Vladimir V. Kiselev

Collective Effects in Condensed Matter Physics

De Gruyter Studies in Mathematical Physics

Volume 44

Vladimir V. Kiselev

Collective Effects in Condensed Matter Physics

DE GRUYTER

Mathematics Subject Classification 2010
82D, 81V80, 74B

Author
Dr. Sci. (Phys.-Math.) Vladimir V. Kiselev
Russian Academy of Sciences
M. N. Mikheev Institute of Metal Physics
Sophia Kovalevskaya str., 18
620108 Yekaterinburg
kiseliev@imp.uran.ru

ISBN 978-3-11-058509-4
e-ISBN (PDF) 978-3-11-058618-3
e-ISBN (EPUB) 978-3-11-058513-1
ISSN 2194-3532

Library of Congress Control Number: 2018934560

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

Typesetting: le-tex publishing services GmbH, Leipzig
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Foreword

To date, numerous phenomena and processes observable on a macroscopic scale have been studied in condensed matter. However, to explain them properly, quantum concepts about the microstructure of the medium are required. The interconnection and mutual complementation of different fields of knowledge is one of the noteworthy features of the current state of science. All the monographs, on the one hand, are devoted to special problems of condensed matter physics. They give a deep and complete exposition, for example, of the physics of semiconductors, superconductivity, or quantum optics intended for experienced specialists in a narrow area of solid state physics. The purpose of other publications is to draw readers' attention to the fundamental problems of quantum macrophysics and its numerous technological applications. The lack of discussion about constructive aspects of the theory, arising difficulties, and possible ways to overcome them is a disadvantage of most works on this theme. To successfully resolve the problems of quantum macrophysics in practice, it is necessary to overview a huge amount of material as quickly and thoroughly as possible, which requires great effort. There is an acute shortcoming in the scientific literature and methodical techniques on quantum macrophysics for students, graduate students, and researchers.

A distinctive feature of this book is that it covers not only selected issues deemed to be a necessary part of the education of a scientific researcher, but it also discusses the ways of a gradual approach to solving the problems considered, i.e., what usually remains behind the scenes. The discussion of specific issues is brought about through complementary levels of theoretical description, quantum, semiclassical, and phenomenological approaches. This makes it possible to more fully address the multifaceted nature of macroscopic quantum effects and, ultimately, come closer to the problems of the current state of science within this field.

This book is based on a course of lectures the author has been delivering for a number of years to students of the Theoretical Physics and Applied Mathematics Department of the Faculty of Physics and Technology at Ural Federal University. The specifics of the existing education system is that third year students (future engineers or scientists) who seek a bachelor's degree should have the knowledge of condensed matter physics. It is this connecting role that is played by the course "Quantum Macrophysics". The exercises are chosen, not only for educational purposes, but also in order to supplement the main material.

This book was first published in the scientific and educational series, Condensed Matter Physics, initiated and edited by V.V. Ustinov, a Member of the Academy of Science and a director of the Institute of Metal Physics of the Ural Branch of the Russian Academy of Sciences. The present edition is intentionally supplemented with a new chapter: Dislocations and Martensitic Transitions. It includes analysis of dislocation systems, solutions of the strongly nonlinear and nonlocal Peierls model for dislocation

https://doi.org/10.1515/9783110586183-201

cores, and the discussion of the soliton mechanism of anomalous acoustic emission near the martensitic phase transition point. The author is very grateful to the administration of the Institute of Metal Physics for creating favorable conditions for work and to D.V. Dolgikh and A.A. Raskovalov for technical assistance in preparing the book.

Contents

Prefactory Notes

1. The Concept of Quasiparticles as Collective Perturbations of a Medium

Quantum macrophysics establishes a relationship between the quantum properties of individual atoms and molecules, as well as the properties of giant associations consisting of different combinations of atoms or molecules. Regularly ordered systems as crystals, including molecular crystals, can serve as an example of these associations.

Macroscopic quantum effects take place not only in crystals. Based on quantum concepts, we have the ability to theoretically describe such macroscopic properties of biological objects as the spread of localized excitations in deoxyribonucleic acid (DNA) and protein molecules, self-trapping in globular proteins, and the spread of motor signals along nerve fibers. Quantum concepts underlie macrophysics of autocatalytic chemical reactions, of the propagation of waves and oscillations in solutions of chemical agents, such as flame propagation, etc. There are numerous other examples of phenomena and processes observable on macroscopic scales. However, for these to be explained, the quantum concepts about the microstructure of medium are required. To highlight their features and red flag the field of knowledge, alternative to classical physics of continuous medium, the term "quantum macrophysics" has been introduced.

Many condensed matter properties observable in reality are inexplicable from the standpoint of classical physics. Nevertheless, they can be described in terms of the quantum theory of many particle systems through the notion of so called collective excitations of a medium [1–16]. It is curious that, in their parameters, the elementary collective excitations of a medium bears a strong resemblance to microparticles. As such, they are often called *quasiparticles*. The prefix "quasi" means "as if" or "almost". It should be emphasized that the concept of quasiparticles works well only under weak external influences on the medium.

Quasiparticles obey the laws of quantum mechanics. The main difference between quasiparticles and material points of classical physics is that identical quasiparticles are principally impossible to distinguish from each other, whereas identical material points can be enumerated and thoroughly followed along their trajectories. In quantum mechanics, microparticles and quasiparticles are described by wave packets. The square modulus of the total wave function of a system of particles determines the probability of detecting the latter in space. When separated in space, the wave packets associated with identical microparticles can be enumerated and watched individually. However, once the wave packets have overlapped, we will no longer be able to tell the microparticles apart. The meaning of the wave function of the particle system implies that there is an equal probability of being any two microparticles in the region of the

https://doi.org/10.1515/9783110586183-202

two overlapping wave packets. That is to say, it is impossible to ascertain which of these microparticles was the first to fall in the region, and which was the second.

According to the laws of quantum mechanics, the total wave function of a system of identical quasiparticles can be either symmetric or antisymmetric depending on the rearrangement of quantum numbers of any two particles. Micro- and quasiparticles with half integer spin have an antisymmetric wave function. Such particles are called *fermions* (e.g., an electron has a spin equal to 1/2). Symmetric wave functions describe identical microparticles with integer spin. Such particles are called *bosons* (e.g., quantums of electromagnetic field as photons have a spin equal to one). This rule is a generalization of the experimental data and is the core of one of the postulates of quantum mechanics.

Even if there is no interaction between identical fermions at all, the antisymmetry property of their total wave function, nevertheless, leads to certain consistency (correlations) in the motion of individual particles. This property manifests itself in the energy distribution of fermions, which is very specific and unusual for classical physics. Namely, for noninteracting fermions, the *Pauli principle* holds: every quantum mechanical state can be occupied by only one fermion. Fermions are "individualists" and this circumstance conditions the macroscopic properties of Fermi systems. For example, the properties of Fermi quasiparticles govern the heat capacity of metals at low temperatures and the conductivity of metals and semiconductors.

Similarly, when bosons do not interact with each other, the symmetry property of the total wave function of a system of the bosons manifests itself in the peculiarities of motion of the bosons and their energy distribution. In this case, the macroscopic properties of the system of bosons are very different when compared to the system of fermions. Identical bosons can simultaneously be in the same quantum mechanical state. Bosons are "collectivists." At the macro level, this feature of bosons exhibits itself, for example, as the self-induced coherent radiation of medium (in lasers), the superconductivity in some metals at low temperatures, and also explains the temperature dependence of the lattice heat capacity of solids.

Our goal is to become familiar with methods for theoretical description of quantum macrosystems and to explain their behavior, at least on a qualitative level. The understanding of the issues involved requires using the results of many areas of physics such as analytical mechanics, field theory, quantum mechanics, thermodynamics, statistical mechanics, and electrodynamics of continuous media. For completeness of the discussion, we shall perform brief mathematical digressions to furnish the reader with information needed for further analysis. However, the subject matter is impossible to expose rigorously. Therefore, in writing the book, we had to find a compromise between the description of the features of analytical apparatus and qualitative explanations of macroscopic quantum phenomena and effects. The next section is explores aforementioned mathematical digressions.

2. The Energy Distribution of Particles: Ideal Fermi and Bose Gases

In classical physics, an ideal gas is a system of N material points whose interaction can be ignored.

Let the gas be enclosed in a container of volume V and placed in an external force field and let the potential energy of interaction between k-th material particle and the external field depend only on the particle coordinates: $U(\vec{r}_k)$. Then each particle moves in accordance with Newton's second law:

$$m\ddot{\vec{r}}_s = -\frac{\partial}{\partial \vec{r}_s} U(\vec{r}_s) \,,$$

where $s = 1 \ldots N$. The corresponding particle energy is:

$$\varepsilon_s = \frac{m\dot{\vec{r}}_s^2}{2} + U(\vec{r}_s) \,.$$

In the frame of classical physics, we can trace in detail the motion of each of the material points of an ideal gas by solving the system of equations of motion with given initial conditions. However, this process involves serious computational difficulties. Fortunately, it is not absolutely necessary to follow this process. When the number of particles in the system is large, experimentally observable parameters can be ascertained, not through intricate paths of the individual particles, but as a result of their collective action. Collective properties do not depend on the initial conditions of the particle motion, and are well described in terms of averages and probabilities.

The course of classical statistical physics shows that if particles of a thermally equilibrium ideal gas, being in an external potential field, are in a state with the energy ε_s, their average number can be determined by the *Boltzmann distribution:*

$$f_s = \exp\left(\frac{\mu - \varepsilon_s}{k_\mathrm{B}}\right) \,, \tag{1}$$

where k_B is the Boltzmann constant and T and ε_s are the gas temperature and the energy of a single particle, respectively. The parameter μ is called the chemical potential and is determined by the normalization condition. The latter expresses the conservation of the total number of particles N in the system:

$$\sum_s f_s = N \,.$$

Here we note an important point. The definition of an ideal gas neglects the potential energy of interaction between its particles. At the same time, the equilibrium distribution of particles of an ideal gas in energy (the Boltzmann distribution) is caused by the interaction of the gas particles with the thermostat walls and their collisions with each other.

The above conception has impelled scientists to use statistical methods for describing quantum mechanical systems of identical particles, which has led to the emergence of classical statistical physics. When the initial wave function of a system consisting of N identical particles is known, the system's evolution is possible to watch by solving the *Schrödinger equation:*

$$i\hbar\frac{\partial}{\partial t}\Psi = \widehat{H}\left(\hat{\vec{p}}_1, \hat{\vec{q}}_1, \ldots, \hat{\vec{p}}_N, \hat{\vec{q}}_N\right)\Psi\,,$$

where t is time, Ψ is the wave function of the system, and $\hat{\vec{p}}_i$, $\hat{\vec{q}}_i$ is the coordinate and momentum operators of the i-th particle. Unlike classical physics, we are currently unable to follow the trajectories of the particles. However, in principle, by solving the Schrödinger equation with the initial wave function, we can detect the particle system at any time.

Nonetheless, the above procedure is rarely feasible, even when the number of the particles is small. This is due to not knowing the initial conditions with the desired accuracy. In the case of a large number of the particles, any attempts to estimate the multiparticle wave function are doomed to failure as, for example, 1 cm^3 of a typical metal (Cu, Ag, Al) contains 10^{23} electrons.

Instead, it is useful to turn to the statistical properties of an ensemble of identical particle systems differing from each other only in the initial conditions of motion of these particles. In doing so, we should employ the ergodic hypothesis. The statistics of the ensemble is assumed to allow the mean time behavior of the given particle system to be predicted as the latter is in a state of thermal equilibrium with its surroundings (we replace averaging over time by the averaging over the ensemble).

Consider a quantum mechanical system consisting of N weakly interacting particles enclosed in a volume V. The physical nature of the particles is of no interest to us, even if the particles were quanta of electromagnetic field (photons). Suppose the density of the particles are small enough, with good approximation, we can regard the system energy as the sum of the energies of the individual particles.

Under the conditions formulated above, the laws of quantum statistical physics give only two of the most probable distributions of particle in energies.

1. For N identical particles with half integer spin, the average number of the particles with the energy ε_s (and a given spin projection if ε_s is spin independent) at a given temperature T is determined by the *Fermi–Dirac distribution:*

$$f_s = \frac{1}{\exp\left(\frac{\varepsilon_s-\mu}{k_B T}\right)+1}\,. \tag{2}$$

Just as in classical physics, the chemical potential μ can be found from the normalization condition:

$$\sum_s f_s = N\,.$$

It is important that the summation is over all admissible quantum mechanical states of the system, including spin ones. In particular, when spin is 1/2, the energy ε_s does not depend on the spin and each summand is present in the sum twice.

2. An ideal gas of N particles with integer spin. At thermal equilibrium, the average number of bosons with the energy ε_s (and a given spin projection, if ε_s does not depend on the spin) is determined by the *distribution of the Bose–Einstein:*

$$f_s = \frac{1}{\exp\left(\frac{\varepsilon_s - \mu}{k_B T}\right) - 1} \,. \tag{3}$$

The latter differs from the Fermi–Dirac distribution only in the sign before unity in the denominator.

The total number of particles in an ideal Bose gas is given by:

$$\sum_s f_s = N \,.$$

The above formula is responsible for the *chemical potential* μ. The summation is also over all admissible quantum mechanical states, including spin ones.

Through the course of statistical physics, we can see that the chemical potential μ formally introduced by us is related to the second law of thermodynamics. Namely, for systems with a variable number of particles, the second law of thermodynamics is written in the following form:

$$T\mathrm{d}S = \mathrm{d}E + p\mathrm{d}V - \mu\mathrm{d}N \,, \tag{4}$$

where T, S, E, p, V, μ, N is the temperature, entropy, internal energy, pressure, volume, chemical potential, and the number of particles in the system, respectively. Recall that the amount of heat $\delta Q = T\mathrm{d}S$ and the work $\delta A = p\mathrm{d}V$ are not total differentials in the mathematical sense of the word. These are a function of the system process. For the system of particles, according to the laws of thermodynamics, the total differentials are but the increment of entropy $\mathrm{d}S$ and internal energy $\mathrm{d}E$. The term $\mu\mathrm{d}N$ in equation (4) allows for the change in the internal energy of the system due to the change in the number of particles in the system.

Using the second law of thermodynamics (4), it is easy to check that, for given T and V, we have:

$$\mu = \left(\frac{\partial F}{\partial N}\right)_{T,V=\mathrm{const}} \,,$$

where $F = E - TS$ is a function called the *free energy* of the system.

Interestingly, in the limit:

$$\exp\left(\frac{\varepsilon_s - \mu}{k_B T}\right) \gg 1 \,,$$

the quantum Fermi–Dirac and Bose–Einstein distributions merge into the classical Boltzmann distribution:

$$f_s = \exp\left(\frac{\mu - \varepsilon_s}{k_B T}\right) \,.$$

Here, we can see the correspondence principle between more general quantum physics and classical physics limited in their ability.

It is known that the classical Boltzmann distribution is often not acceptable to describe quasiparticles at real temperatures and energies. Ideal gases of quasiparticles obey quantum statistics. This causes the exhibiting of their unusual macroscopic properties (from the point of view of experience of everyday life). It should be emphasized that the case refers to systems of micro- or quasiparticles whose mutual interaction can be neglected.

Our next goal is to understand why the approximation of ideal Fermi and Bose gases describes observable macroscopic properties of condensed matters surprisingly well, consisting of a large number of strongly interacting electrons and ions.

At the first sections of the book we tell about the electronic properties of metals and semiconductors. And only after that we examine the dynamics of a crystal lattice. In our view, such a sequence of exposition allows one to explain the fruitfulness of the concept of quasiparticles and the necessity of utilizing special methods to set forth periodic structures. In doing so, we will resort to relatively simple examples. A distinctive feature of this book is that the author not only selected issues required for the education of a scientist, but also considered gradual approaches for solving the problems at hand. In other words, the author discloses what usually remains "behind the scenes." Most of the problems of quantum macrophysics begins with the application of qualitative methods being the most attractive and beautiful trait of this science. It is about building simple models, which correctly take into account the main interactions and peculiarities of the problems, but at the same time, despite their simplicity, provide exact solutions. The author cannot help touching various estimation techniques, the study of the limiting cases when the smallness of a parameter can be used, and the extraction of corollaries from the symmetry properties of a medium, etc. This book discusses the particular issues through mutually complementary levels of the theoretical description such as quantum, quasiclassical, and phenomenological approaches. All of the foregoing topics make it possible to investigate macroscopic quantum effects more thoroughly, pictorially illustrating the universality and equivalency of the theoretical approaches (in the general area of applicability), and, ultimately, come closer to exploring the problems of the current state of modern science.

1 Electrons and Holes in Metals and Semiconductors

1.1 Electrons in Crystalline Solids: The Formulation of a Simplified Single Particle Model

In condensed matters, the outer shells of atoms overlap. Therefore, the valence electrons of the atoms can move from one atom to another. Such an overlapping explains the high bond energy of crystalline solids and their specific mechanical properties. The inner electron shells of atoms in crystals overlap slightly; they are almost the same as isolated atoms.

It is important to note that inner shell electrons, and even almost free (valence) electrons, interact strongly with each other and with ions of the crystal lattice. The potential energy of the interactions is in the order of the kinetic energy of the electrons. In this case, the theory of crystalline solids, at first glance, seems quite impossible to construct. However, there is currently quite a rigorous description of most phenomena in crystals. The reasons for this success can be explained as follows:

I. It turns out that the theoretical description of a system of strongly interacting electrons and ions may often be reduced to the analysis of a simpler problem. Namely, the analysis of the behavior of noninteracting quasiparticles as being fermions in a weak external field needs to be accomplished. The latter represents the average field of the lattice and the other electrons.

There are two main reasons why the strong interaction of the valence electrons between both themselves and the lattice ions leads to the resultant effect, describable by the weak potential:

a) In the permitted zone, due to their mobility in a crystal, the valence electrons diminish the total potential by acting on an individual electron. Firstly, they can screen the positive charge of ions, thereby reducing the total potential. Secondly, the screening of the electron-electron interaction occurs. Due to Coulomb repulsion, every electron pushes the rest of the other electrons away from itself. Therefore, a region that contains the positive uncompensated charge of fixed lattice ions surrounds such an electron. The cloud of the positive charge moves along with it. Every electron "sees" not only the other electrons but the time dependent polarization cloud created by them. As a result, the effective interaction between the electrons is dynamically screened.

Ultimately, it appears that the valence electrons in a crystal have nothing to do with free electrons in a vacuum. They turn into quasiparticles: fermions. The latter differ in mass from electrons in a vacuum and almost do not interact with each other at distances of interatomic order.

b) The second reason is more compelling and associated with the Pauli principle. The interaction between electrons and ions is strongest at small distances. However, the Pauli principle forbids valence electrons from being near the ions, as this area

https://doi.org/10.1515/9783110586183-001

is already occupied by electrons of the ion core. The Pauli principle also limits the number of collisions between conduction electrons. As a consequence, the conduction electrons rarely suffer scattering by each other.

Given the importance of these reasons, we shall return later to discuss them.

II. Analysis of the behavior of electrons in a crystal is simplified by employing a single particle model. If one picks up a quasiparticle weakly interacting with its surroundings, an electron from the whole system, then describes the effect of all the other particles on it through an effective potential energy, the initial many particle problem is reduced to a system of a single particle Schrödinger equation:

$$-\frac{\hbar^2}{2m}\Delta\Psi + U(\vec{r})\Psi = \varepsilon\Psi \ ,$$

where m is the mass of the quasiparticle, correlated to the electron, Ψ being a stationary wave function of the quasiparticle, and ε being the energy.

It turns out that most observable macroscopic properties and the most important phenomena in solids can be explained by the rational choice of the effective potential $U(\vec{r})$ for a single electron in the Schrödinger equation. Many implications follow from the fact that the average field of the ions and other electrons possess the symmetry properties of the crystal lattice; periodicity, in particular.

So, the periodic potential is assumed to model an impact of the periodic field of the crystal ions on the selected electron, as well as periodic effects caused by the action of the rest of the electrons on the given electron.

It should be emphasized that perfect periodicity of $U(\vec{r})$ is an idealization. Real solids are never absolutely pure. They contain impurities and, in the neighborhood of the impurity atoms, the solid is not the same as elsewhere in the crystal. The real crystals have various types of defects. The ions, in fact, continually undergo thermal vibrations around their equilibrium positions. All the disturbances of the periodicity must be taken into account. For example, based on these disturbances, we can explain why the conductivity of a metal is not infinite. However, to construct the theory, the problem is artificially divided in two parts:

1. At the outset, we should consider a hypothetical perfect crystal with absolute periodicity.
2. Furthermore, we should investigate the effect of all possible deviations from the periodicity, treated as small perturbations, on the properties of the crystal. This is necessary in order to describe the real macroscopic properties of crystal bodies.

III. The periodicity of the potential $U(\vec{r})$ is one of the reasons why the electrons experience no scattering by the lattice of a crystal. In a perfect periodic crystal, the free motion of the electron can be described through the electronic wave function. This is because any linear wave in any periodic structure propagates freely and faces the periodic structure only in special cases. In the last case, reflection/scattering takes place. This will be discussed in more detail later.

The behavior of electrons in the periodic potential can be examined not only in metals. Most of our general conclusions also refer to dielectrics and semiconductors.

To go further in the analysis of the properties of a gas of noninteracting quasiparticles, fermions in the periodic potential, we should get familiar with methods of the theoretical description of the crystal lattices in solids.

1.2 The Theoretical Description of the Periodic Structure of Crystals

The existence of a periodic ion lattice is of fundamental significance for modern solid-state physics. The periodical ion lattice underlies all analytical constructions. If such a lattice did not exist, the theory of crystals would hardly reach serious success. Because of the lack of the periodic arrangement of ions, the theory of amorphous solids is still in its infancy. Achievements in the theory of liquids are also poor, although these two forms of matter have a density close to the density of crystalline solids.

To describe infinite crystal structures, a special symbolic language was created. An ideal crystal can be built up in the space by periodically repeating identical structural units. The simplest crystals have a structural unit consisting of a single atom. In crystals of more complex substances, the structural unit may contain a few atoms or molecules. We will describe the crystal structure in terms of a crystal lattice unit cell periodically recurring in space. It is known as an *elementary cell*, which binds a certain group of atoms. This group of atoms is called a *basis*; the basis is repeated in space and forms the *crystalline structure*.

A Bravais Lattice and a Lattice with a Basis

Although the crystal structure is formed by the repetition of real physical objects (a certain group of atoms or ions), for a theoretical description of the periodic structure of a crystal, the abstract concept of a Bravais lattice is used. The latter, in principle, may have no atoms in its nodes. The Bravais lattice concept reflects only geometry of a regularly distributed array of crystal elements whatever these in fact are.

We give two equivalent definitions of the Bravais lattice.

A Bravais lattice (two-dimensional or three-dimensional) is an infinite periodic structure formed by discrete mathematical points and has the exact same spatial order and orientation when viewed from any lattice point [1]).

A Bravais lattice (two-dimensional or three-dimensional) is a set of mathematical points with the radius vectors of the form:

$$\begin{aligned} \vec{R}^n &= n_1\vec{a}_1 + n_2\vec{a}_2 + n_3\vec{a}_3 \quad \text{(for a three-dimensional space)}, \\ \vec{R}^n &= n_1\vec{a}_1 + n_2\vec{a}_2 \quad \text{(for a plane)}, \end{aligned} \tag{1.1}$$

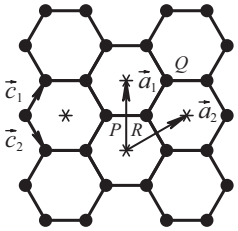


Fig. 1.1: A two-dimensional crystal, formed by atoms located at the vertices of hexagonal "honeycombs".

where $\vec{a}_i$ are any linearly independent vectors and n_i are arbitrary integers. The vectors $\vec{a}_i$ are called *basis vectors* of the Bravais lattice.

To better understand the difference between a real crystal formed by physical objects, atoms, and a Bravais lattice, a two-dimensional crystal rather than a three-dimensional one is easier to start with. For example, the simplest two-dimensional crystal is a layer of atoms adsorbed on the crystal face – a substrate.

Let us consider a two-dimensional crystal formed by atoms located at the vertices of hexagonal "honeycombs" (Figure 1.1). This structure seems to be the same if one looks at it from the points P and Q. However, the views from the point R and the point P differ in a 180 degree turn. It means that, by the first definition, the vertices of the honeycombs do not form the Bravais lattice.

On the other hand, it is easy to check that no two linearly independent vectors connecting the atoms of the crystal, such as $\vec{c}_1$ and $\vec{c}_2$ (Figure 1.1), generate the Bravais lattice. Therefore, by the second definition, the vertices of the two-dimensional honeycombs do not form the Bravais lattice.

In this case, the Bravais lattice is generated by the vectors $\vec{a}_1$ and $\vec{a}_2$ not connecting any real atoms. In Figure 1.1, the asterisks mark the Bravais lattice points and the bold dots indicate real atoms. This example demonstrates the difference between the Bravais lattice whose nodes, in general, contain abstract mathematical points and the natural crystal, the nodes of which are always occupied with atoms or ions.

The term "a Bravais lattice" can be used not only to refer to a set of the points, but also to denote the set of vectors, connecting any of these points with the rest. Any vector $\vec{R}^n$ of (1.1) defines a translation (shift) when the entire aggregation of the system's elements moves as a whole in space at a distance R^n towards the vector $\vec{R}^n$. The translation vectors of the crystal lattice bind different points of the Bravais lattice. In practice, from the problem situation, it is always clear whether there are either points or vectors or translations.

If, being subjected to all translations of the form:

$$\vec{R}^n = n_1\vec{a}_1 + n_2\vec{a}_2 + n_3\vec{a}_3 \quad \text{(for a three-dimensional Bravais lattice)},$$
$$\vec{R}^n = n_1\vec{a}_1 + n_2\vec{a}_2 \quad \text{(for a two-dimensional Bravais lattice)},$$

a volume of space fills up the entire space/plane, never overlaps with itself and leaves no voids, it is referred to as a *primitive (elementary) unit cell*.

