompanied by a large heat release and increases the volume by 20%. The slaked lime is a white powder with a particle size of about 2 microns. With excess water, this well-dispersed powder turns pasty

mass, which after mixing with sand is molded. When mortar drying, a solid low-strength material is formed. It consists of filler (sand) particles interlinked by layer of lime. Subsequent hardening of the material occurs due to interaction of lime layer bonding the filler particles (sand) with carbon dioxide in air, wherein the calcium carbonate is newly formed.

Setting with hardening of mortar with filler is slow process, which begins from the surface. For thick layers it never reaches the center of the detail, that significantly reduces the strength of the material as a whole. Furthermore, setting can not occur under water.

The modern cement has a high content of tricalcium silicate ($3CaO \cdot SiO_2$) due to firing at higher temperatures and the use of fluxes. It is this substance is responsible for more rapid maturing of cement mixture.

The *typical cement* consist of 65 wt.% CaO, 20 wt.% SiO_2, 5 wt.% Al_2O_3, and about 10 wt.% of other substances, including Fe_2O_3. Iron oxide is usually contained in the feed or added on purpose. The feed mixture is then subjected to thin milling, and then is fired in a rotary furnace at a temperature of about 1500°C.

Grinding is *dry* or *wet*, or their combination. In the wet method, limestone and clay are milled in wet state with humidity 50-60 vol.%. In the combined method of grinding raw material components are crushed in the wet state and then dried before firing.

In the firing process, a low-melt alumina and iron oxide form a liquid phase, which dissolves refractory lime and silica reacting with it. As a result, a dicalcium and tricalcium silicates arise, which precipitate from the melt in the form of crystallites. The resulting sunstance was called *clinker*.

The presence of the liquid phase is substantial to progress of clinker formation process, since the reaction rate in a solid phase, such as during firing of limestone containing also clay, is very small. For the production of high-quality cement, it is important full removal free lime after firing, which will destroy the concrete due to hydration process.

The hydration reactions, occurring when mixing cement particles with water, are very complex. To achieve complete hydration and the largest strength, about two parts by weight of cement to one part water is typically mixed. It should be emphasized that the hardening of the concrete is not associated with the process of water evaporation. In fact, it is often necessary to take special measures to prevent the surface evaporation of water and air flow to the hardening concrete before the end of hydration process. However, excess water is also undesirable because it leads to the porosity of the concrete, which reduces its strength, while facilitating the process of mixing and packing of cement.

The most important product of the hydration is hydrate tricalcium disilicate, which contains nanoparticles of dicalcium and tricalcium silicates. In this dried gel, particle sizes are typically of the order of 10 nm.

Table 3.2.2. Chemical reaction (the starting materials and final products) for the hydration of cement powder

Cement component	Content, Wt.%	Reaction	Hydrated cement
Tricalcium silicate	45	$2(3CaO\cdot SiO_2)+6H_2O\rightarrow$ $\rightarrow 3CaO\cdot 2SiO_2\cdot 3H_2O+3Ca(OH)_2$	Hydrate calcium silicate + calcium hydro-oxide
Dicalcium silicate	30	$2(2CaO.SiO_2)+4H_2O\rightarrow$ $\rightarrow 3CaO\cdot 2SiO_2\cdot 3H_2O+3Ca(OH)_2$	Hydrate calcium silicate + calcium hydro-oxide
Tricalcium - aluminate	10	$3CaO\cdot Al_2O_3+12H_2O+Ca(OH)_2\rightarrow$ $\rightarrow 2CaO\cdot Al_2O_3\cdot Ca(OH)_2\cdot 12H_2O$ $3CaO\cdot Al_2O_3+30H_2O+$ $+3(CaSO_4\cdot 2H_2O)\rightarrow$ $\rightarrow 3CaO\cdot Al_2O_3\cdot 2CaSO_4\cdot 36H_2O$	Hydrate tetracalcium aluminate Hydrate monosulfo-aluminat
Tetracalcium-alumo-ferrite	10	$4CaO\cdot Al_2O_3\cdot Fe_2O_3+10H_2O+$ $+2Ca(OH)_2\rightarrow$ $\rightarrow 6CaO\cdot Al_2O_3\cdot Fe_2O_3+12H_2O$	Hydrate calcium alumo-ferrite

Owing to the process of adhesion due to the formation of ionic bonds between the particles in tobermorite paricles in tobermorite gel and particles of other materials in the cement powder, the components of concrete form a solid mass. These specific adhesion forces are very large, including due to fine particles in a gel state, and therefore, the developed specific surface area which reaches approximately $3\cdot 10^5$ m^2/kg. The unique combination of properties of tobermorite gel (high specific adhesion forces) and a large specific surface area causes great cohesion forces between the particles in the hardened cement (concrete).

The most important indicator of the strength of the concrete is compressive strength σ_C. It depends on many factors, notably the properties of the components and the relative content of concrete – cement, sand, coarse filler and water. Concrete strength also depends on the method of mixing of the components, temperature and environment humidity, and of course, the duration of aging.

The tensile strength of concrete is too small and does not have any practical value. The water-to-cement ratio is of great practical value, since an increase in

the proportion of water in the cement paste concrete strength decreases rapidly due to increase its porosity. On the other hand, the less the water content, the more difficult to mix and lay concrete due to the low ductility cement mixture. Therefore, for full hydration (and thus to maximize the strength of the concrete grade) the water-to-cement weight ratio must be equal to 0.4. Thus from 0.25 to them for chemical reaction and 0.15 for the gel formation tobermoritic.

The particle size of the filler and its volume must be selected to fit snugly way to fill the space in the formation of concrete. This allows to obtain a dense material with a minimum amount of expensive cement. Typically, for the closest packing of the concrete particles it is necessary to use multiple fractions of particles with different sizes – from 50 to 0.25 mm (conventionally considered to be the aggregate gross, if the particle size is greater than 5 mm). The amount of cement should be sufficient to fully cover and bind all of the aggregate particles. In a most common conventional concrete mixtures, the volume ratio of cement, sand and aggregate is 1:2:4.

3.2.4. Cermets

Most modern cermets consist of solid crystalline particles of oxides, carbides or nitrides of a refractory metals, which often integrally connected to the metal matrix. In some cases, refractory alloys obtained by methods of powder metallurgy (when compacted powder mixture of high temperature refractory metals or compounds are subjected by heat treatment for intensifying the processes of diffusion) are also classified as cermets.

Cermets originally developed as materials for cutting tools, but their further technical functions were substantially expanded (for example, they have been used for making engine parts, gas turbine blades, etc.). Tool cermets must withstand high local concentrations of stresses (compressive stress near the cutting edge of the tool, as well as tensile stresses at points remote from the cutting edge). They also have to withstand high temperatures, rapid temperature changes (local thermal shocks) and large temperature gradients arising due to friction of cutting edges of the workpiece as the tool must retain its strength and must be opposed to oxidation. Cermets should also possess good resistance to abrasion, which leads to tool wear. Diffusion and adhesive interaction between the tool material (cermet) and the workpiece should be minimized. The criticality of each of these properties and, consequently, the choice of composition for the cutting tool material is dependent on the specific conditions of operation, and the type workpieces.

The first tool cermet was quasi-binary solid composite of tungsten carbide with cobalt (WC + (6-20) wt.%Co). The structure of this type of cermets has the following form. Tungsten carbide is a hard brittle ceramics with a simple hexagonal crystal lattice while cobalt is ductile metal with hexagonal close-

packed structure. Being melted Co wets the carbide phase WC and forms a strong adhesive bond with it. Although generally WC remains fragile in cermet, the presence of cobalt increases achievable plastic deformation and energy absorbed before failure compared to pure tungsten carbide. This cermet is used in cutters of lathes increased cutting speed up to 300 m/min while maintaining high wear resistance and toughness (withstanding of micro-impacts of cutting edges operation). It is also used in the boring tools

Production of WC-Co cermets is performed in several stages. First, a fine powder (with particles sizes of the order of micrometers) of tungsten carbide and cobalt is manufactured and mixed. Then, a mixture manufactured with the needed combination of components that is either compacted by compressing and then subjected to high temperature treatment (sintered) in the furnace or directly subjected to hot pressing (pressure sintering). Sintering is carried out at a temperature below the melting point of cobalt (about 1400°C).

Consider the mechanism of obtaining a lasting cermet, regarding it as quasi-binary alloy. Obviously, when the sintering temperature is approached, tungsten and carbon atoms diffuse from solid WC to cobalt. Thus, despite the fact that the sintering temperature lies below the melting temperature WC (2600°C) and Co (1492 °C), a liquid phase is carried out by peritectic reaction at alloying of components (Co and WC in the presence of free carbon). When the WC content in cobalt approximately approaches 30 wt.%, cermet is transformed into completely liquid state. The resulting melt penetrates into the voids of the compacted mixture, dissolving the increasing volume of WC and allowing to wet the remaining (insolubled part) particles of tungsten carbide, while is reached equilibrium composition of the quasi-binary alloy WC-Co at content of 50 wt.% WC. Sintering of pressed compact results in WC grain growth due to recrystallization and reducing its volume down to approximately 50% (owing to reducing of pores density).

Upon cooling, the liquid phase solidifies at about 1320°C by the eutectic reaction with the release of carbide grains. Further precipitation of carbide particles is observed in the solid state at cooling down to room temperature, when WC is substantially insoluble in cobalt.

After development of WC-Co cermets for manufacturing of cutting tools a number of more complex compositions has been proposed, in which the tungsten carbide is completely or partially replaced by other carbides, such as TiC, TaC, MoC, etc. These cermets have given the best results compared to high-speed cutting steel. When adding titanium carbonitride and further Ta(Nb)C, Mo_2C and VC, to them, obtained ceramics can be used for wells boring.

Obviously, in cermets with low cobalt content (no more than 6 wt.%) carbide particles can not be separated by a solid cobalt “shells”. In this case, it is

considered that carbide crystallites form a continuous three-dimensional backbone. The presence of a thin cementing film at several larger cobalt content well explains the increase in the shear modulus from 500 MN/m^2 in pure sintered WC to 1.7 GN/m^2 in the cermet. Although the plastic deformation of thin film is limited to neighboring solid WC particles, but the film still impedes the movement of cracks through a continuous carbide phase. Experiments show that at even higher cobalt content WC particles enclosed (embedded) into cobalt matrix. Physical properties of tool materials, including cermets, are compared in Table 3.2.3.

Table 3.2.3. Comparison of the physical properties of different tool materials

Properties	Quick-cutting steel Fe – 18Wt.% W – 4Wt. % Cr – 1Wt. % V – 0.7Wt. %C	Cermet WC - Co	Al_2O_3– 1Wt.% MgO	Al_2O_3 – 10Wt.% TiO
Dimension of particles, μm	10	2	3	2
Ultimate bending strength, MN/m^2	3500	1600	600	600
Ultimate compression strength, MN/m^2	4100	4500	3000	3500
Microhardness	740	1800	1570	1660
Maximal operation temperature, °C	600	900	1000	1000

Development of cermet application indicates that the instrument on their base allows significant increase in cutting speed with high durability of cutting tools. This means that the energy component of their cost is significantly reduced. Thus, the application of the tool cermets improves energy efficiency.

In the last 20-30 years considerable efforts have been directed towards the development of cermet-based high-temperature structural materials useful for the manufacture of gas turbine blades and engine parts. Here the resistance to thermal shock, creep strength and toughness at high temperatures are more important properties than the hardness and wear resistance. The development of high-temperature cermets with low-density and high heat and abrasion resistance allowed to use them for the manufacture of engine components and turbine blades, thus improving the efficiency of power machinery and plants. Another promising field of their application is the manufacture of high-temperature cermets for mechanical seals and bushings in automotive engine valves and valves in the hot water lines. The use of cermets for this purposes saves water consumption (and hence energy), which is usually significant when using fiber or rubber seals. An important advantage of cermets is their

resistance to chemicals that can be successfully used also in the valves of pipelines destined for pumping corrosive liquids.

3.2.5. Special ceramics

In this section, the term "special ceramics" is used for different compositions based on crystalline and amorphous-crystalline oxides, carbides and nitrides of very low (close to zero) porosity. These ceramics have a number of unique properties and may be used in nuclear reactors (beryllium and uranium oxides) in special bearings and tools (e.g., boron carbide, which has the highest hardness among all materials known), in the manufacture of heaters, refractory coatings and abrasives (silicon carbide), as well as in the manufacture of gas turbine blades and bearings (silicon nitride). Ceramics based on barium titanate allow an increase in the dielectric constant to 1000-10000 times.

Number of special ceramics are now so large that to hold a full discussion almost is practically impossible. Therefore, in this section we will consider only some of their basic classes, such as electrical, ferroelectric and pyroelectric, magnetic (ferrites) and high-temperature superconducting ceramics and ceramic containers for the storage of radioactive materials and waste.

Electrotechnical ceramics. In electrical engineering, used clean and low porosity alumina ceramics based on corundum (α-Al_2O_3) and sapphire.

In the crystal lattice of corundum oxygen atoms form a slightly distorted hexagonal close-packed lattice, and aluminum atoms occupy octahedral interstitial sites. To comply with the stoichiometry (the desired ratio of the number of aluminum and oxygen atoms) it is necessary that one of the three interstitials in the unit cell were empty.

Microstructure of electrical alumina ceramics, made from argil, has normally low purosity (about 95 wt .% Al_2O_3). Adding silicates and alkaline earth metals in agril leads to the formation of vitreous layers (strata) around corundum crystallites. Such low-purous materials are cheaper, however, they tend to have inferior properties.

Fig. 3.2.1 shows the dependence of grain size on porosity of corundum in agril containing over 99 wt.% Al_2O_3, which is controlled by changing the pressing temperature and pressure. It was found that the porosity of less than 2 % leads to a rapid increase in grain size. This effect is inhibited by the addition of 0.1 Wt.% MgO, which after sintering results in a high alumina density (close to the theoretical density) and fine-grained structure (with dimensions of the order of 1-2 μm).

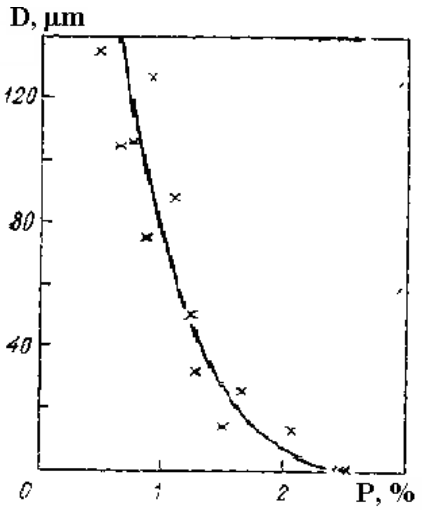


Fig. 3.2.1. Correlation between grain size D and porosity P for alumina of high purity (~ 99.9 Wt.%)

High-densed technically pure alumina (more than 99 Wt.% Al_2O_3) and high-alumina ceramic (containing about 85 Wt.% Al_2O_3) were used in electrical engineering thanks to a successful combination of their electrical, mechanical and chemical properties. Electrical properties of this material are characterized by high electrical resistance, the average values of dielectric constant of the order of 5-10, good dielectric strength (breakdown voltage is about 200-250 kV/cm) and very low dielectric loss at high frequencies ($tg\ \delta < 10^{-3}$). Since the melting temperature of alumina is sufficiently high (~2050°C), its operating temperature may exceed 1000°C. Low dielectric losses and considerable mechanical strength at high temperatures allow the use of alumina in the manufacture of flasks of microwave lamps, circuit boards for electronics and other electronic products. Since the translucent alumina has a high resistance to impact of alkali metals, it is used in sodium lamps with high efficiency at temperatures up to 1000°C.

Fig. 3.2.2 and 3.2.3 show the dependence of dielectric properties in technical alumina containing about 88 Wt.% Al_2O_3, on density and temperature. As can be seen, the permittivity of alumina increases with increasing its density (Fig. 3.2.2). Loss tangent tends to increase, upon heating, and growth with impurity concentration (Fig 3.2.3).

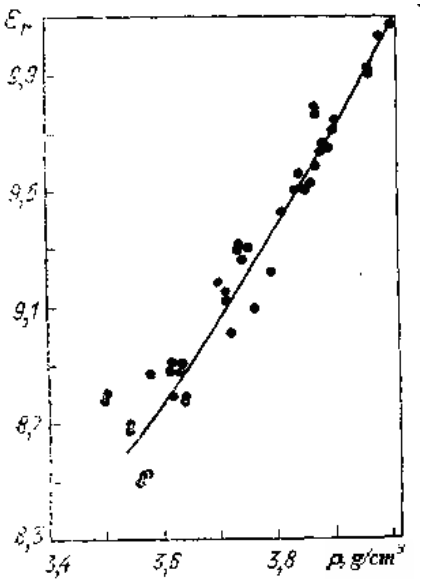


Fig. 3.2.2. Dependence of permittivity ε on density ρ for electrical ceramics from technical alumina containing up to 88 Wt.% $Al2O_3$

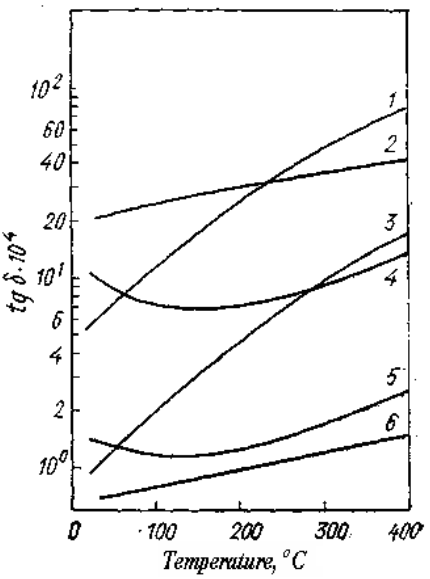


Fig . 3.2.3. Dependence on loss tangent tg δ on temperature for electrical ceramics of alumina Al2O3 with different purity : *1* – 98.5 wt. %, *2* – 88 wt . %, *3* – 99.9 wt. %, *4* – 93 wt. %, *5* – 98.8 wt. %, *6* – monocrystalline sapphire

Note that a classical porcelain ceramics (comprising feldspar, clay and quartz) can be used as high temperature insulators (such as spark plugs, which can be heated up to 2500 °C). As opposite to porcellan, quartz poorly withstands thermal shock (due to the structural phase transition at the temperature of 573°C), and feldspar having ion conductivity gives leakage currents at high temperatures. Increase of agrill in porcelain ceramics (e.g., $Al_2O_3 \cdot SiO_2$, which decomposes during firing on $3Al_2O_3 \cdot SiO_2$ and quartz) and replacement of low-melted feldspar by talc ($3MgO \cdot 4SiO_2 \cdot H_2O$) can reduce the concentration of silica down to 5 wt.% and increase the firing temperature to up to 1425°C (while simultaneous reducing the cost of production).

Increasing the alumina content enables a favorable combination of thermal and electrical properties of the alumina-based electro-ceramics: reducing the coefficient of thermal expansion, improved thermal stability, conductivity properties of the insulator, the enhanced electrical resistance at high temperatures, increased resistance to thermal and mechanical shocks and corrosion resistance and also some growth of resistance to atmosphere of combustion products for engines. This allows the use of such ceramics in spark plugs in vehicles that operate in more severe conditions than most conventional insulators .

Ferroelectric and pyroelectric ceramics. Ferroelectrics belong to dielectric crystals with spontaneous polarization, when the dipole moment (nonlinearly dependent on the electric field intensity) occurs even in the absence of an external electric field or persists after its removal (see Part 1 of this Book). In the first case, initial dipole moment is due to asymmetric arrangement of ions in crystal lattice (along certain crystallographic directions). It exists at all temperatures, so far this type of the crystal lattice exists. In the latter case, the dipole moment is induced by the external electric field and is conserved after its switching off. Unlike the vast number of dielectric materials in ferroelectric materials permittivity is a function of the electric field $\varepsilon = f(E)$ and reaches very high values.

Spontaneous polarization was first discovered in the compound $KNaC_4H_4O_6 \cdot 4H_2O$ and received the name of *Rochelle salt*. Experiments have shown that the spontaneous polarization of Rochelle salt exists only in a limited temperature range – from T_{c1} = -18 to T_{c2} = +24 °C. At $T < T_{c1}$ ions forming electric dipoles in a ferroelectrics, are "frozen" (when difficult to move them in space and redirect by the external field). At $T > T_{c2}$ polarization (dipoles) disappears altogether due to thermal motion (oscillation) of ions.

Note that in the region of ferroelectric state existence some dielectrics also possess *pyroelectric effect*. The latter consists of change in polarization and, as

a result, rise of electromotive force on heating (or cooling) of the material. Manifestation of this effect in ferroelectrics should not be confused with the phenomenon of *pyroelectricity* in pyroelectrics, which has initially a dipole moment without any involuntary discharges.

The most commonly used in the art are ferroelectric barium titanate $BaTiO_3$, lead titanate $PbTiO_3$, lithium niobate $LiNbO_3$ and others. It has been shown experimentally that for the appearance of ferroelectricity we need some internal strain (displacement different atomic sublattices in the crystal) or disordering of the crystal lattice. For example, there is a shift in $BaTiO_3$ sublattices of titanium and barium relative to the oxygen sublattice.

When application of electric field, especially in the process ofceramics preparation, electric moments of the ions (dipoles) in polycrystalline grains are reoriented so that their direction becomes ordered instead of accidental. In this case separation of charges is established in the material – one face accumulate negative and the other – positive charges. Besides the internal deformation, resulting from the displacement of the sublattices, the ferroelectric polarization (e.g., in crystals $NaNO_2$) may be due to the transition of ions from a disordered to an ordered state when an electric field aoolication.

Ferroelectrics have the most widely practical application as dielectric layers in capacitors. This is due to the fact that even in the polarized state ferroelectric material has high insulating properties (high dielectric constant), but at the same time, it is able to accumulate electric charge.

Note that if in the crystal of ferroelectric ceramics ion charge is distributed asymmetrically with respect to its center, the mechanical deformation may lead to a shift of the cations and anions relative to one another. This change in polarization leads to the creation of a significant electrical charge on the surface of such crystals, i.e. to the *piezoelectric effect*. The reverse effect is possible when the deformation of the crystal under the influence of an applied electric field occurs. Such ferroelectric ceramics are called *piezoceramics*.

In some piezoelectric materials mechanical effects may result in electrical voltages up to 10^3-10^4 V and above. Such crystals can be used in devices for igniting the gas supply (household gas lighters) or gasoline vapor (in spark ignition of combustion engines). Inverse piezoelectric effect in piezoelectric ceramics can be used for the manufacture of some ultra-sensitive detectors and gears of movement (such as tunneling and atomic force microscopes), creating of ultrasonic vibrations (such as in quartz watches), etc.

Interaction of light with electric dipoles in ferroelectric ceramics allows to change the refractive index of light beams when an electric field application and thus to control their intensity. The latter property allows to apply such ceramics in light modulators. Pyroelectrics are also used in the manufacture of infrared

detectors with extremely high sensitivity that allows to record extremely small changes in temperature (up to 10^{-6} K).

Magnetic ceramics. This class of ceramic materials include primarily ferrites used in electronics and electrical engineering, in particular at high working frequencies of alternating electric and magnetic fields.

It was shown in Part 1 that the magnetism of a substance depends on the nature of interaction between the magnetic moments of atoms constituting the crystal lattice. Thus neighboring atoms in the lattice of the material may be different in nature. Interaction between these atomic moments, depending on the interatomic distances, leads to their orientation in either one direction (ferromagnetism) or opposite (ferrimagnetism and antiferromagnetism).

The first case – ferromagnetism, for example, is realized in α-iron (see, Part 1). In case of equality of the magnetic moments of neighboring ions and their antiparallel orientation material becomes *antiferromagnetic*, as, for example, in chromium. *Ferrimagnetism* is observed when the magnetic moments in adjacent ions are not equal and oriented oppositely, that leads to their incomplete compensation. Materials of this kind have been called *ferrites*.

Many of ferrites are prepared on the base of natural magnetite Fe_3O_4, which is a double oxide of $FeO \cdot Fe_2O_3$. In this case, neighboring iron ions Fe^{2+} and Fe^{3+} with non-equal magnetic moments are located in two different sublattices. Magnetic properties of magnetite are determined by the mutual orientation of these moments.

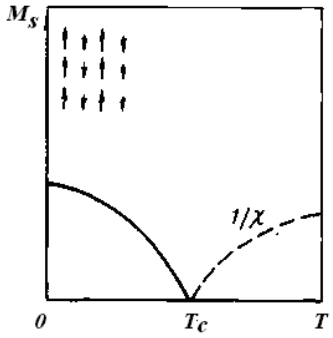


Fig. 3.2.4. The temperature dependence of the magnetic saturation moment MS and inverse magnetic susceptibility χ in ferrimagnetic material. The inset shows schematically the arrangement of the vectors of the magnetic moments

Schematic (antiparallel) arrangement of the vectors of the magnetic moments and the total magnetic moment of saturation in ferritic crystals depending on temperature are shown in Fig. 3.2.4. As shown, when heating it reaches

a critical Curie temperature T_C where ferrimagnet loses magnetic properties and has zero magnetic moment.

Ferrites have significant differences from the metallic ferromagnetic materials because combine high magnetic parameters (low coercivity, high magnetic susceptibility, etc.) with good dielectric properties. Electrical resistivity of iron ferrites reaches values of the order of 10^3 Ω·cm, which is six orders of magnitude greater than that of iron.

Soft magnetic ferrites are characterized by very narrow hysteresis loop and magnetized proportional to the applied magnetic field. The high resistance of ferrites eliminates eddy currents which are responsible for the superheat of transformers cores made from soft iron. Therefore, they are used in the cores of transformers operating at frequencies above 100 kHz. Ferrites containing manganese and magnesium cations can create almost rectangular hysteresis loop that was previously used in the memory cells of computing devices. The most important applications of ferrites in electrical and radio engineering are cores in inductors, transformers, throttles, magnetic antennas and magnetic circuits capable to operate at high and radio frequencies.

Addition of barium to magnetite allowed to create a group of iron-based ferritic materials like $BaO{\cdot}(Fe_2O_3)_6$, which have, on the contrary, a large coercivity exceeding 80 kA/m. Such material (e.g., *magnadyur* $BaFe_{12}O_{19}$) are used for the manufacture of permanent magnets.

Superconducting ceramics. In 1986 ceramics based on lanthanum, barium and copper oxide was firstly obtained. It had an unusually high superconducting transition temperature T_C = 35 K. The T_C of this ceramics not only exceeded the boiling point of liquid helium (4.2 K), but also turned out to be 12 K higher than the best known at that time T_C values for the known superconductors (see, Part 1). In 1987 was obtained ceramics based on yttrium-barium-copper oxide with T_C = 93 K, exceeding the boiling point of liquid nitrogen (77 K). From this time such ceramic materials are called high temperature superconductors (HTSC), in contrast to low temperature superconductors (LTSC) on the basis of metals and metallic alloys and chemical compounds.

It was found that yttrium-barium-copper oxide HTSC ceramics has optimal composition, when the ratio of Y, Ba and Cu atoms is maintained against 1:2:3 respectively. Structural studies have confirmed that this type of high-temperature superconductors has distorted perovskite-like structure with a deficiency of oxygen. Such a structure is characterized by the layered crystal lattice – the lattice constants in two directions are 0.28 nm and in the third – 1.2 nm. Calculations show that the conduction electrons are concentrated in a copper-oxygen layer as a result of overlapping of d-shells of copper atoms with p- shells of oxygen atoms.

Although it was proved that in the high-T_C ceramic materials superconductivity is provided by Cooper electron pairs, the nature of electron-electron attraction forces is not yet clear. It is believed that in addition to the electron-electron attraction over phonons there is another (or others) the mechanism of mutual attraction of electrons.

Apparently, one of the first areas in which the HTSC may have a significant impact is the electronic equipment because Joule losses are almost zero for sufficiently high current densities. This can be used for increasing the densities of elements in integrated circuits up to 10^8 cm^{-2}.

One of the main applications of HTSC can also be high-speed surface transportation. Very real prospect of making motors with HTSC material, which in size is 10 times smaller than a conventional motor of the same power. Real prospect in connection with the use of HTSC acquires such developments as transportation on magnetic suspension, storage of electricity, MHD generators and electric transmission lines.

Ceramics for containers of radioactive materials and waste storage. At the beginning of the Part 2, by defining power industry development as a way of improving the living standards of society, it was stated that one of the limiting factors in the development of nuclear energy is the difficulty of disposal of radioactive waste. The conventional approach to the development of materials for this purpose consists of three stages:

1. The waste is introduced in a relatively chemically resistant insoluble substance;
2. This substance is enclosed in a leakproof containers;
3. Containers are buried in a dry and stable geological structure.

For the first stage were used and applied boron ceramics and glasses. The main requirement for such materials is their high ability to absorb neutrons and γ-rays. Among the elements of the periodic table light elements, such as H, Li, B, are possessed by the most absorbtion of neutrons. But they occur nuclear reactions during the absorption of neutrons that results in secondary γ-radiation. For this reason, the protective material should comprise, on the contrary, the heavier elements, mainly Pb, since the absorption γ-rays obeys an exponential law $N = N_0 \exp(-\alpha d)$, where N_0 – γ-ray quanta density before absorption; N – γ-ray quanta density after absorption; d – density attenuating substances; α – absorption coefficient.

Application of pure lead is impractical due to the considerable yielding of protective wall consisting of lead, even under the influence of its own weight. More effective protective materials from γ-radiation are lead oxide PbO and more complex oxides such $2PbO \cdot PbSO_4$. They have high densities, relatively

high operating temperatures and the manufacturable in the powder technology, and also when it is compressed and sintered. Before molding, these oxides are mixed with boron-containing compounds, such as B_2O_3, B_4C, with borates and borides of metals MBO_3 type of MW or MB_2. Note that metals M which are included in these compounds should give a very low level of secondary γ-radiation. After sintering, such mixtures are transformed into dense ceramics with low porosity.

But, despite all efforts, ceramics from boron- and lead-containing substances has many disadvantages. The main of them is lack of chemical resistance. It is pointed out even lower resistance other known and widely used materials, such as various concrete compositions. For this reason, in most cases, both concrete and B-Pb-ceramics are used more on the second stage in the form of leakproof containers. For the first stage is generally accepted that a boron-silicate glass is holding well radioactive waste. In the late 70-ies of the last century, the exclusive feature of two natural minerals – perovskite CaTiO3 and zirkonlite $CaZrTi_2O_7$ was opened. They were absolutely insoluble in water unlike many others substances. At the same time, Ca atoms in the perovskite and Zr atoms zirconlite may be replaced by the radioactive elements like uranium and thorium with the formation of solid solutions. Such solid solutions (and themselves raw minerals) can be produced using in solid-phase chemical reactions during sintering of ceramics from the oxides of titanium, zirconium and calcium containing additives of radioactive substances. On the basis of these minerals has been developed a special ceramics named "sinrok", successfully passed all the tests on stability.

3.2.6. Inorganic glasses

Inorganic glasses are obtained by supercooling of melts of inorganic oxides, aqueous salt solutions, as well as ultra-rapid quenching of the molten metals. Therefore, in contrast to crystalline solids, glasses are in a non-crystalline (amorphous) state. Often inorganic glasses have the frame-like atomic structure (Fig. 3.2.5a), which makes them similar to polymers. In glasses coordination number are usually conserved (Fig. 3.2.5a), although atoms at the second and more distant coordination spheres are arranged randomly. In other words, there is a conventional short-range order in glasses, but long-range order as in crystals is not observed (periodicity is absent).

Upon cooling, the molten glass passes from the liquid to a viscous state (at temperature $T = T_V$), and then the solid (at $T = T_S$). The temperature range ($T_V - T_S$) in which this transition occurs is called the *glass vitrification interval*. This interval ($T_V - T_S$) for the inorganic glasses depends on the chemical composition and cooling rate, and generally does not exceed 100-200°C.

Transition process liquid ↔ glass during heating and cooling may be reversible if the crystallization occurs when cooling.

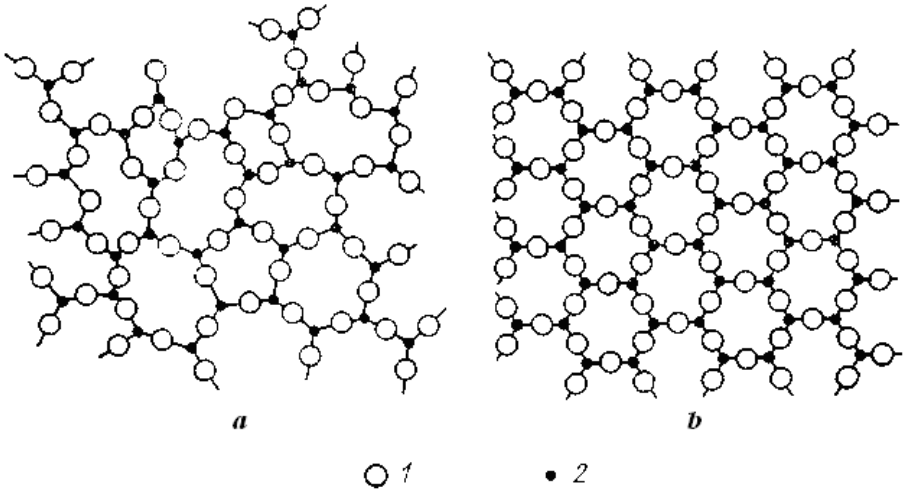


Fig . 3.2.5. A two-dimensional diagram of the pure silica structure in vitreous (*a*) and crystalline (*b*) states: *1* – oxygen ions, *2* – silicon ions. The diagram shows only three of the four oxygen atoms surrounding each Si atom

Inorganic glasses differ by composition and functional performance. By composition, silicate glasses are picked out. They consist of silica and other oxides (primarily, B_2O_3 and P_2O_5).

Physical-chemical properties and application of glasses. The most important property of the glasses is their *transparency* in defined ranges of the electromagnetic irradiation spectrum. Inorganic oxide glasses are characterized by high transmittance I/I_0 (I_0 – Intensity of the incident light on the glass surface, I – intensity of light transmitted through the glass) in the visible wavelength range. For example, for conventional inorganic glasses used in windows transparency is about 0.83-0.90 and for optical glasses – 0.95-0.99.

The transparency exists in those cases when after cooling of glass from a liquid state it does not contain defects, like gas bubbles, cracks, and also internal crystallites and grain boundaries, with sizes close to the light wavelength. Just these defects are sources of electromagnetic radiation scattering in ceramics and composites, produced by sintering. Refractive index of glasses may be varied by suitable additives. High transparency of some inorganic glasses made them indispensable for the glazing of buildings and various vehicles, manufacturing of lamps with wide assortment and purposes, lighting equipment, etc.

Depending on the composition and production conditions, inorganic glasses may acquire the ability to refract, scatter and absorb electromagnetic radiation not only in the visible (light), but also in the ultraviolet, infrared and X-ray ranges of the spectrum. Some inorganic glasses are characterized by photosensitivity. Such glasses are used for the production of the so-called photochromic inorganic glasses allowing to make energy-saving glasses in insulating glazed windows.

From the viewpoint of mechanical properties (response on mechanical load), glass is viscoelastic fluid medium. Which properties of glass predominate – elastic or viscous – depends on the ratio of the loading duration to the relaxation time. As for glasses $G \approx 30$ GN/m^2 and a viscosity at room temperature is $\eta \geq 10^{13}$ Ns/m^2, we obtain for the relaxation time the value $\tau \approx 330$ s. This implies that under normal loads, with short duration (e.g., at a shock), the glass exhibits elastic properties and is easily broken at room temperature.

The modulus of elasticity E for various inorganic glasses ranges 44.2-87.2 GPa. Tensile strength of glass is usually very small – about 100 MN/m^2, and there are significant deviations from the mean value. Glass can withstand large compressive stresses (~ 10 GN/m^2) that is approximately equal to the theoretical strength (E/10) ≈ (7-10) GN/m^2). These values are not depend practically on the composition. Bending strength of inorganic glass is about 30-120 MPa. Low tensile strength is due to the presence of cracks at the surface, which act as a local stress concentrators.

Electrical properties of inorganic glass depend on composition and operating temperatures. As a result, inorganic glass can be insulator, semiconductor or conductor of electricity. Since the charge carriers in inorganic oxide glasses are cations of alkali and alkaline earth metals, their electrical conductivity generally increases with increasing the content of these ions and the temperature increase.

Glass insulators are used in high-voltage power lines. Suitability of electrical inorganic glasses for working in different temperature conditions depends on their composition and is estimated by temperature (TK_{100}), at which the inorganic glass has electrical conductivity $1.00 \cdot 10^{-6}$ S/·m. In electric industry quartz glasses with TK100 value = 600°C are used, while for others – 230-520°C.

Almost perfect insulator is quartz glass. Its permittivity has the lowest value (3.8-4.0) among glasses. Above 200°C, the permittivity increases with the temperature growth for the frequencies up to 50 Hz. Dielectric losses tgδ are mostly low for silicate glasses. For example, quartz glass at 20°C and a frequency of 1010 Hz has tgδ = 0.0001. Dielectric strength (breakdown

voltage) of inorganic glasses in a uniform electric field reaches high values (10^4-10^5 kV/m).

Thermal conductivity of glasses is several orders lower than that of ceramics with the same crystalline structure. For example, for silicate inorganic glasses thermal conductivity is 0,6-1,34 W/(m·K). This is due to the fact that passing of waves, associated with the thermal fluctuations of atoms, through glass is substantially hindered by their disordered structure, in particular due to nonclosedness frame-like net (Fig. 3.2.5a). For the same reasons, *thermal expansion coefficient* of quartz glass (around $5.5 \cdot 10^{-7}$ K^{-1}) is significantly lower than that of metals (~ 10^{-5} K^{-1}).

Thermal stability of glasses, i.e. their ability to maintain mechanical and other characteristics when heated, essentially depends on their composition. Conventional silicate glasses are heat resistant in the temperature range 60-100°C: for Pyrex it is close to 280°C, for quartz glass – about 1000°C. Specific heat of inorganic silicate glasses is in the range of (0.3-1.05) kJ/(kg·K) at room temperature.

Inorganic glasses possess by *chemical resistance* to the action of air, water, acids (HF, H_3PO_4). The most chemical resistance is characteristic for quartz, boron-silicate (no more than 17 wt.% B_2O_3) and alumo-silicate glasses.

3.2.7. Vitrified glasses (Sitalls)

Glass ceramics (in Russian literature the term Sitall is used) are polycrystalline materials prepared by controlled crystallization of glass keeping a small amount of polymer-like phase with frame-like (net) structure in them. This class of materials is characterized by a combination of physical and mechanical properties of conventional glasses and ceramics (see below). Consider briefly the main characteristics of this unique type of special materials.

Glass ceramics are commonly obtained from photosensitive glasses, in which clusters of copper, silver and gold are precipitated under the action of ultraviolet rays, and subsequent heat treatment. Thereafter, heating above a certain temperature results in crystallization of the vitreous mass, when crystallites are nucleated on metallic nanoparticles. Since the number of nucleation centers is very large, and their distribution on the glassis homogeneous, so crystallized ceramics is characterized by fine-grains, homogeneous composition and lack of porosity. The mechanical strength and dielectric resistance of such ceramics are very high .

Glass ceramics commonly obtained from photosensitive glass , which under the action of ultraviolet rays, and subsequent heat treatment fall clusters of copper, silver and gold. Thereafter, heating above a certain temperature results

in the crystallization of the glass, which originates at the metal nanoparticles precipitated . As the number of nucleation sites is very large, and their distribution on the glass evenly, then crystallization occurs when ceramics characterized by fine-grained, homogeneous composition and lack of porosity. The mechanical strength and electric resistance of such ceramics were high.

During heat treatment of vitreous mass properties of glass ceramics formed are changed. It becomes non-transparent (unlike glass) due to light scattering at the interfaces between the crystalline grains and the remaining glassy phase, which have different refractive indices. The thermal expansion coefficient generally decreases, but it is possible to vary coefficient in the wide range from 10^{-7} to 10^{-5} K^{-1}, chousing those to be equal thermal expansion of metal components, with which ceramics is working. When glassy phase disappears, glass-ceramic thermal conductivity sharply increases, but never reaches the values characteristic of the non-transparent crystalline oxides. The softening point is increased from about 500 to 1000°C. Electrical properties are generally improved due to increasing of resistivity and lowering of dielectric loss tangent.

Perhaps the most important effect is at least a two-fold increase in the average mechanical strength – approximately from 100 to 200 MN/m^2 and sometimes even more. Thermal shock resistance is also increased sharply in glass ceramics. Vickers hardness is low (about 700), but the abrasion resistance is extremely high and close to the resistance of sapphire (pure crystalline alumina). Currently, glass ceramics are increasingly used as insulators, due to their high dielectric strength and the ability to choose the coefficient of thermal expansion that is consistent with the thermal expansion of the supporting metal parts.

3.3. Organic materials

Polymers materials and produced on their basis plastics, rubbers and wood are the basis for organic materials.

Polymers are *synthetic* and *natural (organic).* Naturally occurring polymers include natural rubber, cellulose, mica, asbestos, natural graphite. However, the leading group of polymeric materials is synthetic polymers, obtained during the chemical synthesis of small molecule compounds. The possibilities of creating new polymers and changing the properties of existing are very large. Synthesis can produce polymers with varied properties and even create materials with predetermined properties.

3.3.1. Polymers

Polymers called synthetic or natural substances (inorganic and organic, amorphous and crystalline), consisting of «monomeric units» connected by

chemical bonds or a focal length of the macromolecule, which consist of numerous elementary units (monomers) of the same structure. The polymer is a high molecular compound: number of monomer units in the polymer (degree of polymerization) must be sufficiently large. As a result, the molecular weight of monomers may be from several thousand to several million. In many cases, the amount of monomers (units) may be sufficient to refer to a polymer molecule if in the process of the monomeric unit adding molecular properties are not changed. With such large amounts of macromolecules properties of polymers depend not only on the chemical composition of macromolecules, but their mutual arrangement and structure.

Macromolecules of polymer are chains composed of individual repeating structural fragments – monomeric units comprising a few atoms. Cross section of such a chain is a few tenths of a nanometer and length – thousands of nanometers. That is why polymer macromolecules inherent flexibility (which is limited to the size segments – hard areas consisting of several units). Flexibility of macromolecules is one of the distinguishing features of polymers, which determines their applied significance.

Macromolecule of polymers are characterized by strong chemical bonds in the macromolecules themselves and relatively weak (such as Van der Waals forces) between them. Atoms included in the macromolecule often linked by covalent bonds. Along the chain the energy of these bonds in the average is $(250-350)\frac{\text{kJ}}{\text{mol}}$. For example, for the C-C bonds, it is equal to $80\ \frac{\text{kJ}}{\text{mol}}$, for C-O bonds – $79\ \frac{\text{kJ}}{\text{mol}}$ and for C-N bonds – $66\ \frac{\text{kJ}}{\text{mol}}$. Energy of bonds between the chain molecules is electrostatic in nature and usually does not exceed $37\ \frac{\text{kJ}}{\text{mol}}$ (i.e. 10-50 times less). However, in actual polymers such interactions are of great importance due to the large length of chain macromolecules. The strongest intermolecular interactions occur through hydrogen bonds (they are only 4-10 times weaker than covalent).

Macromolecule of polymers may have the same chemical composition, but differ in size. This phenomenon is causing a large spread of physico-chemical characteristics of the material, called the polydispersity.

Macromolecules can be constructed from the same chemical structure of the monomers (*homopolymers* or *polymers*) or dissimilar monomeric units (*copolymers* or *heteropolymers*). Homopolymers include, for example, polyvinyl chloride (in which structural units $(\text{-CH}_2\text{-CHCl-})_n$ are repeated), natural rubber, etc. Sometimes macromolecules of material may consist of

alternating large chemically homogeneous segments (*blocks*) of different compositions, which are called *block copolymers*.

In the process of synthesis, we can add (to "ingraft") to a main molecular chain, composed of identical monomers of one type, some sections of the other monomers. In this case the so-called *graft copolymers* are obtained. When the backbone of the same atoms is built, the polymer is called *gomochained*. If the polymer is built up of atoms of different types, the polymer is called *heterochained*.

Of great importance is the *stereospecificity* of polymer, when all links are arranged in space in a certain order. This gives the material with the improved physical-chemical properties, compared with irregular polymers.

The polymer is formed from monomers by *polymerization* or *polycondensation* reactions. Numerous natural organic compounds belong to polymers: proteins, nucleic acids, polysaccharides, natural rubber and other organic substances. In most cases, this concept relates to organic compounds, but there are a lot of *inorganic polymers*. A large number of polymers are produced by synthetic method on the basis of the simplest compounds of origin natural components through polymerization reactions, chemical reactions, and polycondensation. The names of the polymers are formed from the monomer name with a prefix *poly-*: *poly*ethylene, *poly*propylene, *poly*vinylacetate, etc.

Industrial production of chain polymers was developed in two directions. They were obtained by either processing of natural or organic polymers in artificial polymer materials, or synthesis of artificial polymers from organic low-molecular compounds. In the first case, the production of polymers is based on *cellulose*. Just on this basis *celluloid* was obtained, the first polymeric material. On the basis of Simple and complicated cellulose esters are still the bases for production of the most films, fibers, paints and stickers.

Production of synthetic polymers has began with bakelite resin which is a product of condensation of phenol and formaldehyde, converted by heating in a three-dimensional polymer. Bakelite for a long time has been used for the manufacture of packages in electrical products, batteries, televisions, outlets, etc., and is currently being used as a binder and adhesive substance.

In the mid-20th century industrial production of synthetic rubber and then polystyrene, polyvinyl chloride, polymethyl methacrylate and high-density polyethylene has started. Polystyrene and polyvinylchloride are excellent electrical insulating materials. Without PMMA (organic glass titled «Plexiglas») massive aircraft construction would be impossible.

In the 50s of the 20th century polyester fiber has been developed. Such fabrics as polyester or polyethylene terephthalate, which are then used for

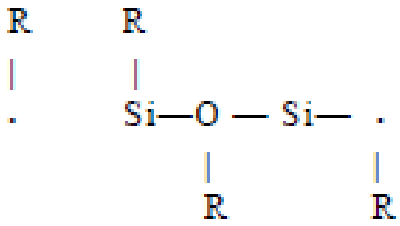
synthetic wool (polypropylene and nitrone) production, are obtained on its basis.

Discovery of catalysts and their rapid industrial development in the mid 50s of the 20th century has led to the emergence of such polymeric materials as polypropylene and low pressure polyethylene and stereoregular polymers with the ability to crystallization. Then polyurethanes − the most common sealants, adhesives and soft porous material (foam) − were prepared, as well as polysiloxanes − organometallic polymers having a high heat resistance and elasticity compared with the organic polymers.

Classification of polymers. For the convenience of studying the interconnection between the composition, structure and properties of polymers it is adopted to classify them according to different features (composition, shape of macromolecules, phase state, the polarity of bonds, resistance to the heat). By chemical composition, all the polymers are divided into organic, organometallic and inorganic.

Organic polymers constitute the mostly comprehensive group of polymers. Natural organic polymers are resins and rubbers.

Organometallic polymers are created synthetically. They contain in the composition of the main chain only atoms of silicon, titanium, and aluminum, which are combined with organic radicals of the methyl, ethyl and phenyl types. Such polymers don't exist in nature. Organosilicon compounds are representatives of this group of polymers. They have the following form:



The silicon and oxygen atoms can form the *Siloxane bond*, which is very strong, reaching $360\frac{\text{kJ}}{\text{mol}}$. Therefore, silicone resins and rubbers are highly heat-resistant, though their flexibility and elasticity is less than that of organic ones. Polymers having titanium and oxygen in a backbone are called *polytitaniumoxanes*.

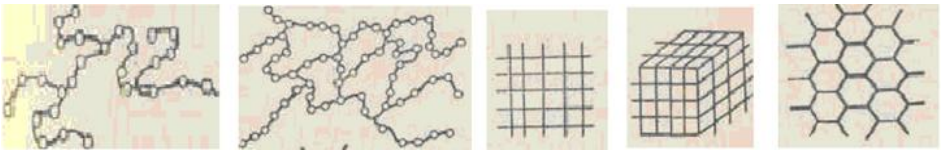
Organic radicals impart strength and elasticity to organometallic polymers, and inorganic atoms give the increased heat resistance.

Silicate glasses and ceramics discussed above are also often referred to synthetic inorganic polymers. Mica and asbestos relate to the natural inorganic polymers. These compounds do not have a carbon skeleton in their composition. They are based on various oxides of silicon, aluminum, magnesium, calcium, etc. Natural graphite, which is a carbon-chain polymer, also relates to inorganic polymers.

Both individual polymers and combinations of their different groups and with other types of materials are used in the specific technical materials. Such materials are called compositional (such as fiberglass).

The peculiarity of the polymeric properties is due to the structure of macromolecules. Under the form of the macromolecules polymers are divided into linear (chain-shaped), branched (a special case − star-shaped), flat, belt (stair-like), comb, spatial or net-like, etc. Linear macromolecules of polymer are long (zigzag or twisted into a spiral) chains (see, Fig. 3.3.1a).

Flexible macromolecules with high strength along the chain and weak intermolecular bonds provide elasticity of the material; its ability to soften when heated and to harden when cooled again. Many of these polymers are soluble in solvents. The packing density of molecules per unit volume affects on the physical, mechanical and chemical properties of linear polymers. In close-packed state the stronger intermolecular attraction appears, that leads to an increase in density, strength, softening temperature and to a solubility decrease.



a) *b*) *c*) *d*) *e*)

Fig. 3.3.1. Schematic representation of different types of polymeric macromolecules: *a* − linear (chain), *b* − branched, *c* − flat (two-dimensional), *d* – netlike (three-dimensional), *e* − hexagonal cellular (graphite)

Linear polymers are most suitable for preparation of fibers and films (e.g., polyethylene, polyamides, etc.).

Branched macromolecules of polymer (see, Fig. 3.3.1b) are actually also being linear, distinguished by the presence of lateral branches. These branches hinder the convergence of the macromolecules, i.e. their close packing. This shape of the macromolecules predetermines the reduced intermolecular interaction and consequently lower strength, lower thermal resistance (lower melting point) and the improved solubility (polyisobutylene). The branched

polymers include, in particular, “graft” polymers, which have unlike composition of the main chain and widely spaced side branches.

Spatial or netlike polymers of different types (flat, three-dimensional and hexagonal) are formed by the connecting («linking») of the macromolecules between themselves in a transverse direction by strong chemical bonds either directly or via chemical elements or radicals. As a result of such a connection of the macromolecules netlike structure with different types of unit cells in nets is formed (see Fig. 3.3.1c, 3.3.1d).

Lamellar polymers that have a two-dimensional planar structure are related to netlike polymers (see Fig. 3.3.1c). Such polymers lose their ability to dissolve, are melted at higher temperatures and have high elasticity (for example, elastic rubber). Netlike polymers can form not only two-dimensional but also three-dimensional cellular nets. Such polymers are characterized by sufficient hardness, as well as thermal and chemical resistance. Spatial netlike polymers underlie many structural nonmetallic materials. Natural graphite is an example of such a polymer having a reticular spatial structure of hexagonal type (see, Fig. 3.3.1d). By the phase state, polymers are divided into amorphous, crystalline and amorphous-crystalline.