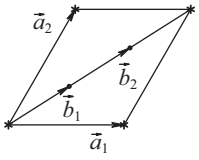


Fig. 1.2: The elementary unit cell of the two-dimensional crystal shown in Figure 1.1; $\vec{b}_2 = 2\vec{b}_1$, $\vec{a}_1 + \vec{a}_2 = 3\vec{b}_1$, $|\vec{a}_1| = |\vec{a}_2|$, the angle between the vectors $\vec{a}_1$ and $\vec{a}_2$ is 60°.

A parallelepiped (parallelogram) built on the vectors $\vec{a}_1, \vec{a}_2, \vec{a}_3$ ($\vec{a}_1, \vec{a}_2$) is always a primitive cell in the three-dimensional space/in the plane. However, such a choice is not the only one possible or the most convenient. We will elaborate on this statement later, but now let us come back to the two-dimensional crystal in the shape of honeycombs.

In this case, such a single unit cell (a parallelogram formed by the vectors $\vec{a}_1$ and $\vec{a}_2$) contains two atoms. The position of any atom in the honeycomb lattice can be written as:

$$\vec{b}_\alpha + \vec{R}^n ,$$

where $\vec{b}_\alpha$ are vectors depicted in Figure 1.2, $\alpha = 1.2$, $\vec{R}^n = n_1\vec{a}_1 + n_2\vec{a}_2$, n_1, n_2 being integers.

This example shows that a physical crystal can be described by giving its underlying Bravais lattice and specifying the arrangement of atoms, molecules, and ions, etc., within a separate primitive unit cell. In order to emphasize the difference between the abstract understanding of points forming the Bravais lattice and the real physical crystal possessing such a lattice, a special term is used: *crystal structure*. When one primitive cell contains several atoms, the crystal structure is always described by the Bravais lattice with a basis. The honeycombs are characterized by a lattice with a basis. The basis atoms of the primitive cell correspond to points with the radius vectors $\vec{b}_1$ and $\vec{b}_2$.

When the primitive cell contains one atom, it can be merged with a node of the Bravais lattice. As a consequence, the situation is simplified due to the merging of the Bravais lattice points and real atom positions.

However, the lattices with a basis should also be used for describing crystal structures whose atoms or ions are located at the points of the Bravais lattice without full translational symmetry. The latter is violated because more than one kind of atom or ion is present. A simple example of that is a NaCl crystal, which is composed of equal numbers of sodium and chlorine ions. These are placed at alternate points of a simple cubic lattice so that each ion of one type is surrounded by six different ions of another type as its nearest neighbors.

An example of a two-dimensional Bravais lattice (Figure 1.3) shows that there is no unique choice of the basis vectors, as well as the primitive cell. This is because the primitive cell can always be built using the basis vectors of the Bravais lattice. If the surface of the primitive cell comprises lattice points (Figure 1.3), all the points except one can be assigned to adjacent cells.

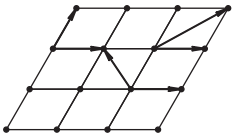


Fig. 1.3: Example of a two-dimensional Bravais lattices without a basis (the lattice nodes are atoms).

In general, each primitive cell has only one point of the Bravais lattice. Therefore, if n is the density of the mathematical points of the Bravais lattice, and V_a is the volume of the unit cell, the following formula holds:

$$V_a n = 1 \quad \text{or} \quad V_a = \frac{1}{n}\,. \tag{1.2}$$

Hence an important conclusion suggests itself. Due to the result (1.2) being valid for any primitive cell, the unit cell volume is independent on the choice of the cell.

It also follows from the definition of a primitive cell, that for two primitive cells of arbitrary shape one of them can be always cut into pieces, which after shifting by the appropriate lattice vectors are join together into another primitive cell.

The Primitive Wigner–Seitz Cell

There are many possible ways of choosing a primitive cell. Choosing a primitive cell in the shape of a parallelepiped has its drawback. The latter is that it does not reflect the full symmetry of the Bravais lattice. We can always choose a primitive cell so that it would possess the symmetry of the Bravais lattice. An example of such a choice is the Wigner–Seitz cell.

The *Wigner–Seitz cell*, centered at some point (a mathematical point of the Bravais lattice), is the region of space that lies closer to that point than to any of the other lattice points.

When being subjected to translations through all lattice vectors, the Wigner–Seitz cell fills up the entire space without any overlapping, i.e., the Wigner–Seitz cell is a primitive cell.

The following simple method offers the construction of the Wigner–Seitz cell. Here are a few steps to do it [6].

1. Choose any point of the Bravais lattice.
2. From the chosen center (the Bravais lattice point is centered), we draw the translation vectors to the nearest lattice nodes.
3. Draw planes perpendicular to these vectors, passing through their centers.

The resulting cell that takes the smallest volume containing the given point bounded by the built planes is the primitive Wigner–Seitz cell.

Note that not all nearest lattice points can be used to construct the Wigner–Seitz cell (Figure 1.4).

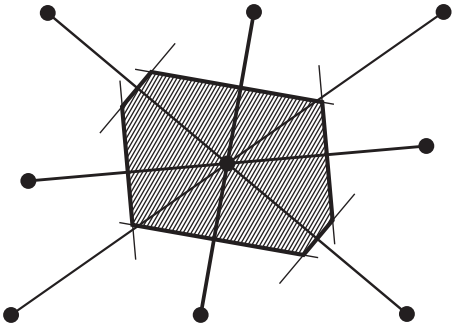


Fig. 1.4: The Wigner–Seitz cell of a two-dimensional Bravais lattice.

All points being equivalent, the Bravais lattice must be infinite in extent. Clearly, actual crystals have finite dimensions. However, if they are large enough, most of the points are so far from the surface as to be unaffected by its existence. The sample is often treated as the finite lattice with N nodes to be the set of points:

$$\vec{R}^n = n_1\vec{a}_1 + n_2\vec{a}_2 + n_3\vec{a}_3 \,,$$

where $0 \le n_1 \le N_1, 0 \le n_2 \le N_2, 0 \le n_3 \le N_3$ and $N = N_1N_2N_3$. Further in this Chapter, we shall justify such a choice of the finite size of the crystal and related boundary conditions.

1.3 A Reciprocal Lattice and the First Brillouin Zone

Let $\vec{R}^n = n_1\vec{a}_1 + n_2\vec{a}_2 + n_3\vec{a}_3$ be the set of points composing the Bravais lattice. Then the set of points corresponding to the wave vectors $\vec{K}$, for which:

$$\exp\left(\mathrm{i}\vec{K}\cdot\vec{R}^n\right) = 1 \,, \tag{1.3}$$

forms a *reciprocal lattice*.

The reciprocal lattice is defined with respect to a particular Bravais lattice. A Bravais lattice, corresponding to the given reciprocal lattice is called a *direct lattice*. If one considers a lattice with a basis, we should apply the reciprocal lattice defined only with regard to the Bravais lattice, leaving the basis vectors aside.

Note that if $\vec{K}_1$ and $\vec{K}_2$ satisfy the relation (1.3), then any sum or difference also meets this relation. Consequently, the reciprocal lattice is itself a Bravais lattice (reciprocal lattice points form a Bravais lattice).

A direct verification shows that the following vectors:

$$\vec{b}_1 = \frac{2\pi\left[\vec{a}_2\times\vec{a}_3\right]}{V_a} \,, \quad \vec{b}_2 = \frac{2\pi\left[\vec{a}_3\times\vec{a}_1\right]}{V_a} \,, \quad \vec{b}_3 = \frac{2\pi\left[\vec{a}_1\times\vec{a}_2\right]}{V_a} \,, \tag{1.4}$$

where $V_a = \vec{a}_1\cdot[\vec{a}_2\times\vec{a}_3]$ is the volume of the elementary unit cell of the direct lattice, generate the three-dimensional crystal reciprocal lattice. In particular, the relation:

$$\vec{b}_i\cdot\vec{a}_j = 2\pi\delta_{ij} \,,$$

where δ_{ij} is the Kronecker symbol, holds true.

Therefore, any reciprocal lattice vector can be represented as:

$$\vec{K} = n_1\vec{b}_1 + n_2\vec{b}_2 + n_3\vec{b}_3 \,, \tag{1.5}$$

where n_i are integers.

It is also not hard to show that if V_a is the volume of the direct lattice cell, the primitive unit cell in reciprocal space has the volume:

$$V_b = \vec{b}_1 \cdot \left[\vec{b}_2 \times \vec{b}_3\right] = (2\pi)^3 / V_a \,. \tag{1.6}$$

The elementary (primitive) Wigner–Seitz cell for the reciprocal lattice is called the *first Brillouin zone* (see Fig. 1.5–1.8). As the name implies, it can be specified as the following Brillouin zones, which are the elementary cells of a different kind. Higher Brillouin zones appear in the theory of electron energy levels in a periodic potential. The terms "the Wigner–Seitz cell" and "the first Brillouin zone" are attributed to the same geometric constructions. The last term is used only to refer to a cell in the reciprocal space.

Fig. 1.5: A one-dimensional direct lattice (a chain of atoms).

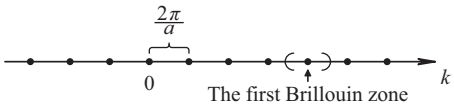


Fig. 1.6: A one-dimensional reciprocal lattice.

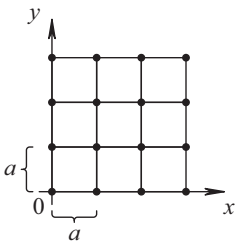


Fig. 1.7: A square direct lattice (atoms are located in the crystal face – a substrate).

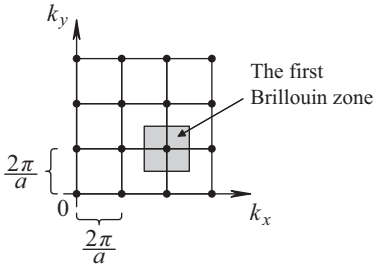


Fig. 1.8: A square reciprocal lattice.

Atomic Planes

There is a close relationship between the vectors of the reciprocal lattice and atomic planes of the direct lattice. A plane passing through mathematical points of a direct Bravais lattice is called an *atomic plane*. *A family of atomic planes* of a lattice is the set of equally spaced parallel atomic planes. We will demonstrate that the reciprocal lattice makes it very easy to classify all possible families of the atomic planes. To do this, we first explain the geometrical meaning of the reciprocal lattice vectors. Let vectors of the form $\vec{R}^m = \vec{a}_1 m_1 + \vec{a}_2 m_2 + \vec{a}_3 m_3$ generate a direct Bravais lattice, m_i are integers; $\vec{a}_1, \vec{a}_2, \vec{a}_3$ being vectors of the basic translations. $\vec{K} = \vec{b}_1 n_1 + \vec{b}_2 n_2 + \vec{b}_3 n_3$ is a fixed vector of the reciprocal lattice, where $\vec{b}_i$ are vectors of the basic translations in reciprocal space. Next, we are interested in only the direction given by the vector $\vec{K}$. Therefore, from all vectors parallel to $\vec{K}$, we choose only that for which the numbers n_1, n_2, n_3 contain no common divisors and give the smallest length of the vector $\vec{K}$.

The vectors $\{\vec{R}_N\}$ define a set of points of the direct lattice, which are solutions of the equation:

$$\vec{R}_N \cdot \vec{K} = 2\pi(m_1 n_1 + m_2 n_2 + m_3 n_3) = 2\pi N = \text{const}\,, \tag{1.7}$$

where N is an integer. For given N and $\vec{K}$, the number m_i runs a certain set of values. From analytic geometry it is known that equation (1.7) defines a plane perpendicular to the vector $\vec{K}$. In our case, this equation is that of the atomic plane passing through the nodes of the direct lattice with a normal line parallel to the vector $\vec{K}$.

Taking into account that:

$$\vec{A} \cdot \vec{B} = \left|\vec{A}\right| \left|\vec{B}\right| \cos(\vec{A}\vec{B}) = \left|\vec{A}\right| \left(\mathrm{Pr}_{\vec{B}}\, \vec{A}\right) = \left|\vec{B}\right| \left(\mathrm{Pr}_{\vec{A}}\, \vec{B}\right)\,,$$

the atomic plane equation can be rewritten as:

$$\left(\mathrm{Pr}_{\vec{K}}\, \vec{R}_N\right) \left|\vec{K}\right| = 2\pi N\,, \tag{1.8}$$

where N is an integer and $\mathrm{Pr}_{\vec{K}}\, \vec{R}_N$ is the projection of the vector $\vec{R}_N$ onto the direction of the vector $\vec{K}$. The nearest integer to N is either $(N + 1)$ or $(N - 1)$. For the same $\vec{K}$, the equation $\vec{R}_{N+1} \cdot \vec{K} = 2\pi(N + 1)$ determines the atomic plane in the direct lattice. The atomic plane is parallel to the original plane and is at a minimum distance from it. The equation of this plane can be written as:

$$\left(\mathrm{Pr}_{\vec{K}}\, \vec{R}_{N+1}\right) \left|\vec{K}\right| = 2\pi(N + 1)\,. \tag{1.9}$$

Let us find the distance between the nearest atomic planes perpendicular to the vector $\vec{K}$ (Figure 1.9):

$$d = \mathrm{Pr}_{\vec{K}}\, \vec{R}_{N+1} - \mathrm{Pr}_{\vec{K}}\, \vec{R}_N = 2\pi / \left|\vec{K}\right|\,. \tag{1.10}$$

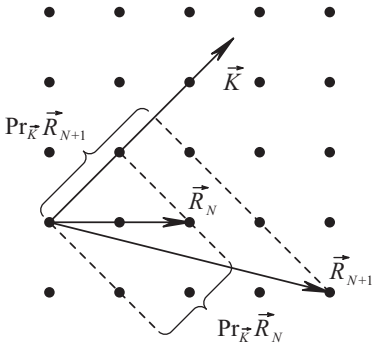


Fig. 1.9: Atomic lines of a two-dimensional lattice.

General Conclusions

1. Any vector $\vec{K}$ of the reciprocal lattice defines a family of atomic planes perpendicular to it.
2. The minimum length of the vectors parallel to the vector $\vec{K}$ is the distance between the nearest planes of the appropriate family of atomic planes (1.10).

The correspondence between the reciprocal lattice vectors and the families of atomic planes gives a convenient way to specify the orientation of the atomic plane. The plane orientation is given by the normal vector to the plane. It is natural to choose the reciprocal lattice vector $\vec{K}$ as a normal to the atomic plane. Out of all vectors parallel to vector $\vec{K}$, the shortest vector should be selected. By doing so, the Miller indices of the plane can be estimated.

Coordinates of the shortest reciprocal lattice vector, perpendicular to the family of atomic planes of the direct lattice in the coordinate system specified by the basic vectors of the reciprocal lattice, are called the *Miller indices*.

If $\vec{K} = \vec{b}_1 n_1 + \vec{b}_2 n_2 + \vec{b}_3 n_3$ is the shortest vector, then n_1, n_2, n_3 are the Miller indices. Therefore, the Miller indices defined must be integers. This is because any reciprocal lattice vector is a linear combination of three basic vectors taken with integer coefficients. The Miller indices depend on the choice of the basic vectors. The set of the indices (n_1, n_2, n_3) can mean a single plane or a family of parallel planes.

1.4 Energy Levels of an Electron in a Periodic Potential and Bloch's Theorem

Consider an infinite crystal periodically translated by vectors of the form $\vec{R}^l = \vec{a}_1 l_1 + \vec{a}_2 l_2 + \vec{a}_3 l_3$, where l_i are integers and $\vec{a}_i$ are the vectors of the basic translations $(i = 1, 2, 3)$.

Bloch's Theorem (a First Formulation)

If $V(\vec{r} + \vec{R}^l) = V(\vec{r})$ is the periodic potential for all $\vec{R}^l$ of the Bravais lattice, then the eigenstates of the Schrödinger equation

$$\left[-\frac{\hbar^2}{2m}\Delta + V(\vec{r})\right]\Psi = \varepsilon\Psi$$

may be selected so that the wave functions should have the form of the plane wave $\exp(\mathrm{i}\vec{k}\cdot\vec{r})$, multiplied by a function with periodicity of the Bravais lattice:

$$\Psi_{n\vec{k}}(\vec{r}) = u_{n\vec{k}}(\vec{r})\exp\left(\mathrm{i}\vec{k}\cdot\vec{r}\right) . \tag{1.11}$$

Here

$$u_{n\vec{k}}\left(\vec{r} + \vec{R}^l\right) = u_{n\vec{k}}(\vec{r}) \tag{1.12}$$

for any $\vec{R}^l$ belonging to the Bravais lattice, n is a natural number, with the energy eigenvalues being $\varepsilon = \varepsilon_n(\vec{k})$.

The relations (1.11) and (1.12) can be written as follows:

$$\Psi_{n\vec{k}}\left(\vec{r} + \vec{R}^l\right) = \Psi_{n\vec{k}}(\vec{r})\exp\left(\mathrm{i}\vec{k}\cdot\vec{R}^l\right) . \tag{1.13}$$

Therefore, the Bloch theorem can be formulated in an alternative way.

Bloch's Theorem (a Second Formulation)

The eigenstates of the operator

$$\widehat{H} = -\frac{\hbar^2}{2m}\Delta + V(\vec{r})$$

can be chosen in such a way that each of them should be associated with a wave vector $\vec{k}$, and for any $\vec{R}^l$ of the Bravais lattice, the equality should be fulfilled:

$$\Psi_{n\vec{k}}\left(\vec{r} + \vec{R}^l\right) = \Psi_{n\vec{k}}(\vec{r})\exp\left(\mathrm{i}\vec{k}\cdot\vec{R}^l\right) .$$

Let us explain this further. The theorem states that, although the potential in the single particle Schrödinger equation is periodic, the eigenwave functions, generally speaking, are not periodic. However, they have an interesting structure: they are similar to the plane wave $\exp(\mathrm{i}\vec{k}\cdot\vec{r})$ describing a free particle with the wave vector $\vec{k}$.

General Conclusions

1. Bloch's theorem introduces a wave vector $\vec{k}$, which is an analogue of the wave vector of the free microparticle in the general problem of motion in a periodic potential.

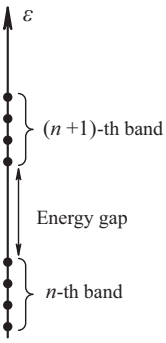


Fig. 1.10: Schematic depiction of the permitted energy values $\varepsilon_n(k)$.

2. The periodic function $u_{n\vec{k}}(\vec{r})$ with the index n modulates any wave $\exp(\mathrm{i}\vec{k}\cdot\vec{r})$. As will be shown below, the Schrödinger equation has a set of discrete solutions for each fixed $\vec{k}$. The index n enumerates these states and is said to be a *band number*. The name is due to the fact that for a fixed n the particle's eigenenergies $\varepsilon_n(k)$, depending on $\vec{k}$, are densely clustered along the energy axis. In doing so, they form a band separated from an n+1 neighboring band by the forbidden energy region; an energy gap (Figure 1.10). Therefore, the eigenwave functions and eigenvalues of the particle's energies in a periodic potential have two indices. One of those is the particle's wave vector $\vec{k}$, another is the number n of the energy band.

The Proof of the Bloch Theorem

In this example, $\widehat{T}r_{R^l}$ is the translation operator for each vector belonging to a Bravais lattice; which shifts the argument of any function $f(\vec{r})$ by $\vec{R}^l$:

$$\widehat{T}_{\vec{R}^l} f(\vec{r}) = f\left(\vec{r} + \vec{R}^l\right) . \tag{1.14}$$

Here, we show how the operators $\widehat{T}_{\vec{R}^l}$ and $\widehat{H} = -\hbar^2/(2m)\Delta + V(\vec{r})$ commute. By virtue of the periodicity of the Hamiltonian system, we have:

$$\widehat{T}_{\vec{R}^l}\left(\widehat{H}\Psi\right) = \widehat{H}\left(\vec{r} + \vec{R}^l\right)\Psi\left(\vec{r} + \vec{R}^l\right) = \widehat{H}(\vec{r})\Psi\left(\vec{r} + \vec{R}^l\right) = \widehat{H}(\vec{r})\widehat{T}_{\vec{R}^l}\Psi(\vec{r}) . \tag{1.15}$$

Since the equation (1.15) is satisfied identically for each function Ψ, the following operator identity holds true:

$$\widehat{T}_{\vec{R}^l}\widehat{H} = \widehat{H}\widehat{T}_{\vec{R}^l} . \tag{1.16}$$

Furthermore, we note that the result of two successive translations never depends on the order of their application, since for any function $\Psi(\vec{r})$, we have:

$$\widehat{T}_{\vec{R}^{l'}}\widehat{T}_{\vec{R}^l}\Psi = \widehat{T}_{\vec{R}^l}\widehat{T}_{\vec{R}^{l'}}\Psi = \Psi\left(\vec{r} + \vec{R}^l + \vec{R}^{l'}\right) = \widehat{T}_{\vec{R}^l + \vec{R}^{l'}}\Psi . \tag{1.17}$$

Therefore, we obtain:

$$\widehat{T}_{\vec{R}^{l'}}\widehat{T}_{\vec{R}^l} = \widehat{T}_{\vec{R}^l}\widehat{T}_{\vec{R}^{l'}} = \widehat{T}_{\vec{R}^l + \vec{R}^{l'}} . \tag{1.18}$$

The relations (1.16) and (1.18) show that the Hamiltonian $\widehat{H}$ and the operators $\widehat{T}_{\vec{R}}$ for all Bravais lattice vectors $\vec{R}$ form a set of commuting operators. According to the commuting operator theorem, the operators $\widehat{H}$ and $\widehat{T}_{\vec{R}}$ have a common system of the eigenfunctions:

$$\widehat{H}\Psi = \varepsilon\Psi\,, \quad \widehat{T}_{\vec{R}}\Psi = c(\vec{R})\Psi\,. \tag{1.19}$$

From (1.18) (1.19), we have:

$$\widehat{T}_{\vec{R}^l}\left(\widehat{T}_{\vec{R}^{l'}}\Psi\right) = \widehat{T}_{\vec{R}^l}\left(c(\vec{R}^{\vec{l}'})\Psi\right) = c(\vec{R}^{\vec{l}'})\widehat{T}_{\vec{R}^l}\Psi = c(\vec{R}^{\vec{l}'})c(\vec{R}^l)\Psi = c\left(\vec{R}^l + \vec{R}^{\vec{l}'}\right)\Psi\,. \tag{1.20}$$

Consequently, for the translation operator eigenvalues, the equality must be fulfilled:

$$c\left(\vec{R}^l + \vec{R}^{\vec{l}'}\right) = c(\vec{R}^{\vec{l}'})c(\vec{R}^l)\,. \tag{1.21}$$

where $\vec{R}^l, \vec{R}^{\vec{l}'}$ are arbitrary vectors belonging to the Bravais lattice.

As a translation vector, we take an arbitrary s-th basic vector of the direct Bravais lattice. Then, we can always write $c(\vec{a}_s)$ as:

$$c(\vec{a}_s) = \exp\left(2\pi\mathrm{i}x_s\right)\,, \tag{1.22}$$

choosing the values of x_s ($s = 1, 2, 3$) in an appropriate way.

By substituting the expression for the Bravais lattice translation vector $\vec{R}^l = \vec{a}_1 l_1 + \vec{a}_2 l_2 + \vec{a}_3 l_3$ into the formula for $c(\vec{R}^l)$, we come up with: $c(\vec{R}^l) = c(\vec{a}_1 l_1 + \vec{a}_2 l_2 + \vec{a}_3 l_3)$. Applying formula (1.21) to the above expression, we find:

$$c(\vec{R}^l) = \left[c(\vec{a}_1)\right]^{l_1}\left[c(\vec{a}_2)\right]^{l_2}\left[c(\vec{a}_3)\right]^{l_3}\,,$$

or taking the presentation (1.22) into account:

$$c(\vec{R}^l) = \exp\left[2\pi\mathrm{i}(x_1 l_1 + x_2 l_2 + x_3 l_3)\right] \equiv \exp\left(\mathrm{i}\vec{k}\cdot\vec{R}^l\right)\,. \tag{1.23}$$

Here $\vec{k} = x_1\vec{b}_1 + x_2\vec{b}_2 + x_3\vec{b}_3$, $\vec{R}^l = \vec{a}_1 l_1 + \vec{a}_2 l_2 + \vec{a}_3 l_3$. In writing formula (1.23), we have used the fact that for the vectors of the direct and inverse lattices satisfy the orthogonality condition:

$$\vec{a}_s \cdot \vec{b}_p = 2\pi\delta_{sp}\,.$$

Thus we have shown that the eigenwave functions Ψ of the Hamiltonian $\widehat{H}$ may be selected so that the equality

$$\widehat{T}_{\vec{R}^l}\Psi(r) = \Psi(\vec{r} + \vec{R}^l) = c(\vec{R}^l)\Psi(\vec{r}) = \exp\left(\mathrm{i}\vec{k}\cdot\vec{R}^l\right)\Psi(\vec{r}) \tag{1.24}$$

should be performed for each vector $\vec{R}^l$ of the Bravais lattice. This is the Bloch theorem.

The wave function must remain bounded under arbitrary translations in the infinite crystal, so the vector $\vec{k}$ of the representation (1.24) must be real.

Born–Karman Boundary Conditions

By imposing appropriate boundary conditions on the wave functions, we can show that the wave vector $\vec{k}$ must be real in a finite sample as well. For a sample of large size, its innerphysical properties are assumed to be independent neither on the shape nor the boundary conditions at its surface. Therefore, for the convenience of calculations, the sample (crystal) should be chosen in the form of a parallelepiped built on the vectors $N_1\vec{a}_1, N_2\vec{a}_2, N_3\vec{a}_3$. As with boundary conditions, the simplest ones need to be used. Born and Karman were first to propose them:

$$\Psi_{n\vec{k}}\left(\vec{r} + N_s\vec{a}_s\right) = \Psi_{n\vec{k}}(\vec{r}) \,, \tag{1.25}$$

where $\vec{a}_s$ is the trio of the basis vectors of the crystal's Bravais lattice, N_s are positive integers,$(N_s \gg 1)$, $s = 1, 2, 3$, $N = N_1N_2N_3$ is the total number of elementary unit cells in the crystal. In this section, there is no summation over twice repeated indices.