As was revealed by X-ray diffraction and electron microscopy analysis, macromolecules in polymers can be located not only randomly, but also can have an ordered structure. Structures resulting from different assemblying (the order in arrangement) of molecules are called *supramolecular*. Ordering in the formation of polymer structures is determined by the flexibility of linear and branched (with short branches) units of macromolecules. Latter gives to macromolecules the ability to change their shape, moving by parts (for example, by twists or turns). Great influence on the structure of the macromolecules provides interatomic chemical bonds in the chain (they define rigidity of the chains) and intermolecular forces (they determine the ability of the macromolecules to sliding relative to each other).

Amorphous polymers are single-phase. They can be constructed of chain molecules, assembled in different ways forming a supramolecular structure. In one types, macromolecules of the amorphous polymer are assembled in structural elements called *batches*. The latter consist of many rows of macromolecules that are able to move relative to the adjacent structural components (batches) to accommodate consecutively after each other, without forming an ordered structure.

Amorphous polymers can also be constructed from chains, rolled into coils, which are called *globules*. Globular structure of some polymers decreases the mechanical properties, increasing the fragility of polymers because of the possibility of bond breaking along the boundaries of the globules. At elevated

temperatures globules unfold, forming linear structures, thereby increasing mechanical properties of polymers.

Crystalline polymers are formed if the structural elements constituting macromolecules (e.g. batches) can rearrange, and phase transition with the formation of spatial lattice (ordering) is possible in them under appropriate conditions.

Crystallizable polymers include polyamides, polypropylene, polyethylene, and others. Unlike metallic crystals, crystallization of polymers is always realized in a certain temperature range. Under normal conditions complete crystallization of polymers doesn't occur. As a result, the actual polymers typically have a biphasic structure consisting of crystalline and amorphous phases. Availability of crystalline phase imparts enhanced heat resistance, rigidity and strength to polymers.

Polymers are usually divided by polarity of their macromolecules units that affect their solubility in different liquids. Polarity of polymer units is determined by the presence of dipoles in their structure. Polymers, whose units have significant polarity, called *polar* or *hydrophilic*. Polymers with nonpolar units called *nonpolar* or *hydrophobic*. Polymers, whose unit contains both polar and nonpolar groups of atoms, are called *amphiphilic*.

Polarity of the substance can be estimated by the average value of the dipole moment *p*, equal to the product of the elementary charge (electron charge) *q* on average distance *l* between the centers of gravity of all positive and all negative charges. It is easy to show that the mean values of the dipole moment $p = ql$ are of the order $5 \cdot 10^{-27}\ \mathrm{C} \cdot \mathrm{cm}$. This value is sometimes called Debye unit. For example, for C-H, C-N, C-O, C-F, C-C1 bonds *p* values in Debye units are respectively equal D = 0.2; 0.4; 0.9; 1.83; 2.05.

Conditions of polarity occurrence in polymers are determined by the presence of atoms forming the monomeric units of polar bonds (groups like C-F, C-C1, C1-OH or F-OH) in groups and asymmetry of their structure. For example, non-polar polyethylene has a symmetrical arrangement of the atomic groups CH_2 in monomers, and therefore the dipole moments of the atoms bonds are mutually canceled out. In polypropylene dipole moments of groups C-H and C-CH_3 are equal and therefore it is nonpolar. In fluoroplastic-4 dipole moment of C-F bond is significant, but the sum of the moments is zero, since they cancel each other out. So fluoroplastic-4 is also nonpolarn. Monomer in polyvinylchloride is asymmetric and therefore dipole moments of C-H (0.2D) and C-O (2.05D) are not mutually compensated.

Polarity significantly affects the properties of the polymers. Thus, the non-polar polymers (mainly hydrocarbon-based) are good high frequency dielectrics. Mechanical properties of non-polar polymers at low temperatures

deteriorate slightly, so that they have good frost resistance (e.g., polyethylene does not become brittle even up to temperatures of -70°C). Polarity, increasing the intermolecular forces of attraction, gives the polymer rigidity and heat resistance, but reduces their dielectric properties (impeding reversal processes). Therefore dielectrics based on polar polymers can operate without loss in only a limited range of low frequencies. Furthermore polar polymers are characterized by low cold resistance (e.g. polyvinylchloride can not be used at temperatures below 10-20°C). With respect to heating all polymers are usually classified as thermoplastic and thermosetting.

Thermoplastic polymers (or *thermoplastics*) are soften when heated, and some of them are even melted and then solidified upon cooling. This process is reversible, i.e. no further chemical transformations of the material does not undergo. Representatives of thermoplastics include polyethylene, polystyrene, polystyrene, polyamides, etc.

Thermosetting polymers are of a linear structure at a first stage of formation and softened at heating (without melting). Then they, due to chemical reactions, are solidified (irreversibly altering their chemical structure) and remain solid. Examples of thermosetting polymers are phenol, glyptal and other resins. Elastic properties of the thermosetting polymers are higher than that of thermoplastics, but they do not practically have fluidity, thereby have a lower fracture stress.

Main characteristics of polymers. Features of the polymers structure have a decisive impact on their physical and chemical properties. Due to the high molecular weight they can not pass in a gaseous state (evaporate) during heating and can not form a low viscosity fluid. Those that possess the thermostable spatial structure even not soften at elevated temperatures. The solubility of polymers decreases with increase of molecular weight. At molecular weight of the order of $(3 \div 4) \cdot 10^5$ and low polarity polymers are soluble in certain solvents

Mechanical properties of polymers (strength, elasticity, impact toughness) not only depend on their structure, but also on the physical state, temperature, humidity, etc. From the point of view of behavior under mechanical stresses polymers can be found in three physical states − glassy, rubbery and viscous-flow. Which of these states is realized, depends on the temperature, which determines the nature of atomic motion and the mobility of macromolecules and, as a consequence, the polymer structure.

In the *glassy (solid) state* atoms that enter into the molecular chain of polymer, can make only oscillatory motion about the equilibrium positions and the motion of their units and the whole movement of macromolecules doesn't occur.

Rubbery (highly elastic) state, inherent only to high polymers, is characterized by their ability to form large reversible changes of their shape at low mechanical loads. Such a state is characterized not only by vibrations of the atoms in the units, but also by themselves units' oscillations. As a result, of the vibrational motion of the units, macromolecules acquire the ability to bend. In the *viscous state*, the whole macromolecule as a whole is mobile (it is able to bend and slide relative to each other) and the behavior of polymer under load is similar to a fluid with very high viscosity.

The ability of transition from one physical state to another, when the temperature changes, is determined by the polymer structure. For example, polymers with linear or branched structure of macromolecules can come from the glassy to viscous-flow state with temperature increase. At the same time, polymers having a three-dimensional spatial structure may stay only in glassy state. The netlike structure with large-dimension structural elements allows it possible to convert polymers from glassy to high-elastic state when temperature increases.

Relaxation properties of polymers. Many of polymers` properties depend on the duration and the rate of mechanical stresses application. This is due to the structural features of macromolecules forming the polymer. Under the applied stresses, both straightening and «unwinding» of chains and movement of macromolecules, batches and other supramolecular structures occurs. Such changes in polymeric structure are called *conformation*. Conformation processes have activational nature and therefore characterized by a certain duration which is called *a relaxation time* τ. Therefore, the relaxation (establishment of equilibrium state) of the loaded polymer is not achieved immediately.

Experimental studies show that at conformational changes relaxation time of polymers in the high-elastic state equals $\tau \approx 10^{-4}$-10^{-6} s, so that, in case of of macromolecules themselves and supramolecular structures displacements, it can reach days and even months. For elastic polymers due to relaxation processes curves $\sigma(\varepsilon)$, measured at increasing and decreasing of load, often do not coincide, i.e. *hysteresis* appears.

All polymers are characterized by tensile strength increase with the loading rate increase (see, Fig. 3.3.2). At the same time, with a decrease in the rate of loading the elastic modulus and the length of the «fluidity» section are increased (influence of inelastic deformation is growing).

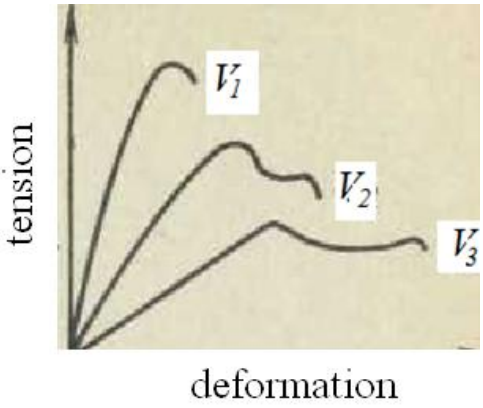


Fig. 3.3.2. Influence of the rate V of mechanical loading on the character of stress-strain curves ($V_1 > V_2 > V_3$)

Aging of polymers. Under the aging of polymeric materials is understood the spontaneous irreversible change (degradation) of the most important technical characteristics, occurring as a result of complex physical and chemical processes developing in material during operation and keeping. Factors that determine the aging of polymers are the humidity, heat, oxygen, ozone, light, etc. The aging process is enhanced in multiple deformation. Aging is significantly less affected as compared with humidity.

Aging is caused by complex chain reactions occurring under the influence of the external factors listed above which leads to the formation of free radicals (rarely of ions). It is accompanied by destruction of the polymer and (or) its restructuring. Quite often, the cause of aging is oxidation of the polymer by atmospheric oxygen or ozone.

3.3.2. Plastics

Plastic materials or plasrics are called synthetic substances obtained on the basis of organic polymeric binders with the fillers. Due to the presence of the polymeric binder, these materials are capable to become soft upon heating and become flexible, that allows imparting them a predetermined shape under pressure. Depending on the nature of a binder polymer substance, type of the filler substance and applied pressure, transition of the plastic product in the solid state is carried out either during its additional heating or when subsequent cooling.

General characteristics of plastics. Mandatory component of most plastics is a binder. The binders used in the preparation of plastics are synthetic resins and cellulose esters.

Filler is the second major component of a significant part of plastics. Powdered, fibrous and other materials both organic and inorganic origin are used as fillers. After impregnation of the filler by a binder, the semifinished product, pressed into a solid mass, is obtained. Fillers impart material high mechanical strength and other specific properties (friction, antifriction, etc.).

To increase the plasticity semifinished product, plasticizers are usually added. Plasticizer are organic substances with high boiling point and low freezing point (e.g. oleic acid, stearin, dibutyl phthalate, etc.). The plasticizer imparts elasticity of plastics, that facilitates imparting the desired shape to products.

Finally, the original composition may contain hardeners or inhibitors of thermosetting binders and dyes (mineral pigments and organic dyes alcoholic solutions, serving for decorative purposes). Inhibitors protect plastic semifinished products against premature (spontaneous) solidification.

The properties of plastics are dependent on such parameters as composition, shape and dimensions of particles of components (binder, filler, plasticizer, inhibitor, etc.), and also their quantitative ratio. Varying these parameters allows to change the characteristics of plastics in a fairly wide range. These issues will be considered in detail in the next Section «Composite Materials» below. By the nature of the binder plastics are divided into thermoplastics, obtained on the basis of thermoplastic polymers, and thermosets, based on thermosetting resins.

Many of plastics, mainly *thermoplastic*, are composed of only a binder (polymer), such as polyethylene, organic glass etc. Just to this reason, this specified type of plastics will be discussed in this section. The rest plastics, which use fillers and which are actually composites, will be examined in the Section on composite materials.

Thermoplastics. The basis of thermoplastic materials are polymers with linear or branched structure, in which composition the plasticizers are sometimes introduced. Among the thermoplastic polimers both nonpolar and polar plastics are quite common.

Let us first consider non-polar thermoplastics, which include polyethylene, polypropylene, polystyrene, and fluoroplastic-4.

Polyethylene, comprising of the chains of the type $(-CH_2-CH_2-)$, is a product of polymerization of ethylene (a colorless gas) and belongs to crystallizable polymers. By density, polyethylene subdivided into low and high density polyethylene. *Low-density polyethylene* is polymerised at high pressure and contains 55-65 vol.% of crystalline phase. *High-density polyethylene* has a higher degree of crystallinity (up to 74-95 vol.%) and is obtained at low pressure.

The higher the polyethylene density and crystallinity, the higher its mechanical strength and heat resistance. Thermal stability of polyethylene does not exceed 60-100°C, and the frost resistance is about 70°C or lower. Polyethylene is chemically resistant and while heating is resistant to water, acetone and alcohol. Under normal (room) temperatures polyethylene is insoluble in any of the known solvents.

Unstabilized polyethylene prones to aging. To suppress aging, stabilizers and inhibitors are introduces. For example, about 2-3% of carbon-black slows the aging process about 30 times. By radiation (ultraviolet, β- and γ-radiation) polyethylene hardens, gaining considerable strength and heat resistance. Polyethylene is used for the manufacture of pipes, cast and extruded containers, covers, valves, etc., as well as films for wire and cable insulation and protection of metals from moisture and corrosive environments.

Polypropylene, comprising the chains of (-CH_2-$CHCH_3$-) type, is derived from ethylene with significant amount of stereoregular structure which is obtained by using organometallic catalysts. It is a non-toxic material with high physical-mechanical properties. Compared to polyethylene, this plastics has higher than polyethylene rigidity and temperature resistance while retaining the shape to a temperature of about 150°C. At the same time, the polypropylene can not be used at temperatures below -20°C. Films of propylene are more durable and gas-tight and fibers are flexible, durable and chemically stable at normal temperatures, compared with polyethylene products. Unstabilised polypropylene like polyethylene is liabled to rapid aging. Polypropylene is used for the manufacture of pipes, structural parts of automobiles, motorcycles, refrigerators, pump housings, various containers, etc. Polypropylene films are used for the same purposes as polyethylene ones.

Polystyrene comprising of (-CH_2-CHC_6H_5-) type chains has an amorphous structure and considerable strength. Due to the high hardness, polystyrene is mechanically processed well, but is prone to cracking. By the dielectric characteristics, it is close to the polyethylene.

Polystyrene is chemically resistant to acids and alkalis and is insoluble in alcohol, gasoline, oil, water. However, being a non-polar, it is soluble in many polar solvents such as benzene or dichloroethane. Due to the presence of the phenyl radical C_6H_5 in macromolecules polystyrene is more resistant to radiation as compared with other thermoplastics. Disadvantages of polystyrene are its low thermal resistance and susceptibility to aging in unstabilized state.

Blockcopolymer of styrene with synthetic rubber gives acrylonitrilebutadienestyrene (ABS) polystyrene, in which the impact resistance is 3-5 times and the elongation is in 10 times higher than that of conventional polystyrene. Such copolymers are characterized by high thermal stability (up to 100-125 °C) and chemical resistance. However the dielectric properties of these

plastics are lower than for the conventional polystyrene. Radio and electrotechnic components, cans and other parts of machines, vessels for liquids and chemicals, films for electrical insulation (e.g. stirofleks) are produced from polystyrene. ABS-plastics are used to manufacture durable pipes, in automobile industry, electronic engineering, etc.

Fluoroplastics (polytetrafluoroethylene fluoroplastic-4), consisting of chains of polytetrafluoroethylene (-CF_2-CF_2-), refer to saturated polymers with macromolecules as zigzag spirals. The presence of the amorphous phase in the rubbery state gives to fluoroplastic-4 the relative softness. Fluoroplastics destruction occurs at temperatures above 415°C. Heating to a temperature not higher than 250°C is slightly affected the mechanical properties of fluoropolymers because their rate of crystallization is small. Therefore, they are characterized by long-term operating at temperatures up to 250°C.

The fluoroplastic-4 glass transition temperature is 120°C. This plastics is not embrittled even at temperatures down to -269°C. Fluoroplastic-4 has high-temperature resistance due to high energy of C-F bonds. Furthermore, due to small size, the fluorine atoms form a dense shell around C-C chain, protecting it from chemicals that provides high chemical resistance of the material. Therefore, fluoroplastic-4 is resistant to solvents, acids, alkalis, oxidizing agents. Practically fluoroplastics are destroyed only by the subjection of molten alkali metals (potassium, sodium) and elemental fluorine. It is not wetted by water.

Fluoropolymers are unstable to irradiation. They belong to the most high-quality dielectrics whose properties vary slightly in a wide range of temperatures. Fluoroplastic-4 has a very low coefficient of friction (about 0.04), which is independent of the temperature up to 327 °C, when the crystalline phase begins to melt. Disadvantage of fluoroplastic-4 is its cold-fluidity (due to recrystallisation), separation of toxic fluorine at high temperatures and the difficulty of processing (due to the low plasticity). Well machined by cutting tools.

Fluoroplastic-4 is used for the manufacture of pipes, parts of valves, gates and pumps (seals, gaskets, bellows), electro- and radiocomponents, anti-friction coatings on metal bearings and bushings, etc. Fluoroplast products are fibers and films that do not burn, chemically stable, allowing their use for the manufacture of tanks, hoses, clothing, diaphragms, etc.

Organic glass (PMMA) relates to transparent thermoplasts based on esters of acrylic and methacrylic acids, which have an amorphous structure. Its transparency in the visible region of light reaches 92%, and in the ultraviolet range − 75% (in silicate glass is only 0.5 %). At the same time, this material has

approximately 2 times lower density (about 1.18 $\frac{g}{cm^3}$) as compared with the inorganic silicate glass, at the sufficient strength. It is highly resistant to environmental factors. At 80°C the organic glass begins to soften, and at temperatures 105-150°C plasticity occurs, that allows forming different parts from it. Mechanical properties of the organic glass are temperature dependent. The disadvantage of organic glass is the low surface hardness (it is easily scratched). Organic glass is resistant to the action of dilute acids and alkalis, hydrocarbon fuels and lubricants, but is soluble in esters and ketones, organic acids, aromatic and chlorinated hydrocarbons. Plexiglass aging is slow in natural conditions.

Increase of heat resistance and toughness of PMMA is achieved by stretching in different directions in the plastic state (this process is called *orientation of glass*). Orienting increases the toughness several times and increases the resistance to cracking. The surface hardness and heat resistance of the organic glass is achieved by copolymerization or "graft" polymerization with other polymers, obtaining partially crosslinked structure (heat stable glass), using laminated multi-layered glasses («triplexes») with organic glasses (such as polyvinylbutyral), which are glued by transparent film.

Criterion, which determines the suitability for use of organic glasses, is not only their transparency, strength and surface hardness, but also the appearance of small cracks on the surface and inside the material. The reason for the latters are internal stresses occurring during cooling of the material after molding, because of low thermal conductivity (large temperature gradients), and a high linear expansion coefficient.

Organic glasses are used in aircrafts and automotive industry. Illuminating engineering parts, optical lenses, etc. are made of organic glasses. On the basis of PMMA, self-solidificated plastics are produced, from which you can make stamps, casting patterns and abrasive tools.

Polyvinylchloride is a polar amorphous polymer composed of macromolecules with the chemical formula (-CH_2-CHC_1-CH_2-). Plastics on its basis have good electric insulating properties, are resistant to chemical and atmospheric agents and non-flammable.

Unplasticized solid polyvinylchloride was called viniplast. Viniplast have high resiliency and mechanical strength. Their disadvantages are weak long-term strength, low operating temperature under load (no more than 60 − 70°C), large coefficient of linear expansion $(6.5\text{-}8.0)\cdot 10^{-5}\ K^{-1}$, low temperature embrittlement (T_{embr} = -10°C).

Introduction of plasticizer in polyvinylchloride increases its frost resistance down to -50 °C and a softening point up to 160-195°C.

Pipes for corrosive gases, liquids and water, protective coatings for electrical wiring, parts of ventilation systems, heat exchangers, protective coatings for metal containers, facings for galvanizing baths and in bildings are manufactured from unplasticized polyvinylchloride. Film materials are used for wire and cable insulation, conservation of engines, manufacturing of security facilities when working with radioactive substances. Tubes, printed rolls, gaskets, etc., prepared from viniplasts, are used to cover the tissues (e.g. conveyor belts).

Polyamides are a group of plastics, composed of macromolecules that include groups of amide (-NH-CO-) and methylene (-CH_2-) repeated from 2 to 10 times. The general formula of the polyamides is:

-NH-CO-(CH_2)-N-CO-(CH_2)-.

Polyamides (caprone, nylon, etc.) are crystallizable polymers in which the individual links of macromolecules are arranged so that hydrogen bonds are formed between the groups CO and NH, belonging to different rings. This contributes to the formation of a regular structure and increase the melting temperature up to 210-264°C. At uniaxial orientation, polyamide fibers, filaments and films of high quality are obtained.

Properties of different types of polyamides are very close. The tensile strength for the films is about $(10-50)\frac{\text{N}}{\text{mm}^2}$ and for fibers – $600\frac{\text{N}}{\text{mm}^2}$. Elongation is in the range from 10 to 350%. Polyamides have a low coefficient of friction (less than 0.05) and can work for a long time to wear. Also, polyamides have a high impact resistance and able to absorb the vibration. They are resistant to high humidity, as well as to alkalis, gasoline and alcohol.

The disadvantages of polyamides are their hygroscopicity and liability to aging due to the oxidizability (especially in the processing of resins). Resistance of polyamides to light is increased by using the stabilizers. To enhance antifriction properties, fillers (e.g. graphite) are introduced.

Antifriction coatings of metals, gears, bushings, bearings, bolts, nuts, pulleys, oil pipes, hydraulic seals, wheel centrifugal pumps, turbines, turbodrills, etc. are manufactured from polyamides. Polyamides are widely used in the electrical industry.

Polyurethanes contain the urethane group macromolecules (-NH-COO-), where the oxygen atoms give the flexibility and elasticity to polymers. These materials are characterized by a high resistance to the atmosphere. Frost resistance of polyurethanes reaches -70ºC. The upper limit of the operating temperature is 120-170°C, although under explosion of high humidity it is reduced to 100-110°C.

Properties of polyurethane are close enough to the properties of polyamides. The produced from polyurethane fibers and film materials, which have low

hygroscopisity, are chemically stable and have low thermal conductivity. Depending on the used fillers, polyurethanes may have different properties, be solid, resilient and even thermoset.

Polyethyleneterephthalates (lavsan, mylar) are polyesters with the chemical formula:

-CH2-CH2-OCO-COO …

Polyethyleneterephthalate is a crystalline polymer, but under rapid cooling of the melt, amorphous polymer can be produced, which upon heating above 80°C again begins to crystallize.

The presence of oxygen in the main chain of the material gives good frost resistance (up to -70°C) to the material. Availability of the benzene ring increases the melting point to 255-257°C that raises thermal stability, including those at high humidity conditions. The mechanical tensile strength is about $17.5\ \frac{N}{mm^2}$, elastic modulus − of the order of $3500\ \frac{N}{mm^2}$. At the orientation of the polymer its strength increases. Polyethylene terephthalate is a good electric insulator and has a relatively high chemical resistance. It is used to produce fabrics, films, gears, brackets, belts, etc.

Polycarbonate (diflon) refers to polyesters of carbonic acid. It relates to crystallizable polymers, which can be converted into a glassy state and becomes transparent upon melting and subsequent cooling. Properties of polycarbonates are peculiar − they are characterized by strength, rigidity and flexibility at the same time. By tensile strength, this material is close to viniplast and has a high toughness. Samples retain their dimensions and flexibility at low temperatures during prolonged heating up to the softening temperature.

Polycarbonate is chemically resistant to salt solutions, dilute acids and alkalis, fuels, oils, but is destroyed in concentrated alkali. It is not aged after prolonged action of light, heat and vacuum, has good resistance to thermal shocks. Polycarbonate has limited resistance to ionizing radiation.

Flexible and durable films, the gears, bearings, car parts, radio components, etc. are made of polycarbonate. It can be used in cryogenics, including in contacting with liquid gases.

Polyformaldehyde refers to linear polymers which macromolecules ($-CH_2-O-$) contain oxygen. It has a high content of the crystalline phase (up to 75%) and, therefore, extremely dense packaging, which gives a high stiffness, hardness, impact resistance and flexibility simultaneously, as well as resistance to moisture (water), mineral oils and gasoline. This plastic can be used in a temperature range from -40 to +130°C. The combination of these properties allows manufacturing gears, bearings, valves, automobile parts, etc. from polyformaldehyde. Phenylon is used for manufacturing bearings, sealing details

of shutoff devices (valves, etc.), gears, electric and radio components, as well as different kinds of films, fibers and even paper (nomex).

Widely used aromatic heterocyclic polymers include linear *polyimides*, whose macromolecules chain contain imide rings and aromatic rings connected by flexible bonds of (-O-) and (-CO-) types. Depending on the structure, polyimides can be both thermoplastic and thermosetting. They are distinguished by quite a wide range of operating temperatures (from -200 to +300°C).

Polyimide products have high mechanical properties and good resistance to creep. Their strength is not inferior to the lavsan polyester strength. They are characterized by good electrical properties, and are resistant to radiation. Polyimides are resistant to solvents, oils, and weak acids and bases, but are destroyed by prolonged exposure of boiling water and water vapor. They can also operate very long time in a high vacuum, including high temperatures. Polyimide molded parts have a low coefficient of friction and resistance to abrasion.

In the form of films, polyimides are used for wire and cable insulation, printed circuit boards, electronic-vacuum thermal insulation. Pressed polyimide materials may be used for the manufacture of structural elements, including an antifriction and electrically destination.

3.3.3. Wooden materials

Wood is one of the most common materials for a long time used as a structural material (primarily in bilding and for walling). It is used both in natural form and in a variety of wood materials.

The advantages of wood as a construction material are sufficiently high mechanical strength and high specific strength (due to the low density), as well as good resistance to shock and vibration. Thermophysical properties of wood are characterized by low thermal conductivity, and a coefficient of linear thermal expansion is 2-3 times less than in steel. Wood has a high chemical resistance to a number of acids, salts, oil and gases. Important properties of wood are ease of machining (including bending), the ability to gluing, possibility of rapid mechanical connection by nails and screws.

Along with these advantages, wood has a number of drawbacks limiting its use as a structural material. We can note the following shortcomings: hygroscopicity, which is the reason for the lack of stability of shape, size and strength properties in the material that are changed with the changes in wet atmosphere; propensity to the defeat of fungal diseases; lack of fire resistance; low modulus of elasticity; anisotropy of mechanical properties due to the fibrous structure; heterogeneity of the structure, as a result of which the material

properties are different not only within the same breed, but within a single trunk.

Timber structure. Wood has a fibrous structure and consists of organic matter: 43-45% of cellulose ($S_6N_{10}O_5$), 19-29% of lignin, and the rest is the carbohydrates with low molecular weight and other components. The portion of inorganic components in timber does not exceed 1-2%.

Physical properties of the timber. Moisture, bulk density and changes in the size and shape of wooden products are among the main physical properties which play a major role for wood as a structural material.

Moisture content of wood is estimated as a percentage amount of water in it presented as $M = [(m-m_o)/m_o]\cdot 100\%$, where m is sample mass (in grams) with this humidity, m_0 – weight of the sample (grams) dried at temperature (100±5)ºC.

The water contained in the wood, is of two kinds: free (capillary) water filling internal voids and *bound (hygroscopic) water* in the cell membranes. When drying, wood loses first the free water and then the bound one.

For different types of wood the maximum number of related moisture in the industrial wood ranges from 23 to 30%. In freshly cut wood humidity can reach 50-100%. In greewood humidity can approach 10-20%. Room-dry wood has lied a long time in the indoor environment has a moisture content not exceeding 7-10% and for absolutely dry wood humidity is zero.

Humidity, meet the conditions of the workroom, is called *workroom humidity*. Humidity 15% is accepted to be the standard for humidity of wood. It represents an average humidity of air-dried wood. Therefore, for comparison, all properties of wood are typically defined by the standard timber samples with 15% humidity. Workroom humidity should be close to operational one or should be 2% lower (otherwise the wood will dry out).

The change of size and shape of wood (shrinkage, swelling, warping) is usually due to changes in humidity. During the drying of wood free moisture is removed first, so that the cell size does not change (only reduced the weight of water). After reaching the saturation point fiber walls of wood fibers begin to lose moisture and related dimensions of the product during the drying process are reduced. That is why expressed as a percentage decrease in linear dimensions and volume of wood during drying is called the shrinkage of wood. Shrinkage depends on the direction relative to the axis of the tree trunk: the lowest shrinkage occurs along the fibers, and the largest – the tangential direction.

Shrinkage plays a negative role because it serves the cause of internal stresses in the wood, causing cracks and warping, which should be considered in the manufacture of parts.

Wood of different breeds has the same chemical composition, so that the density of matter, forming cell walls, shall be equal to 1.54 kg/m^3. For practical purposes it is important to know the volumetric mass *m*, which depends on the moisture content and the coefficient of volumetric shrinkage. Average density of all the timber is 0.98 kg/m^3. Lighter species are pine, spruce, fir, linden, aspen, alder, very heavy – hornbeam, pear, boxwood. The greater bulk density, the denser the wood and the better it resists mechanical stress.

Due to anisotropic (fiber-like) structure of wood, its mechanical properties depend on direction of measurements (Fig. 3.4.1 and 3.4.2), the humidity, and other factors. Therefore, to be specific, the characteristics of the mechanical properties of wooden structures in the calculations relate to the strength of wood without defects at the constant humidity $W = 15\%$.

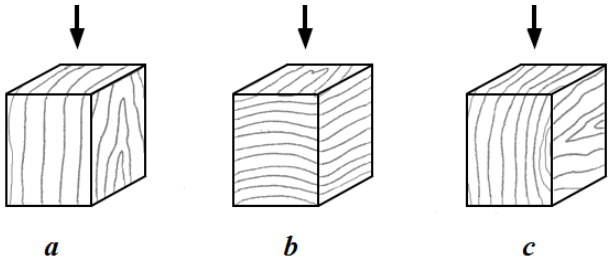


Fig. 3.4.1. The scheme of testing of wood on strength: *a* – along fibers, *b* – across fibers (radial), *c* – across fibers (tangential)

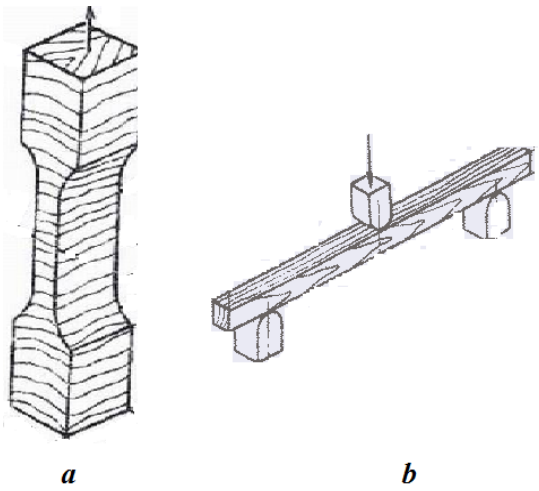


Fig. 3.4.2. The scheme of testing of wood on tensile strength (*a*) and bending strength (*b*)

The average yield strength of the timber along the fibers are in the following ranges: in the radial direction $\sigma_{FR} \approx (934.2 \div 54.9)\frac{N}{mm^2}$; in the tangential direction $\sigma_{FR} \approx (76 \div 160)\frac{N}{mm^2}$ (in some cases up to $270\frac{N}{mm^2}$); yield strength across fibers (face direction) σ_F is (6 – 30) times lower, than along. Shear resistance in the plane of the fibers (longitudinal direction – chipping) is small and amounts $\left(\frac{1}{6} \div \frac{1}{8}\right)\sigma_{FR}$, and resistance to bending σ_{Bend} is (1.5-2.0) times higher than σ_{FR}. Modules of elasticity in tension and compression are approximately equal, wherein in the longitudinal direction their values are (10-30) times greater than transversal. Along the fibers $E \approx (1.17 \div 1.58) \cdot 10^4 \frac{N}{mm^2}$.

Durability of wood depends on the loading rate: the loading rate less, the tensile strength lowers. At shock loading the resistance to bending of gummi-woods (ash, oak) is (1.5-3) times higher than the brittle soft-woods (pine, spruce, fir).

Over time, the wood resistance to mechanical stresses gradually decreases, reaching a certain limit of long-term resistance at which the wood element can operate indefinitely long time. For all kinds of the timber stressed state, the sustained resistance value is taken equal to $2/3$ of the ultimate strength. When vibrating loads, we must take into account fatigue (or endurance) of wood. The endurance strength σ_{Fat} is always less than the static tensile strength σ_{Stat}. The ratio $\sigma_{Fat}/\sigma_{Stat}$ at bending load amounts values from 0.24 to 0.38 for different wood.

Wood is easily flammable by the fire at temperatures of about (330 ÷ 470) ^oC. To improve its fire resistance, it is necessary to use impregnation by special chemicals – flame retardants (antipyrines) and flame retardant coating paints. As the flame retardants, ammonium salts and salts of phosphoric or boric acids are used. Fireproofing silicate paints based on liquid glass and perchlorovinylic coatings are nonflammable and and have low values of heet conductance.

3.4. Composite Materials

Conventionally used metallic and nonmetallic materials are reached, in a great measure, their limits of structural strength. However, the development of modern power technology requires the development of materials, working safely in a complex combination of force and temperature fields, under the impact of aggressive media, radiation, high vacuum and high pressures. Frequently the requirements for materials can be controversial. It is possible to meet this challenge by the use of composite materials.

General characteristics and classification. Tree-dimensional heterogeneous system consisting of strongly differing in properties, mutually insoluble (in the limit − chemically non-interacting) components whose structure allows the use of the advantages of each is called *composite material (CM)* or *composite*. Unlike many heterogeneous alloys, an overwhelming number of CMs is not in thermodynamic equilibrium. The principle of CM constructing people borrowed from nature. Typical composite materials are tree trunks, stems of plants, human and animal bones.

CM allows having a predetermined combination of manifold (often mutually exclusive of each other in the equilibrium state) properties that can not be achieved in metallic alloys: the high specific strength and rigidity, heat resistance, wear resistance, thermal properties, etc. As a result, the use of CM allows creating previously unavailable, fundamentally new design. Due to CM a new qualitative leap has become possible, which allowed to increase the engine power while reducing their weight and size, to lower weight of various power constructions, to improve the weight efficiency of the vehicles and aero/spacecrafts, to increase energy efficiency of building envelope, etc.

Important characteristics of CM operating in the specified applications are specific strength σ_B/ρ and specific rigidity E/ρ, where σ_B is temporary tensile strength, E − modulus of elongation, ρ – density of the material. As can be seen from Fig. 3.4.1, composite materials surpass all known structural alloys by these parameters.

CMs used in power constructions typically consist of two components – a relatively ductile material (matrix) and more hard and durable dispersed components (fillers). CM properties depend on the properties of the matrix, fillers, and the bonding strength between them.

The matrix binds the composition in the monolithic material, giving it form, and serves to transmit external loads to the durable fillers. Depending on the matrix material, we can distinguish the following types of CM: with metal matrix or metal composite materials (MCM), with polymer matrix − polymer composite materials (PCM), and with ceramic matrix − ceramic composite materials (CCM).

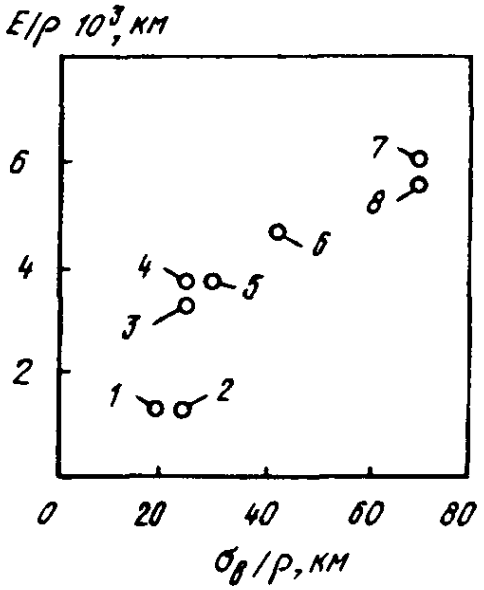


Fig. 3.4.1. The relationship of specific strength and specific modulus of some metallic materials and CM reinforced with fibers (up to 50 vol. %): *1* – aluminum; *2* − titanium and steel; *3* − titanium reinforced with beryllium wire; *4* − titanium reinforced with SiC fibers; *5* − titanium reinforced with SiC /B/W fibers; *6* − aluminum, reinforced with boron fibers; *7* − epoxy resin reinforced with graphite fibers; *8* − epoxy resin reinforced with boric fibers

A leading role in CM strengthening plays filler, which is often called hardeners. Last have high values of strength, hardness and elastic modulus. By type of reinforcing fillers all the CMs are divided into *dispersion-hardened*, *fibrous* and *layered* (see Fig. 3.4.2) composites.

a) *b*) *c*)

Fig. 3.4.2. Schematic representation of dispersion-hardened (*a*), fiber (*b*) and layered (*c*) CMs

Uniformly distributed small refractory particles of carbides, oxides, nitrides, and other solids that do not react with the matrix and does not dissolve in it up to the melting point are artificially introduced to matrix to make dispersion-reinforced CM. The finer the particles of the filler and less the distance between them, the stronger the CM becomes. Unlike fiber, in the dispersion-strengthened CM matrix is a main bearing element. Ensemble of dispersed filler particles reinforces the composite material (suppressing plastic deformation) due to supressing dislocations motion under the action of load. Effective resistance to the movement of dislocation is created up to the melting temperature of the matrix, allowing dispersion strengthened CM to characterize by high values of heat resistance and creep resistance.

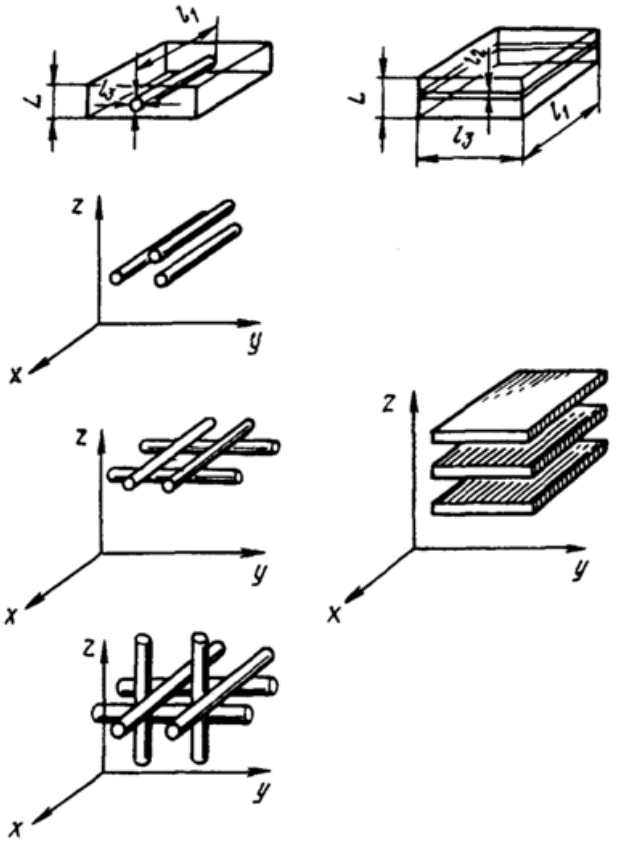
Fibers of different forms (armature) are used as fillers in fibrous CM: yarns, ribbons, different mesh netting. Reinforcement of fibrous CM can be obtained through uniaxial, biaxial and triaxial modes (see Fig. 3.4.3a). The strength and stiffness of these materials are determined by properties of the reinforcing fibers, which perceive the main load. Reinforcing with fibers gives a stronger CMs, but dispersed hardening is technologically easier and cheaper.

Layered composite materials (see Fig. 3.5.3b) are prepared from alternating layers of matrix and filler material (sandwich structure). Layers of such filler in the CM can have different orientations. It is also possible to use alternating layers of various fillers with different mechanical properties. For layered compositions typically non-metallic materials are used. Besides ensuring the strength and solidity of the constructions, matrix must have the needed flexibility and operational temperature range, for which the CM is designed.

Dispersion-hardened CMs. At the dispersion-hardening, solid filler particles block processes of dislocation motion in the matrix. Efficiency of hardening, for a minimum of interaction with the matrix, depends on the type of particles, their shape and volume concentration, as well as the uniformity of their distribution in matrix. Refractory particles of dispersed phases of type Al_2O_3, SiO_2, BN, SiC, etc. having a low density and a high modulus of elasticity are applied. Dispersion-reinforced CMs are usually obtained by powder metallurgy techniques. Their important advantage consists of isotropic properties in different directions.

In industry, dispersion-hardened CMs on the base of aluminum and nickel substrates are widespread. Typical representatives of this kind of materials are composite materials such as SAP (sintered aluminum powder), which consists of an aluminum matrix reinforced by dispersed particles of aluminum oxide. Aluminium powder is prepared by spraying the molten metal, followed by grinding in a ball mill in the presence of oxygen to a size less than 1 micron. With increase in duration of grinding, powder becomes smaller and the content of alumina increases there. Further production technology of issues and

semifinished products from SAP includes cold pressing, pre-sintering, hot pressing, rolling or extrusion of sintered aluminum billet through draw dies (forms) for finished products, which can then be subjected to further heat treatments.



a) *b*)

Fig. 3.4.3. Schematic representation of the filamentous (a) and layered (b) CM reinforcement

CM from SAP is satisfactorily deformed in the hot-pressed state with any content of hardener. CM with Al_2O_3 content not exceeding (6-9) vol.% can be deform at room temperature. Foils with thickness up to 30 μm can be obtained from such CM by cold drawing. These materials are well-processed by cutting and have high corrosion resistance.

SAP composites typically contain (6-23) vol.% of Al_2O_3. By increasing the volume content of alumina their strength increases. At room temperature strength characteristics of CM of SAP type reach the following values: σ_B = (280-420) MPa and $\sigma_{0.2}$ = (220-340) MPa.

CM of SAP type has a high heat resistance and surpass by this perfomance all the deformable aluminum alloys. Even at 500°C they have σ_B least (60-110) MPa. Heat resistance of SAP is explained by retarding of the recrystallization process by dispersed particles of aluminum oxide. The strength characteristics of SAP-type CMs are very stable. SAP tests are demonstrated that their properties have been almost unchanged for several years both at room temperature and under heating up to 500°C. At 400°C the strength of SAP is approximately 5 times higher than the strength of aged aluminum alloys. Effect of test temperature on the mechanical properties of SAP-type CMs is shown in Table 3.4.1.

SAP-type CMs are used in power engineering to produce parts with a high specific strength and corrosion resistance, which operate at temperatures up to 300-500°C. Piston rods, compressor blades, fuel cladding tubes and heat-exchangers are made of them.

Table 3.4.1. Strength characteristics of SAP type CM at different temperatures of tests

T_{test}, °C	20	100	200	300	400	500
$\sigma_{0,2}$, MPa	265	235	190	155	120	105
σ_B, MPa	380	315	235	175	130	105
δ, %	7.0	6.5	5.0	3.5	2.0	2.5

Fiber CMs. The most important criterion for selecting a matrix material for fiber in CM is operating temperature. For CM applied at temperatures below 200°C, the polymer matrix is used. Fiberglass, which is reinforced with short glass fibers in a matrix of polyester resin, belongs to such CM. Fibreglasses are used for production of car bodies, as well as certain types of devices. Thermosetting polymers are also used as matrixes, because they have cross-bonds between main chains forming a rigid three-dimensional network structure. Such polymers belongs to epoxy resins which have a higher heat resistance due to lateral bonds. At manufacturing of such CMs, fibers are subjected to surface treatment improving adhesion and are drawn through a bath of the polymeric resin. The resin binds the fibers in the ribbon, which is then collected in a layered (like plywood) sheet material or is wound in a more complicated shape. The material assembled or wound into sheets is baked by

heat treatment. Layers can be packaged with alternating direction of fibers (forming a cellular structure of CM), which provides the material stiffness.

The disadvantage of this simple layered CM is the possibility of their delamination due to the absence of transverse reinforcement in each layer and between the layers as a crack formed in the bulk of the CM is easily spreaded between layers. To eliminate these drawbacks CM with a woven reinforcement are made.

For operation at higher temperatures the metal matrixes with a low density (aluminum, magnesium, titanium) are used. Metallic CMs has several advantages over polymer. Apart from the higher operating temperature, they are characterized by a better isotropy, greater stability of the properties during performance and higher resistance to erosion.

Plasticity of metallic matrixes gives the needed toughness to structure facilitating rapid equalization of local mechanical loads. High thermal conductivity of metallic CMs protects them from local overheating, which is especially important for such products as missiles tips, leading edges of the wings and engines of aircrafts, etc. High electrical conductivity of metallic CMs allows their use for screening of electromagnetic radiation, lightning, reduces the risk of static electricity. An important advantage of the metallic CMs is their high manufacturability when forming, heat treatment, compounds forming.

For the enhancement oft operating temperatures, some types of ceramics are used as the matrix material. At the same time, their lack of plasticity is partially neutralized by the reinforcing fibers, inhibiting crack propagation in ceramics.

Using of amorphous carbon as matrix material and fibers of crystalline carbon (graphite) as a reinforcing material allows to create a composite enduring heat to 2500°C. Lack of carbon matrix is in possible oxidation and mass loss from the surface of the product by gas stream (ablation). To prevent oxidation and ablation such CMs are covered with a thin layer of silicon carbide.

In addition, to improve strength and rigidity, the basic requirement imposed on fibers for the CM is a good wetting of the fiber material by molten matrix substance in composite manufacturing process. The weak interaction of fiber with matrix material and its high oxidation resistance are also important conditions.

For reinforcement of metallic CMs continuous fibers are usually used: carbon (C), boron (B), alumina (Al_2O_3), silicon carbide (SiC), boron carbide (B_4C), boron nitride (BN), titanium diboride (TiB_2), silica (SiO_2). As fibers for the manufacture of CMs, fine metallic wires made of steel, tungsten, titanium, molybdenum and beryllium is often used. Less commonly are used filamentary crystals specially grown of various materials.

Greatest spreading for metallic CMs received the reinforcement by discrete or continuous fibers of carbon and boron, which properties are shown in Table 3.4.2.

Table 3.4.2. Properties of reinforcing fibers for CMs

Fiber type	Basis	$\rho, \frac{g}{cm^3}$	σ_B, GPa	E, GPa	$\alpha, 10^{-6} K^{-1}$
C	PAN Viscose Pitch	1.7 – 2.0 1.6 – 1.8 2.0	1.7 – 3.2 0.6 – 3.2 1.1 − 2,1	170 − 517 400 − 525 380 − 700	–1.0 ÷ –1.5 – –1.3
SiC	on W on C	3.15 3.05	3.1 – 3.4 3.45	420 − 450 400 − 420	3,8 ÷ 5,0
B	on W on C	249 2.25	3.52 3,32	400 380	4.8 ÷ 5.0
B-SiC	on W	2.50	2.9	400	4.9
Al_2O_3	–	3.90	3.9 – 4.3	380 − 500	8.5

Composites are manufactured by different methods. For example, boron-based fibers are generally prepared by precipitating it on the heated refractory metallic filaments (wires). The deposition is carried out from the gas phase upon dissociation of metal halides, such as boron trichloride BCl_3, on the tungsten filaments with 12 μm diameters, heated to 1100-1200°C. During the deposition (within 1-2 min) boron diffuses into the tungsten wire forming the fiber of tungsten borides. Overall diameter of boric fiber reaches 100-150 microns. The structure of the composite material presents strips of 0.5 mm thick comprising boron fibers in an aluminum matrix, see Fig. 3.4.5.

The raw material for high-strength and high-modulus carbon fibers serve such polymers with a high carbon content, as polyacrylonitrile (PAN) and least − pitch and viscose. PAN type fiber is extruded through draw dies of small diameter and then subjected to pyrolysis in an inert atmosphere at a temperature of about 2000°C. Carbon fibers are produced in the form of filaments comprising up to 10000 elementary fibers with a diameter of about 7 μm. Modulus and strength of these fibers remain unchanged when heated up to 600°C.

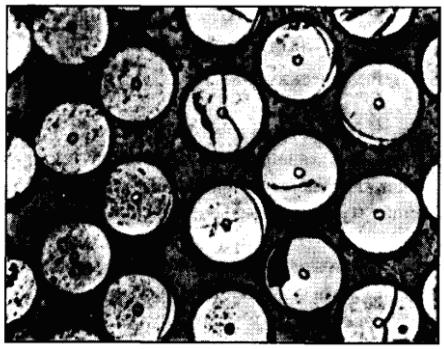


Fig. 3.4.4. Microphotograph of aluminum-boron type CM

Another method for preparing of fiber CMs is the impregnation of the fiber bundles by liquid melts of metals with low melting temperatures (e.g., aluminum, and magnesium), plasma spraying and hot pressing occasionally followed by hydro-extrusion or rolling of workpieces. When reinforcing, continuous fibers in sandwich-like compositions are used. They consist of alternating layers of aluminum foils and fibers, which are produced by rolling, hot pressing, explosion welding and diffusion welding.

Bars and pipes, reinforced by high-strength fibers, are prepared by casting of liquid metallic phase. When producing a polymer CM, fiber bundle is continuously pulled through a bath of liquid material (aluminum, magnesium or gum), where it is impregnated under pressure. When leaving the impregnating bath, fibers are connected and passed through the die forming rod or tube. This method provides a high content of fibers in the composite (up to 85 %), their homogeneous distribution in the cross section and continuity of the process.

For many kinds of fibers manufacturing processes of coating is used to provide better wettability, optimal interaction with the fiber matrix, as well as resistance to oxidation (to prevent oxides formation). For example, boron fibers are protected from chemical interaction with the titanium and aluminum melts by creating a diffusion barrier of silicon carbide or boron at the surface. Because of high oxidizing capability of the carbon fibers, special coatings on their surface are also applied and refining processes carried out in a protective atmosphere.

Carbon fibers have a negative thermal coefficient of linear expansion, whereby it becomes possible for the fibers to obtain a corresponding laying close to zero expansion coefficient of this CM.

Silicon carbide fibers produced by chemical vapor deposition onto a substrate of carbon or boron fibers. These fibers have good heat resistance, oxidation resistance and little react with the metal.

Layered CM. Plastics, metals or ceramics are matrix materials for composites with a laminated structure. Polymeric fibers, ribbons of natural and glass fabrics and other materials are used as fillers. For example, in construction, engineering, etc. laminates made from resins that are reinforced with glass fibers or polymers are widely used.

Layered CM includes ablative materials produced on the basis of phenol-formaldehyde resins with carbon or glass fibers. Glass fiber, which provides thermal protection and high mechanical properties of the articles (e.g. thin-walled tubes, plugs, etc.) at a multilayer applying, is often used in these materials.

There are composites, in which layered bonding are aluminum, titanium, copper, nickel and cobalt sheets and foils, and layers that define special properties and applications − ceramics, intermetallic compounds or other metal materials.

Layered CM based on ceramic matrices is used in systems operating under extreme conditions (e.g., thermal insulation covering in spacecraft). Components of this type CM are often ceramics, carbon, metals such as aluminum oxide, pyrolytic carbon, carbides, oxides, nitrides of the composition with aluminum, copper, titanium, nickel, cobalt, tantalum and iron.

Properties and application of CM. Physico-mechanical properties of the main components of CM are shown in Tables 3.4.2 and 3.4.3.