According to the *Born–Karman boundary conditions*, an electron reaches the boundary of the sample without reflecting but "disappears" and, at the same time, reenters the opposite face of the crystal.

We demonstrate that the boundary conditions (1.25) are consistent with the Bloch theorem. We apply the Bloch theorem (1.13) to the boundary condition (1.25):

$$\Psi_{n\vec{k}}(\vec{r} + N_s\vec{a}_s) = \exp\left(\mathrm{i}N_s\vec{a}_s \cdot \vec{k}\right)\Psi_{n\vec{k}}(\vec{r}) = \Psi_{n\vec{k}}(\vec{r}) \,. \tag{1.26}$$

Then we have:

$$\exp\left(\mathrm{i}N_s\vec{a}_s \cdot \vec{k}\right) = 1 \,. \tag{1.27}$$

Given $\vec{k} = x_1\vec{b}_1 + x_2\vec{b}_2 + x_3\vec{b}_3$, $\vec{a}_s \cdot \vec{b}_p = 2\pi\delta_{sp}$, from (1.27), we obtain:

$$\exp\left[2\pi\mathrm{i}N_s x_s\right] = 1 \,.$$

So, the conditions must be met:

$$x_s = \frac{m_s}{N_s} \,.$$

Here, m_s are arbitrary integers. Therefore, the allowed Bloch wave vectors are real and have the form:

$$\vec{k} = \sum_{i=1}^{3} \frac{m_i}{N_i}\vec{b}_i \,. \tag{1.28}$$

From (1.28) it follows that any $\vec{k}$-space volume corresponding to each allowed wave vector $\vec{k}$ is equal to the volume of a small parallelepiped with the edges $\vec{b}_i/N_i$:

$$\delta V_{\vec{k}} = \frac{\vec{b}_1}{N_1} \cdot \left[\frac{\vec{b}_2}{N_2} \times \frac{\vec{b}_3}{N_3}\right] = \frac{V_b}{N} \,. \tag{1.29}$$

Since $V_b = \vec{b}_1 \cdot (\vec{b}_2 \times \vec{b}_3)$ is the volume of the elementary unit cell of the reciprocal lattice, formula (1.29) means that the amounts of allowed wave vectors in one unit cell and of cells in the crystal (N) are equal.

In the direct and reciprocal lattices, the volumes of the unit cells are related by:

$$V_b = \frac{(2\pi)^3}{V_a} \,. \tag{1.30}$$

Therefore, formula (1.29) can be rewritten as:

$$\delta V_{\vec{k}} = \frac{V_b}{N} = \frac{(2\pi)^3}{V_a N} = \frac{(2\pi)^3}{V} \,, \tag{1.31}$$

where $V = NV_a$ is the volume of the crystal.

Thus, the greater the volume of the crystal V is, the closer the allowed wave vectors in k-space are to each other.

General Remarks about Bloch's Theorem

1. We have shown that the stationary Schrödinger equation with the periodic potential $V(\vec{r})$:

$$-\frac{\hbar^2}{2m}\Delta\Psi + V(\vec{r})\Psi = \varepsilon\Psi \tag{1.32}$$

has the solution:

$$\Psi_{n\vec{k}}(\vec{r}) = u_{n\vec{k}}(\vec{r}) \exp\left(\mathrm{i}\vec{k}\cdot\vec{r}\right) \,,$$

where $u_{n\vec{k}}(\vec{r} + \vec{R}) = u_{n\vec{k}}(\vec{r})$, $\vec{R}$ is the vector of the crystal's Bravais lattice.

Now we can also write the nonstationary Schrödinger equation:

$$-\frac{\hbar^2}{2m}\Delta\Psi + V(\vec{r})\Psi = \mathrm{i}\hbar\frac{\partial\Psi}{\partial t} \tag{1.33}$$

the solution of which is:

$$\Psi(\vec{r}, t) = \sum_{\vec{k}'} g(\vec{k}') u_{n\vec{k}'}(\vec{r}) \exp\left[\mathrm{i}\vec{k}'\cdot\vec{r} - \frac{\mathrm{i}}{\hbar}\varepsilon_n(\vec{k}')t\right] \,. \tag{1.34}$$

Each electron is described by a superposition of the wave functions (1.34) which form the wave packet:

$$\Psi(\vec{r}, t) = \sum_{\vec{k}'} g(\vec{k}') u_{n\vec{k}'}(\vec{r}) \exp\left[\mathrm{i}\vec{k}'\cdot\vec{r} - \frac{\mathrm{i}}{\hbar}\varepsilon_n(\vec{k}')t\right] \,. \tag{1.35}$$

In formula (1.35), all the allowed wave vectors $\vec{k}'$ reside near the average value of $\vec{k}$:

$$g(\vec{k}') \approx 0 \quad \text{if} \quad \left|\vec{k}' - \vec{k}\right| \geq \Delta k \,.$$

It is known from quantum mechanics that, when localized in k-space in a region with a characteristic size of the order Δk, a wave packet is localized in a normal coordinate space in a region size of:

$$\Delta R \sim \frac{1}{\Delta k} \,. \tag{1.36}$$

This region moves as a whole with a group velocity, which is calculated by the formula:

$$\vec{V} = \frac{1}{\hbar}\frac{\partial \varepsilon_n(\vec{k})}{\partial \vec{k}} \,. \tag{1.37}$$

The velocity (1.37) is given by the gradient $\varepsilon_n(\vec{k})$ of the dispersion relation at the point with the radius vector $\vec{k}$. This is because the harmonic components of the wave packet have the wave vectors close to the vector $\vec{k}$.

We have come up with an interesting outcome. It turns out that, in the periodic crystal, the electron has a steady state of being in which it moves with a constant velocity without changing its energy. Also, we cannot regard collisions between electrons and fixed ions as a mechanism for governing the change in the electron's velocity, since equation (1.33) (the solution of which is the function (1.34)) completely takes into account the electron-lattice interaction. Such a result is incompatible with classical ideas; it is a consequence of the wave nature of electrons.

In the periodic lattice of scattering centers, the wave propagates without damping due to the interference of scattered waves. Since the resultant wave (1.35) corresponding to the electron in the crystal does not decay, the conductivity of an ideal crystal should be infinite. The cause of electrical resistance of metals is crystal lattice defects or impurities or lattice ion vibrations. All of them disrupt the periodicity of the potential energy of the electron.

2. Bloch's theorem introduces the wave vector $\vec{k}$ into theory. The quantity $\hbar\vec{k}$ is often called the *quasimomentum of an electron*. For a free electron, the quasimomentum and momentum are related: $\vec{p} = \hbar\vec{k}$. However, this formula does not hold for an electron in a crystal: $\vec{p} \neq \hbar\vec{k}$. This is clear from the following considerations. The existence of the connections $\varepsilon = \varepsilon_n(\vec{k})$, $\vec{p} = \hbar\vec{k}$ would mean that the energy and momentum of the electron can be measured simultaneously. In other words, the energy and momentum operators, corresponding to these values commute: $[\widehat{H}, \widehat{\vec{p}}] = 0$. In turn, the condition $[\widehat{H}, \widehat{\vec{p}}] = 0$ would mean that the momentum of the electron is an integral of motion. This is impossible, since the momentum is related to the invariance of the electron Hamiltonian $\widehat{H}$ with respect to arbitrary translations. Under the discontinuous periodic potential, the electron Hamiltonian $\widehat{H}$ does not possess full translational invariance, despite the preservation of its invariance with respect to discrete translations, reflecting the symmetry of the crystal lattice. We have arrived at a contradiction because our initial assumption of the connection $\vec{p} = \hbar\vec{k}$ is incorrect. For an electron in a crystal this connection is $\vec{p} \neq \hbar\vec{k}$.

By direct verification, we can see that, for the electron in an inhomogeneous crystal field, the following stands: $\vec{p} \neq \hbar\vec{k}$. We act by the momentum operator $\widehat{\vec{p}} = -\mathrm{i}\hbar\vec{\nabla}$ on the Bloch electron wave function:

$$\Psi_{n\vec{k}} = u_{n\vec{k}}(\vec{r}) \exp\left(\mathrm{i}\vec{k}\cdot\vec{r}\right) \,.$$

As a result, we get:

$$\widehat{\vec{p}}\Psi_{n\vec{k}} = \hbar\vec{k}\Psi_{n\vec{k}} - \mathrm{i}\hbar \exp\left(\mathrm{i}\vec{k}\cdot\vec{r}\right)\vec{\nabla}u_{n\vec{k}} \neq \hbar\vec{k}\Psi_{n\vec{k}} \,.$$

This means that the quasimomentum of the electron in a crystal has nothing in common with its momentum: $\vec{p} \neq \hbar\vec{k}$.

To comprehend the meaning of the term "quasimomentum," the reaction of Bloch electrons to an external electromagnetic field should be considered. We will come back to this issue later.

3. We can always choose the wave vector $\vec{k}$ of the Bloch function so that it should lie within the first Brillouin zone. If the wave vector $\vec{k}'$ does not fall into the first Brillouin zone, it can always be represented as the sum of the wave vector $\vec{k}$ of the first Brillouin zone and any reciprocal lattice vector $\vec{K}$:

$$\vec{k}' = \vec{k} + \vec{K} ,$$

where $\vec{K} = \vec{b}_1 n_1 + \vec{b}_2 n_2 + \vec{b}_3 n_3$.

As an example, consider a square lattice. In Figure 1.11, the vector $\vec{k}'$ lies outside the first Brillouin zone. The vector $\vec{k}'$ can be paired with the vector $\vec{k}$ being inside the first Brillouin zone if one subtracts the reciprocal lattice vector $\vec{K}$ from $\vec{k}'$. The wave vector of the point A at the zone boundary can be translated to the point A' on the opposite edge of the same zone by adding some vector $\vec{K}$. Then the question arises of whether we can think that both points A and A' belong to the first zone? The answer is that they are said to be identical, so we consider only one of them.

Suppose we have the Bloch function with the wave vector $\vec{k}'$ outside the first Brillouin zone:

$$\Psi_{n\vec{k}'}(\vec{r}) = u_{n\vec{k}'}(\vec{r}) \exp\left(\mathrm{i}\vec{k}' \cdot \vec{r}\right) . \tag{1.38}$$

We replace $\vec{k}'$ by $(\vec{k} + \vec{K})$ and rewrite formula (1.38) in the following form:

$$\Psi_{n\vec{k}'}(\vec{r}) = u_{n\vec{k}'}(\vec{r}) \exp\left(\mathrm{i}\vec{k}' \cdot \vec{r}\right) = \exp\left(\mathrm{i}\vec{k} \cdot \vec{r}\right) \left[\exp\left(\mathrm{i}\vec{K} \cdot \vec{r}\right) u_{n\vec{k}'}(\vec{r})\right] = \exp\left(\mathrm{i}\vec{k} \cdot \vec{r}\right) u_{n\vec{k}}(\vec{r}) . \tag{1.39}$$

Here we have introduced the designation $u_{n\vec{k}}(\vec{r}) = u_{n\vec{k}'}(\vec{r}) \exp(\mathrm{i}\vec{K} \cdot \vec{r})$.

By virtue of the relation $\vec{K} \cdot \vec{R} = 2\pi m$ (m is an integer), the function $\exp(\mathrm{i}\vec{K} \cdot \vec{r})$ as well as $u_{n\vec{k}'}(r)$ is a periodic function in the Bravais lattice. So, $u_{n\vec{k}}(r)$ is periodic too:

$$u_{n\vec{k}}(\vec{r} + \vec{R}) = u_{n\vec{k}}(\vec{r}) .$$

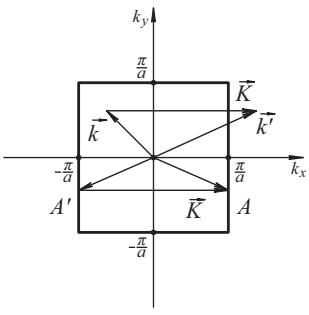


Fig. 1.11: The first Brillouin zone for a plane square lattice.

Instead of the original Bloch function with the wave vector $\vec{k}'$, we have obtained an equivalent function with the wave vector $\vec{k}$ in the first Brillouin zone:

$$\Psi_{n\vec{k}} = u_{n\vec{k}}(\vec{r}) \exp\left(\mathrm{i}\vec{k}\cdot\vec{r}\right) , \quad u_{n\vec{k}}(\vec{r}+\vec{R}) = u_{n\vec{k}}(\vec{r}) .$$

4. Now we answer the question why is the zone number discrete?

The solution of the Schrödinger equation (1.32) needs to be sought in the form of Bloch's wave:

$$\Psi_{\vec{k}}(\vec{r}) = u_{\vec{k}}(\vec{r}) \exp\left(\mathrm{i}\vec{k}\cdot\vec{r}\right) .$$

Substitute the expression directly into equation (1.32) and we find that the function $u_{\vec{k}}(\vec{r})$ is a solution to the eigenvalue problem:

$$\widehat{H}_{\vec{k}} u_{\vec{k}} \equiv \left[\frac{\hbar^2}{2m}(-\mathrm{i}\vec{\nabla}+\vec{k})^2 + V(\vec{r})\right] u_{\vec{k}}(\vec{r}) = \varepsilon_{\vec{k}} u_{\vec{k}}(\vec{r}) , \tag{1.40}$$

with the boundary condition:

$$u_{\vec{k}}(\vec{r}+\vec{R}) = u_{\vec{k}}(\vec{r}) .$$

The operator $\widehat{H}_{\vec{k}}$ in (1.40) is Hermitian: $\widehat{H}^{+}_{\vec{k}} = \widehat{H}_{\vec{k}}$. The superscript "+" indicates the Hermitian conjugate operation.

Due to the periodic boundary conditions, equation (1.40) can be solved within a single elementary unit cell of the direct lattice. Thus, we have the eigenvalue problem for a fixed finite volume. Any such problems have a discrete spectrum. This explains the origin of the index n of the function $\varepsilon_n(\vec{k})$ and $u_{n\vec{k}}(\vec{r})$ ($n = 1, 2, \ldots$).

With the operator $\widehat{H}_{\vec{k}}$ being Hermitian, its eigenvalues $\varepsilon = \varepsilon_n(\vec{k})$ are real. The eigenvalue problem (1.40) contains the wave vector $\vec{k}$ only as a parameter. So, the function $\varepsilon_n(\vec{k})$ is assumed to depend continuously on $\vec{k}$. Recall that the wave vectors of the type $\vec{k} = \sum_{i=1}^{3} m_i/N_i \vec{b}_i$ can take different values under the Born–Karman boundary conditions if the size of the crystal changes. In the limit of the infinite crystal, these values involve a dense set in the $\vec{k}$- space.

5. Since the vectors $\vec{k}$ and $(\vec{k}+\vec{K})$ are physically equivalent, the energy $\varepsilon_n(\vec{k})$ must be a periodic function in reciprocal space:

$$\varepsilon_n(\vec{k}+\vec{K}) = \varepsilon_n(\vec{k}) ,$$

with the periods $\vec{b}_1, \vec{b}_2, \vec{b}_3$ of the basic translations. Recall that any reciprocal lattice vector has the form $\vec{K} = n_1\vec{b}_1 + n_2\vec{b}_2 + n_3\vec{b}_3$, where n_1, n_2, n_3 are integers.

Thus, we come to the description of the energy levels of the electron through a family of the continuous functions of $\varepsilon_n(\vec{k})$, each of which has a periodicity of the reciprocal lattice. These functions determine the *band structure* of solids.

Each function of $\varepsilon_n(\vec{k})$ is periodic and continuous in $\vec{k}$. Consequently, it has upper and lower limits and, therefore, all the energy levels of $\varepsilon_n(\vec{k})$ for a given n lie between these two limits. Zones with different values of n can be separated by energy gaps but may overlap.

6. The Bloch function:

$$\Psi_{n\vec{k}}(\vec{r}) = u_{n\vec{k}}(\vec{r}) \exp\left(\mathrm{i}\vec{k}\cdot\vec{r}\right)$$

is the solution of the Schrödinger equation:

$$\widehat{H}\Psi_{n\vec{k}} = \varepsilon_n(\vec{k})\Psi_{n\vec{k}}\,, \quad \widehat{H} = -\frac{\hbar^2}{2m}\nabla + V(\vec{r})\,. \tag{1.41}$$

Since the eigenvalues of ε_n ($\varepsilon_n^* = \varepsilon_n$) and the operator $\widehat{H}$ are real ($\widehat{H}^* = \widehat{H}$), from (1.41) we have:

$$\widehat{H}\Psi^*_{n\vec{k}} = \varepsilon_n(\vec{k})\Psi^*_{n\vec{k}}\,.$$

Hereinafter, the symbol "*" denotes the operation of complex conjugation. The Bloch function determines the wave vector $\vec{k}$:

$$\Psi_{n\vec{k}}(\vec{r}) = u_{n\vec{k}}(\vec{r}) \exp\left(\mathrm{i}\vec{k}\cdot\vec{r}\right)\,.$$

Let us examine the complex conjugate of this function:

$$\Psi^*_{n\vec{k}} = u^*_{n\vec{k}} \exp\left(-\mathrm{i}\vec{k}\cdot\vec{r}\right)\,.$$

It is clear that, to the wave function $\Psi^*_{n\vec{k}}$, there corresponds the wave vector $(-\vec{k})$. Therefore, the following chain of equalities holds true:

$$\widehat{H}\Psi^*_{n\vec{k}} = \varepsilon_n(\vec{k})\Psi^*_{n\vec{k}} = \varepsilon_n(-\vec{k})\Psi^*_{n\vec{k}}\,.$$

Consequently, the eigenvalues of the electron energy operator are invariant under the replacement $\vec{k} \to -\vec{k}$:

$$\varepsilon_n(\vec{k}) = \varepsilon_n(-\vec{k})\,.$$

7. Crystalline lattices of metals have considerable symmetry, owing to which extremes of the functions $\varepsilon_n(\vec{k})$ are at the boundaries of the Brillouin zone. For lattices of special symmetry there is also an additional extremum at the center of the Brillouin zone. We can make sure of this on particular examples.

The Fermi Surface

The ground state of a system of Bloch electrons in a crystal can be constructed according to the following principles:

1. The principle of minimum energy (electrons fill up allowed energy levels successively, starting with the lower available level).
2. The Pauli Principle.

Single-electron energy levels of $\varepsilon_n(\vec{k})$ depend on two quantum numbers n and $\vec{k}$ and meet the condition: $\varepsilon_n(\vec{k}) = \varepsilon_n(\vec{k} + \vec{K})$. Wave vectors that differ by a reciprocal lattice

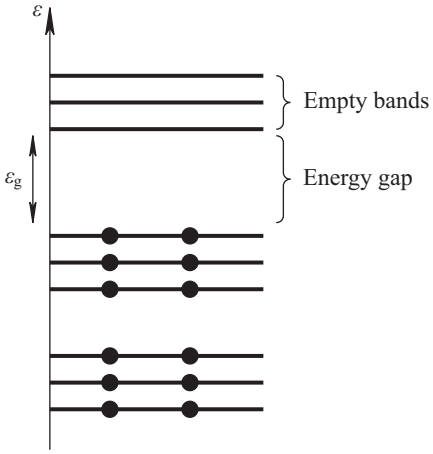


Fig. 1.12: The distribution of the available electrons among energy levels. The bands, empty and fully filled up with the electrons, are depicted here. The electrons are denoted by bold dots.

vector are physically equivalent and correspond to the same energy value. Therefore, when taking each energy level into account only once, we must restrict the values $\vec{k}$ to the first Brillouin zone. In addition, the Pauli exclusion principle reads that no more than two electrons may occupy the same energy level of $\varepsilon_n(\vec{k})$ simultaneously. To occupy the same energy level, two electrons with quantum numbers n, $\vec{k}$ must have opposite spins. Consequently, this stands for different quantum states.

By virtue of the above principles, the electrons distributed in the energy levels can form the following configurations:

1. Some bands may appear to be completely filled with electrons and the remaining may appear to be empty. (Figure 1.12). The energy difference ε_g between the "top" of the highest occupied band and the "bottom" of the lowest empty band is referred to as a *forbidden band* or an *energy gap*.

A solid satisfying the condition $\varepsilon_g \gg k_B T$ is an insulator. However, if $\varepsilon_g \approx k_B T$, it is an intrinsic semiconductor.

The number of levels in the band is equal to the number of elementary cells (N) in a crystal. Every level has two electrons. Consequently, the maximum number of energetic states in every band is $2N$. To obtain a semiconductor or an insulator, it is necessary that the number of electrons per unit cell should be even. Otherwise, not all quantum mechanical states will be filled up with electrons in the upper of the bands. This is a necessary condition for semiconductors and dielectrics to be formed, however, it is not sufficient due to possible overlapping of the bands.

2. Some bands may appear to be partially filled (Figure 1.13). Such crystalline solids are metals.

A crystal with an even number of valence electrons per unit cell should be regarded as having both overlapping and nonoverlapping energy zones. Each case needs to be considered separately. If the zones overlap, we may obtain two or more

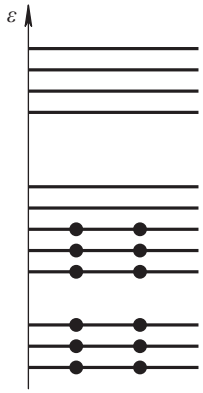


Fig. 1.13: The distribution of electrons among energy levels. There are partially filled zones.

partially filled ones instead of a full band, inherent in a dielectric. This results in exhibiting the metal properties by the crystal.

The energy of the highest occupied level is called the *Fermi energy* ε_F. In three-dimensional crystals, the Fermi energy range can cover some zones. For every partially filled band in the k-space there is a surface which separates occupied states from unoccupied states. Such a surface is called the *Fermi surface*.

Analytically, the Fermi surface is evaluated by the equation:

$$\varepsilon_n(\vec{k}) = \varepsilon_F \,. \tag{1.42}$$

The Fermi surface is not necessarily 1-connected; it can exist in the form of individual sections, having different values of the index n in formula (1.42). In this case, the number of such sections lying around the partially filled zones is always small.

The shape of the Fermi surface is responsible for most of the electronic properties of metals as an electric current occurs by changing the number of occupied electron states near the Fermi surface.

Because $\varepsilon_n(\vec{k})$ is a periodic function in reciprocal space, the solution of (1.42), which defines the Fermi surface, has the same property. The Fermi surface is periodic with the periods $\vec{b}_1, \vec{b}_2, \vec{b}_3$ in the k-space.

There are different ways to depict the Fermi surface geometrically. If one draws the Fermi surface for all admissible vectors $\vec{k}$, a periodic surface will be produced. This is called the *repetitive band method*. For clarity, one can illustrate a periodic Fermi surface within the first Brillouin zone. In this case, each physically different energy electron level is represented by only one point at the Fermi surface; no level will be lost or counted twice. Such a representation of the Fermi surface is characteristic for the so called reduced zone scheme.

Analogously, any periodic function can be represented, such as $\sin x$, either on the real axis (the repetitive zone scheme) or within one of the periods (the reduced zone scheme). We will return to this subject later.

1.5 Electrons in a Weak Periodic Field

Let us consider the problem of the motion of a single electron in a crystal in the presence of a weak periodic potential. In this case, the electron motion is almost free. As such, we can gain a complete understanding of the energy spectrum of the electron and the Fermi surface through the perturbation theory. Metals of groups I-IV in the periodic table have *s* and *p* valence electrons, and their inner atomic orbitals are completely filled up with electrons. Therefore, such metals are well suited for building the nearly free electron model to qualitatively explain the behavior of the electrons within them.

A metallic state is one of the most important states of matter. Chemical elements clearly prefer to be in the metallic state: more than two-thirds of them are metals. Therefore, in spite of being more applicable to metals than to semiconductors, it is difficult to overestimate the significance of the nearly free electron model.

Before proceeding to the basic subject matter, we will take a little mathematical excursus.

Periodic Function Expansion into Plane Waves for Multiple Measurement Cases

Any function possessing the periodicity property can be expanded into plane waves, which in turn form a complete set of functions. In a one-dimensional space, a function $f(x)$ with the above property $f(x) = f(x + L)$ can be represented as a Fourier series:

$$f(x) = \sum_K f_K \exp(\mathrm{i}Kx) \ ,$$

where $K = 2\pi n/L$, n is an integer:

$$f_K = \frac{1}{L} \int_0^L \exp(-\mathrm{i}Kx) f(x) \mathrm{d}x \ .$$

The situation is similar in the three-dimensional case. If the function $f(\vec{r})$ has the same periodicity as the Bravais lattice:

$$f(\vec{r} + \vec{R}) = f(\vec{r}) \ ,$$

where the vectors $\vec{R} = n_1\vec{a}_1 + n_2\vec{a}_2 + n_3\vec{a}_3$ define points of the Bravais lattice, it can be written down in the form of expansion:

$$f(\vec{r}) = \sum_{\vec{K}} f_{\vec{K}} \exp\left(\mathrm{i}\vec{K} \cdot \vec{r}\right) \ , \tag{1.43}$$

Here, $\vec{K} = \vec{b}_1 m_1 + \vec{b}_2 m_2 + \vec{b}_3 m_3$, with $\vec{b}_1, \vec{b}_2, \vec{b}_3$ being the reciprocal lattice vectors;

$$f_{\vec{K}} = \frac{1}{V_a} \int_{V_a} \mathrm{d}^3\vec{r} \exp\left(-\mathrm{i}\vec{K} \cdot \vec{r}\right) f(\vec{r}) \ . \tag{1.44}$$

The integration in (1.44) is carried out over a volume V_a of each elementary cell of the direct lattice. Strictly speaking, the choice of the cell is of no consequence because the integrand in (1.44) is periodic. The integral of the periodic function of the elementary cell is cell independent. In fact, any primitive cell can be split into parts. By translating the latter by the Bravais lattice vectors, we can obtain other primitive cells. Moreover, when translated, the periodic function $\exp(\mathrm{i}\vec{K}\cdot\vec{r})f(\vec{r})$ remains unchanged.