Table 3.4.3. Properties of the CM components

Fiber type	$\rho, \frac{\text{g}}{\text{cm}^3}$	σ_t, MPa	E, GPa	$\frac{\sigma_t}{\rho}$, km	$\frac{E}{\rho}, 10^3$ km	$\alpha, 10^{-6}\,\text{K}^{-1}$	T_{max}, °C
Matrix of Al – alloy	2-63-2,8	250-573	69-73	20	2.5	11-13	150
C	–	850/70	360/35	90	20	1.0-3.6	–
B	2.1-2.3	1800/330	250/140	70	10	6	500-540
SiC	2.65-2.85	1600/350	230/140	56	7	6,1	300
B-SiC	2.9-2.7	1400/320	220/180	50	–	–	–
Al_2O_3	2.8-3.4	1200	260/140	34	7	–	–

Good compatibility of the matrix with the reinforcing element, the high strength properties of the boron fiber and a satisfactory ductility of the matrix

material define the specific values of higher strength and stiffness of MCM (ratio of tensile strength and elastic modulus to density) in combination with good processability and structural reliability of products made of this material.

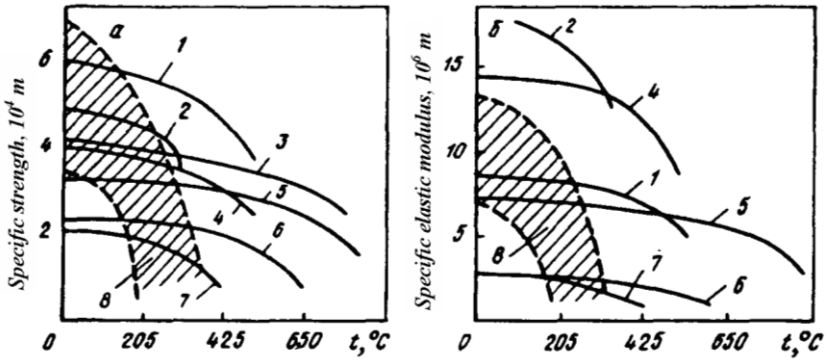
To create a metal CM having lower density magnesium is applied. Composite materials based on magnesium are 30% lighter than reinforced aluminum alloys. Metal magnesium based CM have good specific properties, stable temperature coefficient of linear expansion over a wide temperature range, that is achieved by combinations of fiber and matrix properties and can be adjusted depending on the particular conditions of use. Such materials can be prepared in the form of castings, including flat plates, tubes, rods and specially shaped products. Systems carbon fiber-aluminum and carbon fiber-magnesium are promising for use in aerospace devices due to the high values of the specific strength and stiffness, small temperature coefficient of linear expansion and relatively high thermal conductivity (for heat dissipation).

Metals with high plasticity and strength are good combined with a high-strength and rigid fibers of low density and ductility, forming a CM with higher stiffness and low weight. The titanium, reinforced with silicon carbide or boron fibers may provide as an example of such a combination. However, such systems have reduced fatigue strength due to residual stresses, the high reactivity of titanium matrix, leading to chemical interaction between the fibers and the matrix at elevated temperatures in the manufacturing process. Properties of metal CM based on titanium matrix are presented in Table. 3.4.4.

Table 3.4.4. Properties of metal CM based on titanium alloys

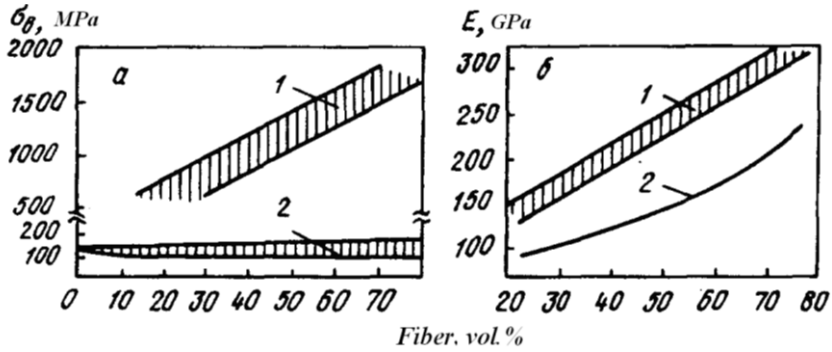
Properties	Titanium alloys	Reinforcing fibers		
		B	SiC	B-SiC
$\rho, \ \frac{g}{cm^3}$	4,5	3,3 − 3,5	3,8 − 4	3,7 − 3,9
σ_t, MPa	500 − 1200/113	1500/550230	1720/650	1400/550
E, GPa	27	43	250/200	290/200
$\frac{\sigma_t}{\rho}, km$	2.6	6.5	46	37
$\frac{E}{\rho}, 10^3 km$	9	–	7.5	7.5
$\sigma_{0,10}, MPa$	–	650	–	–
$T_{\max}, °C$	490	–	700	4.5-5.7

Influence of test temperature on specific strength and stiffness of composite materials is shown in Fig. 3.4.6. Fig. 3.4.7 lists typical properties of the composite material based on aluminum, boron fiber reinforced (coated with silicon carbide).



a *b*

Fig. 3.4.6. The dependence of the specific resistance (*a*) and the specific rigidity (*b*) of materials on temperature: *1* – B/Al; *2* – YB/Mg; *3* – B/Ti; *4* – YB/Al; *5* – SiC/Ti; *6* – Ti; *7* – Al; *8* – the shaded area corresponds to the data of different authors for carbon fiber/polymer resin type CM



a *b*

Fig. 3.4.7. The dependence of tensile strength (*a*) and modulus of elasticity (*b*) of the boron-aluminum composite material along (*1*) and transverse (*2*) from the axis of the reinforcement fiber on volume content of boron

The strength and modulus of elasticity and also impact resistance of materials for unidirectional composite materials based on aluminum,

magnesium and titanium increases as in the composition of fiber volume content. For very high temperatures, such as in the combustion chambers of jet engines, a system containing molybdenum and tungsten wire in a matrix of titanium and superalloys is used. The greatest strength at 1093°C has the W-Re-Hf-C alloy wire that is 6 times higher than the strength of nickel or cobalt superalloys at the same temperature.

Great prospects are opening with the development of three-dimensional reinforcement processes of metal CM. Particularly for the metal CM three-dimensional reinforcement provides a significant gain in toughness. Fiber CM of Al_2O_3/Al type with three-dimensional reinforcement absorbs almost the same energy at impact as pure aluminum. Furthermore, the thickness reinforcement provided by three-dimensional fiber structure prevents delamination and inhibits crack propagation.

The disadvantages of metal CM are their relatively high cost (it several times higher than the cost of polymer composites) and the complexity of the fabrication technology. Therefore, it is necessary to reduce the cost of metal CM through improved technology that will use their unique properties in a number of designs, such as in aerospace engineering. Details of boronplastic and boronaluminum are used by many companies in the manufacture of horizontal and vertical stabilizers, rudders, elements empennage, rotor blades, etc.

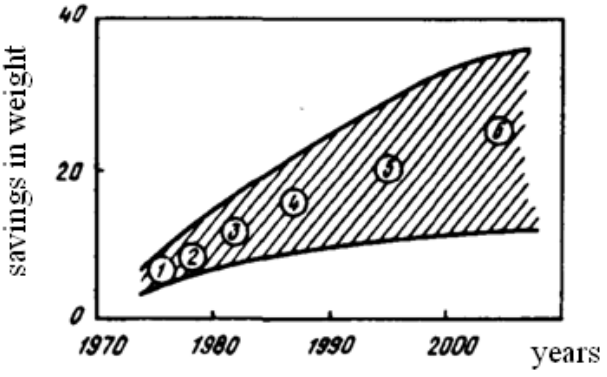


Fig. 3.4.8. The potential savings in weight of aerospace structures using different construction materials: *2* − aluminum alloys; *3* − 10 vol. % Ti, 2 vol. % Al, V-Fe; *4* – Al-Li; *5* – powder materials; *6* – metal CM

Areas of using of metal CM continuously expanding. In addition to improving the performance of many critical parts the use of metal CM can lead to a weight savings up to 20 − 30 % (see, Fig. 3.4.8). For example modern Al-Li alloys have this rate at 10 − 15% level.

As can be seen from Fig. 3.4.8, details of the structures of boronaluminum CM give weight reduction by 30-40%, providing more prolonged and higher fatigue strength when heated to 500°C as compared with titanium alloys. Application of boronaluminum in rocket and space technology allows reducing aircraft weight by 20-50%, which significantly increases the payload. The use of metal CM from alumina reinforced with a mixture of short $A1_20_3/SiO_2$ fibers with diameter of about 3 microns and length of up to 10 microns in different ratios allowed to increase strength and modulus of elasticity by increasing the mass fraction of fibers and to improve the wear resistance of $A1_20_3$ during the growth fraction of SiO_2 fibers.

This type of CM can be used instead of expensive nickel alloys for the manufacture of linings pistons, allowing raising the temperature in the combustion chamber of the engine and its output. By increasing the durability of the pistons, maintenance-free vehicle mileage increases to 300 thousand kilometers. The correct choice of materials to save weight of aircraft of space technology plays a crucial role as weight saving reduces the cost of the spacecraft launch into orbit for the amount that exceeds expenses for the cost of the structural elements of these materials.

4. Modern methods of evaluation of parameters for energy effective materials

There are several physical parameters of energy effective materials that should be controlled and investigated. The most important group of parameters are: recombination, optical and thermal ones.

4.1. The recombination parameters mean: life time of carriers, diffusion coefficient of carriers, velocities of surface recombination, diffusion length of carriers.

4.2. The optical parameters mean: energy gap of semiconductors, optical absorption coefficient spectra both in the band to band region as also in the free electrons absorption region, contribution of a surface to bulk absorption.

4.3. The thermal parameters: thermal conductivity, specific heat, thermal diffusivity and thermal effusivity.

There are several methods of evaluation of these parameters: photoacoustic methods with the microphone and piezoelectric detection, PTR and PCR methods, microwave conductivity decay μ- PCD method, photoluminescence

imaging method PL, carrier density imaging CD, kinetics of a photocurrent and others.

4.1. Recombination parameters – measurement

There are a few noncontact methods of determination of recombination parameters: modulated free carriers absorption method MFCA, photothermal infrared radiometry method PTR, photocarrier radiometry PCR method, photoacoustic method PA. There are also contact methods of evaluation of the recombination parameters.

4.1.1. Modulated Free Carriers Absorption method MFCA

The idea of this method is the following. The beam of light of the energy of photons bigger, than the energy gap of the investigated semiconductor, illuminates the sample. This beam of light is modulated in intensity. As a result of the absorption of this light electrons are generated in the conduction band and their concentration changes periodically. The same spot of a sample is illuminated with the beam of the infrared IR radiation whose intensity is constant in time. As a result of absorption of this IR radiation, by free electrons in the conduction band, the intensity of the IR radiation going through the sample, changes periodically. The information about, among others, the lifetime of carriers is extracted from the amplitude and phase frequency characteristics of this MFCA signal that is detected by a photodiode. The schematic diagram of the experimental set-up is presented in Fig. 4.1.1.

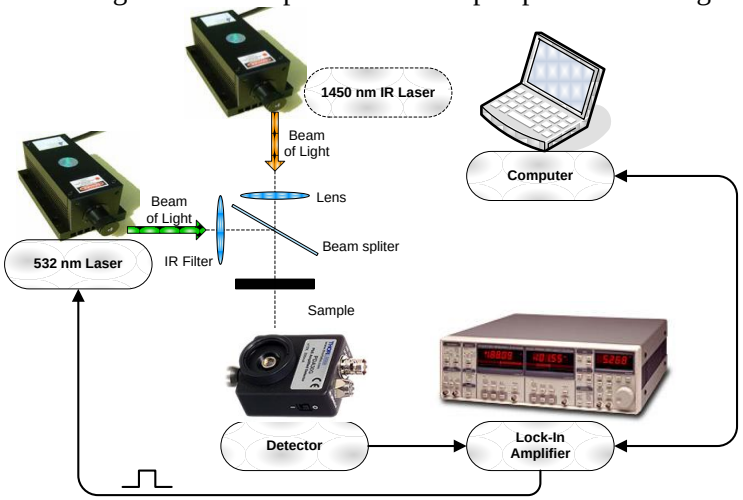


Fig. 4.4.1. Schematic diagram of the experimental set-up for the measurements of the recombination parameters with the MFCA method

The presented above set-up consists of two semiconductor lasers. One of them emitting the light at 532 nm modulated in intensity is the pumping laser generating carriers in the conduction band of a semiconductor. The other laser, emitting a constant in intensity beam of light at 1450 nm (IR), illuminates the sample in the same place as the first one. Intensity of the IR light going through the sample is detected with a fast detector. This signal was next measured with a dual channel phase sensitive amplifier.

Spatial distribution of carriers generated with the pumping beam of light is presented below.

$$\delta n(x)=K(f,\tau)\left[B_1(f,\tau,D,V_1,V_2)\mathrm{e}^{\frac{x}{Le(f,\tau)}}+B_2(f,\tau,D,V_1,V_2)\mathrm{e}^{\frac{-x}{Le(f,\tau)}}-\mathrm{e}^{-\beta x}\right] \quad (4.1.1)$$

where: τ – lifetime of carriers, D – diffusion coefficient of carriers, β – optical absorption coefficient of the pumping light, f – frequency of modulation of the intensity of the pumping beam of light, V_1 and V_2 –velocities of the surface recombination, x – spatial coordinate.

The diffusion length of carriers is described by the formula below.

$$Le(f,\tau)=\left[\frac{D\tau}{1+\mathrm{i}\omega(f)\tau}\right]^{\frac{1}{2}} \quad (4.1.2)$$

The MFCA signal is proportional to the integral of the spatial distribution of carriers along the thickness of the sample.

$$MFCA(f,\tau,V_1,V_2)=a\cdot\int_0^d\delta n(x)\cdot\partial x \quad (4.1.3)$$

Theoretical basics of the MFCA method, presented above, were applied successfully for the interpretation of the frequency amplitude and phase characteristics obtained for the series of silicon wafers. They were next modified by the spatial separation of the pumping and IR beams of light and extended by investigations of the influence of the width of the IR beam of light on the MFCA signal characteristics. In practice two MFCA methods can be, and are, applied for the determination of the recombination parameters of semiconductors: the first one is based on the measurement of the frequency characteristics of the amplitude and phase of the MFCA signal and the second one is based on the measurement of the MFCA signal as a function of the distance of the pumping and IR beam of light at the constant frequency of modulation. In the most advanced approach, when two beams of light are spatially separated, the computations of the spatial distributions of carriers in a 3D model are necessary for numerical interpretation of the MFCA frequency characteristics. Results of investigations comparing the MFCA method with

other methods using microwaves (μ-PCD) and photoconductance decay measurements proved correctness and usefulness of the MFCA method for determination of the lifetime of carriers and the velocity of the surface recombination of carriers.

Example theoretical and experimental MFCA characteristics obtained for the silicon wafer are presented in Fig. 4.1.2.

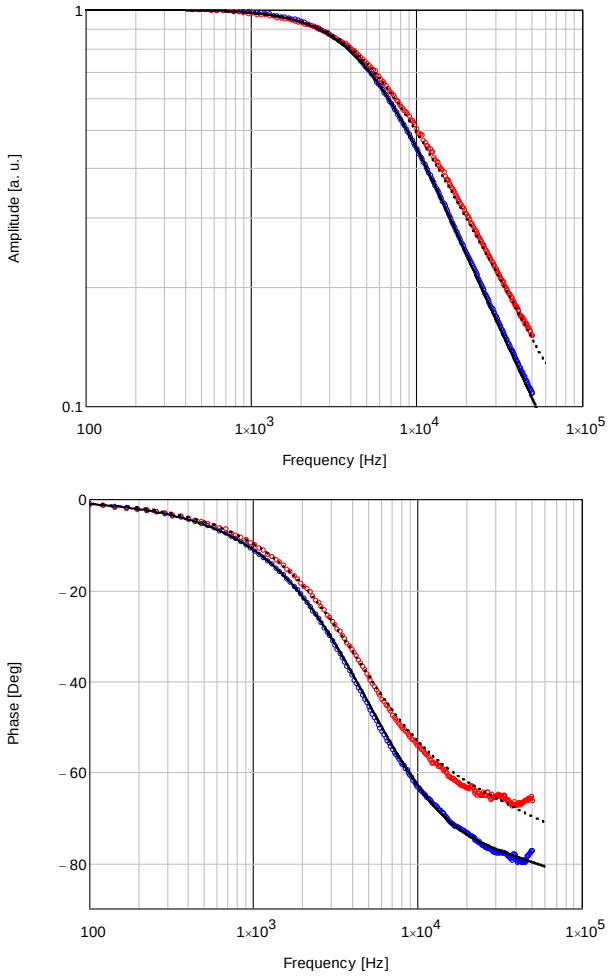


Fig. 4.1.2. Experimental (circles) and theoretical (lines) frequency amplitude (*a*) and frequency phase characteristics of the MFCA signal of a silicon sample for the following fitting parameters: d=0.058 cm, D=15 cm/s, V_g=500 cm/s, V_b=1564 cm/s, τ=71 μs, D=15 cm^2/s. Dashed line – the matt side of a sample is illuminated, solid line – a mirror side of a sample is illuminated

The MFCA method can be also used for imaging of the lifetime of carriers on the semiconductor wafer. The amplitude of the MFCA signal for a given frequency of modulation depends on the lifetime of carriers. In general when the lifetime of carriers increases the amplitude of the MFCA signal also increases. Theoretical characteristics illustrating this idea is presented in Fig. 4.1.3.

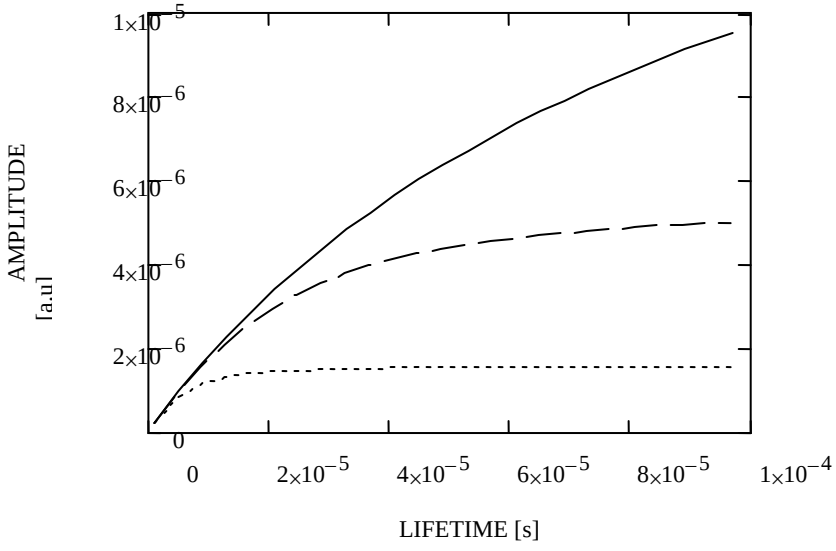


Fig. 4.1.3. Dependence of the amplitude of the MFCA signal on the lifetime of carriers for given frequencies of modulation of the pumping beam of light: 1000 Hz – a solid line, 5000 Hz – a dashed line, 20 000 Hz – a dotted line

4.1.2. A photoacoustic method

Another method of determination of the recombination parameters of semiconductors is based on the photoacoustic (PA) effect. The sample in the open photoacoustic cell is illuminated with the intensity modulated beam of light. As a result of such illumination three effects are observed in the sample. The first effect is the result of a fast intraband relaxation of electrons to the bottom of a conduction band. As a result of this process a temperature spatial distribution arises described as $\Theta_T(x)$. In the second step electrons at the bottom of the conduction band diffuse in the crystal and relax with the lifetime τ creating another temperature spatial distribution described as $\Theta_{NRR}(x)$. The third effect is connected with the diffusion of electrons to the illuminated and dark

sides of the sample and their surface nonradiative relaxation described with the surface velocity of relaxation $\Theta_{SR}(x)$. In general the total temperature spatial distribution arises in the sample. $\Theta_{TOTAL}(x)= \Theta_T(x)+ \Theta_{NRR}(x)+ \Theta_{SR}(x)$, where x is the spatial coordinate. If d is the thickness of the sample the temperature of the dark side of the sample is given by the formula: $\Theta_{TOTAL}(d)= \Theta_T(d)+ \Theta_{NRR}(d)+ \Theta_{SR}(d)$.

The periodical temperature of the dark side of the sample heats up the gas in the photoacoustic cell. As a result a periodical overpressure in the cell is observed and measured by the microphone. In the experiment a frequency characteristics of the amplitude and phase of the photoacoustic signal are measured and the recombination parameters are extracted from the fitting of theoretical curves to experimental frequency characteristics. The schematic diagram of the experimental set-up used for the frequency domain amplitude and phase photoacoustic characteristics is presented in Fig. 4.1.4.

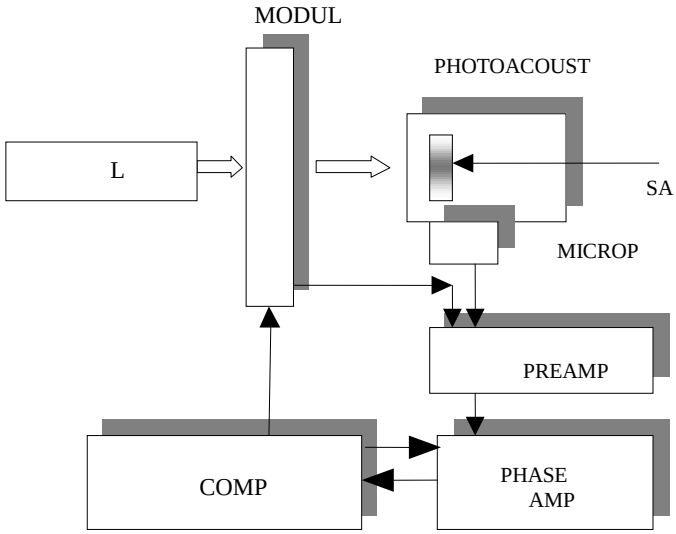


Fig. 4.1.4. A schematic diagram of the experimental set-up for the frequency domain photoacoustic characteristics

The laser beam is modulated in intensity by a modulator and next illuminates the sample in the photoacoustic cell. The microphone detects periodical changes of the pressure in the cell. Next the signal is amplified by the preamplifier and measured by the lock in amplifier. All measurements are computer controlled.

The numerical interpretation of the experimental photoacoustic characteristics needs the theoretical model of the photoacoustic effect describing the influence of plasma waves on the temperature spatial distribution in the sample and the photoacoustic (PA) signal. The experimental frequency characteristics of the PA signal, measured in the transmission configuration, for example for Ge samples of different thickness, have shown that it is possible to determine the set of the following material parameters from the fitting of theoretical frequency amplitude and phase PA curves to experimental characteristics: thermal diffusivity, velocity of surface recombination and the lifetime of carriers.

This method was applied sucessfully for investigation of plasma waves and determination of the recombination parameters for a series of semiconductors: Si,Ge, GaAs, InSb and $PtSb_2$. The thermal diffusivity, lifetime, diffusion coefficient of carriers, surface recombination velocity were also determined with the photoacoustic method for $CdInGaS_4$ crystals. In this case frequency amplitude PA characteristics of $CdInGaS_4$ samples of a different thickness from 80 μm to 250 μm were measured in a transmission configuration.From the fitting of theoretical curves, in a model of plasma waves, to experimental characteristics several recombination and thermal parameters of the crystal were determined.

Temperature of the dark side of the sample, in a transmission configuration, ($x=L$) is described by formulae which are a special case of the spatial temperature distribution $T(x)$ for $x=L$ and for a big value of the optical absorption coefficient of the sample for the illuminating light.

$$T(L) = T_1(L) + T_2(L) + T_3(L)$$

$$T_1(L) = B \cdot \frac{h \cdot \nu - E_g}{2 \cdot \sigma} \cdot \frac{1}{\sinh(\sigma \cdot L)}$$

$$T_2(L) = B \cdot \frac{E_g \cdot F}{\tau \cdot \sigma^2} \cdot \frac{1}{m^2 - 1} \times$$

$$\times \left(\frac{m \cdot \frac{D}{L_D} \cdot \sinh\left(\frac{L}{L_D}\right) - m \cdot V \cdot \cosh\left(\frac{L}{L_D}\right)}{\sinh(\sigma \cdot L)} + m \cdot V \cdot \coth(\sigma \cdot L) - \frac{D}{L_D} \right)$$

$$T_3(L) = B \cdot \frac{E_g \cdot F}{\sigma} \times$$

$$\times \left(\frac{V}{\sinh(\sigma \cdot L)} \cdot \left(\frac{D}{L_D} \cdot \cosh\left(\frac{L}{L_D}\right) - V \cdot \sinh\left(\frac{L}{L_D}\right) \right) + \frac{D}{L_D} \cdot \frac{V}{\tanh(\sigma \cdot L)} \right)$$

$$B = \frac{I_0}{k \cdot h \cdot \nu}, L_D = \left(\frac{D \cdot \tau}{1 + i \cdot 2 \cdot \pi \cdot f \cdot \tau} \right)^{1/2}, \sigma = \left(\frac{i \cdot 2 \cdot \pi \cdot f}{\sigma} \right)^{1/2}$$

$$F = \frac{1}{\left(\frac{D}{L_D} + V \right)^2 \cdot e^{\frac{L}{L_D}} - \left(\frac{D}{L_D} - V \right)^2 \cdot e^{-\frac{L}{L_D}}}$$

$$m = \frac{1}{L_D \cdot \sigma}$$

where: E – energy of the absorbed photon, E_g – energy gap of a semiconductor.