At this point it would be appropriate to prove the identity:

$$\frac{1}{V_a}\int\limits_{V_a} \mathrm{d}^3\vec{r}\,\exp\left[\mathrm{i}(\vec{K}-\vec{K}')\cdot\vec{r}\right] = \delta_{\vec{K},\vec{K}'}\,. \tag{1.45}$$

The expression under the integral sign in (1.45) has the same periodicity as the Bravais lattice. Therefore, a shift of the integration region by an arbitrary vector $\vec{d}$ does not change the integral. By integrating $\exp[\mathrm{i}(\vec{K}-\vec{K}')\cdot\vec{r}]$ over a mixed cell of $V_{a'}$, we can arrive at an integral over the original cell of V_a of $\exp[\mathrm{i}(\vec{K}-\vec{K}')\cdot(\vec{r}+\vec{d})]$:

$$\int\limits_{V_{a'}} \mathrm{d}^3\vec{r}\,\exp\left[\mathrm{i}(\vec{K}-\vec{K}')\cdot\vec{r}\right] = \int\limits_{V_a} \mathrm{d}^3\vec{r}\,\exp\left[\mathrm{i}(\vec{K}-\vec{K}')\cdot(\vec{r}+\vec{d})\right]\,. \tag{1.46}$$

As long as the result is cell independent, $V_{a'}$ can be replaced by V_a in formula (1.46) and rewritten in the form of the equality:

$$\left(\exp\left[\mathrm{i}(\vec{K}-\vec{K}')\cdot\vec{d}\right]-1\right)\int\limits_{V_a} \mathrm{d}^3\vec{r}\,\exp\left[\mathrm{i}(\vec{K}-\vec{K}')\cdot\vec{r}\right] = 0\,.$$

This expression must be satisfied by any $\vec{d}$, however, it is also correct for $\vec{K}\neq\vec{K}'$ provided that the integral vanishes:

$$\int\limits_{V_a} \mathrm{d}^3\vec{r}\,\exp\left[\mathrm{i}(\vec{K}-\vec{K}')\cdot\vec{r}\right] = 0\,. \tag{1.47}$$

Thus, we have proved that the identity (1.45) holds true for $\vec{K}\neq\vec{K}'$. Moreover, it remains valid for $\vec{K}=\vec{K}'$ as well, because $(1/V_a)\int_{V_a}\mathrm{d}^3\vec{r}=1$.

Using (1.45), it is easy to verify the validity of the representation (1.44) for the coefficients $f_{\vec{K}}$ in the expansion (1.43). To this end, we should multiply the equality (1.43) by $\exp(-\mathrm{i}\vec{K}'\cdot\vec{r})$ and integrate the result obtained over $\vec{r}$ within the unit cell of V_a.

Perturbation Theory in the Case of a Weak Periodic Potential

1. In the Schrödinger equation for an electron in a crystal:

$$-\frac{\hbar^2}{2m}\Delta\Psi + U(\vec{r})\Psi = \varepsilon\Psi\,, \tag{1.48}$$

the potential energy has the periodicity of the Bravais lattice: $U(\vec{r} + \vec{R}) = U(\vec{r})$, so we represent the function $U(\vec{r})$ in the form of a Fourier series:

$$U(\vec{r}) = \sum_{K} U_{\vec{K}} \exp\left(\mathrm{i}\vec{K} \cdot \vec{r}\right) , \tag{1.49}$$

where $\vec{K}$ is the reciprocal lattice vector. The Fourier coefficients $U_{\vec{K}}$ are related to $U(\vec{r})$ by:

$$U_{\vec{K}} = \frac{1}{V_a} \int\limits_{V_a} \mathrm{d}^3\vec{r} \exp\left(-\mathrm{i}\vec{K} \cdot \vec{r}\right) U(\vec{r}) . \tag{1.50}$$

In formula (1.50), the integration is over the elementary unit cell V_α of the crystal's Bravais lattice.

2. We can always replace the potential energy by an additive constant value. Then we choose the value so that the mean value of the potential U_0, taken for one elementary unit cell, should vanish:

$$U_{\vec{K}=0} = \frac{1}{V_a} \int\limits_{V_a} \mathrm{d}^3\vec{r} U(\vec{r}) = 0 . \tag{1.51}$$

3. The potential energy is a real function, so the equality:

$$U_{-\vec{K}} = U^*_{\vec{K}} , \tag{1.52}$$

holds true.

4. When a crystal has a central symmetry, a suitable choice of the coordinate system yields: $U(\vec{r}) = U(-\vec{r})$. Then (1.50) and (1.52) imply the chain of equalities:

$$U_{\vec{K}} = U_{-\vec{K}} = U^*_{\vec{K}} . \tag{1.53}$$

Thus, for crystals with an inversion center, $U_{\vec{K}}$ is the real quantity. Note that the one-dimensional lattice always has a center of inversion (see Figure 1.15).

Imagine that the sample is in the shape of a parallelepiped (Figure 1.14), built on the vectors $\vec{L}_i = N_i\vec{a}_i$. Here $i = 1, 2, 3$. It is worth pointing out that there is no summation

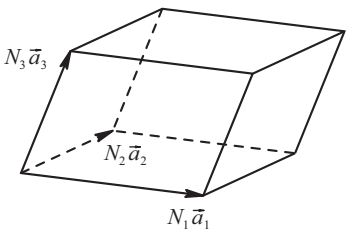


Fig. 1.14: A parallelepiped shaped crystal.

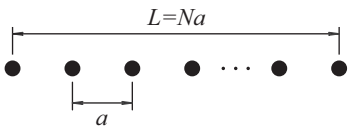


Fig. 1.15: The chain of N atoms is a one-dimensional crystal.

over the twice repeated index i. N_i are positive integers that define the lengths of the sides of the parallelepiped.

For $U(\vec{r}) = 0$ and the periodic *Born–Karman conditions:*

$$\Psi\left(\vec{r} + \vec{L}_i\right) = \Psi(\vec{r}) \,, \tag{1.54}$$

the solution of the Schrödinger equation (1.48) is trivial:

$$\Psi_{\vec{k}} = \frac{1}{\sqrt{V}} \exp\left(\mathrm{i}\vec{k} \cdot \vec{r}\right) \,, \quad \varepsilon_n(\vec{k}) = \frac{\hbar^2 \vec{k}^2}{2m} \,.$$

Here $V = \vec{L}_1 \cdot [\vec{L}_2 \times \vec{L}_3]$ is the volume of the crystal; the allowed values of the wave vector have the form:

$$\vec{k} = \sum_{i=1}^{3} \frac{m_i}{N_i} \vec{b}_i \,, \tag{1.55}$$

where m_i are arbitrary integers. For the sample's large sizes (for large N_i), the wave vectors (1.55) completely fill the entire k-space.

Our goal is to use perturbation theory to calculate changes of wave functions and energy levels of an electron in a weak periodic field $U(\vec{r}) \neq 0$. Calculations and final results for three-dimensional and one-dimensional crystals are close. Certainly, the one-dimensional case is atypical in many respects; it differs from the two-dimensional and three-dimensional ones. For example, the one-dimensional case does not imply the introduction of the Fermi surface – there are no overlapping bands. However, this problem makes complex geometric concepts more pictorial. Therefore, we give calculations and figures for the one-dimensional crystal. And then we come back to the three-dimensional problem and discuss a real situation.

A one-dimensional crystal is a chain of N atoms of length $L = Na$, where a is the distance between neighboring atoms (Figure 1.15).

The normalized wave function Ψ_k and levels of the energy $\varepsilon_0(k)$ of a free electron in the one-dimensional crystal can be written as:

$$\begin{aligned} \Psi_k(x) &= \frac{1}{\sqrt{L}} \exp\left(\mathrm{i}kx\right) \,, \\ \varepsilon_0(k) &= \frac{\hbar^2 k^2}{2m} \,, \quad k \equiv k_s = \frac{s}{N} \frac{2\pi}{a} \,. \end{aligned} \tag{1.56}$$

The function (1.56) satisfies the Born–Karman boundary condition $\Psi(x + L) = \Psi(x)$ if s is an arbitrary integer.

The energy of a free particle as a function of the wave number has the form of a parabola (Figure 1.16). The dots in the parabola are the allowed values of the electron energy. For large N, the distances Δk between adjacent points along the axis k decrease; $\Delta k = (2\pi)/(Na) \to 0$ as $N \to \infty$. Therefore, when it comes to the electron energy, we will often draw continuous curves, ignoring the fact that the curves are sets of closely spaced discrete points.

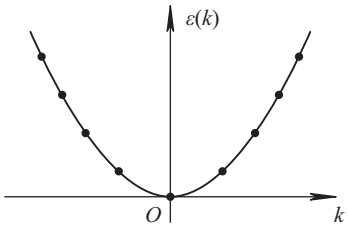


Fig. 1.16: The energy of a free electron.

The periodicity condition for the one-dimensional crystal has a simple form:

$$U(x + na) = U(x) ,$$

where n is an arbitrary integer. For the wave vectors of the reciprocal lattice: $K \equiv K_s = 2\pi s/a$, s is the integer.

Next, we expand the potential $U(x)$ in a Fourier series in a fashion that facilitates the further passage from the one-dimensional to three-dimensional case:

$$U(x) = \sum_K U_K \exp(\mathrm{i}Kx) = \sum_n U_{K_n} \exp\left(\mathrm{i}\frac{2\pi}{a}nx\right) ,$$

$$U_K = \frac{1}{a}\int_0^a \mathrm{d}x \exp(-\mathrm{i}Kx)\, U(x) .$$

To make allowances for the energy $\varepsilon_0(k)$ of the free electron due to the periodic potential $U(x)$, we need to use quantum mechanical perturbation theory for the stationary Schrödinger equation:

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\Psi}{\mathrm{d}x^2} + U(x)\Psi = \varepsilon\Psi .$$

Considering the weak potential $U(x)$ as a perturbation, in accordance with the perturbation theory we have:

$$\varepsilon(k) = \varepsilon_0(k) + \langle\Psi_k\left|\widehat{U}\right|\Psi_k\rangle + \sum_{k',k'\neq k} \frac{\left|\langle\Psi_k\left|\widehat{U}\right|\Psi_{k'}\rangle\right|^2}{\varepsilon_0(k) - \varepsilon_0(k')} + \cdots , \tag{1.57}$$

where $|\Psi_k\rangle$ are the vectors of a free electron state.

Let us estimate the matrix elements $\langle\Psi_k|\widehat{U}|\Psi_{k'}\rangle$ for the one-dimensional crystal ($k \equiv k_s$, $k' \equiv k_{s'}$):

$$\begin{aligned}\langle\Psi_k\left|\widehat{U}\right|\Psi_{k'}\rangle &= \int_0^{L=Na} \mathrm{d}x\Psi^*_{k_s} U(x)\Psi_{k_{s'}} = \\ &= \frac{1}{L}\int_0^{L=Na} \mathrm{d}x \exp\left[\frac{2\pi\mathrm{i}x}{Na}(s'-s)\right]\sum_n U_{K_n}\exp\left(\frac{2\pi\mathrm{i}n}{a}x\right) = \\ &= \frac{1}{L}\sum_n\left(\int_0^{L=Na}\mathrm{d}x\exp\left[\frac{2\pi\mathrm{i}x}{Na}(s'-s+nN)\right]\right)U_{K_n} .\end{aligned} \tag{1.58}$$

The problem reduces to the calculation of the integral:

$$\frac{1}{L}\int_0^{L=Na} \mathrm{d}x \exp\left[\frac{2\pi \mathrm{i}x}{Na}(s'-s+nN)\right] = \begin{cases} 0\,, & \text{if} \quad s'-s+nN \neq 0 \\ 1\,, & \text{if} \quad s'-s+nN = 0\,. \end{cases}$$

Since the condition $s' - s + nN = 0$ is equivalent to the equality $k_{s'} - k_s + K_n = 0$, we write the final result in the form:

$$\langle \Psi_k \left| \widehat{U} \right| \Psi_{k'} \rangle = \sum_K \delta_{(k-k'),K} U_K\,. \tag{1.59}$$

A similar outcome holds in the three-dimensional case. The passage to the three-dimensional crystal corresponds to the formal replacement: $k \to \vec{k}$, $k' \to \vec{k}'$, $K \to \vec{K}$ in (1.59).

We have chosen the point of reference for the potential energy of the crystal so that the matrix element $\langle \Psi_k | \widehat{U} | \Psi_k \rangle = U_0$ vanishes (1.51). Therefore, the first nonzero correction to the energy of the free electron due to the periodicity of the potential energy appears as:

$$\sum_{k',k'\neq k} \frac{\left|\langle \Psi_k \left| \widehat{U} \right| \Psi_{k'} \rangle\right|^2}{\varepsilon_0(k) - \varepsilon_0(k')} = \sum_{K\neq 0} \frac{|U_K|^2}{\varepsilon_0(k) - \varepsilon_0(k-K)}\,.$$

Finally, the electron energy in the periodic crystal is:

$$\varepsilon(k) = \varepsilon_0(k) + \sum_{K\neq 0} \frac{|U_K|^2}{\varepsilon_0(k) - \varepsilon_0(k-K)}\,. \tag{1.60}$$

The perturbation theory has a right to exist when each subsequent term of the expansion is less than the previous one. In this given case, this requirement is not met in the region of the point $k = K/2$ because $\varepsilon_0(K/2) = \varepsilon_0(-K/2)$ and the denominator in the second summand in (1.60) becomes small for $k \approx K/2$. For such values of k, the perturbation theory is inapplicable and in need of modification.

For the three-dimensional case, calculations and a result are analogous to the one-dimensional case:

$$\varepsilon(\vec{k}) = \varepsilon_0(\vec{k}) + \sum_{\vec{K}\neq 0} \frac{|U_{\vec{K}}|^2}{\varepsilon_0(\vec{k}) - \varepsilon_0(\vec{k}-\vec{K})}\,.$$

The perturbation theory is not applicable to values of the wave vector:

$$\varepsilon_0(\vec{k}) = \varepsilon_0(\vec{k}-\vec{K})\,. \tag{1.61}$$

The three-dimensional case differs from the one-dimensional one in that $\vec{k}$ is a vector rather than a number. Hence, the condition (1.61) gives a constraint not only to the magnitude but also to the direction of the wave vector.

To analyze the condition (1.61), we need to write it explicitly:

$$\frac{\hbar^2\vec{k}^2}{2m} = \frac{\hbar^2}{2m}(\vec{k}-\vec{K})^2\,.$$

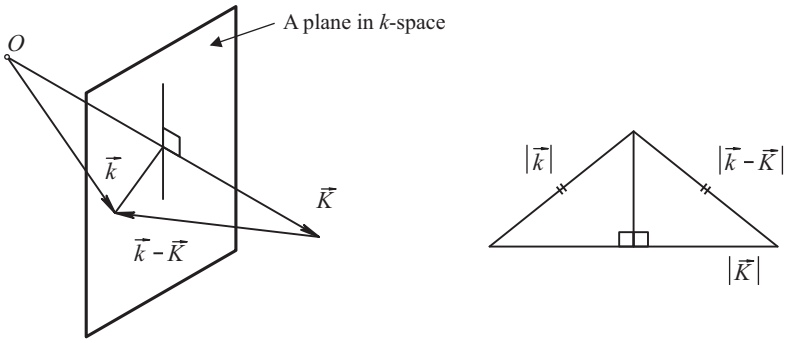


Fig. 1.17: Geometric ratios for the plane (1.62) in *k*-space.

This yields two equations:

$$\text{a)} \quad |\vec{k}| = |\vec{k} - \vec{K}| \,, \qquad \text{b)} \quad \vec{k} \cdot \vec{K} = \frac{K^2}{2} \,. \tag{1.62}$$

It is known from analytical geometry that, for a fixed $\vec{K}$, a plane perpendicular to the vector $\vec{K}$ in *k*-space corresponds to equation (1.62) b. In the reciprocal space, the wave vector $\vec{k}$ of an electron ties the starting point O to any other point of this plane. From equation (1.62) a, it follows that the plane is not only orthogonal to the vector $\vec{K}$ but it also accurately divides it in half. It is easy to see that the last assertion is true by using the hypotenuse leg theorem (theorem of the equality of right-angled triangles with the same cathetus and hypotenuse) (Figure 1.17).

Recall that $\vec{K}$ is a vector of the reciprocal crystal lattice. Suppose $\vec{K}$ is the shortest in a discrete set of parallel vectors of the reciprocal crystal lattice. This would make a plane passing through its middle the first Brillouin zone boundary. As a result, the perturbation theory is inapplicable, for example, near the boundaries of the Brillouin zone.

Inside the first Brillouin zone $\varepsilon_0(\vec{k}) \neq \varepsilon_0(\vec{k} - \vec{K}/2)$, the perturbation theory considered here is valid. The energy levels of the electron in the weak potential differ little from the energy levels of a free electron. This is one of the reasons why the first Brillouin zone is the most convenient primitive unit cell in reciprocal space to determine the electron wave vector.

When lengths of the vectors $\vec{K}$ in reciprocal space are arbitrary, there are, except for the of the first Brillouin zone boundaries, other planes of the type:

$$\vec{k} \cdot \vec{K} = K^2/2 \,. \tag{1.63}$$

The theory of perturbations can also be utilized for them. Let us elucidate the physical meaning of the condition (1.63).

It would be appropriate at this point to recall that the electron wave vector is related to the de Broglie wavelength, correlated to the electron: $|\vec{k}| = 2\pi/\lambda$. Let $\vec{K}_{\min}$ be

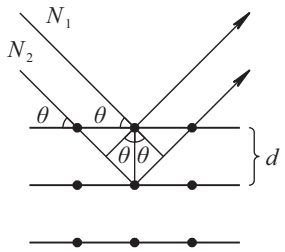


Fig. 1.18: Bragg reflection of x rays from a family of atomic planes separated by a distance d.

the shortest vector, parallel to the vector $\vec{K}$. Then $|\vec{K}| = n|\vec{K}_{\min}|$, where n is a natural number. Earlier, we have shown that $|\vec{K}_{\min}| = 2\pi/d$, where d is the distance between successive atomic planes of the direct lattice, with these being orthogonal to the vector $\vec{K}$. From this it follows that $|\vec{K}_{\min}| = 2\pi n/d$. We denote the angle between the vectors $\vec{k}$ and $\vec{K}$ by $(\pi/2-\theta)$. Next, we transform the condition (1.63), taking the comments made into account:

$$\left|\vec{k}\right|\left|\vec{K}\right|\cos\left(\frac{\pi}{2}-\theta\right) = \left|\vec{K}\right|^2/2 \quad \Rightarrow \quad \frac{2\pi}{\lambda}\sin\theta = \frac{1}{2}\frac{2\pi}{d}n \quad \Rightarrow \quad 2d\sin\theta = n\lambda\,.$$

The condition of inapplicability of the perturbation theory in this case has turned out to be exactly equivalent to the *Wulff–Bragg formula:*

$$2d\sin\theta = n\lambda\,. \tag{1.64}$$

The above formula specifies the conditions of x ray diffraction in the crystal (Figure 1.18).

Formula (1.64) is the maximum condition for the x ray beams N_1 and N_2 to interfere. The path difference of these beams is equal to $2d\sin\theta$. In x ray crystallography, the incidence angle θ from the reflection plane is measured rather than the complementary angle $(\pi/2 - \theta)$ (as in optics). Vector $\vec{k}$ defines the direction of the incident electron beam. The vector $\vec{K}$ is directed along a normal plane to the atomic planes, which are the reflection planes. In 1921–1923 Davisson and Germer were the first to observe the diffraction of electrons in crystals by the same laws as for x ray beams.

Overall Conclusions

1. This version of perturbation theory is valid only for electron wave vectors lying inside the first Brillouin zone. When the wave vectors lie near the Brillouin zone boundaries, there is a coherent reflection from the atomic planes of the complex wave, matched to the motion of the electron. In general, the perturbation theory is inapplicable when the Wulff–Bragg condition is met. Under this condition, an electron moving in the lattice intensively reflects from the crystalline planes. Its wave function undergoes significant changes.

2. In the vicinity of the Bragg planes, defined by the condition $\vec{k} \cdot \vec{K} = \vec{K}^2/2$, an electron has two close energies. It is this circumstance that makes the traditional perturbation theory unsuitable. Consequently, to describe the behavior of the electron with such wave vectors, it is necessary to exploit other versions of perturbation theory (in the presence of degeneracy).

We have found out that there are "dangerous" planes in the reciprocal space, i.e., planes, near which the electron energy varies greatly. The question arises of what happens to the energy of the electron when its wave vector is close to these planes? To reach an answer, we should seek a solution of the Schrödinger equation (1.48) in a combination of two wave functions with close energies $\varepsilon(k) \approx \varepsilon(k - K)$:

$$\Psi = A_1 \Psi_k + A_2 \Psi_{k-K} \,. \tag{1.65}$$

Here, A_1, A_2 are constants. Substituting (1.65) into the Schrödinger equation:

$$\left[-\frac{\hbar^2}{2m} \frac{\mathrm{d}^2}{\mathrm{d}x^2} + U(x) \right] \Psi = \varepsilon \Psi \,,$$

we get:

$$A_1[\varepsilon_0(k) - \varepsilon]\Psi_k + A_2[\varepsilon_0(k - K) - \varepsilon]\Psi_{k-K} + U[A_1\Psi_k + A_2\Psi_{k-K}] = 0 \,. \tag{1.66}$$

Then, we successively multiply equation (1.66) by Ψ_k^* and Ψ_{k-K}^* and integrate it over x within the limits $0 \le x \le L = Na$. Next, using the condition of orthogonality and normalization of the wave functions, as well as the matrix elements calculated earlier for the operator $U(x)$, we arrive at a linear homogeneous system which determines the coefficients:

$$\begin{cases} A_1(\varepsilon_0(k) - \varepsilon) + U_K A_2 = 0 \\ U_k^* A_1 + A_2[\varepsilon_0(k - K) - \varepsilon] = 0 \,. \end{cases} \tag{1.67}$$

The fact that the determinant of the system is equal to zero (the condition of the existence of a nontrivial solution) gives a quadratic equation to calculate the eigenvalues of the energy of an electron near the Bragg planes:

$$\varepsilon(k) = \frac{1}{2}\left[\varepsilon_0(k) + \varepsilon_0(k - K)\right] \pm \sqrt{\left[\varepsilon_0(k) - \varepsilon_0(k - K)\right]^2/4 + |U_K|^2} \,. \tag{1.68}$$

The expression (1.68) takes a particularly simple form for the electron wave vectors lying directly in the Bragg plane:

$$\varepsilon\,(K/2) = \varepsilon_0\,(K/2) \pm |U_K| \,. \tag{1.69}$$

We see that formula (1.68) is correct enough, as it contains no specifics. The perturbation theory modified by us is applicable whenever $|U_k| \ll \varepsilon_0(K/2)$.

Formula (1.69) shows that there are two possible values of the quasiparticle's energy in the Bragg plane. One of them is greater and another is less than the energy

of a free electron. To select a proper solution, it is necessary that the electron energy should be close to $\varepsilon_0(k)$ and far away from the so called "dangerous" planes. Given the quasicontinuity of the function $\varepsilon(k)$ near the value $k = K/2$, we should take the minus sign when $k < K/2$ and the plus sign when $k > K/2$ to make allowances. As a result, in the "dangerous" planes the electron energy abruptly rises, and any arbitrarily small potential forms a gap.

For the three-dimensional crystal, we obtain the same expression for the electron energy (1.68) by replacing the wave number k by the wave vector $\vec{k}$ and the number K by the reciprocal lattice vector $\vec{K}$. To simplify calculations for the three-dimensional crystal, we choose a coordinate system in reciprocal space so that the projection k_z of the electron wave vector should be directed along the vector $\vec{K}$. To do so, we introduce a new variable:

$$k_n = k_z - K/2 \ .$$

We express the energy of a free electron in the three-dimensional crystal in a convenient form for further analysis:

$$\varepsilon_0(\vec{k}) = \frac{\hbar^2}{2m}\left[k_\perp^2 + k_n^2 + K^2/4 + Kk_n\right] \ ,$$
$$\varepsilon_0(\vec{k} - \vec{K}) = \frac{\hbar^2}{2m}\left[k_\perp^2 + k_n^2 + K^2/4 - Kk_n\right] \ ,$$

where $\vec{k}_\perp$ is the projection of the electron wave vector on the plane with the normal $\vec{K}$. Having transformed formula (1.68), we come to an expression for the electron energy near the Bragg plane:

$$\varepsilon(\vec{k}) = \frac{\hbar^2}{2m}\left[k_\perp^2 + k_n^2 + \left(\frac{K}{2}\right)^2\right] \pm \sqrt{\left(\frac{\hbar^2}{2m}Kk_n\right)^2 + |U_{\vec{K}}|^2} \ .$$

It is easy to see that $(\partial\varepsilon)/(\partial k_n)|_{k_n=0} = 0$, i.e., the function $\varepsilon(\vec{k})$ has extrema at the Brillouin zone boundaries. Now we calculate the second derivative of the function $\varepsilon(\vec{k})$ at the extremum:

$$\left.\frac{\partial^2\varepsilon}{\partial k_n^2}\right|_{k_n=0} = \frac{\hbar^2}{m} \pm \frac{\hbar^2}{m}\cdot\frac{2\varepsilon_0(\vec{K}/2)}{|U_{\vec{K}}|} \approx \pm\frac{\hbar^2}{m}\cdot\frac{2\varepsilon_0(\vec{K}/2)}{|U_{\vec{K}}|} \neq 0 \ .$$

In doing so, we have taken into account that $|U_k| \ll \varepsilon_0(K/2)$.