Example amplitude PA characteristics, in a transmission configuration, computed in a plasma wave model described above, are presented in Fig. 4.1.5. Computations were performed for the following set of parameters: α = 0.03 cm^2/s, D = 10 cm^2/s, V = 200 cm/s, τ = 100 μs solid line, 50 μs dotted line, 1 μs dashed line.

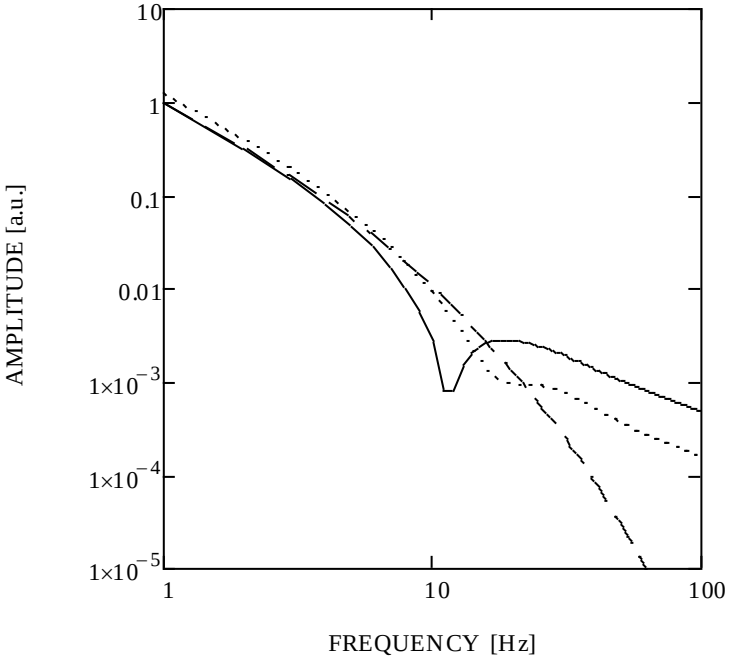
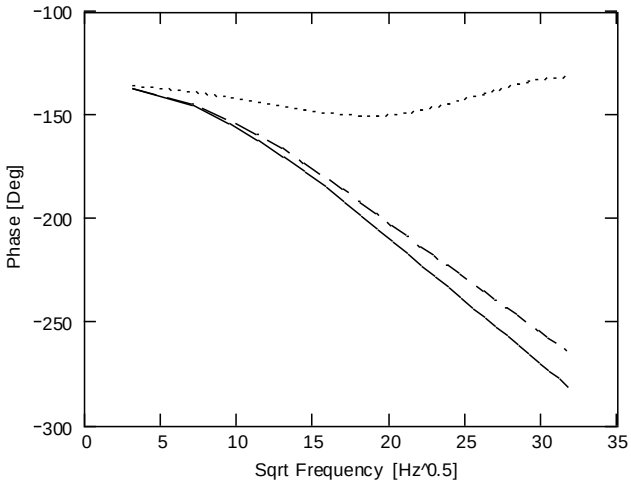
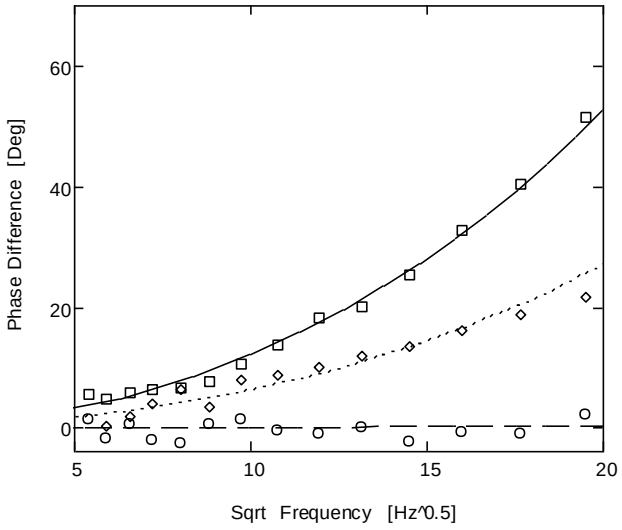


Fig. 4.1.5. Frequency amplitude PA characteristics computed for three lifetimes of carriers: solid line τ=100 μs, dotted line τ=50 μs, dashed line τ=1 μs



a)



b)

Fig. 4.1.6. Phase characteristics *a*) and phase difference characteristics *b*) of the PA signals versus the frequency of modulation. Lines are theoretical curves, symbols are experimental points. Boxes – τ=0.140·10^{-3}s, d=0.058 cm, L=0.046 cm, triangles – τ=0.050·10^{-3}s, d=0.049 cm, L=0.027 cm, Circles – τ=0.001·10^{-3}s, d=0.075 cm, L=0.0038 cm, V_g=1500 cm/s, V_b=100 cm/s, D=15 cm^2/s, α=0.95 cm^2/s

The above method of determination of the lifetime and the diffusion coefficient of carriers was applied for silicon samples and for thin layer samples of CdTe and RuS_2. This method was also applied for investigations of the recombination parameters of several narrow band gap samples such as: PbTe, PbTe:B, PbSe:B, SnTe:B.

A different photoacoustic method of determination of the lifetime of carriers is based on two measurements of the frequency phase characteristics of a PA signal in a transmission configuration. If a measured sample has two different surfaces i.e. matt and mirror then the difference of these two phase characteristics carries information about the lifetime of carriers. The experiments were performed on several Si crystals. The PA results were confirmed by the measurements of the kinetics of the photocurrent. PA phase characteristics exhibited the lifetimes: 150 μs, 50 μs, 1 μs of silicon samples. The example PA phase and phase difference characteristics obtained for Si samples exhibiting different lifetimes of carriers are presented in Fig. 4.1.6.

The frequency domain photocurrent phase, theoretical and experimental, characteristics performed for the same set of samples are presented in Fig.4.1.7.

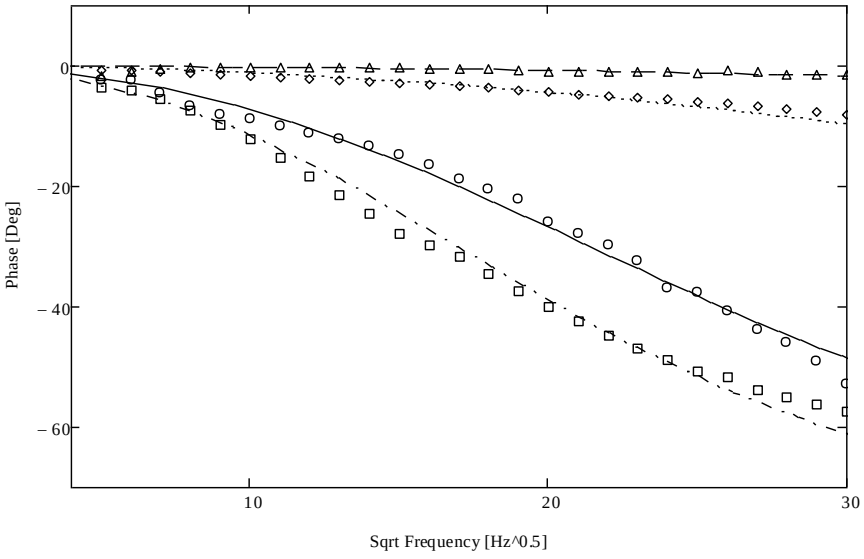


Fig. 4.1.7. Phase frequency dependences of the photocurrent versus the frequency of modulation : triangles – $\tau < 5$ μs, diamonds τ=35 μs, circles – τ=180 μs, boxes – τ=320 μs. Solid, dashed, dashed-dot and dotted lines are theoretical curves of the photocurrent versus the frequency of modulation

4.1.3. A Photothermal Infrared Radiometry method PTR

The recombination parameters can be also determined from the PTR frequency characteristics. The idea of a photothermal infrared radiometry method is based on the illumination of the sample with an intensity modulated beam of light from the laser of the energy of photons bigger than the energy gap of the semiconductor. The intensity of light is modulated by the acoustooptic modulator in the frequency range from 1 Hz to 100 kHz. The beam of light is absorbed in the sample and generates both thermal and plasma waves in it. Periodical changes of the temperature of the sample cause emission of a periodical infrared IR radiation according to the Planck's distribution. This IR radiation is focused on the photodiode detector sensitive for the 12μm IR radiation. The signal from the photodiode is measured by the lock-in amplifier, as a function of the frequency of modulation, and collected in the computer. The PTR signal can be detected from the illuminated side of the sample, called a reflection configuration, or from the dark side of the sample, called a transmission configuration.

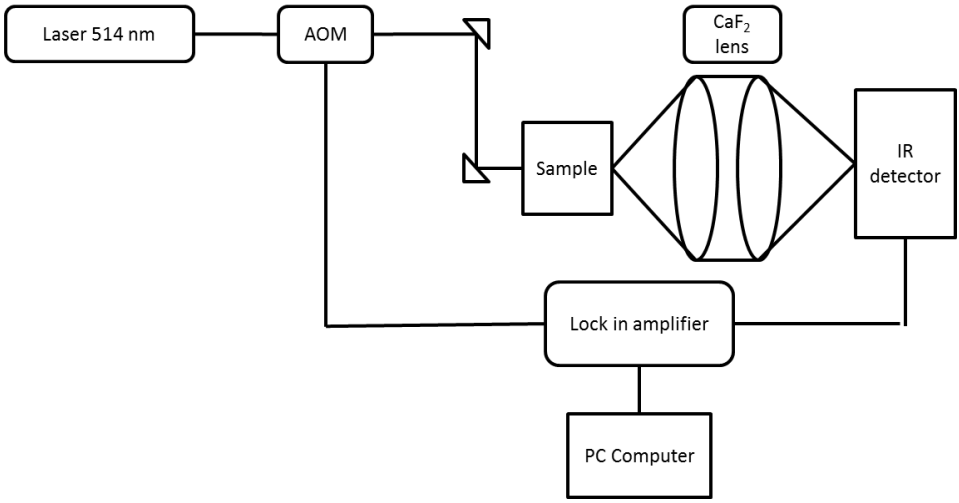


Fig. 4.1.8. A schematic diagram of the experimental set-up for the PTR measurements

From the amplitude and phase PTR characteristics, measured in the range of small frequencies of modulation, the optical and thermal parameters of a sample can be determined. From the amplitude and phase PTR characteristics, measured in the range of high frequencies, the recombination parameters of a sample can be determined. In general the photothermal infrared radiometry

signal is a sum of two components: thermal and plasma ones. It can be written as:

$$PTR(f) = A \cdot \int_0^l T(x, f, l, \lambda, D_T, D, \tau, V_1, V_2) \cdot \partial x +$$

$$+ \int_0^l \Delta n(x, f, l, \tau, D_T, D, \tau, V_1, V_2) \cdot \partial x$$

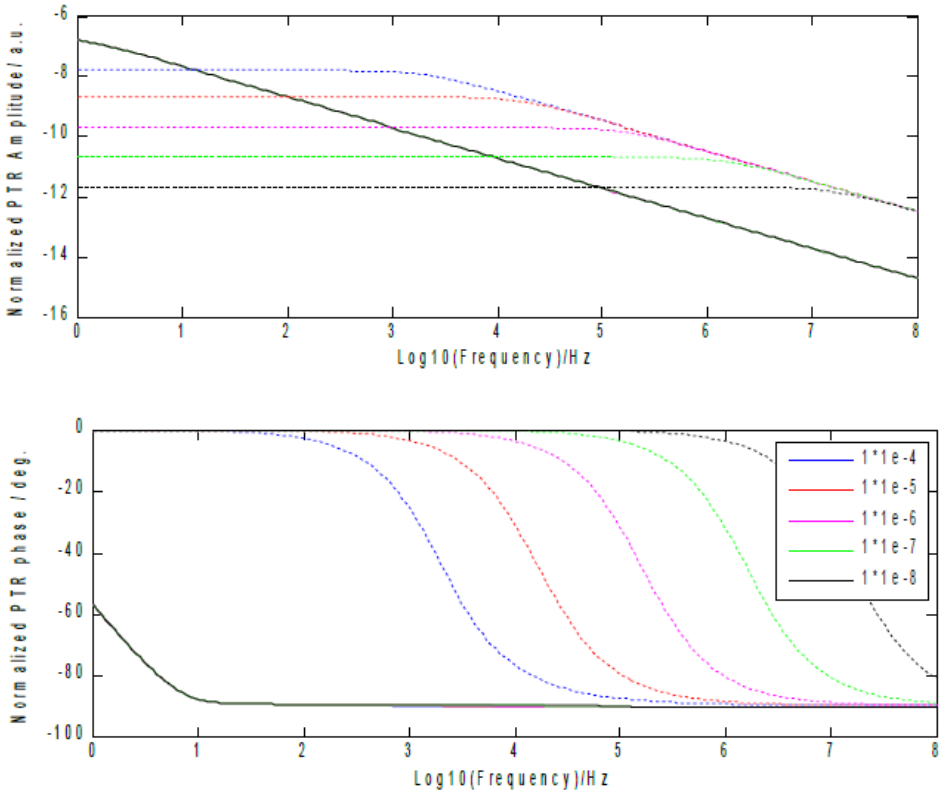
Where: l is a thickness of the sample, $T(x, f)$ is the periodical spatial temperature distribution, $\Delta n(x, f)$ is a periodical spatial concentration of electrons distribution, x is the spatial coordinate, f is the frequency of modulation, λ is a thermal conductivity, D_T is a thermal diffusivity, D is a carrier diffusivity, τ is the lifetime of carriers, V_1, V_2 are velocities of the surface recombination of carriers.

The first component, called a thermal component, describes that part of a PTR signal that is the results of a total thermal waves generated in the sample causing the emission of the periodical IR radiation. $T(x) = T_T(x) + T_{NRR}(x) + T_{SR}(x)$. The first component $T_T(x)$ is the result of a fast intraband relaxation of electrons to the bottom of a conduction band. The second component describes electrons at the bottom of the conduction band that diffuse into the volume of the crystal and relax with the lifetime τ creating another temperature spatial distribution described as $T_{NRR}(x)$. The third component $T_{SR}(x)$ is connected with the diffusion of electrons to the illuminated and dark sides of the sample and their surface nonradiative relaxation described with the surface velocity of relaxation.

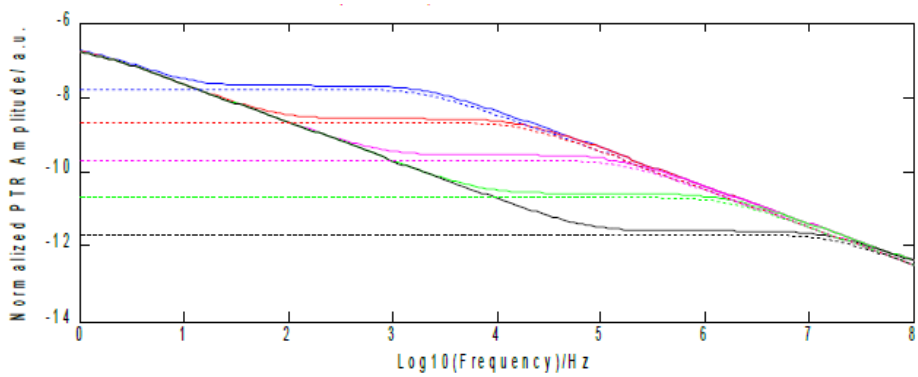
The second component, called a plasma component, describes that part of a PTR signal that is the result of the modulation of the constant IR emission, connected with a constant temperature of the sample T_0, by absorption of the IR radiation by a periodically changing concentration of carriers. That second component is analogical to the MFCA signal being proportional to the integral of the carriers concentration over the thickness of the sample. The figures below illustrate different cases of the PTR signal.

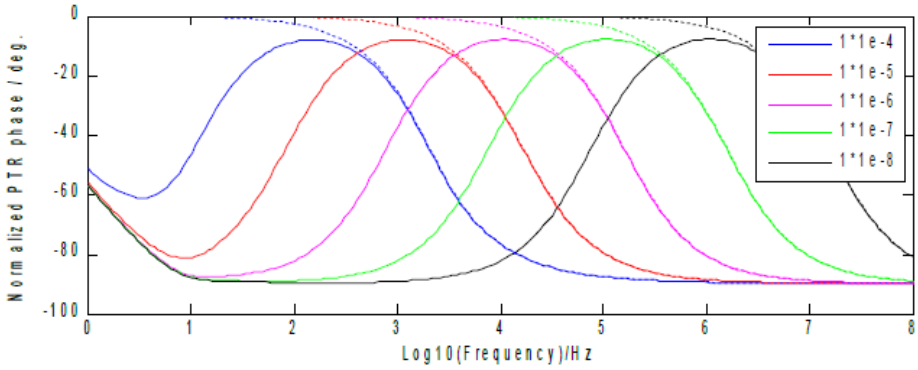
The example experimental and theoretical PTR amplitude and phase frequency domain characteristics measured in the front illumination obtained for nonimplanted and implanted silicon samples are presented in the figure below. From the fitting procedure of theoretical curves to experimental amplitude and phase characteristics several recombination parameters of the sample can be determined. The most important parameters are the lifetime of carriers and the velocity of the surface recombination.

Case 1 – solid thick line – only the thermal component, dashed lines – only the plasma component for different lifetimes of carriers.



Case 2 – resulting PTR signal, small contribution of the thermal component.





Case 3 – resulting PTR signal, big contribution of a thermal component.

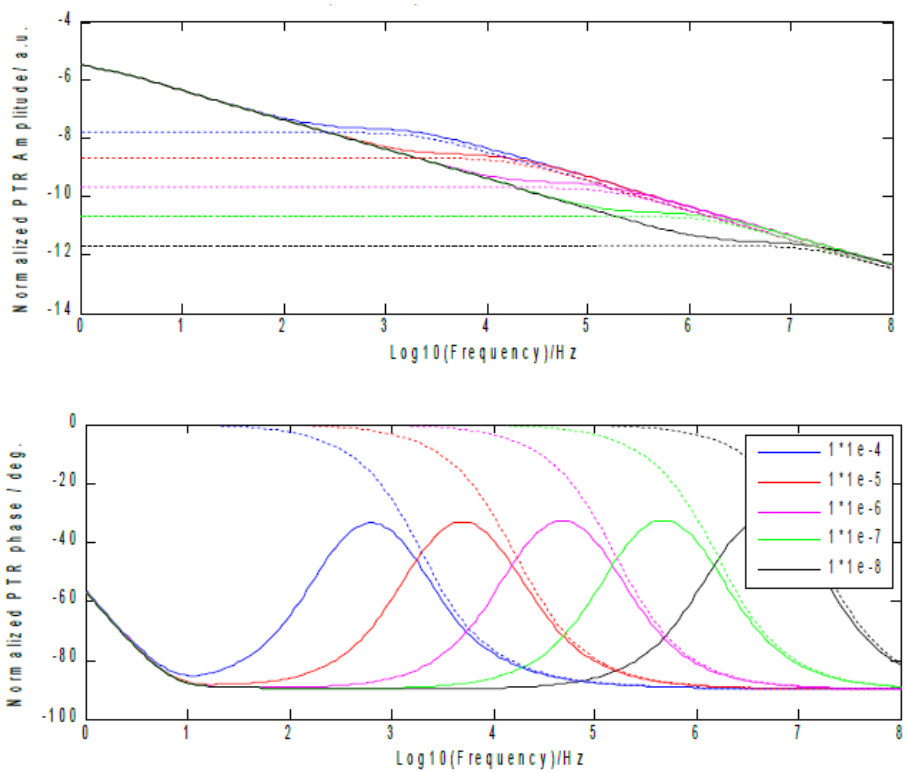
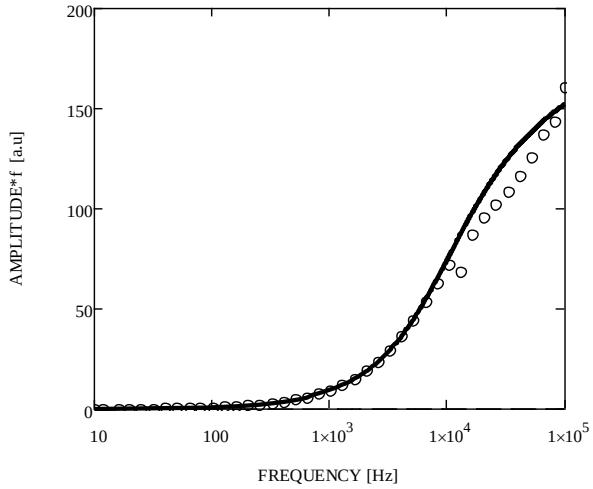
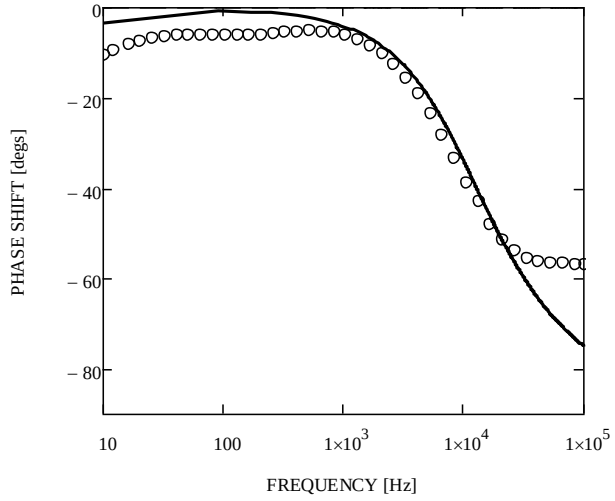


Fig. 4.1.9. Theoretical predictions of the frequency domain PTR amplitude and phase characteristics



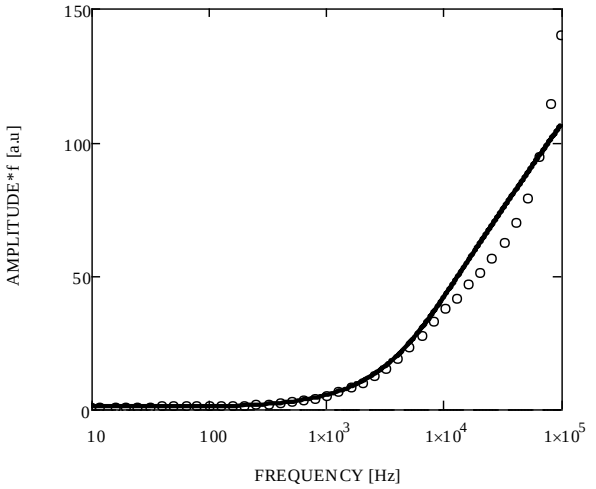
a)



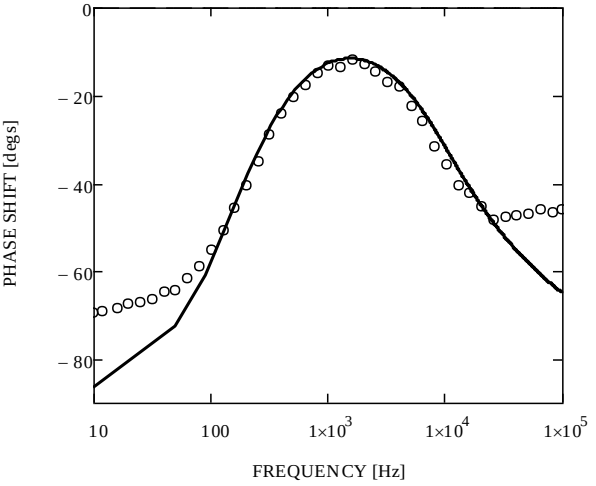
b)

Fig. 4.1.10. PTR normalized amplitude a) and phase b) frequency characteristics of nonimplanted *n*-type silicon sample. Fitting parameters: D=12 cm^2/s V_1=800 cm/s, V_2=1000 cm/s, A=10^{-6}, τ=18 μs. l=410 um, ρ=3-5 Ωcm. Circles are experimental results, solid lines are theoretical curves

Example characteristics of implanted silicon PTR characteristics are presented in the figure below.



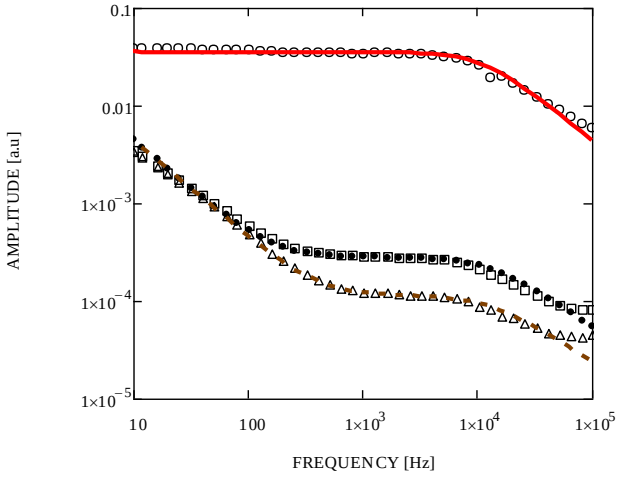
a)



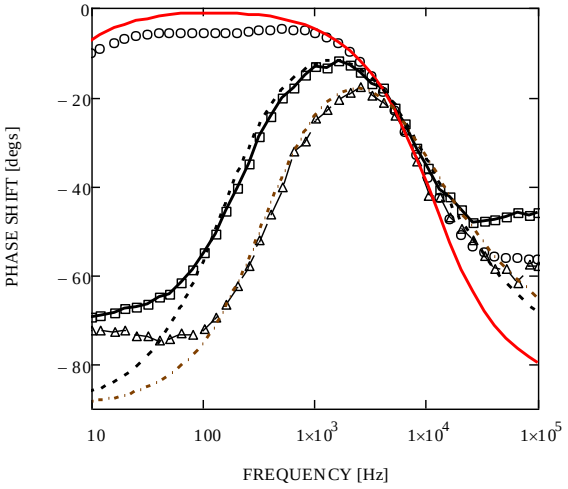
b)

Fig. 4.1.11. PTR normalized amplitude a) and phase b) frequency characteristics of O^{+6} implanted n-type silicon sample. Fitting parameters: D=12 cm^2/s, V_1=2500cm/s, V_2=1000 cm/s, A=150·10^{-6}, τ=18 μs, dose 5·10^{13} ions/cm^2, 90 keV. Circles are experimental results, solid lines are theoretical curves

These characteristics show what changes are introduced to PTR characteristics as a result of ion implantation. Collected PTR amplitude and phase frequency characteristics are shown in Fig. 4.1.12.



a)



b)

Fig. 4.1.12. Amplitude and phase frequency characteristics of nonimplanted and O^{+6} implanted silicon of n-type. Fitting parameters: $A1=3\cdot10^{-6}$ Vg1=400 cm/s, $A2=180\cdot10^{-6}$ Vg2=1800 cm/s, $A3=360\cdot10^{-6}$ Vg1=2500 cm/s τ=18 μs

Implantation of silicon with oxygen ions did not change the lifetime of carriers but the velocity of the surface recombination increased from 400 cm/s to 2500 cm/s.

4.1.4. Other methods of determination of the lifetime of carriers. Life measurements by μ-PCD (microwave photoconductivity decay)

A pulse laser illuminating the sample, of the energy of photons bigger than the energy gap of the matrial, creates excess carriers. The increased time dependent concentration of carriers change the conductivity of the semiconductor.The sample is at the same time illuminated by the beam of microwaves. Microwave reflection is dependent on the conductivity of the semiconductor. The time constant of the decay in conductivity and as a result of the intensity of the reflected microwave beam is measured. The dependence of the intensity of the reflected beam of the microwaves is described by the formula, where τ is the lifetime of photoexcited carriers.

$$I_R(t) = I_R(0) \cdot \exp\left(\frac{-t}{\tau}\right)$$

The slope of a photoconductive decay (1/τ) in time is measured for each point and is the measure of the lifetime as it is illustrated in the figure below. The disadvantage of this method results from the necessity of the measurements of this slope for each point of the map.

The equipment for this type of mesurements is built by Semiconductor Physics Laboratory, Inc. (Semilab). It enables mapping of the lifetime of carriers on the silicon plates.

The second method of the measurement of the lifetime of carriers is based on the measurement of the luminescence intensity of silicon. So that method is limited to this material. It is called a PCR – Photo Carrier Radiometry method. In this method excess carriers are excited with light of the energy of photons $h\nu > E_g$ (810 nm). Photoluminescence (PL) of silicon is measured, i.e. the portion of recombination that is a radiative recombination, $h_\nu \sim E_g$ (~1150 nm). The intensity of this silicon luminescence is measured by the CCD camera cooled to -75°C. The longer is the lifetime of the excess carriers the bigger is the intensity of the silicon luminescence. There is a linear dependence of the lifetime of carriers and the intensity of the edge silicon luminescence. The advantage of this method results from the simultaneous mesurement of the luminescence for all points of the map. So this method is much faster than the by μ-PCD method.

This First mapping of the lifetime of carriers for Si was done by S. Ostapenko, I. Tarasov, J. Kalejs, et al. in 1999. Comparison of the the spatial distributions of the lifetime done by these two methods shows a very good correlation.

The third method of the measurement is the Carrier Density Imaging (CDI). In this method a hot plate is a black body source of the infrared (IR) radiation. The camera collects the picture of the IR radiation of the source after passing through the sample and being the result of the IR absorption of the infrared radiation of the hot plate.

4.2. The optical parameters

The optical parameters mean: energy gap of semiconductors, optical absorption coefficient spectra both in the band to band region as also in the free electrons absorption region, contribution of a surface to bulk absorption. For this purpose several methods can be applied: optical, photoacoustic, photothermal PTR, piezoelectric photothermal PPT.

4.2.1. A transmission method

The optical absorption coefficient spectrum of semiconductors is very important for applications in semiconducting processing and solar cells. Typically in the optical methods measurements the absorbance spectra of the sample are measured. The absorbance spectrum is described by formula below where the $I_0(\lambda)$ is the intensity of light illuminating the sample, $I(\lambda)$ is the intensity of the transmitted light through the sample.

$$Abs(\lambda) = \log\left(\frac{I_0(\lambda)}{I(\lambda)}\right) \tag{4.2.1}$$

In the second step the transmission spectrum of the sample iss computed from the experimental absorbance spectrum according to formula (4.2.2).

$$TR(\lambda) = 10^{-Abs(\lambda)} \tag{4.2.2}$$

The transmission spectrum is connected with the optical absorption coefficient spectrum and the optical reflection coefficient as:

$$TR(\lambda) = \frac{(1-R)^2 \cdot \exp(-\beta(\lambda) \cdot l)}{1 - R^2 \cdot \exp(-2 \cdot \beta(\lambda) \cdot l)} \tag{4.2.3}$$

The optical reflection coefficient R can be determined from the value of a transmission for the transparency region when $\beta(\lambda)=0$. It is expressed by the formula

$$R = \frac{1 - TR(\lambda)}{1 + TR(\lambda)}. \qquad (4.2.4)$$

Optical absorption spectra of crystals were computed from the transmission spectra according to formula

$$\beta(\lambda) = -\ln\left(\frac{-(1-R)^2 + \left((1-R)^4 + 4R^2(TR(\lambda))^2\right)^{1/2}}{2R^2 TR(\lambda)} \right) \cdot l^{-1}. \qquad (4.2.5)$$

where l is a sample thickness. This procedure is applied for semiconductors both for the measurements of the optical absorption spectra in the band to band absorption region as also in the free carrier absorption region. Example characteristics illustrating this computational procedure are presented in figures below. For example optical characteristics of CdSe crystal of the thickness l=0.12 cm are taken.

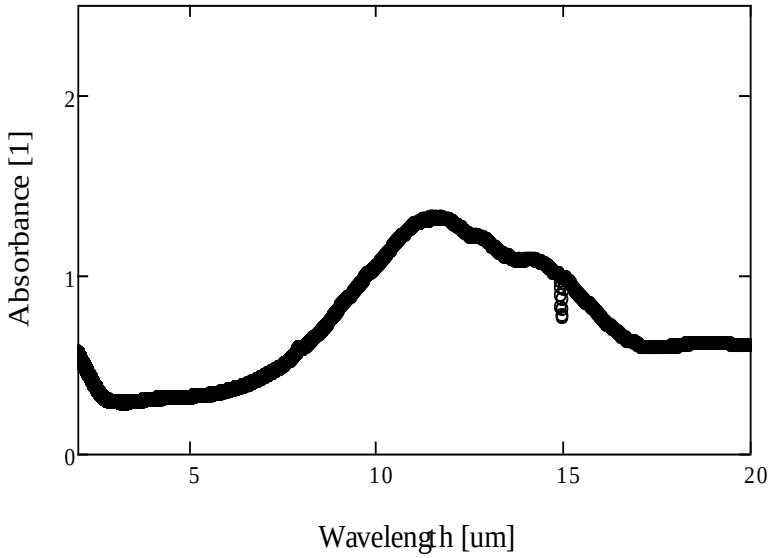


Fig. 4.2.1. Absorbance spectrum of the CdSe sample

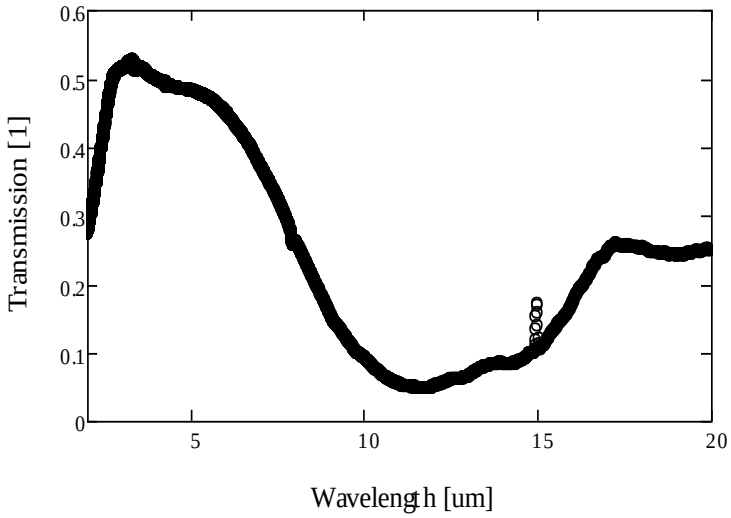


Fig. 4.2.2. Transmission spectrum of the CdSe sample

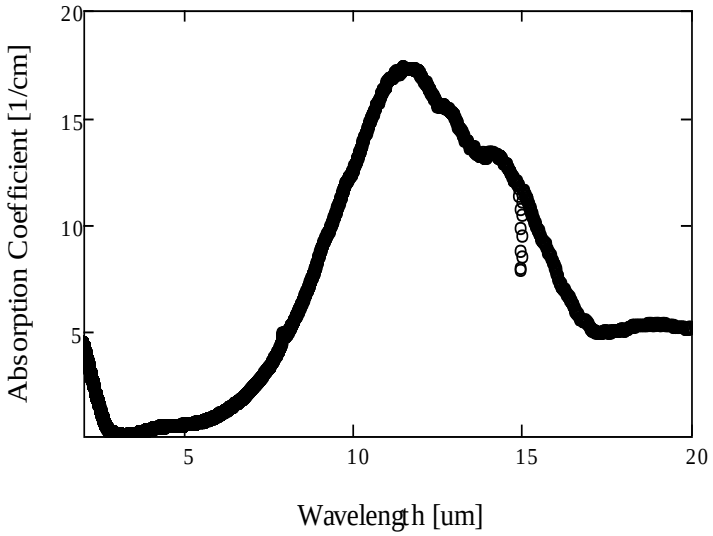


Fig. 4.2.3. Optical absorption coefficient of the CdSe sample in the IR region being the result of the free carrier absorption

The same procedure can be applied for the computations of the optical absorption coefficient spectra of semiconductors in the band to band absorption region near the energy gap of semiconductors. In this case the energy gap is determined from the Tauc characteristics. For example for semiconductors with a direct band gap it takes the form:

$$(\beta(\lambda)\cdot h\nu)^2 = A^2\cdot(h\nu - E_g) \tag{4.2.6}$$

where: hν is the energy of the photon, E_g is the energy gap of a semiconductor and β(λ) is the optical absorption coefficient spectrum of the sample described by formula (4.2.5). The energy gap is determined from the intersection point of the characteristics with the energy axis i.e where β(λ) reaches zero. The optical absorption coefficient spectrum can be recomputed for the effective absorption coefficient used for the photothermal methods.

This approach, as illustrated in Figs. 4.2.1 - 4.2.3, was applied for several new materials for solar cells and power electronic materials like SiC or even for different types of silicon which is still one of the basic materials for the photovoltaics. More recently, studies of nanostructured Si have become a rich area in searching for high efficiency and low cost solar cell solutions. Thus the optical constatnts of Si are of great importance and have been extensively studied with particular interests in the wavelength between 300 nm and 1100 nm. The objective of the studies presented in many papers was to do the experimental research of the absorption coefficient of silicon materials in the range between 500 nm and 1000 nm. The absorption coefficient was determined by measuring the transmittance of thin Si wafers at 300K using a monochromator and an integrating sphere. Other methods that are used for determination of the optical parameters of semiconducting materials are photoacoustics and piezoelectric methods.

4.2.2. Photoacoustic spectra

In the photoacoustic spectroscopy a sample is illuminated with the intensity modulated beam of light at the constant frequency and the amplitude and phase of the photoacoustic signal are measured by the microphone as a function of the wavelength of the illuminating light. The photoacoustic spectra bring information about the optical absorption coefficient spectrum of the sample. The experimental set-up for the measurements of the photoacoustic spectra of solid state samples is presented in Fig. 4.2.4.

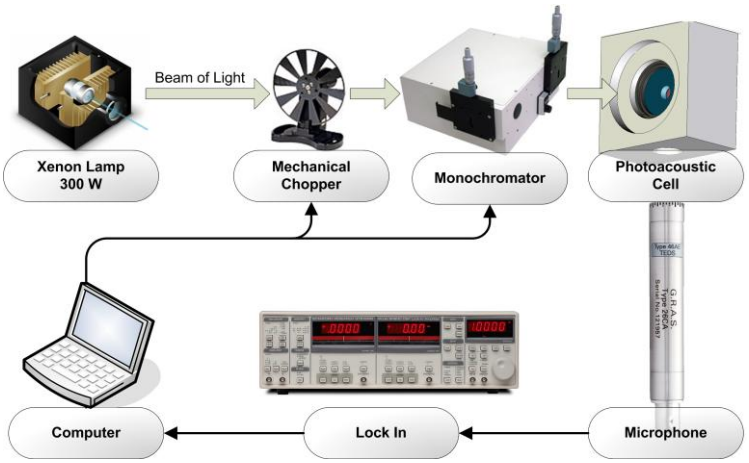
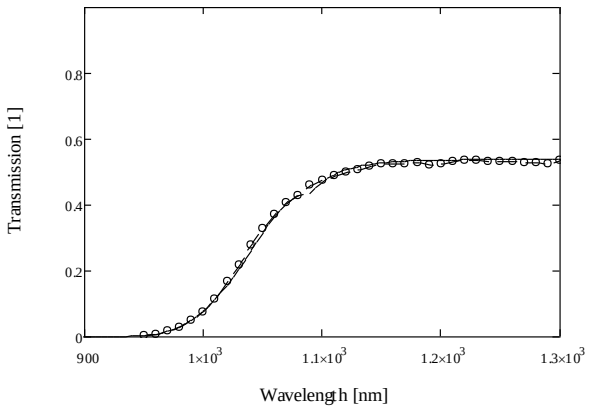


Fig. 4.2.4. The experimental set-up for the measurements of the photoacoustic spectra

The example photoacoustic transmission a) and amplitude b) spectra measured on the silicon sample of the thickness l=370 μm are presented in Fig. 4.2.5.

a)



b)

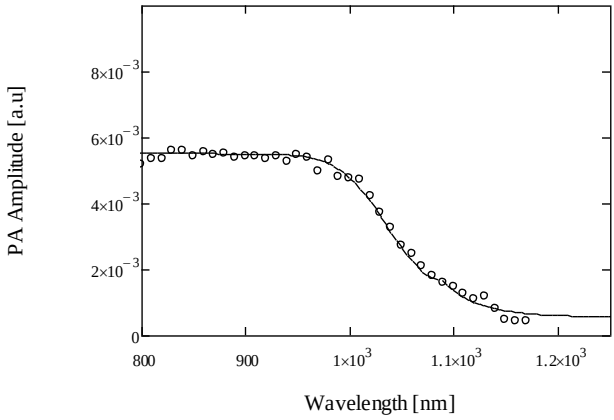


Fig. 4.2.5. The example photoacoustic transmission a) and amplitude b) spectra measured on the silicon sample of the thickness l=370 μm

The optical absorption spectrum of silicon computed both from the transmission and amplitude photoacoustic spectra is presented in Fig. 4.2.5.

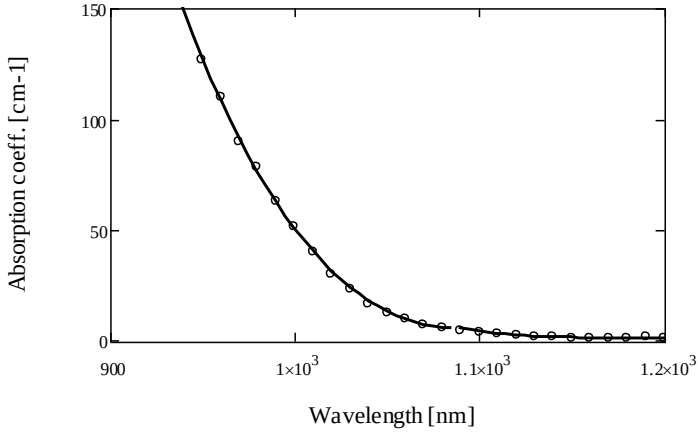


Fig. 4.2.6. Optical absorption coefficient spectrum of silicon computed from the photoacoustic spectra

This spectrum obtained from the photoacoustic amplitude spectrum agrees very well with the optical absorption coefficient spectrum obtained from the transmittance spectrum of thin silicon samples presented in the literature.

4.2.3. Piezoelectric photothermal spectra

In the piezoelectric spectroscopy a sample is also illuminated with the intensity modulated beam of light at the constant frequency and the amplitude and phase of the piezoelectric signal are measured by the piezoelectric transducer as a function of the wavelength of the illuminating light. The piezoelectric spectra bring information about the optical absorption coefficient spectrum of the sample. The piezoelectric signal that is measured is a complex function of the optical absorption coefficient spectrum β(λ):

$$S \; - \left(\frac{1}{l} \int_0^l T_{L/R}(x,\beta(\lambda))dx - \frac{6}{l^2} \int_0^l \left(\frac{l}{2} - x \right) T_{L/R}(x,\beta(\lambda))dx \right) \qquad (4.2.7)$$

a) b)

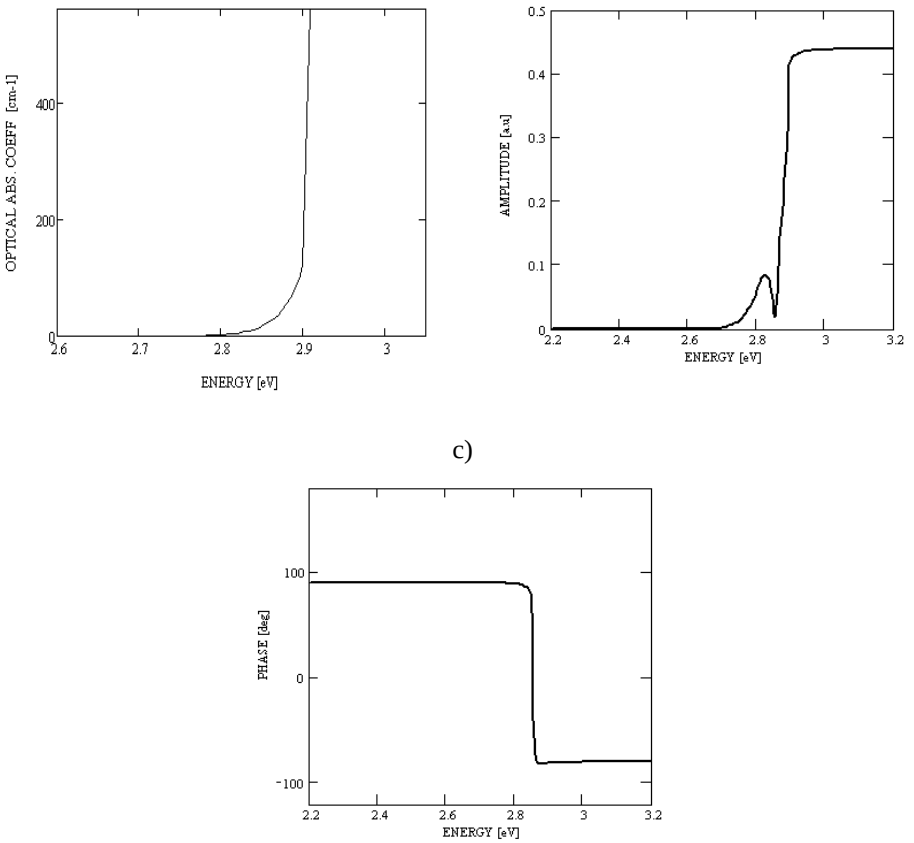


Fig. 4.2.8. a) Optical absorption coefficient spectrum taken for computation, b) resulting theoretical amplitude piezoelectric spectrum in the rear configuration, c) resulting theoretical phase piezoelectric spectrum in the rear configuration

The example piezoelectric spectra showing the correlation of the optical absorption coefficient spectrum and the resulting piezoelectric photothermal spectra are presented in the figures below. As it can be seen from the above spectra, piezoelectric spectra bring information ,among others, about the energy gap of the semiconductors. This method was applied for the measurements of the energy gap of several new materials for the photovoltaics. The piezoelectric photothermal technique was also used for the evaluation of the surface quality of semiconductor samples and in general for the optical characterization of semiconductor materials.

4.3. The thermal parameters: thermal conductivity, specific heat, thermal diffusivity and thermal effusivity

One of the methods applied for determination of the thermal parameters of solid state materials is the photoacoustic method. The experimental set-up for the frequency domain photoacoustic measurements is presented in Fig. 4.3.1.

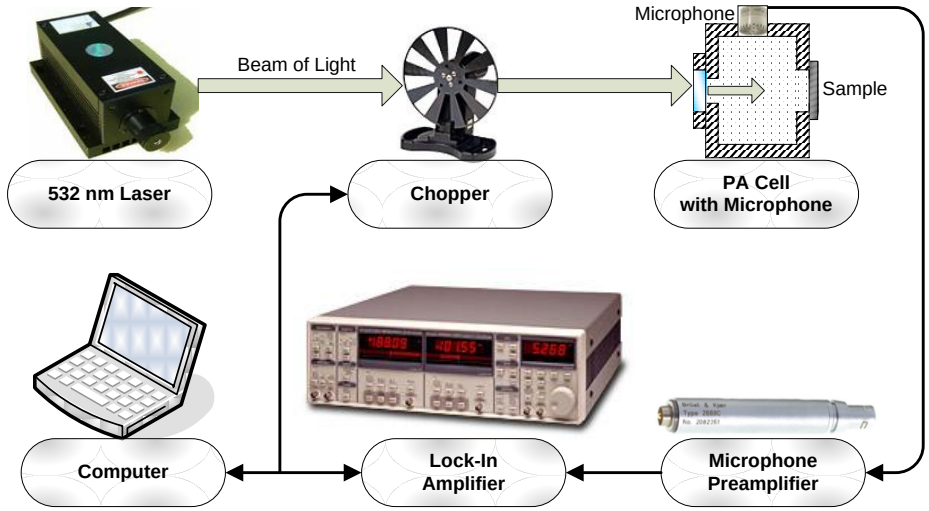


Fig. 4.3.1. Experimental set-up for the measurements of the thermal parameters of samples

The idea of this method is based on the illumination of the sample with the intensity modulated beam of light. As a result of the absorption of light the periodical temperature field arises in the sample. The temperature of the nonilluminated side of the sample is in contact with the gas in the airtight photoacoustic cell. Periodical temperature of the surface of the sample heats the

gas in the cell causing periodical changes of the pressure registered by the microphone connected with the cell. Thermal diffusivity of the material of the sample is extracted from the slope of the experimental frequency amplitude or phase photoacoustic characteristics measured in the transmission experimental configuration. The thermal diffusivity α of the material is expressed as:

$$\alpha = \frac{\lambda}{\rho \cdot c}$$

where: λ – thermal conductivity of a sample, ρ – density of the material and *c* is its specific heat. When ρ and *c* of the material are known then the thermal conductivity is calculated from the thermal diffusivity.

The thermal diffusivity of the material is determined from the frequency dependence of the photoacoustic signal (amplitude or phase) in the transmission configuration. The photoacoustic signal is connected with the photoacoustic signal by the formula below:

$$P_R(f,\alpha) = \frac{I_R}{\lambda \cdot \sigma(\alpha, f)^2} \cdot \frac{1}{\sinh\left(l \cdot \sigma(\alpha, f)\right)}$$

$$\sigma(\alpha, f) = \frac{1+i}{\mu} \qquad \mu = \sqrt{\frac{\alpha}{\pi \cdot f}}$$

This is the most frequently used method of determination of the thermal diffusivity of thermally thin samples used in photoacoustics. There are also other photoacoustic methods of determination of the thermal diffusivity of solid samples. One of them is the Sf/Sr and a phase lag methods. In this approach the ratio of the amplitudes of the photoacoustic signals and the phase difference of the photoacoustic signals in the front and rear configurations are measured and from the fitting of theoretical curves to experimental characteristics a thermal diffusivity of the samples is extracted.

References

1. W. Shockley, The theory of p-n junctions in semiconductors and p-n junction transistors, Bell Syst. Tech. J., vol. 28, 1949, p. 435
2. C.T. Sah, R.N. Noyce, W. Shockey, Carrier Generation and Recombination in p-n junction and p-n junction characteristics, Proc.IRE, 45, 1975, p. 1228.
3. G. Hall, Ionization Energy losses of highly relativistic charged particles in thin Silicon layers, Nucl. Inst. and Meth., vol. 220, 1984, p. 356–362.
4. J.F. Bak, A. Burenkov, J.B.B. Petersen, E. Uggerhøj, S.P. Møller, P. Siffert, Large Departures from Landau distributions for highenergy particles traversing thin Si and Ge targets, Nucl. Phys., B288, 1987, p. 681-716.
5. Review of Particle Properties – Particle Data Group, Phys. Lett., vol. B170, 1986, p. 1-350.
6. E. Laegsgaard, Position-sensitve semiconductor detectors, Nucl. Inst. and Meth., vol. 162, 1979, p. 93.
7. B. Hyams, U. Koetz, E. Belau, R. Klanner, G. Lutz, E. Neugebauer, A. Wylie, J. Kemmer, A silicon counter telescope to study short-lived particles in high-energy hadronic interactions, Nucl. Inst. and Meth., vol. 205, 1983, p. 99-105.
8. C. Colledani, W. Dulinski, R. Turchetta, F. Djama, A. Rudge, P. Weilhammer, A submicron precision Silicon telescope for beamtest purposes, Nucl. Inst. and Meth., vol. A372, 1996, p. 379-384.
9. ATLAS, Technical Proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN, CER: LHCC, p. 94-43.
10. W.S. Boyle, G.E. Smith, Charge coupled semiconductor devices., Bell Syst. Tech. Journal, vol. 49, 1970, p. 587-593.
11. R.H. Walden, R.H. Krambeck, R.J. Strain, J. McKenna, N.L. Schryer, G.E. Smith, The buried channel charge coupled device, Bell System Technical Journal, vol. 51, 1972, p. 1635-1640.
12. C.J.S. Damerell, F.J.M. Farley, A.R. Gillman, F.J. Wickens, Charge-coupled devices for particle detection with high spatial resolution, Nucl. Inst. and Meth., vol. 185, 1981, p. 33-42.