Consequently, the quasicontinuous function $\varepsilon(\vec{k})$ has either local maxima or local minima at the boundaries along directions perpendicular to the Brillouin zone boundaries (Figure 1.19).

Since the velocity of the electron is $\vec{V} = (1/\hbar)(\partial\varepsilon/\partial\vec{k})$, the foregoing assumes different meaning. The electron velocity component perpendicular to the Brillouin zone boundary vanishes at the same boundary.

Let us turn to the one-dimensional model. The function $\varepsilon(k)$ illustrated in Figure 1.20 takes into account changes in the electron energy at all Bragg planes. This

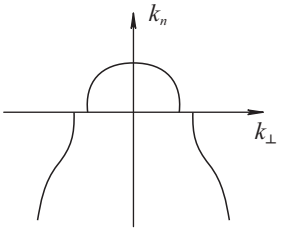


Fig. 1.19: The surface $\varepsilon(\vec{k}) = \text{const}$ intersects the plane $k_n = 0$ in k-space.

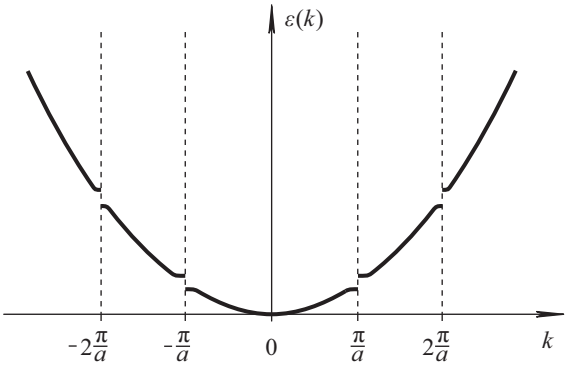


Fig. 1.20: The energy spectrum of an electron in a one-dimensional crystal in the extended zone scheme.

method of representing electron energy levels is called the *extended zone scheme*. In the extended zone scheme, the energy spectrum of the electron in the crystal bears a strong resemblance to the spectrum of a free electron (compare Figures 1.16 and 1.20).

Figure 1.20 shows a parabola, although $\varepsilon(k)$ is a periodic function with the period $2\pi/a$. Translations along the axis k by $2\pi n/a$ (n is an integer) superpose the remaining parabolas and that depicted in Figure 1.20. Therefore, all of these parabolas are equivalent to each other.

Suppose the electron energy levels are given more preferably by means of the wave numbers k of the first Brillouin zone. Then we can translate the segments of the curves in Figure 1.20 by shifting along the reciprocal lattice vectors. In doing so, we enumerate all the allowed energy levels without labeling them twice. Figure 1.21 is the result of the above translations. Such a representation of the energy spectrum of the electron is referred to as the *reduced zone scheme*.

Note that the appearance of the extrema in the center of the Brillouin zone, as in Figure 1.21, is due to the symmetry of a one-dimensional lattice: $U(x) = U(-x)$ (a one-dimensional crystal formed by atoms of one kind always has a center of inversion).

It should be emphasized that the electron energy in the k-space is periodic. For this purpose, we continue Figure 1.21 periodically throughout the entire k-space. The resulting Figure 1.22 clearly displays that each level with a given k can also be de-

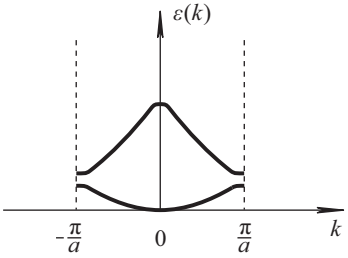


Fig. 1.21: The image of the electron energy levels in the reduced zone scheme.

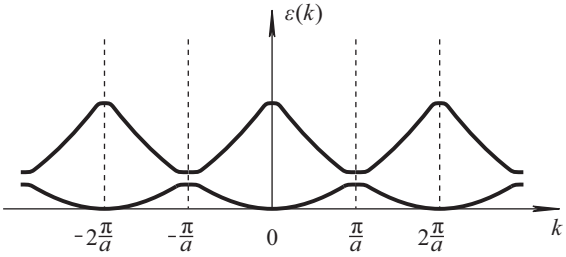


Fig. 1.22: The image of the allowed values of the electron energy in a crystal in the repeated zone scheme.

scribed by other wave vectors, which differ from k by any reciprocal lattice vectors. This type of representation of the energy of an electron in a crystal is known as the *repeated zone scheme*. The repeated zone scheme is the most general representation, but a description in this fashion, in contrast to the reduced-zone scheme, is redundant. Each energy level for all physically equivalent wave numbers is shown many times in Figure 1.22: $\varepsilon(k) = \varepsilon(k + 2\pi n/a)$, where n is an integer.

We have shown that the weak periodic potential acts mainly on levels of free electrons whose wave vectors lie near the Bragg planes. The periodic potential causes energy gaps in the electron spectrum to emerge. Let us come back to the algebraic system (1.67) that determines the coefficients A_1 and A_2 to analyze the view of the wave function of an electron near the Bragg planes:

$$\Psi = A_1\Psi_k + A_2\Psi_{k-K} \,. \tag{1.70}$$

Because of the linear dependence of the equations, it suffices to take one of them, for example, below:

$$A_1(\varepsilon_0(k) - \varepsilon) + U_K A_2 = 0 \,. \tag{1.71}$$

For $k = K/2$ we have $\varepsilon = \varepsilon_0(K/2) \pm |U_k|$, therefore, from (1.71) we find:

$$A_1 = \pm\frac{U_{\vec{K}}}{|U_{\vec{K}}|}A_2 \,.$$

By formally replacing $K \to \vec{K}$ such a form can be generalized to a three-dimensional crystal.

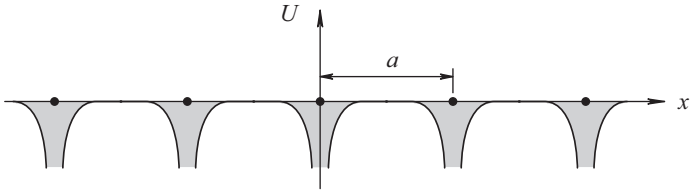


Fig. 1.23: Changes in the potential energy of an electron in a one-dimensional crystal. The bold dots mark centers of the ion cores.

For crystals with a center of inversion, the Fourier component of the potential $U_{\vec{K}}$ is real, so:

$$A_1 = \pm \operatorname{sign} U_{\vec{K}} A_2 \ . \tag{1.72}$$

At the same time, it should be kept in mind that a one-dimensional crystal, which is distinct from a three-dimensional one, always has a center of inversion. Therefore, in the one-dimensional case, formula (1.72) is always true.

For the one-dimensional model, the change in the potential energy of the electron in the field of ionic cores is shown in Figure 1.23. Near the ion cores, the potential energy must be negative, since it reduces to the Coulomb interaction energy between electric charges of the opposite sign. Outside the ionic cores, the electron behaves as free, and therefore the potential energy is $U \approx 0$.

Suppose that $U_K < 0$ in formula (1.72), then we have $A_1 = \pm A_2$. As a consequence, the relation (1.70) yields two wave functions, one of which corresponds to the upper edge of the energy gap and another to the lower edge:

$$|\Psi_+|^2 \sim \left|\sin \frac{1}{2} Kx\right|^2 \quad \text{for} \quad \varepsilon = \varepsilon_0 \left(\frac{K}{2}\right) + |U_K| \ ,$$

$$|\Psi_-|^2 \sim \left|\cos \frac{1}{2} Kx\right|^2 \quad \text{for} \quad \varepsilon = \varepsilon_0 \left(\frac{K}{2}\right) - |U_K| \ .$$

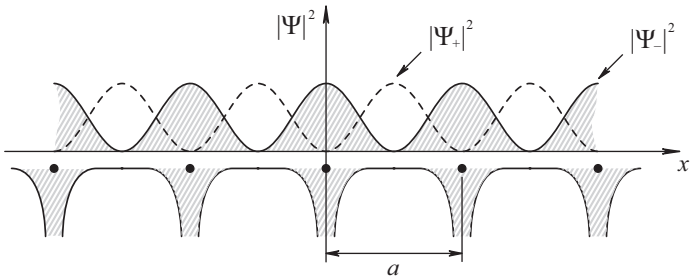


Fig. 1.24: Probability densities of finding an electron in a crystal under the Wulff–Bragg condition. The solid curve shows $|\Psi_+|^2$, the dashed line is $|\Psi_-|^2$. The bold dots indicate centers of the ion core.

General Conclusions

1. When the Wulff–Bragg conditions are met, the traveling de Broglie waves mapped to the electron cannot propagate through the lattice. Their reflection from the Bragg planes and then their interference in the crystal lead to arising the standing de Broglie waves.
2. Recall that the square of the modulus of the wave function determines the probability density of finding an electron in coordinate space. The function $|\Psi_-|^2$ has antinodes at places where ions arrange themselves. The probability of finding the electron is higher near the ions. Due to this, the Coulomb electron-ion interaction energy diminishes. Hence the total energy of the electron in the crystal also decreases. By contrast, the function $|\Psi_+|^2$ has both antinodes far away from the ion cores and nodes at places where the ions are. Thus, a state which is described by the function corresponds to higher electron energy.

The scheme set forth here is the key to understanding the origin of energy gaps in the electron energy spectrum in a crystal [2, 3].

1.6 The Fermi Energy, Surface, Temperature, and Thermal Layer for a Gas of Free Electrons

To introduce a number of new terms and simplify further geometric constructions, it is useful to ignore periodicity of potential energy of an electron in a crystal for a while. The next step considers the formal problem of a gas of noninteracting quasiparticles correlated to electrons. That is to say, we should look upon a gas of free fermions in a box with the Born–Karman periodic boundary conditions.

Comments on the Simplified Model

1. Even within this model, we can understand why the Fermi surface can be entered only for two- and three-dimensional crystals. In the one-dimensional case, this cannot be done.
2. Knowing the Fermi surface of free electrons in the box, we can construct the Fermi surface for real electrons in a metal, taking into account the influence of a weak periodic potential on the electron.
3. The results of the free electron model help numerically estimate parameters characterizing the properties of metals.
4. Even so, such a model can explain why many parameters of metals are weakly temperature dependent and have the same values as $T = 0\,\mathrm{K}$.

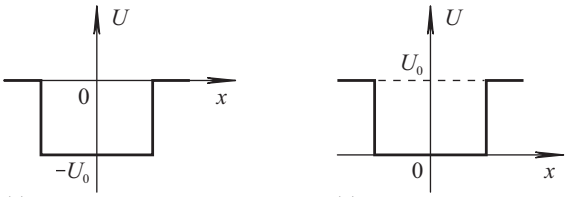


Fig. 1.25: Potential well: (a) The original form; (b) the transformed view, with the well bottom being at the zero level.

Imagine an ideal gas of quasiparticles, electrons, in a potential box with the Born–Karman boundary conditions. The box will partially simulate the interaction between the electrons and positive ions of the crystal lattice. Although the potential energy of a single electron in this model is constant outside and inside the crystal, the constant values are different. Outside the crystal, the potential energy is $U = 0$, and inside it is $U = -U_0 < 0$ due to the electron-positive ion interaction. To lift the electron to the crystal surface, we must perform work against the attraction forces of the positive ions.

Within the quantum mechanical problem, it is usually more convenient to measure the electron's energy from the bottom of a potential well, so we will treat the well bottom as a zero level (Figure 1.25). This transformation is quite admissible because the potential is defined up to a constant. Then the energy of the free electrons are:

$$\varepsilon(\vec{k}) = \frac{\hbar^2 \vec{k}^2}{2m} \,. \tag{1.73}$$

By virtue of the Born–Karman boundary conditions, the values of the wave vector $\vec{k}$ vary discretely. A reciprocal space volume per one allowed wave vector $\vec{k}$ is equal to $(2\pi)^3/V$. In an infinite crystal of volume V, allowed wave vectors quasicontinuously fill up the entire reciprocal space.

If $\mathrm{d}^3\vec{k}$ is the volume element in the reciprocal space, we find the number of allowed vectors in the volume by dividing $\mathrm{d}^3\vec{k}$ by the infinitesimal volume containing one vector:

$$\mathrm{d}^3\vec{k} : \frac{(2\pi)^3}{V} = \frac{V\mathrm{d}^3\vec{k}}{(2\pi)^3} \,. \tag{1.74}$$

For free electrons, their energy and the wave vector $\vec{k}$ are related by equation (1.73). Each energy state can be occupied with two electrons differing in spin projection values. Therefore, the allowed energy values in the volume $\mathrm{d}^3\vec{k}$ are double the permitted values of $\vec{k}$:

$$\mathrm{d}n_{\vec{k}} = \frac{2V\mathrm{d}^3\vec{k}}{(2\pi)^3} \,. \tag{1.75}$$

Here $\mathrm{d}n_{\vec{k}}$ is the number of allowed energy states of the electron in the volume $\mathrm{d}^3\vec{k}$.

However, by the laws of quantum mechanics, electrons cannot fill up some allowed energy levels. In the case of thermodynamic equilibrium, the number of electrons with a given spin projection value and energy $\varepsilon(\vec{k})$ is determined by the Fermi–Dirac distribution:

$$f = \frac{1}{\exp\left[\left(\varepsilon(\vec{k}) - \mu\right)/k_{\mathrm{B}}T\right] + 1} . \tag{1.76}$$

The quantity f shows how many electrons with a given spin projection value are in the state with the energy $\varepsilon(\vec{k})$ at a temperature T. With the volume $\mathrm{d}^3\vec{k}$ being small, all energies are identical within it. Suppose that the electrons are in thermodynamic equilibrium and fill the quantum mechanical states in the volume $\mathrm{d}^3\vec{k}$. As a result, the total number of electrons $\mathrm{d}N$ can be computed as the product of the number of electrons in a given energy state f and the number of states $\mathrm{d}n_{\vec{k}}$:

$$\mathrm{d}N = f\mathrm{d}n_{\vec{k}} = \frac{2Vf\mathrm{d}^3\vec{k}}{(2\pi)^3} . \tag{1.77}$$

Consequently, we get the total number of electrons N in a crystal by integrating the expression (1.77) over the entire $\vec{k}$-space:

$$N = \frac{2V}{(2\pi)^3}\int \mathrm{d}^3\vec{k} f . \tag{1.78}$$

As a matter of fact, this is the normalization condition for calculating the chemical potential μ as a function of crystal temperature and electron density $n = N/V$, with the function being independent of the crystal volume, i.e., $\mu = \mu(T, n)$.

For a Fermi gas, the function $f(\varepsilon)$ is a fuzzy step $\sim k_{\mathrm{B}}T$ centered around the value $\varepsilon = \mu$.

$$\text{If} \quad \varepsilon < \mu - k_{\mathrm{B}}T , \quad \text{then} \quad f(\varepsilon) \approx 1 .$$
$$\text{If} \quad \varepsilon > \mu - k_{\mathrm{B}}T , \quad \text{then} \quad f(\varepsilon) \approx 0 .$$

It is interesting that for any $T > 0$ as $\varepsilon = \mu$, we have $f(\varepsilon = \mu) = 1/2$.

As $T \to 0$, the graph of the function $f(\varepsilon)$ is transformed into a rectangular step (Figure 1.27).

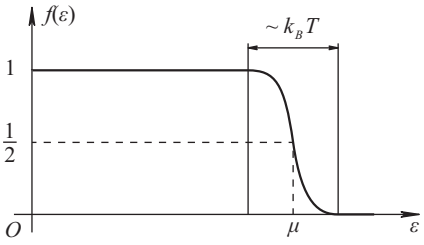


Fig. 1.26: The Fermi–Dirac distribution function.

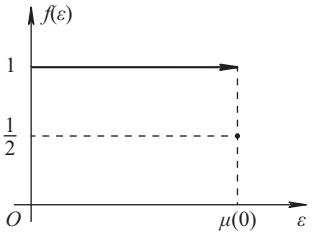


Fig. 1.27: The form of the Fermi–Dirac distribution function as $T \to 0$.

This means that, in accordance with the Pauli principle, the electrons fill all of the energy levels in the crystal ground state at $T = 0$ K. One electron occupies each quantum mechanical state up to the Fermi energy.

Notes

1. Figure 1.28 shows the energy levels filled with electrons in the potential well (in a crystal). The energy required to remove an electron from the crystal is $\Delta\varepsilon = U_0 - \varepsilon_F$ (the depth of the well minus the energy of the most energetically capable electron). It can be seen that the quantum problem differs from the outcome of classical physics, where the energy to remove one of noninteracting particles from the potential well is U_0 at $T = 0$ K.

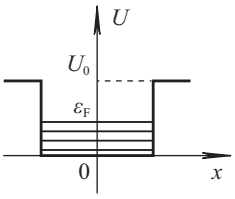


Fig. 1.28: Potential well energy levels filled with electrons.

2. As we will soon show, the chemical potential of metals is weakly temperature dependent and, therefore, even at room temperature it is:

$$\mu(T) \approx \varepsilon_F \, .$$

Yet, under accurate calculations one should not equate the chemical potential and the Fermi energy as this can lead to errors.

Suppose N is the number of noninteracting electrons in the ground state of a system at $T = 0$ K. Then, the allowed energy levels of the electrons are located inside a sphere:

$$\varepsilon(k) = \frac{\hbar^2 k^2}{2m} \leq \varepsilon_F \, . \tag{1.79}$$

The surface of the sphere is the Fermi surface of radius:

$$k_F = \frac{1}{\hbar}\sqrt{2m\varepsilon_F} \, . \tag{1.80}$$

If $T = 0$ K, the k_F and ε_F values can be calculated. In the formula for the electron density in the crystal:

$$n = \frac{N}{V} = \frac{2}{(2\pi)^3}\int \mathrm{d}^3\vec{k} f \, , \tag{1.81}$$

we replace the function f by the rectangular step. In doing so, the range of integration over $\vec{k}$-space will be restricted:

$$n = \frac{2}{(2\pi)^3}\int\limits_{|k|\leq k_F} \mathrm{d}^3\vec{k} = \frac{1}{4\pi^3}\cdot\frac{4}{3}\pi k_F^3 = \frac{1}{3\pi^2}k_F^3 \, . \tag{1.82}$$

The expression (1.82) solves for the radius k_F of the *Fermi sphere* as a function of the electron density:

$$k_F = \left(3\pi^2 n\right)^{1/3} . \tag{1.83}$$

For almost all metals, the electron density is $n \sim 1/a^3$, where a is a lattice constant. The interatomic distances in crystals are $a \sim 10^{-8}$ cm. To calculate the radius k_F, the estimate $k_F \sim \pi/a$ is often used and therefore the radius is $k_F \sim \pi \cdot 10^8\ \mathrm{cm}^{-1}$.

Knowing the connection between k_F and ε_F, the Fermi energy can be expressed in terms of the concentration and the mass of the electrons:

$$\varepsilon_F = \frac{\hbar^2}{2m}(3\pi^2 n)^{2/3} . \tag{1.84}$$

For concentrations of metals, the Fermi energy lies in the range: $\varepsilon_F \sim 1.5 \div 15$ eV.

Now, we introduce the notion of the *Fermi velocity*:

$$V_F = \frac{\hbar k_F}{m} \sim 10^8\ \mathrm{cm/s} .$$

This is a fairly high speed; it is 1% of the speed of light. From the point of view of classical physics, the result is impossible to understand. This is because at $T = 0\,\mathrm{K}$ the electron must completely stop moving. According to the laws of classical physics, even at room temperature, the speed of the thermal motion of the electron is:

$$V \sim \sqrt{\frac{3k_B T}{m}} \sim 10^7\ \mathrm{cm/s} .$$

In order to estimate the width of the thermal smearing of the Fermi step (Figure 1.26), the *Fermi temperature* should be introduced:

$$T_F \equiv \frac{\varepsilon_F}{k_B} \sim \left(10^4 \div 10^5\right) K .$$

Given the room temperature is about 300 K, the relation $k_B T/\varepsilon_F = T/T_F$ yields:

$$k_B T \sim (10^{-2} \div 10^{-3})\varepsilon_F \ll \varepsilon_F .$$

The region of the thermal smearing is far less than ε_F. Therefore, only a small portion of all electrons in a metal can change their quantum mechanical states under the influence of temperature and external fields. This portion amounts to $k_B T/\varepsilon_F$. However, it is these electrons which govern the conductivity, specific heat, thermal conductivity, and other properties of metals. The foregoing macroscopic properties of metals cannot be explained by classical physics.

Since the change in the number of occupied energy states occurs near the Fermi surface, most of the properties of metals are determined by its shape. Therefore, it is important to know what shape of the Fermi surface real metals have.

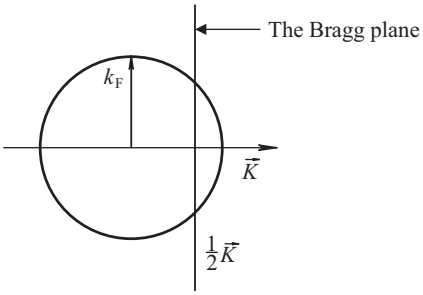


Fig. 1.29: The Fermi surface intersects the Bragg plane.

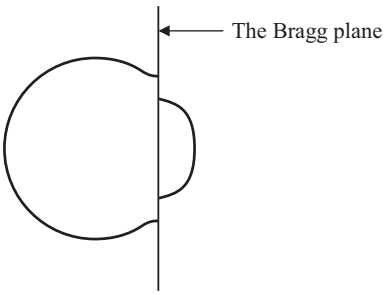


Fig. 1.30: The Fermi surface, deformed near the Bragg plane.

1.7 Method of Constructing the Fermi Surface of a Weak Potential: The Second and Subsequent Brillouin Zones

Let us get familiar with the method of constructing the Fermi surface in a three-dimensional (or two-dimensional) crystal within the nearly free electron model. It is as follows.

1. At the outset, we construct the Fermi surface for free electrons. It is a sphere centered at the point $\vec{k} = 0$ in reciprocal space. The Bragg planes cross the sphere. They are perpendicular to the reciprocal lattice vectors $\vec{K}$ and pass through their middles (Figure 1.29).
2. It should be taken into account that whenever the sphere intersects the Bragg plane, it becomes deformable in a certain manner. In this case, each Bragg plane forms energy gaps and local extrema (Figure 1.30).
   When we consider the influence of all the Bragg planes, we get an image of the Fermi surface in the extended zone scheme.
3. In order to obtain the above image in the reduced zone scheme, the separate parts of the Fermi surface should be translated by the reciprocal lattice vectors. Therefore, we move them into the first Brillouin zone.

The procedure set forth previously needs systematization. For this purpose, we should introduce the concept of higher Brillouin zones whose boundaries meet the Wulff–Bragg condition:

$$\vec{k} \cdot \vec{K} = \vec{K}^2/2 \,. \tag{1.85}$$

Here, it is worth recalling that this formula is satisfied by the vectors $\vec{k}$. Their ends lie in the plane perpendicular to the vector $\vec{K}$ and dividing the latter in half. This is not necessarily the first Brillouin zone boundary. The vector $\vec{K}$ is an arbitrary reciprocal lattice vector:

$$\vec{K} = n_1\vec{b}_1 + n_2\vec{b}_2 + n_3\vec{b}_3 \,. \tag{1.86}$$

Now we can give a new definition of the first Brillouin zone.

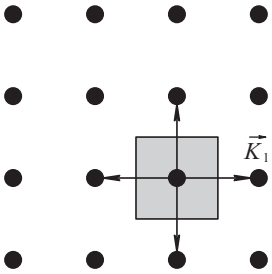


Fig. 1.31: The first Brillouin zone for a two-dimensional lattice (shaded area).

One of the points in the reciprocal lattice should be designated as the origin. The *first Brillouin zone* is a set of points which can be reached from the starting point without crossing any Bragg plane.

Let us explain constructing the first Brillouin zone by using the example of a two-dimensional square lattice in k-space. The first Brillouin zone is constructed by taking the shortest length of the vector $\vec{K}_1$ and three other vectors of the same length (Figure 1.31). The above vectors, being reciprocal lattice vectors, connect the origin point to their nearest neighbors. The next step is to draw straight lines perpendicular to these vectors and passing through their middles. The region bounded by the straight lines is the first Brillouin zone. In a three-dimensional crystal, they are the Bragg planes. As can be seen, the new definition of the first Brillouin zone is equivalent to the previous one.

The *second Brillouin zone* is a set of points which can be reached from the first zone by crossing only one Bragg plane. To outline such a zone, it is necessary, along with the vector $\vec{K}_1$ and vectors equivalent to it in length, to use the vector $\vec{K}_2$, which is greater in length, and the same long vectors (Figure 1.32). Furthermore, we draw new straight lines through the middles of all the vectors. These lines intersect with each other and the lines used for the construction of the first Brillouin zone. At the intersection of the lines, one needs to take only those areas that can be reached from the first Brillouin zone by crossing only one of the lines. These areas are the second Brillouin zone.

It is important to note that each Brillouin zone is a primitive unit cell of the reciprocal lattice. Therefore, all of the Brillouin zones have the same volume and can be combined with each other by translating their separate segments by reciprocal lattice vectors.

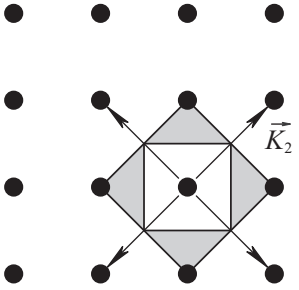


Fig. 1.32: The second Brillouin zone.

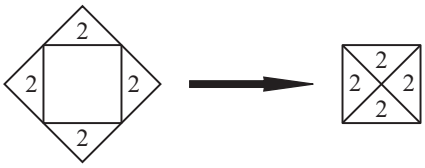


Fig. 1.33: The reduction of the second Brillouin zone to the first one.

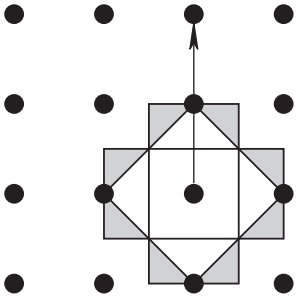


Fig. 1.34: The third Brillouin zone.

As an example, Figure 1.33 presents the reduction of the second Brillouin zone to the first one.

Generally, the *n-th Brillouin zone* is a set of reciprocal lattice points that lie outside the $(n-1)$-th zone, and which can be reached from $(n-1)$-th zone by crossing one Bragg plane. It is in such a way that the third (Fig. 1.34) and other orders can be constructed.

Let us consider an example where the Fermi surface (sphere) touches the boundaries of the second Brillouin zone (Figure 1.35). The first zone is completely filled with electrons. The second zone does not have enough electrons.

Now, the segments of the second zone, filled with electrons, should be reduced to the first one (Figure 1.36). Figure 1.36 displays shaded and unshaded regions; the boundary between them is the Fermi surface (in a two-dimensional case, it is a line) in the reduced zone scheme.

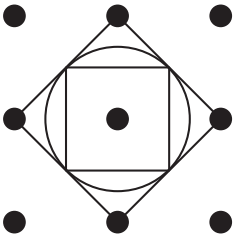


Fig. 1.35: The Fermi surface touches the second Brillouin zone boundaries.

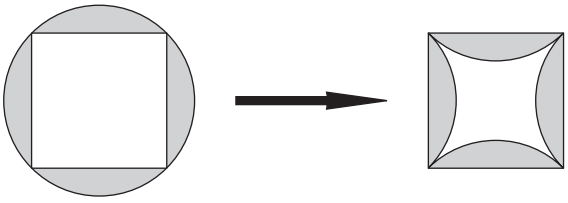


Fig. 1.36: The reduction of the segments of the second zone, filled with electrons, to the first Brillouin zone.

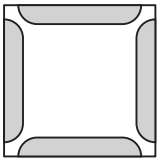


Fig. 1.37: The Fermi surface of nearly free electrons in the reduced zone scheme.

To go over from the Fermi surface for free electrons to the Fermi surface for nearly free electrons, we should take into account the following:

1. The interaction of the electrons with the periodic potential of a crystal makes energy gaps appear at the zone boundaries.
2. The Fermi surface crosses the Brillouin zone boundaries perpendicularly.
3. At its corners, the Fermi surface very sharply "feels" the influence of the transcrystalline potential.
4. The total volume enclosed by the Fermi surface depends only on the electron concentration and does not depend on the features of the electron-lattice interaction.

Given these remarks, the Fermi surface in the reduced zone scheme takes the form shown in Figure 1.37. For an illustration, compare Figures 1.36 and 1.37.

Of course, without detailed calculations, it is impossible to predict how the parts of the Fermi surface near the Brillouin zone boundaries will be transformed. For example, the problem in hand may have another possible scenario: the electrons travel into the third Brillouin zone through the corners of the second Brillouin zone (Figure 1.37). Then, we will also have to draw the Fermi surface pieces in the third Brillouin zone.

The Fermi surface is periodic in k-space and, to reflect this fact, it is sometimes useful to use the repeated zone scheme (Figure 1.38).

Of all the schemes presented above, the extended zone scheme, at first glance, seems to be the simplest. However, it is not suitable for analytical calculations.

To compute the heat capacity, conductivity and other properties of metals, the reduced zone scheme is especially convenient. In this scheme, one Brillouin zone includes all various states of an electron in a crystal, and none of them is considered twice. It is the circumstance that facilitates all integrations over $\vec{k}$.

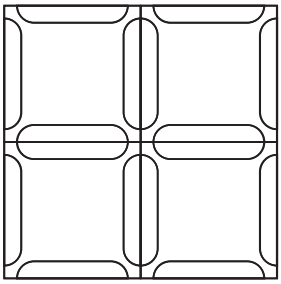


Fig. 1.38: The Fermi surface in the second Brillouin zone in the repeated zone scheme.

The repeated zone scheme is used to describe the dynamics of an electron in a crystal theoretically, for example, when the electron wave vector $\vec{k}$ approaches the Brillouin zone boundary under the influence of an external electromagnetic field. In this case, the calculations require the vector to be continuous. This cannot be done in the reduced zone scheme. Once the wave vector $\vec{k}$ has overstepped the Brillouin zone boundary, it must be returned back via translation by a reciprocal lattice vector. This procedure creates substantial analytical difficulties. In what follows, we discuss the nontrivial dynamics of an electron in a crystal.

1.8 Electronic Specific Heat of Normal Metals

Using the model of an ideal gas of fermions in a periodic field, we derive an expression for the electronic specific heat of metals [4]. Although we call the Fermi particles forming the gas electrons, they are in fact not like real electrons. Here, we are saying about quasiparticles, particle like objects, into which turn real electrons due to their interaction with the crystal lattice and other electrons. Note that the differences between the quasiparticles and the electrons are:

1. The quasiparticles do not interact with each other.
2. The quasiparticles and electrons have different masses.

Since the quasiparticles in a crystal do not interact with each other, the number of states occupied by the quasiparticles in the volume element $\mathrm{d}^3\vec{k}$ of the reciprocal lattice is the same as that of the free electron gas (fermions).

The number of allowed energy levels for electrons in the volume element $\mathrm{d}^3\vec{k}$ is given by:

$$\mathrm{d}n_{\vec{k}} = \frac{2V\mathrm{d}^3\vec{k}}{(2\pi)^3}\,, \tag{1.87}$$

and the number of electrons in the state of thermodynamic equilibrium and which fill up these energy levels can be estimated by the formula:

$$\mathrm{d}N = \mathrm{d}n_{\vec{k}} f\,. \tag{1.88}$$

Here, f is the Fermi–Dirac distribution function.

However, there are differences between fermions in a crystal and a gas of free fermions.

1. For fermions in the crystal, all possible states can be restricted by the first Brillouin zone (for free fermions in the potential well, such restrictions do not exist).
2. The energy spectrum of electrons in a crystal forms zones so that the energies of the electrons with different n are separated by large energy gaps ε_{g}. The potential well does not have such gaps in the energy spectrum of free electrons.

If one takes the wave vectors of electrons from the first Brillouin zone, we get several functions $\varepsilon_n(\vec{k})$ for the same $\vec{k}$. These differ in the zone number n. In this case, the lower branches of the energy spectrum of $\varepsilon_n(\vec{k})$ are completely filled with electrons. To calculate the thermal properties of a metal, it suffices to consider the upper energy branch partially filled with electrons and their behavior. Electrons of completely filled zones do not affect the thermal properties of a metal, since their excitation requires very large energy of the order of the gap width ε_g.

Hereinafter, for the sake of simplicity, the band partially filled with electrons is assumed to be only one. The zone number is not labeled by n, i.e., we replace $\varepsilon_n(\vec{k})$ by $\varepsilon(\vec{k})$. The bottom of the partially filled band is the level to begin measuring the electron gas energy.

Given the above remarks, the energy of the Fermi gas is given by:

$$W = 2V \int \frac{\varepsilon(\vec{k}) f \mathrm{d}^3\vec{k}}{(2\pi)^3}\,, \tag{1.89}$$

where the region of the integration over the wave vector $\vec{k}$ is the first Brillouin zone.

The specific heat of the electron gas at constant volume is obtained by differentiating the energy (1.89) over temperature:

$$c = \frac{1}{V}\left(\frac{\partial W}{\partial T}\right)_V = 2\int \varepsilon(\vec{k})\frac{\partial f}{\partial T}\frac{\mathrm{d}^3\vec{k}}{(2\pi)^3}\,. \tag{1.90}$$

The chemical potential μ can be determined from the condition of conservation of the total number of particles:

$$N = 2V \int f \frac{\mathrm{d}^3\vec{k}}{(2\pi)^3}\,. \tag{1.91}$$

Here, N is the number of electrons in the upper partially filled band. These electrons are called conduction electrons.

Since $N = \text{const}$, $V = \text{const}$, we have:

$$\frac{\partial}{\partial T}\frac{N}{V} = 2\int \frac{\partial f}{\partial T}\frac{\mathrm{d}^3\vec{k}}{(2\pi)^3} = 0\,. \tag{1.92}$$

The integration is carried out also over physically different $\vec{k}$ belonging to the first Brillouin zone, rather than the entire k-space, as it was for the free electrons.

Given that $\partial f/(\partial T)$ has the form:

$$\frac{\partial f}{\partial T} = -\frac{\partial f}{\partial \varepsilon}\left[\frac{\varepsilon-\mu}{T} + \frac{\mathrm{d}\mu}{\mathrm{d}T}\right]\,, \tag{1.93}$$

from (1.90) and (1.92) the system of equations follows:

$$\begin{aligned} c &= -2\int \varepsilon \frac{\partial f}{\partial \varepsilon}\left[\frac{\varepsilon-\mu}{T} + \frac{\mathrm{d}\mu}{\mathrm{d}T}\right]\frac{\mathrm{d}^3\vec{k}}{(2\pi)^3}\,, \\ 2&\int \frac{\partial f}{\partial \varepsilon}\left[\frac{\varepsilon-\mu}{T} + \frac{\mathrm{d}\mu}{\mathrm{d}T}\right]\frac{\mathrm{d}^3\vec{k}}{(2\pi)^3} = 0\,. \end{aligned} \tag{1.94}$$

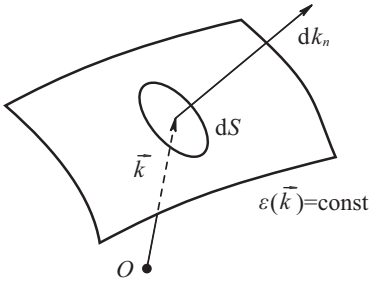


Fig. 1.39: The transition from the integration over the volume element in reciprocal space to the integration over the constant-energy surface and the normal to it.

The integrals over $\vec{k}$ can be converted as follows. Let $\varepsilon(\vec{k}) = \text{const}$ be the surface of constant energy in k-space. Then, the integration over $\mathrm{d}^3\vec{k}$ can be divided into the integration over the constant-energy surface and the normal to it (Figure 1.39). If $\mathrm{d}S$ is an element of the isoenergetic surface, then $\mathrm{d}^3\vec{k} = \mathrm{d}k_n\mathrm{d}S$, where $\mathrm{d}k_n$ stands for the integration over the normal to the element.

It is worth keeping in mind that the gradient in $\vec{k}$ of the function $\varepsilon(\vec{k})$ or $\vec{\nabla}_{\vec{k}}\varepsilon(\vec{k})$ is a vector perpendicular to the surface $\varepsilon(\vec{k}) = \text{const}$, and therefore:

$$\mathrm{d}\varepsilon = \left(\vec{\nabla}_{\vec{k}}\varepsilon \cdot \mathrm{d}\vec{k}\right) \equiv \left|\vec{\nabla}_{\vec{k}}\varepsilon\right| \mathrm{d}k_n \quad \Rightarrow \quad \mathrm{d}k_n = \mathrm{d}\varepsilon / \left|\vec{\nabla}_{\vec{k}}\varepsilon\right| . \tag{1.95}$$

The above outcome allows us to rewrite (1.87) in the form:

$$\frac{\mathrm{d}n_{\vec{k}}}{V} = \frac{\mathrm{d}^3\vec{k}}{4\pi^3} = \frac{\mathrm{d}\varepsilon\mathrm{d}S}{4\pi^3\left|\vec{\nabla}_{\vec{k}}\varepsilon\right|} = \frac{\mathrm{d}\varepsilon}{4\pi^3}\frac{\mathrm{d}S}{\hbar v} , \tag{1.96}$$

where $\vec{v} = (1/\hbar)\nabla_{\vec{k}}\varepsilon$ is the group velocity of the wave packet describing an electron.

We denote through $\nu(\varepsilon)\mathrm{d}\varepsilon$ the quantity:

$$\nu(\varepsilon)\,\mathrm{d}\varepsilon = \frac{1}{V}\int\limits_{\varepsilon\le\varepsilon(\vec{k})\le\varepsilon+\mathrm{d}\varepsilon} \mathrm{d}n_{\vec{k}} = \frac{\mathrm{d}\varepsilon}{4\pi^3}\int\limits_{\varepsilon(\vec{k})=\varepsilon=\text{const}} \frac{\mathrm{d}S}{\hbar v} . \tag{1.97}$$

The function $\nu(\varepsilon)$ is called the level density or the density of states. This is because the quantity $\nu(\varepsilon)$ yields the number of the energy levels of an electron per unit volume of a crystal in the range of energy values from ε to $(\varepsilon + \mathrm{d}\varepsilon)$.

To clarify the physical meaning of this quantity, we should calculate $\nu(\varepsilon)$ for the free electron gas, using the advance knowledge:

$$\vec{v} = \frac{\hbar\vec{k}}{m} , \quad \varepsilon(\vec{k}) = \frac{\hbar^2k^2}{2m} , \quad k = \sqrt{\frac{2m\varepsilon}{\hbar^2}} . \tag{1.98}$$

By applying formulas (1.97) and (1.98), we find that:

$$\nu(\varepsilon) = \frac{1}{4\pi^3}\int\limits_{\varepsilon(\vec{k})=\varepsilon=\text{const}} \frac{\mathrm{d}S}{\hbar v} = \frac{1}{4\pi^3}\frac{m}{\hbar^2k}4\pi k^2 = \frac{mk}{\pi^2\hbar^2} = \frac{m}{(\pi\hbar)^2}\sqrt{\frac{2m\varepsilon}{\hbar^2}} . \tag{1.99}$$

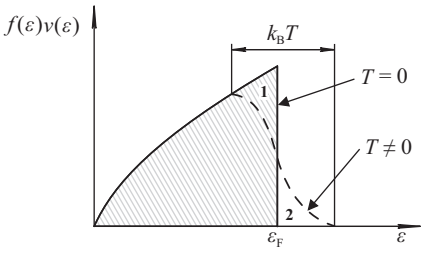


Fig. 1.40: The ε-dependent function $f(\varepsilon)\nu(\varepsilon)$. The solid curve corresponds to states occupied by an electron at absolute zero. The shaded area gives the number of electrons per unit volume of a metal. The dashed line describes the nature of filling the states by electrons at some finite temperatures $T > 0$ as $k_B T \ll \varepsilon_F$. As the system temperature rises from 0 K to T, a part of the electrons pass from the zone 1 to 2 due to thermal excitation. The areas under the dotted and solid curves are the same.

The quantity $\nu(\varepsilon)\mathrm{d}\varepsilon$ governs the number of permitted states, but our concern is the number of states occupied by the electrons. Therefore, we introduce another characteristic $\mathrm{d}N/V$, the density of electrons being in a state of thermodynamic equilibrium and filling the energy levels in the range from ε to $(\varepsilon + \mathrm{d}\varepsilon)$:

$$\frac{\mathrm{d}N}{V} = f(\varepsilon)\,\nu(\varepsilon)\,\mathrm{d}\varepsilon\,. \tag{1.100}$$

The function $f(\varepsilon)\nu(\varepsilon)$ describes the nature of filling the electron levels. The plot of its dependence on ε is shown in Figure 1.40.

Recall that the function $\varepsilon_n(\vec{k})$ is periodic in reciprocal space and bounded above and below for each value of n. In addition, it is differentiable. Hence, it follows that, in each Brillouin zone, the function has at least one local minimum and one local maximum. The extremum points are determined by the extremum condition $\nabla_{\vec{k}}\varepsilon = 0$.

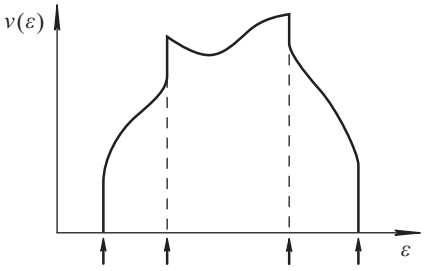


Fig. 1.41: Van Hove singularities in the density levels (indicated by arrows perpendicular to the axis ε).

When $\nabla_{\vec{k}}\varepsilon(\vec{k})$ vanishes, a singularity for the density levels appears in the integrand of formula (1.97). It can be shown that similar features may be integrable in the three-dimensional case and provide the finite values of $\nu(\varepsilon)$ for the density of states. However, they still lead to an infinite growth of the quantity $\mathrm{d}\nu(\varepsilon)/\mathrm{d}\varepsilon$ (Fig. 1.41). Such singularities are called *Van Hove singularities*.

Later, we will see that the value of the derivative $d\nu(\varepsilon)/d\varepsilon$ for $\varepsilon = \varepsilon_F$ involves low temperature expansion of various observables, particularly in the expansion of the chemical potential of the electron gas in powers of temperature. Consequently, we can conclude that if the Fermi surface has points where $\nabla_{\vec{k}}\varepsilon(\vec{k}) = 0$, we should expect anomalies in the low temperature properties of metals. I.M. Lifshitz was the first to predict their existence [5]. Furthermore, for simplicity, we assume that the state density $\nu(\varepsilon)$ has no Van Hove singularities at the Fermi surface of the electrons.

The transition from the integration over $d^3\vec{k}$ to the integration over ε makes the system (1.94) take the form:

$$\begin{aligned} c &= -\int_0^\infty d\varepsilon \;\; \varepsilon\nu(\varepsilon)\frac{\partial f}{\partial \varepsilon}\left[\frac{\varepsilon-\mu}{T}+\frac{d\mu}{dT}\right], \\ &\int_0^\infty d\varepsilon\left[\frac{\varepsilon-\mu}{T}+\frac{d\mu}{dT}\right]\nu(\varepsilon)\frac{\partial f}{\partial \varepsilon} = 0 . \end{aligned} \tag{1.101}$$

Both formulas contain $\partial f/\partial\varepsilon$. However, if $k_B T \ll \mu$, what, as we have already said, is always true for metals, then $\partial f/\partial\varepsilon$ is different from zero only in a small region of energies of the order of $k_B T$ near the value μ (see Figure 1.26). This makes it possible to formulate a general method to approximately calculate integrals of the type:

$$\int d\varepsilon F(\varepsilon)\frac{\partial f}{\partial \varepsilon} . \tag{1.102}$$

Since $\partial f/\partial\varepsilon$ is nonzero only near $\varepsilon = \mu$, we expand the function $F(\varepsilon)$ in the integrand in (1.102) in a Taylor series in powers of $(\varepsilon - \mu)$:

$$F(\varepsilon) = F(\mu) + (\varepsilon-\mu)F'(\mu) + \frac{1}{2!}(\varepsilon-\mu)^2 F''(\mu) + \cdots . \tag{1.103}$$

Certainly, this can be done only in the case when F does not vary too rapidly in the vicinity of the point $\varepsilon = \mu$.

Now, we calculate:

$$\frac{\partial f}{\partial \varepsilon} = -\frac{1}{4k_B T}\frac{1}{\mathrm{ch}^2\left[(\varepsilon-\mu)/2k_B T\right]} . \tag{1.104}$$

In view of the rapid decrease of this function as the latter moves away from the point $\varepsilon = \mu$, the limits of integration over ε in formula (1.102) can be taken from $-\infty$ to $+\infty$. By virtue of parity of function (1.104), the odd terms of the expansion $F(\varepsilon)$ vanish after integrating formula (1.102). The remaining terms can be expressed through the integrals, which are easily calculable:

$$\int_{-\infty}^{\infty} d\varepsilon \frac{\partial f}{\partial \varepsilon} = -1 . \tag{1.105}$$

$$\int_{-\infty}^{\infty} d\varepsilon (\varepsilon-\mu)^2 \frac{\partial f}{\partial \varepsilon} \approx -(k_B T)^2\left(2\int_{-\infty}^{\infty}\frac{\xi^2 d\xi}{\mathrm{ch}^2\xi}\right) = -(k_B T)^2\frac{\pi^2}{3} . \tag{1.106}$$

In computing (1.106), we have taken into account that:

$$2\int_{-\infty}^{+\infty} \mathrm{d}\xi \frac{\xi^2}{\mathrm{ch}^2\xi} = \frac{\pi^2}{3} .$$

As a result, we have:

$$\int \mathrm{d}\varepsilon F(\varepsilon) \frac{\partial f}{\partial \varepsilon} = -F(\mu) - \frac{(\pi k_\mathrm{B} T)^2}{6} F''(\mu) + \cdots . \tag{1.107}$$

Applying the rule of approximate calculation of the integrals (1.107) to the system (1.101), we find:

$$\begin{aligned} c &\approx \mu\nu(\mu)\frac{\mathrm{d}\mu}{\mathrm{d}T} + \frac{\pi^2 T k_\mathrm{B}^2}{3}\frac{\mathrm{d}}{\mathrm{d}\mu}\left[\mu\nu(\mu)\right] , \\ \nu(\mu)\frac{\mathrm{d}\mu}{\mathrm{d}T} &+ \frac{\pi^2 T k_\mathrm{B}^2}{3}\frac{\mathrm{d}}{\mathrm{d}\mu}\nu(\mu) \approx 0 . \end{aligned} \tag{1.108}$$

In deriving these equations, we have left out the higher order corrections to the coefficient beside $\mathrm{d}\mu/\mathrm{d}T$ because $\mathrm{d}\mu/\mathrm{d}T$ is considered to be a small quantity:

$$\frac{1}{k_\mathrm{B}}\frac{\mathrm{d}\mu}{\mathrm{d}T} \sim \frac{k_\mathrm{B} T}{\varepsilon_\mathrm{F}} \ll 1 . \tag{1.109}$$

During the calculations, we substantiate this assumption and the estimate given. Suppose that $\mu(T)$ differs little from its value at $T = 0\,\mathrm{K}$, i.e., from ε_F: $\mu \approx \varepsilon_\mathrm{F}$. Then, in the last equation of the system (1.108), we can replace the functions $\nu(\mu)$ and $\nu'(\mu)$ by their values $\nu(\varepsilon_\mathrm{F})$ and $\nu'(\varepsilon_\mathrm{F})$, respectively. After that, the equation acquires a simple form:

$$\nu(\varepsilon_\mathrm{F})\frac{\mathrm{d}\mu}{\mathrm{d}T} + \frac{\pi^2 T k_\mathrm{B}^2}{3}\nu'(\varepsilon_\mathrm{F}) = 0 . \tag{1.110}$$

It can now be easily integrated:

$$\mu(T) = \varepsilon_\mathrm{F} - \frac{1}{6}(\pi k_\mathrm{B} T)^2 \frac{\nu'(\varepsilon_\mathrm{F})}{\nu(\varepsilon_\mathrm{F})} . \tag{1.111}$$

Here, the summand ε_F gives the main contribution as $(\pi k_\mathrm{B} T)^2 \nu'(\varepsilon_\mathrm{F})/(6\nu(\varepsilon_\mathrm{F}))$, which is a correction term.

To estimate the latter, we consider a free electron gas for which

$$\nu(\varepsilon) \sim \sqrt{\varepsilon} \quad \Rightarrow \quad \frac{\nu'(\varepsilon_\mathrm{F})}{\nu(\varepsilon_\mathrm{F})} = \frac{1}{2\varepsilon_\mathrm{F}} . \tag{1.112}$$

In this case, the relative magnitude of the correction

$$\frac{(k_\mathrm{B} T)^2}{\varepsilon_\mathrm{F}} \frac{\nu'(\varepsilon_\mathrm{F})}{\nu(\varepsilon_\mathrm{F})} \sim \frac{(k_\mathrm{B} T)^2}{\varepsilon_\mathrm{F}^2} = \left(\frac{T}{T_\mathrm{F}}\right)^2 \tag{1.113}$$

amounts to about 10^{-4}, even at room temperature. This estimate justifies our presuppositions:

$$\mu \approx \varepsilon_F , \quad \frac{1}{k_B}\frac{d\mu}{dT} \sim \frac{k_B T}{\varepsilon_F} \ll 1 .$$

Note that, at $T \neq 0$ K, the chemical potential $\mu(T)$ in (1.111) is less than ε_F because $\nu'(\varepsilon_F)/\nu(\varepsilon_F) > 0$ as a rule.

Substituting the second equation of (1.108) into the first one, we find:

$$c = -\frac{\pi^2 k_B^2 T}{3}\mu\nu'(\mu) + \frac{\pi^2 k_B^2 T}{3}(\mu\nu(\mu))' = \frac{\pi^2 k_B^2 T}{3}\nu(\mu) . \tag{1.114}$$

Given that $\mu \approx \varepsilon_F$ in the main approximation, we finally obtain the following equation:

$$c = \frac{\pi^2 k_B^2 T}{3}\nu(\varepsilon_F) . \tag{1.115}$$

The quantitative estimate of the electron heat capacity can be made by an example of the free electron gas. For free electrons:

$$\nu(\varepsilon_F) = \frac{m}{(\pi\hbar)^2}\sqrt{\frac{2m\varepsilon_F}{\hbar^2}} , \tag{1.116}$$

$$\varepsilon_F = \frac{\hbar^2}{2m}\left(3\pi^2 n\right)^{\frac{2}{3}} . \tag{1.117}$$

When we rewrite formula (1.117) in the form of:

$$\sqrt{\varepsilon_F} = \left(\frac{\hbar^2}{2m}\right)^{3/2}\frac{3\pi^2 n}{\varepsilon_F} \tag{1.118}$$

and plug it into (1.116), we get:

$$\nu(\varepsilon_F) = \frac{3}{2}\frac{n}{\varepsilon_F} . \tag{1.119}$$

Using (1.119), we transform expression (1.115) for the electron heat capacity of a metal:

$$c = \frac{\pi^2}{2}\left(\frac{k_B T}{\varepsilon_F}\right) n k_B . \tag{1.120}$$

Comparing (1.120) with the result of classical physics for the specific heat of an ideal gas

$$c = \frac{3}{2} n k_B , \tag{1.121}$$

we see that the Fermi–Dirac statistics leads to a decrease in the specific heat of the electron gas due to the multiplier $\sim k_B T/\varepsilon_F$ that is proportional to temperature. The multiplier's magnitude order is 10^{-2}, even at room temperature. The quantum theory

thus explains the absence of the observed electron contribution to the heat capacity of metals at room temperature, which was a blatant discrepancy between theory and experiment in classical physics. Classical physics could not give an answer to the question: "Why do electrons have the ability to move freely through a metal, but contribute very little to the heat capacity at the same time?" Only the Pauli principle and the Fermi–Dirac distribution provide an answer to this question.