13. C.J.S. Damerell, et al., Design and performance of the SLD vertex detector: a 307 Mpixel tracking system, Nucl. Inst. and Meth., vol. A400, 1997, p. 287-343.

14. P. Middelkamp, Tracking with active pixel detectors, Thesis at Fachbereich Physik Bergische Univ., Germany, 1996.

15. E.H.M. Heijne, P. Jarron, Development of Silicon pixel detectors: An introduction, Nucl. Inst. and Meth., vol. A275, 1989, p. 467-471.

16. G. Vanstraelen, Monolithic integration of solid-state particle detectors and readout electronics on high resistivity Silicon, Thesis at the Katholieke Univ. Leuven, Belgium, 1990.

17. E. Gatti, P. Rehak, Semiconductor Drift Chamber – An Application of a Novel Charge Transport Scheme, Nucl. Inst. and Meth., vol. 255, 1984, p. 608-614.

18. W.L. Hansen, E.E. Haller, G.S. Hubbard, Protective Surface Coatings on Semiconductor Nuclear Radiation Detectors, IEEE Trans. on Nucl. Sci., vol. NS-27, No. 1, 1980, p. 247-251.

19. G.F. Knoll, Radiation Detection and Measurements (second edition), John Wiley and Sons, New York, 1989.

20. J. Lin, E.A. Henry, R.A. Meyer, Detection Efficiency of Ge(Li) and HPGe Detectors for Gamma Rays up to 10 MeV, IEEE Trans. on Nucl. Sci., vol. NS-28, No. 2, 1981, p. 1548-1550.

21. F.E. Cecil, F. J. Wilkinson, R.A. Ristinen, R. Rieppo, Experimental Determination of Absolute Efficiency and Energy Resolution for Nal(TI) and Germanium Gamma-Ray Detectors at Energies from 2.6 to 16.1 MeV, Nucl. Instr. and Meth., vol. A234, 1985, p. 479-482

22. AF. Sanchez-Reyes, M.I. Febrián, J. Baró, J. Tejada, Absolute Efficiency Calibration Function for the Energy Range 63–3054 keV for a Coaxial Ge(Li) Detector, Nucl. Instr. and Meth., vol. B28, 1987, p. 123-127

23. R.H. Pehl, F.S. Goulding, D.A. Landis, M. Lenzlinger, Accurate determination of the ionization energy in semiconductor detectors, Nucl. Instr. and Meth., vol. 59, 1988, p. 45-55.

24. G.H. Nakano, D.A. Simpson, W.L. Imhof, Characteristics of Large Intrinsic Germanium Detectors Operated at Elevated Temperatures, IEEE Trans. on Nucl. Sci., vol NS-24, No. 1, 1977, p. 68-72.

25. D. Venos, D. Srnka, J. Slesinger, D. Zanoucky, J. Stehno, N. Severijins, A. Van Geert, Performance of HPGe Detectors in the Temperature Range 2–77 K, Nucl. Inst. & Meth. in Phys. Res., vol. A365, 1995, p. 419-423.

26. M. Martini, T.A. McMath, Trapping and Detrapping Effects in Lithium Drifted Germanium and Silicon Detectors, Nucl. Inst. & Meth., vol. 79, 1970, p. 259-276.

27. G. Sanderson, A Comparison of Ge(Li) Well and N-Type Coaxial Detectors For Low Energy Gamma-Ray Analysis of Environmental Samples, 1979.

28. L.A. Beach, G.W. Phillips, Development of a Rugged HPGe Detector, Nucl. Inst. & Meth., vol. A242, No. 3, 1986, p. 520-524.

29. R. Keyser, T. Twomey, S. Wagner, Benefits of Using Super-Large Germanium Gamma-Ray Detectors for the Quantitative Determination of Environmental Radionuclides, Radioactivity and Radiochemistry, vol. 1, No. 2, 1990, p. 46-56.

30. R.M. Lindstrom, D.J. Lindstrom, L.A. Slaback, J.K. Langland, A Low-Background Gamma Ray Assay Laboratory for Activation Analysis, Nucl. Instr. and Meth., vol. A299, 1990, p. 425-429.

31. T.H. Becker, E.E. Gross, R.C. Trammell, Characteristics of High-Rate Energy Spectroscopy System with Time-Invariant Filters, IEEE Trans. on Nucl. Sci., vol. NS-28, No. 1, 1981, p. 598.

32. D. Smith, G.E. Derbyshire, R.C. Farrow, A. Sery, A Solid State Detector for Soft Energy Extended X-Ray Absorption Fine Structure Measurements, Rev. Sci. Instrum., vol. 66, No. 2, 1995, p. 2333-2335.

33. L. Britton, T.H. Becker, T.J. Paulus, R.C. Trammell, Characteristics of High-Rate Energy Spectroscopy Systems Using HPGe Coaxial Detectors and Time-Variant Filters, IEEE Trans. on Nucl. Sci., vol. NS-31, No. 1, 1984, p. 455-460.

34. T.W. Raudorf, R.C. Trammell, S. Wagner, R.H. Pehl, Performance of Reverse Electrode HPGe Coaxial Detector After Light Damage by Fast Neutrons, IEEE Trans. on Nucl. Sci., vol. NS-31, No. 1, 1984, p. 253-257.

35. R.M. Keyser, T.R. Twomey, P. Sangsingkeow, Advances in HPGe Detectors for Real-World Applications, ORTEC, 1999.

36. R.M. Keyser, T.R. Twomey, P. Sangsingkeow, Matching Ge Detector Element Geometry to Sample Size and Shape: one does not fit all!, Proceedings of the 1998 Winter Meeting of the ANS, 1998.

37. ANSI/IEEE Std 325-1996, IEEE Standard Test Procedures for Germanium Gamma- Ray.

38. D. Dietzel, J. Gibkes, S. Chotikaprakhan, B.K. Bein, J. Pelz, "Radiometric Analysis of Laser Modulated IR Properties of Semiconductors, International Journal of Thermophysics, vol. 3, vol. 24, 2003, pp. 741–755.

39. B. Li, X. Li, W. Li, Q. Huang, X. Zhang, Accurate determination of electronic transport properties of semiconductor wafers with spatialy resolved photo-carrier techniques, Journal of Physics: Conference Series, vol. 214, 2010, p. 012013.

40. X. Zhang, B. Li, C. Gao, Analysis of free carrier absorption measurement of electronic transport properties of silicon wafers, European Physics Journal Special Topics, vol. 153, 2008, p. 279-281.

41. Q. Huang, B. Li, X. Liu, Influence of probe beam size on signal analysis of modulated free carrier absorption technique, Journal of Physics: Conference Series, vol. 214, 2010, p. 012084.

42. W. Li, B. Li, Analysis of modulated free-carrier absorption measurement of electronic transport properties of silicon wafers, Journal of Physics: Conference Series, vol. 214, 2010, p. 012116.

43. A. Mandelis, J. Balista, D. Shaughnessy, Infrared photocarrier radiometry of semiconductors: Physical principles, quantitative depth profilometry, and scanning imaging of deep subsurface electronics defects, Physical Review, vol. B67, 2003, p. 205208.

44. J. Schmidt, Measurement of differential and actual recombination parameters on crystalline silicon wafers, IEEE Transactions on Electron Devices, vol. 46, No. 10, 1999, pp. 2018–2025.

45. T. Creazzo, B. Redding, E. Marchena, S. Shi, D.W. Prater, Free-carrier absorption modulation in silicon nanocrystal slot waveguides, Optics Letters, 2010, vol. 35, No. 21, 2010, p. 3691-3693.

46. Q. Huang, B. Li, Self-eliminating instrumental frequency response from free carrier absorption signals for silicon wafer characterization, Rev. Sci. Instrum., vol. 82, 2011, p. 043104.

47. Q. Huang, B. Li, Electronic transport characterization of silicon wafers by combination of modulated free carrier absorption and photocarrier radiometry, J. Appl. Phys., vol. 109, 2011, p. 023708.

48. Ł. Chrobak, M. Maliński, Zastosowanie zjawiska modulacji absorpcji na swobodnych nośnikach do nieniszczących badań materiałów półprzewodnikowych, Elektronika 12, 2012, p. 110-113.

49. Ł. Chrobak, M. Maliński, Badania parametrów rekombinacyjnych materiałów krzemowych z wykorzystaniem nieniszczącej techniki MFCA opartej na zjawisku modulacji absorpcji na nośnikach swobodnych, Zeszyty Naukowe WEiI PK, No.3, 2011, p. 13-16.

50. M.D. Dramicanin, Z.D. Ristovski, P.M. Nikolic, D.G. Vasiljevic, D.M. Todorovic, Photoacoustic investigation of transport in semiconductors: Theoretical and experimental study of a Ge single crystal, Physical Review, vol. B51, No. 20, 1995, p. 14226-14232.

51. M. Todorovic, P.M. Nikolić, Investigation of carrier transport processes in semiconductors by the photoacoustic frequency transmission method, Opt. Eng., vol. 36, No. 2, 1997, p. 432-445.

52. Q. Shen, T. Toyoda, Photoacoustic characterizatrion of thermal and electronic transport properties of CdInGaS4 in a transmission detection configuration, Jpn. J Appl.Phys., vol. 39, 2000, p. 3164-3168.

53. A. Martin, I. Reich, P. Diaz, H. Vargas, Influence of carrier recombination on the thermodiffusion, thermoelastic and electronic strain photoacoustic signal generation mechanisms in semiconductors, Analytical Sciences, vol. 17, 2001, p. 284-287.

54. J. Bernal-Alvarado, M. Vargas-Luna, Photoacoustic spectroscopy of semiconductor powders and thin films, Analytical Sciences, vol.17, 2001, p. 309-311.

55. K.T. Radulovic, P.M. Nikolić, D. Vasiljevic-Radovic, D.M. Todorovic, S.S. Vujatovic, A.I. Bojicic, V. Blagojevic, D. Urosevic, A contribution of carrier transport processes to the photoacoustic effects in doped narrow gap semiconductors, Review of Scientific Instruments, vol. 74, No.1, 2003, p. 595-597.

56. L. Bychto, M. Maliński, Comparison of plasma waves models in photoacoustics, JDP IVFrance, vol. 129, 2005, p. 213-216.

57. Ł. Chrobak, M. Maliński, A. Patryn, Theoretical and experimental studies of a plasma wave contribution to the photoacoustic signal for Si samples, Acta Acustica United with Acustica, vol. 95, 2009, p. 60-64.

58. M. Maliński, Ł. Chrobak, Determination of the life time of excess carriers in silicon with photoacoustic and photocurrent methods, Journal of Physics: Conference Series, vol. 214, 2010, p. 012075.

59. Ł. Chrobak, M. Maliński, A. Patryn, Influence of Plasma Waves on the Photoacoustic Signal of Silicon Samples, International Journal of Thermophysics vol.32(9) 2011 p. 1986-1997.

60. A. Salinick, A. Mandelis, H. Ruda, C. Jean, Relative sensitivity of photomodulated reflectance and photothermal infrared radiometry to thermal and carrier plasma waves in semiconductors, J. Appl. Phys., vol. 82, No. 4, 1997, p. 1853-1859.

61. A. Salnick, C. Jean, A. Mandelis, Noncontacting photothermal radiometry of SiO_2/Si MOS capacitor structures, Solid State Electronics, vol. 41, No. 4, 1997, p. 591-597.

62. A. Mandelis, Laser infrared photothermal radiometry of semiconductors: principles and applications to solid state electronics, Solid State Electronics, vol. 42, No. 1, 1998, p. 1-15.

63. A. Salnick, A. Mandelis, C. Jean, Noncontact measurement of transport properties of long-bulk-carrier-lifetime Si wafers using photothermal radiometry, Appl. Phys. Lett., vol. 69, No. 17, 1996, p. 2522-2524.

64. M.E. Rodriguez, A. Mandelis, G. Pan, L. Nicolaides, J.A. Garcia, Y. Riopel, Computational aspects of laser radiometric multiparameter fit for carrier transport property measurements in Si wafers, J. of Electrochemical Society, vol. 147, No. 2, 2000, p. 687-689.

65. A. Salnick, A. Mandelis, F. Funak, C. Jean, Monitoring of ion implantation in Si with carrier plasma waves using infrared photothermal radiometry, Appl.Phys. Lett., vol. 71, No. 11, 1997, p. 1531-1533.

66. A. Mandelis, J. Batista, C. Shaughnessy, Infrared photocarrier radiometry of semiconductors: Physical principles, quantitative depth profilometry, and scanning imaging of deep subsurface electronic defects, Physical Review, vol. B67, 2003, p. 205208.

67. A. Mandelis, A. Othonos, C. Christofides, J. Boussey–Said, Non-contacting measurements of photocarrier lifetimes in bulk – and polycrystalline thin-film Si photoconductive devices by photothermal radiometry, J.Appl.Phys, vol. 80, No. 9, 1966, p. 5332-5341.

68. M.E. Rodriguez, A. Mandelis, F. Rabago, L. Nicolaides, Photothermal characterization of B – implanted Si (shallow) samples, Analytical Sciences, vol. 17, 2001, p. 277-280.

69. M. Nestoros, Y. Karmiotis, C. Christofides, Two layer model for photothermal radiometry applied on semiconducting thin films, J. Appl. Phys., vol. 82, No. 12, 1997, p. 6220-6227.

70. F. Firszt, A. Wronkowska, A. Wronkowski, S. Łęgowski, A. Marasek, H. Męczyńska, M. Pawlak, W. Paszkowicz, J. Zakrzewski, K. Strzałkowski, Growth and optical characterization of $Cd_{1-x}Be_xSe$ and $Cd_{1-x}Mg_xSe$ crystals, Cryst. Res. Technol., vol. 40, No. 4/5, 2005, p. 386-394.

71. A. Wronkowska, A. Wronkowski, H. Arwin, F. Firszt, S. Łęgowski, H. Męczyńska, J. Szatkowski, Characterisation of $Cd_{1-x}Mg_xSe$ solid solutions by spectroscopic ellipsometry, Vacuum, vol. 63, 2001, p. 233-239.

72. Wang, X. Liu, Z. M. Zhang, Absorption coefficients of crystalline silicon at wavelengths from 500 nm to 1000 nm, Int. Journal of Thermophysics, vol. 34, 2013, p. 213-225.

73. M. Maliński, L. Bychto, Photoacoustic Studies of the Absorption Coefficient of CdTe, Proceedings 5th Therminic Workshop Rome, 1999, p. 225-230.

74. M. Maliński, L. Bychto, F. Firszt, J. Szatkowski, J. Zakrzewski, Determination of the Optical Absorption Coefficient of $Zn_{1-x-y}Mg_yBe_xSe$ Mixed Crystals from the PAS Experiments- Improved Approach, Analytical Sciences, vol. 17, 2001, p. 133-136.

75. M. Maliński, L. Bychto, S. Łęgowski, J. Szatkowski, J. Zakrzewski, Photoacoustic Studies of $Zn_{(1-x)}Be_xSe$ Mixed Crystals, Microelectronics Journal., vol. 32, 2001, p. 903-910.

76. Ł. Chrobak, M. Maliński, Transmission and absorption based photoacoustic methods of determination of the optical absorption spectra of Si samples-comparison, Solid State Communications, 2009, p. 1600-1604.

77. M. Maliński, Ł. Chrobak, Numerical analysis of the photoacoustic spectra of silicon samples with differently treated surfaces, Optics Communications, vol. 283, 2010, p. 1004-1007.

78. M. Maliński, Ł. Chrobak, Photoacoustic Operation Modes for the Determination of the Absorption Spectra of SiGe mixed Crystals, Opto Electronics Review, vol. 18, No. 2, 2010, p. 19-25.

79. M. Malinski, Ł. Chrobak, L. Bychto, Photoacoustic investigations of the optical absorption spectra of porous silicon layers on silicon backing, Solid State Communications, vol. 150, 2010, p. 424-427.

80. M. Malinski, Ł.B. Chrobak, L. Bychto, T. Okupski, Investigations of the Optical Absorption Spectra of Porous Silicon Layers on the Silicon Backing by the Nondestructive Photoacoustic Method, Thin Solid Films, vol. 519, 2010, p. 394-398.

81. M. Maliński, Ł. Chrobak, The Photoacoustic Spectroscopic Investigations of the Surface Damage of Silicon Samples, Opto - Electronics Review, vol. 19, No. 1, 2011, p. 43-47.

82. F. Firszt, K. Strzałkowski, A.J. Zakrzewski, S. Łęgowski, H. Męczyńska, A. Marasek, Photoelectric and photothermal investigations of $Zn_{1-x-y}Be_xMn_ySe$ solid solutions, Cryst. Res. Technol., vol. 42, No. 12, 2007, p. 1352-1358.

83. D.M. Todorović, J. Zakrzewski, M. Maliński, T. Grozdić, F. Firszt, Photoacoustic spectra of $Zn_{1-x}Be_xTe$ near the energy gap, European Physical Journal –Special Topics, vol. 153, 2008, p. 263-266.

84. K. Yoshino, H. Mikami, K. Imai, M. Yoneta, T. Ikari, Optical characterization of native defects in ZnSe substrate, Physica B: Physics of Condensed Matter, vol. 302, 2001, p. 299-306.

85. A. Fukuyama, S. Sakamoto, S. Sonoda, P. Wang, K. Sakai, T. Ikari, Piezoelectric photothermal study of the optical properties of microcrystalline silicon near the bandgap, Thin Solid Films, vol. 511-512, 2006, p. 112-116.

86. K. Sakai, S. Tada, A. Fukuyama, T. Ikari, Characterization of deep levels In n-type 4H-SiC single crystals by means of a piezoelectric photothermal and a photoluminescence spectroscopy, Physica B, vol. 340-342, 2003, p. 137-140.

87. A. Memon, A. Fukuyama, S. Sato, T. Ikari, Surface states and band-to-band non-radiative transitions in silicon single crystal investigated by piezoelectric photo thermal spectroscopy, materials Science and Engineering, vol. B102, 2003, p. 12-15.

88. Y. Akaki, N. Ohryoji, A. Fukuyama, Y. Yoshino, S. Kawakita, M. Imaizumi, S. Niki, K. Sakurai, S. Ishizuka, T. Ohshima, T. Ikari, Piezoelectric photothermal investigations of proton irradiation induced defects in CuInSe2 epitaxial films, Thin Solid Films, vol. 480-481, 2005, p. 250-253.

89. A. Memon, M. Maliński, A. Fukuyama, T. Ikari, Effect of surface states on piezoelectric photothermal spectra of silicon single crystals, Japanese Journal of Applied Physics, vol. 43, No. 5A, 2004, p. 2397-2401.

90. W. Ding, Y. Nakano, R. Yamamoto, K. Sakai, H. Nakazawa, A. Fukuyama, T. Ikari, Temperature dependence of the optical gap of diamond - like carbon films investigated by a piezoelectric photothermal spectroscopy, Energy Procedia, vol. 10, 2011, p. 66-70.

91. M. Maliński, Photoacoustics and photoacoustic spectroscopy of semiconducting materials, TU of Koszalin Publishing 2004, 2004, ISSN 0239-7129.

92. J. Zakrzewski, Piezoelectric photothermal spectroscopy in the volume and on the surface of AII-BVI semiconductors, UMK Toruń Publishing, 2013, ISBN 978-83-231-2972-1.

93. M. Maliński, L. Bychto, Determination of the thermal parameters of materials by the photoacoustic amplitude measurements, Molecular & Quantum Acoustics, vol. 18, 1997, p. 179-185.

94. M. Maliński, L.Bychto, Photoacoustic Determination of the Thermal Parameters of Layers in Front and Rear Surface Illumination Methods, Molecular & Quantum Acoustics, vol. 18, 1997, p. 169-177.

95. Z. Suszyński, M. Maliński, L. Bychto, Thermal Parameters Measurement Method of Electronics Materials, IEEE Transactions CPMT-Part A, vol. 21, No. 3, 1998, p. 424-433.

96. A. Patrin, N. Abrosimov, M. Maliński, L. Bychto, The Influence of the Composition of Si-Ge Mixed Crystals on Thermal Diffusivity-Photoacoustic Approach, Solar Energy Materials and Solar Cells, vol. 72, 2002, p. 579-587.

97. M. Maliński, L. Bychto, A. Patryn, J. Gibkes, J. Bein, J. Pelzl, Investigations of the optical and thermal parameters of porous silicon layers with the two wavelength photoacoustic method, Journal de Physique IV France, vol. 129, 2005, p. 241-243.

98. Q. Shen, T. Toyoda, Characterization of thermal properties of Porous silicon film/silicon using photoacoustic technique, J. of Thermal Analysis and Calorimetry, vol. 69, 2002, p. 1067-1073.

99. Q. Shen, T. Takahashi, T. Toyoda, Characterization of optical and thermal properties of porous silicon using photoacoustic technique, Analytical Sciences, vol. 17, 2001, p. 281-283.

100. A. Irudayaraj, R. Srinivasan, P. Kuppusami, E. Mohandas, S. Kalainathan, K. Ramachandran, Photoacoustic measurement of thermal properties of TiN thin films, Journal of Material Science, vol. 43, No. 3, 2008, p. 1114-1120.

101. J. Baltazar-Rodrigues, J.C. de Lima, C.E.M. Campos, T.A. Grandi, Effects of photoacoustic measurements on a nanostructured ZnSe mechanically alloyed, J.Phys: Condensed Matter, vol. 20, 2008, p. 465205

102. K. Behzad, W.M.M. Yunus, Z.A. Talib, A. Zakaria, A. Bahrami, Effect of preparation parameters on physical, thermal and optical properties of n-type porous silicon, Int. J.Electrochem. Sci., vol. 7, 2012, p. 8266-8275.

103. M. Maliński, Photoacoustics and photoacoustic spectroscopy of semiconducting materials, TU of Koszalin: Poland Publishing, 2004, ISSN 0239-7129.

104. M. Pawlak, M. Maliński, Influence of the Ar^{8+} and O^{6+} ion implantation on the recombination parameters of p and n type implanted Si samples investigated by means of the photothermal infrared radiometry, Infrared Physics & Technology vol. 63, 2014, p. 6-9.

105. M. Pawlak, M. Maliński Minority carrier recombination lifetimes in n-type CdMgSe mixed crystals measured by means of the photothermal infrared radiometry, Opto-Electronics Review,vol. 22 (1), 2014, pp. 31-35.

106. M. Pawlak, M. Maliński, F. Firszt, J. Pelzl, A. Ludwig, A. Marasek, A linear relationship between the Hall carrier concentration and the effective absorption coefficient measured by means of photothermal radiometry in IR semi-transparent n-type CdMgSe mixed crystals, Measurement Science and Technology, vol. 25, 2014, p. 035204.

Contributors

Alexander K. Fedotov, Professor, Head of Department of Energy Physics. Belarusian State University, Faculty of Physics, 2 off. 336, Nezavisimosti Avenue, 220030, Minsk, Belarus, tel. +375 29 6326075, fedotov@bsu.by

Mirosław Maliński, Dr. habil. sc., professor KUT, Faculty of Electronics and Computer Sciences, Koszalin University of Technology, 2, Śniadeckich Str., 75-452 Koszalin, Poland, tel. +48 94 3478706, mmalin@tu.koszalin.pl

Aleksy Patryn, Professor, electronics & physics, Faculty of Electronics and Computer Sciences, Koszalin University of Technology, 2, Śniadeckich Str., 75-452 Koszalin, Poland, tel. +48 94 3478719, aleksy.patryn@tu.koszalin.pl

Paweł Żukowski, Dr. habil. sc., professor LUT, Head of Department of Electrical Devices and High Voltage Technology, Faculty of Electrical Engineering and Computer Sciences, Lublin University of Technology, 38a, Nadbystrzycka Str., 20-618 Lublin, Poland, tel. +48 81 5384327, p.zhukowski@pollub.pl, pawel@elektron.pol.lublin.pl

Nikolay Drozdov, Ph.D., associate professor, Department of Energy Physics, Faculty of Physics, Belarusian State University, 2 off. 234, Nezavisimosti Avenue, 220030, Minsk, Belarus, tel. +375 29 6602675, drozdov@bsu.by

Tomasz N. Kołtunowicz, Ph.D. Eng., assistant professor, Department of Electrical Devices and High Voltage Technology, Faculty of Electrical Engineering and Computer Sciences, Lublin University of Technology, 38a, Nadbystrzycka Str., 20-618 Lublin, Poland, tel. +48 81 5384713, t.koltunowicz@pollub.pl

Alexander V. Mazanik, Ph.D., associate professor, head scientific researcher, Department of Energy Physics, Faculty of Physics, Belarusian State University, 2 off. 342, Nezavisimosti Avenue, 220030, Minsk, Belarus, tel. +375 29 2769624, amazanik1@rambler.ru

Mikhail Tivanov, Ph.D., associate professor, Faculty of Physics, Belarusian State University, 2 off. 334, Nezavisimosti Avenue, 220030 Minsk, Belarus, tel. +375 29 5580813, tivanov@bsu.by

Konrad Kierczyński, M.Sc. Eng., assistant, Department of Electrical Devices and High Voltage Technology, Faculty of Electrical Engineering and Computer Sciences, Lublin University of Technology, 38a, Nadbystrzycka Str., 20-618 Lublin, Poland, tel. +48 81 5384328, k.kierczynski@pollub.pl