In deriving the expression for the electron heat capacity of a metal, we needed no specific information about the Fermi surface. Consequently, the electron specific heat is proportional to temperature for all metals. This quantum result can be understood qualitatively even without resorting to any calculations.

When we heat a sample from absolute zero, not all the electrons receive the energy $\sim k_B T$, as they should, according to the classical theory of gases. Only those electrons suffer thermal excitation, for which the energies are disturbed over a range of the width $k_B T$ near the Fermi level. The amount of excess energy received by the electrons is of the order of $k_B T$ as shown in Figure 1.26. This allows us to evaluate the heat capacity of the conduction electron gas. Suppose n is the total number of electrons per unit volume of a metal, then only the part T/T_F of these electrons belonging to the Fermi thermal layer can undergo thermal excitation when the temperature rises from zero to T. In other words, nT/T_F governs the heat capacity of a metal rather than n. Every electron of nT/T_F has the excess thermal energy of the order of $k_B T$. Therefore, the total (thermal excitation) energy ΔE of the electrons is in the order of:

$$\Delta E \sim \frac{nT}{T_F} k_B T \,. \tag{1.122}$$

The electron heat capacity of metals can be obtained in the usual way – by taking the derivative of ΔE over the absolute temperature:

$$c = \frac{\partial \Delta E}{\partial T} \sim n k_B \frac{T}{T_F} \,. \tag{1.123}$$

The result (1.123) coincides with the exact calculation (1.120) up to a factor of the order of unity.

Let us return again to formula (1.115) for the electron heat capacity of a metal to rewrite it differently. Given that:

$$\nu(\varepsilon_F) = \frac{m k_F}{(\pi\hbar)^2} \,,$$

we find:

$$c = \frac{k_B^2 T m k_F}{3\hbar^2} \,. \tag{1.124}$$

The quantity k_F has the same order for all metals: $k_F \sim \pi/a$, where a is the lattice period, which is known to be a little different for different substances. Therefore, at a given temperature, the electron heat capacity depends only on the effective mass m of quasiparticles (electrons). However, the effective mass can vary strongly from metal to metal.

The effective mass forms as a result of the following interactions:

1. Interaction of conduction electrons with the periodic potential of a fixed (rigid) crystal lattice. The effective mass of an electron in this potential field is referred to as the *band effective mass*. It may even be less than the mass of a free electron.
2. Interaction of conduction electrons with lattice vibrations. An electron tends to distort the lattice around itself. The electron travels through a metal, so as if it drags ions encountered along their path. The effect makes itself felt through an increase in their effective mass.
3. Interaction of conduction electrons between themselves. A moving electron is surrounded by a cloud of the perturbed electron gas, which also causes its effective mass to increase as well.
4. Interaction of conduction electrons with atomic nuclei. In this case, an essential role is played by a great mass of quasiparticles – electrons belonging to weakly overlapping unfilled inner shells of atoms in transition metals.

Let us explain the last statement. When N separated atoms are brought together to form a complete crystal, the atom-atom interaction results in splitting each energy level of the isolated atoms into a band of closely spaced sublevels. Such a band, due to the weak overlap of the inner atomic shells, has a small width of the order of $\Delta\varepsilon$. It can be shown that the mass of quasiparticles in this band is $m \sim 1/\Delta\varepsilon$. The width $\Delta\varepsilon$ being small, the mass of the quasiparticles associated with the electrons is large. This is the case of the so called strong coupling of electrons with the atomic nuclei.

Of course, an experiment measures not electrons but the total heat capacity of a metal. The total heat capacity is the heat capacity of both electrons and a crystal lattice. At room and higher temperatures, the total heat capacity is mainly contributed, not by electrons, but thermal vibrations of crystal lattice ions. However, at temperatures well below room temperature, a part of the heat capacity associated with the lattice vibrations decreases by the cubic law $c \sim T^3$. At very low temperatures, it becomes less than the linearly temperature dependent electron gas heat capacity. Therefore, the total heat capacity of metals at low temperatures can be represented as:

$$c_{\text{full}} = AT + BT^3 \ . \tag{1.125}$$

Having plotted the dependence of c_{full}/T on T^2 (Figure 1.42), we can see that, at low temperatures, the graph is a straight line. Using the graph, we can know both the coefficients A and B.

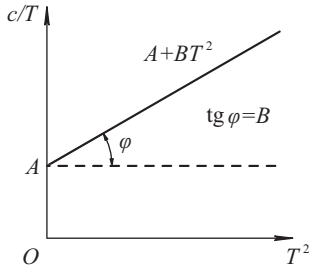


Fig. 1.42: Graph of the dependence of c_{full}/T on T^2.

1.9 Screening of the Coulomb Field of External Electric Charges in Metals (the Thomas–Fermi Model) and Semiconductors

As used herein, a single electron model is underlain by an assumption that many particle Coulomb interactions between electrons and ions in a crystal are almost completely screened. Let us give a quantitative estimate of the screening effect [7].

For this purpose, we will examine the motion of electrons in a simplified model of a metal, or in the so called jelly model (the Sommerfeld model). Such a model replaces some ions of a crystal by a positive fixed charge background uniformly distributed through the crystal. Here, the ionic charges are assumed to be distributed uniformly only because their real periodic arrangement makes the task daunting. At the same time, the basic collective effects of charge screening are easier to estimate by applying the jelly model. The positive background provides the electroneutrality condition. With this model being at equilibrium, the densities of the positive and negative charges are equal to a constant n_0 everywhere inside the sample.

To comprehend what happens with the field of any noncompensated charge in metals, we should consider the following simple problem. Suppose Q is an external charge placed inside a metal at a point with the radius vector $\vec{r} = \vec{0}$. This will result in a spatial redistribution of electrons in the metal. The constant electron density n_0 is replaced by the variable $n(\vec{r})$ and an electric field with the strength $\vec{E}$ appears. The latter creates a force $\vec{F}$ acting on a single electron in the metal:

$$\vec{F} = -|e|\vec{E} .$$

It is known from electrostatics that:

$$\vec{E} = -\vec{\nabla}\varphi ,$$

where the potential φ is the solution of the equation:

$$-\Delta\varphi = 4\pi\rho .$$

Here, ρ is the total uncompensated charge density:

$$\rho = \rho_{\text{ext}} + \rho_{\text{ind}} .$$

It includes the density of the external (imbedded) charge:

$$\rho_{\text{ext}} = Q\delta(\vec{r})$$

and the charge density of the redistributed electrons:

$$\rho_{\text{ind}} = -|e|\left(n(\vec{r}) - n_0\right) .$$

Thus, to find the resulting electric field $\vec{E}$ inside the metal, we need to solve the equation:

$$-\Delta\varphi = 4\pi\left(Q\delta(\vec{r}) - |e|\left[n(\vec{r}) - n_0\right]\right) . \tag{1.126}$$

This equation is not closed because the distribution of $n(\vec{r})$ depends on the potential φ.

In principle, the electron density $n(\vec{r})$ can be expressed in terms of single particle wave functions Ψ_i which are solutions of the Schrödinger equations:

$$\left(-\frac{\hbar^2}{2m}\Delta - |e|\,\varphi\right)\Psi_i = \varepsilon_i\Psi_i ,$$

$$n(\vec{r}) = \sum_{i=1}^{N} |\Psi_i|^2 .$$

However, such a consequent way to calculate $n(\vec{r})$ is complex. The calculations can be simplified by assuming that the potential $\varphi(\vec{r})$ is a smoothly varying function of $\vec{r}$. Suppose the characteristic scale L of changes in the function $\varphi(\vec{r})$ are much larger than the de Broglie wavelength of an electron:

$$L \sim \left|\frac{\varphi}{\partial\varphi/\partial\vec{r}}\right| \gg \frac{2\pi}{k_{\text{F}}} \sim a . \tag{1.127}$$

Here, a is an interatomic distance. Now the energy dependence of the electron on its wave vector can be defined at every point in space. Consequently, it would be reasonable to admit that the dependence is determined within the quasiclassical approximation:

$$\varepsilon(\vec{k},\vec{r}) = \varepsilon_0 - |e|\,\varphi(\vec{r}) , \quad \varepsilon_0 = \frac{\hbar^2k^2}{2m} , \tag{1.128}$$

where $\vec{k}$ is an electron wave vector and m is its mass.

Recall that the electron density in the absence of the external charge can be calculated using the Fermi–Dirac distribution:

$$n_0(\mu) = 2\int \frac{\mathrm{d}^3\vec{k}}{(2\pi)^3}\frac{1}{\exp\left[\left(\varepsilon_0 - \mu\right)/k_{\text{B}}T\right] + 1} . \tag{1.129}$$

Since the external charge changes the energy electron (1.128) when it appears, the density (1.129) also changes:

$$n(\mu) = 2\int \frac{\mathrm{d}^3\vec{k}}{(2\pi)^3}\frac{1}{\exp\left[\left(\varepsilon_0 - \mu - |e|\,\varphi\right)/k_{\text{B}}T\right] + 1} = n_0\left(\mu + |e|\,\varphi\right) . \tag{1.130}$$

The chemical potential values can be considered to coincide in formulas (1.129) and (1.130), under the condition that the potential φ is different from zero only in a finite spatial region, outside which a perturbation of the electron density is small.

We assume that:

$$\left|e\varphi(\vec{r})\right| \ll \mu \approx \varepsilon_{\mathrm{F}} \,. \tag{1.131}$$

In this case, ρ_{ind} can be expanded in a Taylor series in $|e|\varphi(\vec{r})$. Then we restrict ourselves to only the first term of the expansion:

$$\rho_{\mathrm{ind}} = -|e|\left[n_0(\mu + |e|\,\varphi) - n_0(\mu)\right] \approx -e^2\frac{\partial n}{\partial \mu}\varphi(\vec{r}) \,. \tag{1.132}$$

Using the above method to calculate integrals approximately, it is easy to show that, for a free electron gas up to room temperature and above, the relation is true:

$$\frac{\partial n_0}{\partial \mu} \approx \nu(\varepsilon_{\mathrm{F}}) \,. \tag{1.133}$$

Here, $\nu(\varepsilon_{\mathrm{F}})$ is the density of states at the Fermi surface.

As a result, equation (1.126) for computing the potential φ becomes a simple closed form:

$$-\Delta\varphi = 4\pi Q\delta(\vec{r}) - \lambda^2\varphi \,, \tag{1.134}$$

where

$$\lambda^2 = \frac{4me^2k_{\mathrm{F}}}{\hbar^2\pi} \,. \tag{1.135}$$

Let us estimate the constant λ. For this, we introduce the Bohr radius:

$$a_{\mathrm{B}} = \frac{\hbar^2}{me^2} \,, \tag{1.136}$$

and use the estimate for the radius of the Fermi sphere:

$$k_{\mathrm{F}} \sim \frac{\pi}{a} \,, \tag{1.137}$$

where a is an interatomic distance. Plugging formulas (1.136) and (1.137) into (1.135), and given that $a_{\mathrm{B}} \sim a$, we find:

$$\lambda^2 = \frac{4k_{\mathrm{F}}}{\pi a_{\mathrm{B}}} \sim \frac{1}{a^2} \quad \Rightarrow \quad \lambda \sim a^{-1} \,. \tag{1.138}$$

By directly inserting (1.139) into (1.134), it is not difficult to make sure that for $\vec{r} \neq 0$ the solution of the Poisson equation has the form:

$$\varphi = \frac{Q}{r}\exp(-\lambda r) \,. \tag{1.139}$$

Let us show that the singularity of the solution (1.139) at the point with the radius vector $\vec{r} = \vec{0}$ is consistent with the singularity of the term $4\pi Q\delta(\vec{r})$ in equation (1.134). To do this, we integrate equation (1.134) over the spherical volume V by radius $r_0 \to 0$ that surrounds the origin.

Now, using the Gauss–Ostrogradsky theorem, we transform the integral in the left-hand side of equation (1.134) into the integral over the surface of a sphere with radius r_0:

$$-\int\limits_V \Delta\varphi \mathrm{d}^3\vec{r} = -\int\limits_V \mathrm{div}\vec{\nabla}\varphi \mathrm{d}^3\vec{r} = -\oint\limits_S \vec{\nabla}\varphi \cdot \mathrm{d}\vec{S} \,.$$

In the main approximation for $r_0 \to 0$, we have:

$$\vec{\nabla}\varphi \approx \vec{\nabla}\frac{Q}{r} = -\frac{\vec{r}}{r^3}Q \,.$$

Therefore, in the limit $r_0 \to 0$, we find:

$$-\int\limits_V \Delta\varphi \mathrm{d}^3\vec{r} = \oint\limits_S \frac{Q}{r^3}\vec{r} \cdot \mathrm{d}\vec{S} = \oint\limits_S \frac{1}{r^2} Q \mathrm{d}S \underset{r_0\to 0}{=} 4\pi Q \,.$$

After integrating a small volume V, the right-hand side of equation (1.134) yields:

$$4\pi Q \overbrace{\int\limits_V \mathrm{d}^3\vec{r}\delta(\vec{r})}^{1} - \lambda^2 \int\limits_V \mathrm{d}^3\vec{r}\varphi \,. \tag{1.140}$$

We show that the contribution from the second term in (1.140) in the limit $r_0 \to 0$ is zero. To this end, we write the volume element in spherical coordinates: $\mathrm{d}^3\vec{r} = r^2 \mathrm{d}r\mathrm{d}\Omega$. Integrating over the solid angle $\mathrm{d}\Omega$ gives a constant. We compute $\int \varphi \mathrm{d}^3 r$ with accuracy up to this constant. In the limit $r_0 \to 0$, we can hold that $\varphi \approx 1/r$, therefore:

$$\int \varphi \mathrm{d}^3\vec{r} = \mathrm{const} \cdot \int\limits_0^{r_0} r\mathrm{d}r \overset{r_0\to 0}{\longrightarrow} 0 \,.$$

It is clear that the integration of both sides of equation (1.134) in the limit $r_0 \to 0$ leads to the same results. Consequently, the singularities of equation (1.134) and the solution (1.139) are consistent.

Let us analyze the solution obtained:

$$\varphi = \frac{Q}{r}\exp(-\lambda r) \,, \quad \lambda \sim \frac{1}{a} \,.$$

Within the sphere of radius of the order of the interatomic distances (for $\lambda r < 1$), the potential $\varphi(\vec{r})$ of the point charge Q in a metal is the same as in a vacuum: $\varphi \approx Q/r$. Simultaneously, outside the sphere specified (for $\lambda r > 1$), we have $\exp(-\lambda r) \approx 0$. Therefore, $\varphi \approx 0$. That is to say, the free electrons redistributed completely shield the electric field of the embedded charge in a metal already at distances of the order of the interatomic one.

Here, it is worth emphasizing that, despite the quasiclassical nature of the calculations, the finding takes the electron energy distribution into account in accordance

with the Fermi–Dirac quantum statistics. Let us discuss what would change if we carried out the calculations within classical physics.

The density of a classical ideal electron gas in an external potential field $U(\vec{r})$ is described by the Boltzmann distribution:

$$n(\vec{r}) = n_0 \exp\left(-\frac{U(\vec{r})}{k_B T}\right) .$$

In this case $U(\vec{r}) = -|e|\varphi(\vec{r})$, and therefore:

$$n(\vec{r}) = n_0 \exp\left(\frac{|e|\,\varphi}{k_B T}\right) . \tag{1.141}$$

Suppose $|e\varphi(\vec{r})| \ll k_B T$. We can now expand the exponential factor in formula (1.141) in a Taylor series in the small parameter $|e|\varphi/k_B T$ and restrict ourselves to the first two terms of the expansion:

$$n \approx n_0 \left(1 + \frac{|e|\,\varphi}{k_B T}\right) .$$

In the long run:

$$\rho_{\text{ind}} = -|e|\left[n(\vec{r}) - n_0(\vec{r})\right] \approx -\frac{e^2 n_0}{k_B T}\varphi(\vec{r}) .$$

Thus, it can be concluded that classical physics still gives equation (1.134) to calculate the potential φ. However, the length of screening of the external charge by the plasma electrons turns out to be different:

$$\lambda^{-1} = \sqrt{\frac{k_B T}{4\pi n_0 e^2}} . \tag{1.142}$$

It is interesting and important that the Fermi–Dirac distribution for electrons and holes in semiconductors reduces to the Boltzmann distribution. This means that the screening of interactions between the electric charges in semiconductors is well described by classical physics. Later we will justify this statement.

Thus, we have demonstrated that electrons in a metal screen the distribution of an embedded external electric charge. However, the screening also affects the electron-electron interactions because every electron can be regarded as an external charge.

To further discuss the dynamic screening of the electron-electron interactions, we should be familiar with basic collective excitations of the electron gas in metals with properties of quasiparticles (*plasmons*).

1.10 Plasmons and Dynamic Screening of the Electron-Electron Interactions in Metals

We start from the model where the ions of the lattice constitute a uniformly distributed background of the positive charge.

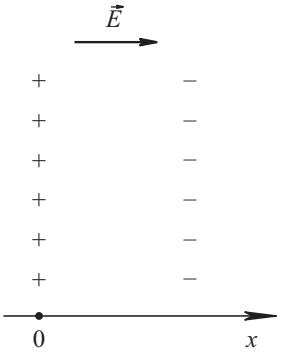


Fig. 1.43: Local change in the equilibrium density of electrons in a metal.

Suppose a local change in the density of the electrons has occurred as shown in Figure 1.43. After the electrons left the region, the positively charged fixed ions formed. Conversely, where the electrons reached, the negative charge became excessive. As a result, due to the imbalance of charges, an electric field $\vec{E}$ appears. It acts with the force $\vec{F} = -|e|\vec{E}$ on every electron and tries to return it to its former equilibrium position with the coordinate $x = 0$. Under the influence of this force, the electrons move towards the initial equilibrium position and flow through it by inertia. Thus, an electric field reappears and it again tries to get the electron back to its equilibrium position with the coordinate $x = 0$. The process takes a periodic nature. This is electron plasma oscillations, which will be described below.

To simplify the analysis, we accept that the motion of each center of mass wave packet corresponding to an electron is described by Newton's second law:

$$m\ddot{x} = -|e|\,E\,, \tag{1.143}$$

where m is the mass of an electron, E is the component of the electric field along the axis Ox, and x is the coordinate of the electron. The validity of this approximation is discussed later.

Electrostatic induction $\vec{D}$ in a metal is determined by the Maxwell equation:

$$\operatorname{div}\vec{D} = 4\pi\rho_{\text{external}}\,. \tag{1.144}$$

In the given case, the metal has no external charges ($\rho_{\text{external}} = 0$) and the problem is one-dimensional. Equation (1.144) acquires the form:

$$\frac{\partial D}{\partial x} = 0\,. \tag{1.145}$$

Here, D is the component of the vector $\vec{D}$ directed along the axis Ox: $\vec{D} = (D, 0, 0)$. Since the field $\vec{D} \to 0$ as $x \to \pm\infty$, a solution of equation (1.145) is:

$$\vec{D} \equiv 0\,. \tag{1.146}$$

However, we should note that:

$$\vec{D} = \vec{E} + 4\pi\vec{P}\,, \tag{1.147}$$

where $\vec{E}$ is the electric field strength in a metal and $\vec{P}$ is the dipole moment per unit volume of medium. All the vectors in the problem at hand have only x-components. Using the equations (1.146) and (1.147), we express the electric field via the electric dipole moment $\vec{P}$ per unit volume of the medium:

$$\vec{E} = -4\pi\vec{P}\,.$$

The electric dipole moment $\vec{P}$ can be easily calculated:

$$\vec{P} = n_0\vec{p}\,.$$

Here, n_0 is the concentration of electrons in a metal, and $\vec{p} = (-|e|x, 0, 0)$ is the dipole moment appearing due to the shift from the equilibrium position of a single electron.

Thus, the x-component of the electric field that returns the electron to its equilibrium position can be written as follows:

$$E = 4\pi\,|e|\,n_0 x\,. \tag{1.148}$$

Given formula (1.148), the equation of motion of the electron (1.143) is equivalent to the equation of a harmonic oscillator:

$$\ddot{x} = -\omega_{\mathrm{p}}^2 x\,, \tag{1.149}$$

where ω_{p} is the plasma oscillation frequency equal to:

$$\omega_{\mathrm{p}}^2 = \frac{4\pi n_0 e^2}{m}\,. \tag{1.150}$$

The solution of (1.149) is:

$$x = A\sin(\omega_{\mathrm{p}} t + \varphi)\,, \tag{1.151}$$

where A, φ are constants of integration. Formula (1.151) describes the longitudinal collective oscillations of an electron gas. When oscillated, the electron plasma condenses and discharges with respect to the fixed lattice ions. Such collective oscillations must occur due to any violation of the homogeneity of the electron distribution in a metal; they try to restore the violated electroneutrality of the electronion system.

The consistent quantum mechanical description of the electron density oscillations in a metal leads to the quantization of their energies: $E_n = \hbar\omega_{\mathrm{p}}(n + 1/2)$, where $n = 0, 1, 2, \ldots$.

By consequently solving the problem, the electron-electron interactions can be treated as a result of the exchange of the plasma oscillations with the energy $\hbar\omega_{\mathrm{p}}$ by virtual quanta. The plasma oscillation quanta play the same role as that of photons

in a vacuum. Like photons, they possess the properties of quasiparticles, so they are called plasmons.

Moving in a metal, the electron is continuously exchanged by quanta of the energy $\hbar\omega_p$ with the electrons surrounding it. Therefore, its energy is determined with accuracy up to the energy of an individual quantum: $\Delta E = \hbar\omega_p$. A time τ over which there is a change of the electron energy can be estimated from the uncertainty relation:

$$\Delta E\tau \sim \hbar\,.$$

Note that the virtual electron energy change occurs over a period of time so small that the virtual electron transitions are impossible to study experimentally. At the same time, an important consequence of the foregoing pattern is to relate the range of forces acting between the electrons and the quantum energy $\hbar\omega_p$ in a metal. In fact, for the time

$$\tau = \frac{\hbar}{\Delta E} = \frac{\hbar}{\hbar\omega_p} = \frac{1}{\omega_p}\,,$$

the virtual energy exchange between two electrons happens, and the latter can move apart at a distance not exceeding:

$$r_0 = 2\tau V_F\,,$$

where $2V_F$ is the relative electron velocity. When the electrons move in opposite directions, the distance is maximal (their velocity value is close to the Fermi velocity). We have:

$$\begin{aligned} r_0 &= \frac{2V_F}{\omega_p} = \frac{2\hbar k_F}{m\omega_p}\,, \quad \omega_p = \sqrt{\frac{4\pi n_0 e^2}{m}}\,, \quad n_0 = \frac{1}{3\pi^2}k_F^3 \quad \Rightarrow \\ r_0^2 &= \frac{4\hbar^2 k_F^2}{m^2}\cdot\frac{m}{4\pi n_0 e^2} = 3\pi\left(\frac{\hbar}{me^2}\right)\frac{1}{k_F}\,. \end{aligned} \tag{1.152}$$

The virtual plasmon cannot be spread at a distance greater than r_0. Consequently, the two electrons exchanging by plasmon cannot interact if the distance between them greatly exceeds r_0. Thus, the range of forces between the electrons in the plasma of a metal is of the order of r_0. Let us estimate r_0:

$$k_F \sim \frac{\pi}{a}\,, \quad a_B = \frac{\hbar}{me^2} \sim a \quad \Rightarrow \quad r_0^2 = 3\pi\left(\frac{\hbar}{me^2}\right)\frac{1}{k_F} \sim \frac{3\pi a_B a}{\pi} \sim 3a_B a \sim a^2\,.$$

In other words, the electrons in a metal have no interaction at distances greater than interatomic.

The resulting expression for r_0 up to a factor of the order of unity reproduces the previous formula for the length of the screening of interactions between charges in the electron plasma of a metal. As can be seen, two different approaches give the same result.

It is interesting that, with a quantum mechanical standpoint, the electron-electron interaction in a vacuum is also associated with the virtual photon exchange.

A similar calculation of the range of the interaction forces between two electrons in a vacuum yields the estimate:

$$r_0 \sim \frac{c}{\omega} ,$$

where ω is the frequency of a virtual photon and c is the speed of light. Since the frequency of the virtual photons can be made arbitrarily small, the range of the interaction forces between the electrons in a vacuum is endless. In full accordance with Coulomb's law $F \sim 1/r^2$, the interaction force between electrons in a vacuum vanishes only when $r \to \infty$. In a metal, the frequency of each plasmon cannot be less than ω_p, therefore the range of the interaction forces between the electrons is finite.

The concept of interaction of one-sort quasiparticles through intermediate fields of other quasiparticles is common in quantum field theory, quantum theory of solids, and particle physics.

Real, not virtual plasmons, can appear in the electron plasma of a metal when fast charged particles (or particle flux) have great energy flying through the metal. The energy required to excite a plasmon is very high and amounts to about 10 eV.

1.11 The Pauli Principle and Suppression of Electron-Electron Collisions in Metals

Within the model of dynamic screening of electron-electron interactions, the rate of electron-electron collisions in a metal remains high because the Coulomb interaction turns out to be strong enough; even when given the screening. In other words, the model cannot explain a large mean free path of electrons in a metal, which, as a rule, is 10^{-4}cm. In metals, another mechanism exists to suppress electron-electron collisions; it is more efficient and associated with the Pauli principle [3].

To illustrate how the Pauli principle affects the scattering frequency, we consider the next N-electronic state. Imagine the Fermi sphere was filled by $N-1$ electrons, and one electron had the energy $(\varepsilon_1 + \varepsilon_F)$, where $\varepsilon_1 > 0$. For further analysis, the level of ε_F is convenient to start measuring the electron energy. Using the free electron model, we now analyze features of the elastic interaction between this electron and an electron of the Fermi sphere at a temperature of absolute zero.

Obviously, the energy of the second electron is less than ε_F. We designate it as $(\varepsilon_F + \varepsilon_2)$, where $\varepsilon_2 < 0$. After interacting, the two electrons have energy higher than ε_F, because at $T = 0$ K, all energy levels inside the Fermi sphere are occupied. The trivial case, when the electrons exchange by their initial places, is not our concern because there is no scattering. The electron energy after the collision can be represented in the form $(\varepsilon_F + \varepsilon_3)$ and $(\varepsilon_F + \varepsilon_4)$, where $\varepsilon_3 \geq 0$ and $\varepsilon_4 \geq 0$.

In elastic collisions, the law of conservation of energy holds true:

$$\varepsilon_1 + \varepsilon_2 = \varepsilon_3 + \varepsilon_4 .$$

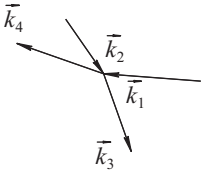


Fig. 1.44: Elastic collisions of electrons.

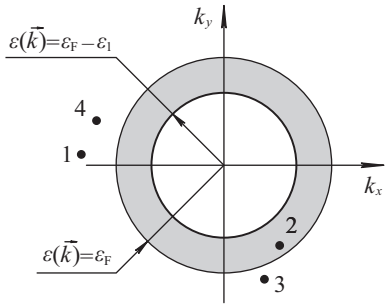


Fig. 1.45: Positions of the electrons with respect to the Fermi sphere (points 1 and 2) and positions of the electrons after the collision (points 3 and 4).

From the law of conservation and the condition $\varepsilon_3 \geq 0$, $\varepsilon_4 \geq 0$, we have:

$$|\varepsilon_2| \leq \varepsilon_1 \,. \tag{1.153}$$

Consequently, the appropriate "targets" for the first electron are not all the electrons inside the Fermi sphere, but only those that meet the condition (1.153). That is to say, those are located inside the spherical layer of a thickness of ε_1. Figure 1.45 shows this layer as a darkened bagel. The number of those electrons is $\varepsilon_1/\varepsilon_F$ times less than the total number of electrons.

When electrons collide elastically, the law of conservation of momentum also holds. Given that, for free electrons, the momentum is linearly related to the wave vector ($\vec{p} = \hbar\vec{k}$), we write the law of conservation of momentum in the form of the law of conservation of the total wave vector of the electrons (see Figure 1.44):

$$\vec{k}_1 + \vec{k}_2 = \vec{k}_3 + \vec{k}_4 \,. \tag{1.154}$$

It can be shown that (1.154) contributes to reducing the number of electrons capable of scattering by $\varepsilon_1/\varepsilon_F$ times. Thus, the number of electrons that may interact with the electron having the energy $(\varepsilon_F + \varepsilon_1) > \varepsilon_F$ is $(\varepsilon_1/\varepsilon_F)^2$ times less than the total number of electrons inside the Fermi sphere.

Consider a case when the temperature is not absolute zero. Here, ε_1 is the same order of magnitude as $k_B T$, which is far less than the Fermi energy. Let us estimate what part of the electrons, in this case, is capable of scattering at room temperature ($T = 300\,\mathrm{K}$). Supposing that $\varepsilon_F = 8\,\mathrm{eV}$, we obtain:

$$(k_B T/\varepsilon_F)^2 \sim \left(8.6 \cdot 10^{-5}\,\frac{\mathrm{eV}}{\mathrm{K}} \cdot 300\,\mathrm{K}/8\,\mathrm{eV}\right)^2 \sim 10^{-4} \,. \tag{1.155}$$

The interaction between the electrons is screened at interatomic distances of the order of $a \sim 10^{-8}$cm. Therefore, the size of an electron is of the order a. Its cross section

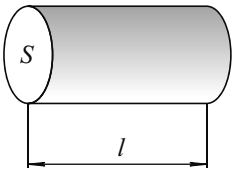


Fig. 1.46: A cylinder that "sweeps" the electron moving between collisions.

in this case is of the order of a^2. Given the fact that not every electron can be targeted, we find the area of the effective cross section of the electrons:

$$S \sim a^2 \left(k_B T/\varepsilon_F\right)^2 \ . \tag{1.156}$$

We estimate the mean free path of the electron, as is done in the course of general physics. The total concentration of electrons is assumed to be:

$$n \sim a^{-3} \ . \tag{1.157}$$

Every electron moves almost freely between collisions. We find the volume of the cylinder that "sweeps" the electron moving freely between two successive collisions (Figure 1.46):

$$V_{\text{cyl}} = Sl \ ,$$

where l is the mean free path of the electron. Since only one electron moves inside the volume V_{cyl}, we have:

$$Sln = 1 \ ,$$

or it can be written as:

$$l = 1/Sn \ . \tag{1.158}$$

Inserting formulas (1.156) and (1.157) into (1.158), we obtain the estimated relation for the mean free path of an electron in a metal:

$$l \sim \left(\varepsilon_F/k_B T\right)^2 a \gg a \ .$$

Using formula (1.155) and the value of $a \sim 10^{-8}$cm at room temperature, we find that $l \sim 10^{-4}$cm. Thus, the Pauli principle answers the central question of the theory of metals: "Why do electrons move almost freely in a metal?"

1.12 The Concept of the Mean Free Path of Electrons: Electrical and Thermal Conductivities of Metals and the Wiedemann–Franz law

The upper energy band partially filled by electrons governs high electrical and thermal conductivities of metals. As a comparison, the thermal conductivity of dielectrics and semiconductors is determined by their lattice. These phenomena cannot be ex-

plained by thinking that the potential of electrons in a metal is purely periodic. This is because, in such a potential, wave packets corresponding to the electrons move with constant group velocity and without any barriers. Due to comprising impurities or defects, the crystal lattice in metals is imperfect and responsible for resistance and thermal conductivity. Moreover, ion oscillations can disrupt the crystal periodicity.

We will not discuss here the full quantum mechanical theory of different electron scattering mechanisms in metals. It turns out that we can understand much, and even derive explicit formulas for the electrical and thermal conductivities of metals, by assuming that some scattering mechanisms exist and that they are completely characterized by the mean time τ (the *relaxation time approximation*) between collisions of electrons with scatterers. The probability of undergoing the collision per unit of time is $1/\tau$. The time τ is called the *relaxation time*. The electron mean free path l, or an average distance traveled by an electron in a metal between collisions, is related to τ by the following equation:

$$l = V_{\mathrm{F}}\tau\,. \tag{1.159}$$

Note that the parameter τ may be temperature dependent.

Based on this assumption, we find, first of all, the diffusion coefficient for electrons in metals. Imagine that the electron density is nonuniform along the axis Oz. The concentration gradient along the direction of the axis gives rise to the mean particle flux. Consider the plane $z = z_0$ (Figure 1.47). If all the electrons move in the same direction, the number of electrons passing through a unit area of the plane $z = z_0$ per unit of time would be equal to:

$$j = nV_z\,. \tag{1.160}$$

Supposing that the value of the modulus of the velocity vector for all the electrons is equal to V_{F}, we obtain an expression for the z-component of the velocity vector:

$$V_z = V_{\mathrm{F}}\cos\theta\,. \tag{1.161}$$

Here, it is worth recalling that the diffusion flux j_{D} is the average number of particles passing through a unit area per unit of time. The electrons diffuse with different velocities in different directions. So, to calculate the diffusion flux, we should average

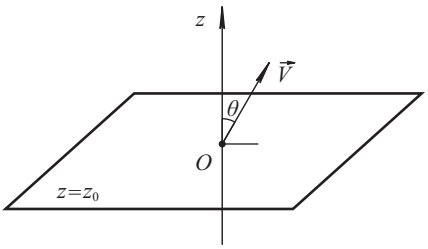


Fig. 1.47: The motion of the electrons through the plane $z = z_0$.

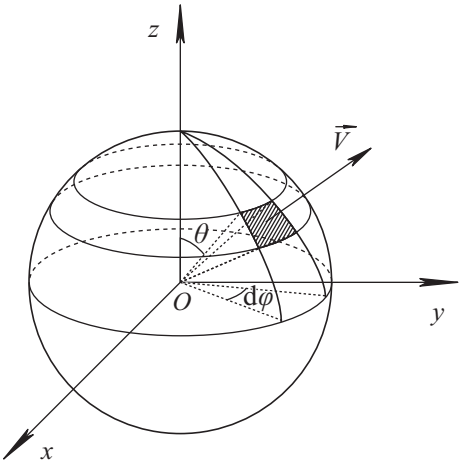


Fig. 1.48: A solid angle.

the electron flux density (1.160) in the velocity directions:

$$j_{\mathrm{D}} = \langle nV_z \rangle \ . \tag{1.162}$$

Suppose the electron velocity vector lies within a solid angle $\mathrm{d}\Omega = \sin\theta \mathrm{d}\theta \mathrm{d}\varphi$. By definition, the solid angle $\mathrm{d}\Omega$ is the area of a segment in a unit sphere. In Figure 1.48, this area is shaded. For small angles, it looks like a rectangle. One side of the rectangle is equal to $\sin\theta \mathrm{d}\varphi$, and the other is $\mathrm{d}\theta$. The entire sphere corresponds to the total solid angle:

$$\int_0^{\pi} \sin\theta \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\varphi = \int \mathrm{d}\Omega = 4\pi \ .$$

Using the standard formula of probability theory for finding a mean value and assuming that, due to the chaotic thermal motion of the electrons their different velocity directions are equally probable, we get the expression:

$$j_{\mathrm{D}} = \frac{\int nV_z \mathrm{d}\Omega}{\int \mathrm{d}\Omega} = \frac{\int nV_z \mathrm{d}\Omega}{4\pi} \ . \tag{1.163}$$

Note that the value of the electron density n cannot be taken at the point $z = z_0$ because the electron moves freely between collisions. The last collision suffered by the electron occurs before the plane $z = z_0$ at the distance $(z_0 - V_z\tau)$ from it. Therefore, the concentration should be taken at this point:

$$n(z_0 - V_z\tau) = n(z_0 - V_{\mathrm{F}}\tau\cos\theta) = n(z_0 - l\cos\theta) \ . \tag{1.164}$$

Substituting formulas (1.161) and (1.164) into the expression for j_{D} (1.163), we get:

$$j_{\mathrm{D}} = \frac{V_{\mathrm{F}}}{4\pi} \int_0^{\pi} n\,(z_0 - l\cos\theta)\cos\theta\sin\theta \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\varphi = \frac{V_{\mathrm{F}}}{2} \int_{-1}^{1} n\,(z_0 - lx)\,x\mathrm{d}x \ .$$

Furthermore, we expand the function $n(z_0 - xl)$ in a Taylor series in powers of xl and limit ourselves to the first two summands of the expansion. Obviously, the first summand in the integrand is an odd function of x and, therefore, does not contribute to the general integral. The integral of the second term is easily calculated:

$$j_D = \frac{V_F}{2}\int_{-1}^{1}\left(n(z_0) - lx\left.\frac{\partial n}{\partial z}\right|_{z=z_0} + \cdots\right)x\mathrm{d}x \approx -\frac{lV_F}{2}\left.\frac{\partial n}{\partial z}\right|_{z=z_0}\int_{-1}^{1}x^2\mathrm{d}x = -\frac{lV_F}{3}\left.\frac{\partial n}{\partial z}\right|_{z=z_0}.$$

Finally, we find:

$$j_D = -\frac{lV_F}{3}\left.\frac{\partial n}{\partial z}\right|_{z=z_0}. \tag{1.165}$$

The *diffusion coefficient D* is the factor preceding the derivative of concentration:

$$D = \frac{lV_F}{3} = \frac{V_F^2\tau}{3}, \tag{1.166}$$

where l is the length of the electron mean free path ($l = V_F\tau$).

We can consider charge and heat transport processes in a metal using the above approach. They can be treated as special cases of diffusion caused by an electrostatic potential gradient or temperature.

Conductivity is, in essence, the diffusion of electrons under an external force $\vec{F} = -|e|\vec{E} = |e|\vec{\nabla}\varphi$, where φ is the potential of an electric field. Without limiting the generality, the axis Oz can be always codirected along the strength vector of the electric field. Therefore, we can accept that $\vec{E} = (0, 0, E)$ (Figure 1.49). This leads to the expression for the electric current density acquiring the form:

$$j = |e|\, j_D .$$

Recall the formula for the electron concentration and apply it to our case:

$$n = \int_{-\infty}^{+\infty} f(\varepsilon - |e|\,\varphi)\nu(\varepsilon)\mathrm{d}\varepsilon, \tag{1.167}$$

where $f(\varepsilon)$ is the Fermi–Dirac distribution function, $(\varepsilon - |e|\varphi)$ is the energy of an electron in an electric field, and $\nu(\varepsilon)$ is the density of the allowed energy levels for electrons in a metal. The given expression involves the z-dependence only through the potential φ. In formula (1.167), we expand the function $f(\varepsilon - |e|\varphi)$ in a Taylor series in powers of $-|e|\varphi$. After that, we differentiate the whole expression over z:

$$\frac{\partial n}{\partial z} \approx -|e|\,\frac{\partial\varphi}{\partial z}\int_{-\infty}^{+\infty}\frac{\partial f(\varepsilon)}{\partial\varepsilon}\nu(\varepsilon)\mathrm{d}\varepsilon . \tag{1.168}$$

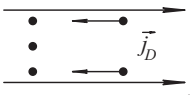


Fig. 1.49: The motion of electrons in an electric field.

In formula (1.168), we have kept only the first nonzero term. Now, we evaluate the integral (1.168) approximately by expanding the function $\nu(\varepsilon)$ in the neighborhood of the point $\varepsilon = \mu \approx \varepsilon_F$ in a Taylor series. Here,μ is the chemical potential. Confining ourselves to the first term, we obtain:

$$-\int_{-\infty}^{+\infty} \frac{\partial f(\varepsilon)}{\partial \varepsilon} \nu(\varepsilon) \mathrm{d}\varepsilon \approx \nu(\varepsilon_F) .$$

As a result, we have:

$$\frac{\partial n}{\partial z} = -|e| \nu(\varepsilon_F) E .$$

Thus, the electric current density is given by:

$$j = -|e| D \frac{\partial n}{\partial z} = De^2 \nu(\varepsilon_F) E . \tag{1.169}$$

The factor preceding the strength of the electric field E in (1.169), by definition, is the *coefficient of conductivity* σ. In this case, we can neglect the anisotropy of the sample because we are using the model of nearly free electrons.

The relationship between the conductivity and diffusion coefficients:

$$\sigma = e^2 D \nu(\varepsilon_F) , \tag{1.170}$$

is referred to as the *Einstein relation*.

Plugging the diffusion coefficient D in explicit form, we get another formula for the conductivity:

$$\sigma = \frac{1}{3} e^2 V_F^2 \tau \nu(\varepsilon_F) . \tag{1.171}$$

Similarly, we can find an expression for the thermal conductivity. The heat flux q along the axis Oz represents the energy passing through a unit area of the plane $z = z_0$ per unit of time and averaged over the electron velocity directions:

$$q = \frac{1}{4\pi} \int \frac{W}{V} V_z \mathrm{d}\Omega , \tag{1.172}$$

where W/V is the energy of electrons per unit volume of a metal.

Suppose that the energy $W = W(T)$ depends on the coordinates only because the latter are temperature dependent, then we expand it in a Taylor series, restricting ourselves to the first two summands. As previously noted, the temperature should be taken, not at the point $z = z_0$, but at the point shifted by $V_z \tau = l \cos\theta$:

$$W(T(z_0 - l\cos\theta)) = W(T(z_0)) - \frac{\partial W}{\partial T} \frac{\partial T}{\partial z}\bigg|_{z=z_0} l\cos\theta + \cdots .$$

Substitute formula (1.161) and the expression obtained above in (1.172). As a result of the integration we find:

$$q = \frac{1}{4\pi V} \int \left(W(T(z_0)) - \frac{\partial W}{\partial T} \frac{\partial T}{\partial z}\bigg|_{z=z_0} l\cos\theta \right) V_F \cos\theta \mathrm{d}\Omega =$$
$$= -\frac{1}{3} l V_F \left(\frac{1}{V} \left(\frac{\partial W}{\partial T} \right)_{V=\mathrm{const}} \right) \frac{\partial T}{\partial z}\bigg|_{z=z_0} = -\frac{1}{3} V_F^2 \tau c_V \frac{\partial T}{\partial z}\bigg|_{z=z_0} ,$$

where $c_V = (1/V)((\partial W)/(\partial T))_{V=\text{const}}$ is the specific heat of a metal at constant volume. *The coefficient of thermal conductivity* κ is equal to the multiplier of the temperature derivative:

$$\kappa = \frac{1}{3} V_F^2 \tau c_V \,. \tag{1.173}$$

It is curious that Drude has been the first to derive a similar formula for κ within the framework of classical physics in the model of a free electron gas of metals. However, Drude has used the mean square thermal velocity V_T^2 of the electrons instead of V_F^2. Classical physics, of course, does not give correct values for either V_F^2 or for the specific heat c_V. Striking success came to Drude due to him twice "successfully" mistaking a hundred times and these errors compensated for each other. At room temperature, the real electron contribution to the specific heat is one hundred times less than that calculated in the framework of classical physics. The real mean square electron velocity V_F^2 is one hundred times greater than V_T^2. This amusing incident, however, stimulated the rapid development of the free electron model for metals, first within classical physics, and after the errors were opened and contradictions were accumulated, in the framework of quantum theory.

Using the quantum mechanical expression for the electron heat capacity of a metal:

$$c_V = \frac{\pi^2 k_B^2 T}{3} \nu(\varepsilon_F) \,,$$

we have arrived at the final formula for the coefficient of thermal conductivity of metals:

$$\kappa = \frac{\pi^2}{9} k_B^2 T V_F^2 \tau \nu(\varepsilon_F) \,. \tag{1.174}$$

Comparing the expressions for σ (1.171) and for κ (1.174), we get an interesting result:

$$\frac{\kappa}{\sigma T} = \frac{k_B^2 T V_F^2 \tau \nu(\varepsilon_F) \pi^2/9}{e^2 V_F^2 \tau \nu(\varepsilon_F) T/3} = \frac{\pi^2 k_B^2}{3e^2} \,. \tag{1.175}$$

As a matter of interest, the right-hand part of the expression is a combination of fundamental constants and includes no specific characteristics of a particular metal. The resulting relation has no relaxation time τ, whose temperature dependence is unknown to us without detailed calculations. Formula (1.175) is called the *Wiedemann–Franz law*, and the constant in its right-hand side is referred to as the *Lorentz constant*.

In the case of free electrons we are aware that:

$$\nu(\varepsilon_F) = \frac{3n}{2\varepsilon_F} \,, \quad \varepsilon_F = \frac{m V_F^2}{2} \,.$$

Then the formulas for σ and κ take the form:

$$\sigma = \frac{e^2 n \tau}{m} \,, \quad \kappa = \frac{\pi}{3} \frac{n \tau T}{m} \,.$$

Although these formulas do not apply to real metals, they are useful for estimations.

It would be instructive to reproduce Drude's conclusion for the conductivity within classical physics through the free electron model. Drude believed that under an electric field $\vec{E}$, an electron in a metal moves as a material particle according to Newton's second law:

$$m\frac{\mathrm{d}\vec{V}}{\mathrm{d}t} = -|e|\,\vec{E} - \frac{m\vec{V}}{\tau}\,.$$

The second term in the right-hand side is some frictional force, which simulates a decrease in the electron momentum as a result of electron scatter interactions.

In a steady-state $(\mathrm{d}\vec{V})/(\mathrm{d}t) = \vec{0}$ and, therefore, the drift velocity of electrons in a conductor is:

$$\vec{V}_{\mathrm{drift}} = -\frac{|e|\,\tau\vec{E}}{m}\,.$$

Then, for the conduction current density, we have:

$$\vec{j} = -|e|\,n\vec{V}_{\mathrm{drift}} = \frac{e^2 n\tau}{m}\vec{E} \equiv \sigma\vec{E}\,,$$

$$\sigma = \frac{e^2 n\tau}{m}\,.$$

It turns out, the formula for the conductivity coefficient and that found earlier are identical. Here, the Principle of Correspondence between classical and quantum physics manifests in a delicate and controversial manner. The coincidence of some of the results is not accidental. Many features of the dynamics of wave packets corresponding to electrons in a metal can indeed be described in terms of quasiclassical equations of motion. Only these equations seldom coincide with Newton's equations.

Notes

1. For metals with impurities, the concept of the relaxation time τ works well at low temperatures ($T < 10\,\mathrm{K}$). In this temperature range, the main scattering mechanism is elastic electron impurity interactions, with the relaxation time τ as well as the conductivity σ being temperature independent. Simultaneously, the thermal conductivity is directly proportional to temperature. The Wiedemann–Franz law is well satisfied.

2. As the temperature increases, the electron scattering by thermal vibrations of the lattice ions begins to emerge more and more rapidly. As a single parameter, the relaxation time becomes insufficient to describe two different phenomena, such as the heat and charge transfer processes. In this case, the σ- and κ-dependencies on temperature are complex. In the temperature range (10–100 K), the inelastic scattering of electrons by the ion lattice vibrations prevails. The Wiedemann–Franz law is violated.

3. At higher temperatures, the scattering of electrons by the ion lattice vibrations again becomes elastic, and the Wiedemann–Franz law holds once again.

4. The electron-electron scattering processes are important only as far as very pure metals at low temperatures are concerned. Under these conditions, the relaxation time can be entered as a single parameter to describe the heat and charge transfer processes. Then, the above theory can be applied to them. The electron-electron

scattering takes place against the background of a periodic lattice. It can be shown that, in this case, the law of conservation of the total wave vector of the electrons as a consequence of the law of conservation of momentum loses its validity and transforms into a statement:

$$\vec{k}_1 + \vec{k}_2 = \vec{k}_3 + \vec{k}_4 + \vec{K} ,$$

where $\vec{K}$ is an arbitrary reciprocal lattice vector. When scattered, the electrons throw over their wave vectors into other Brillouin zones. Therefore, the electron-electron collisions are not elastic. Although these are inelastic, and the relaxation time depends on temperature ($\tau \sim T^{-2}$), the Wiedemann–Franz law is satisfied with pure metals as well.

Let us summarize. Thermal conductivity and resistance for most normal metals (not superconductors) are controlled mainly by electrons, which are in the upper partially filled energy band. The σ- and κ-dependencies on temperature are complex and reflect various electron scattering mechanisms at different temperatures. There is a finite temperature range where the electron scattering by vibrating lattice ions cannot be described through the relaxation time as a single parameter. Typical curves of the electrical resistance $R \sim 1/\sigma$ and thermal conductivity κ as a function of temperature are shown in Figures 1.50 and 1.51.

Analysis of experimental data leads to the following general conclusions.

1. A crystal lattice behaves differently at $T > T_D$ and $T < T_D$. The temperature T_D (the *Debye temperature*) is characteristic of the lattice and will be introduced later.

2. The theory set forth above involves the relaxation time τ as a single parameter. Such a simplified theory is true for only the low temperature region where τ is a constant and the higher temperature range where $\tau \sim T^{-1}$. Figures 1.50 and 1.51 display how the resistance R and the thermal conductivity coefficient κ depend on temperature. As can be seen, the temperature dependence of the resistance ($R \sim T$) is linear and the thermal conductivity coefficient ($\kappa = \text{const}$) is constant at $T \geq T_D$. In this region, the electron scattering by the vibrating lattice ions is a dominating factor.

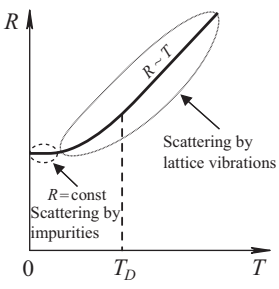


Fig. 1.50: The R-dependence on temperature T.

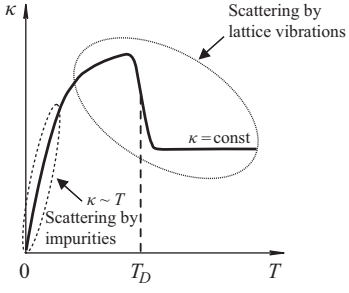


Fig. 1.51: The κ-dependence on temperature T.

1.13 The Semiclassical Dynamics of Electrons in a Crystal

The Sommerfeld theory for free electrons has been generalized by Bloch and applied to the case of an arbitrary periodic potential. Such a generalization was called a *semiclassical model* [1, 5].

Consider a wave packet constructed from the Bloch wave functions for an electron in a crystal:

$$\Psi_n(\vec{r}, t) = \sum_{\vec{k}'} g(\vec{k}') u_{n\vec{k}'}(\vec{r}) \exp\left(\mathrm{i}\vec{k}'\vec{r} - \frac{\mathrm{i}}{\hbar}\varepsilon_n(\vec{k}')t\right) \tag{1.176}$$

where $g(\vec{k}') \approx 0$ for $|\vec{k} - \vec{k}'| > \Delta k$ and $\vec{k}$ is the center of the wave packet in the reciprocal space. All $\vec{k}'$ are inside an area centered at the point with the radius vector $\vec{k}$ and of radius $\Delta k \ll |\vec{k}|$. Assume $|\vec{k}|$, in turn, to be much smaller than the Brillouin zone size:

$$\Delta k \ll k \ll \frac{1}{a} . \tag{1.177}$$

Under such circumstances, the whole wave packet moves integrally with the group velocity:

$$\vec{V}_n = \frac{1}{\hbar}\frac{\partial \varepsilon_n(\vec{k})}{\partial \vec{k}} , \tag{1.178}$$

and its width in the real space has the form:

$$\Delta r \sim \frac{1}{\Delta k} . \tag{1.179}$$

It follows from (1.178) and (1.179) that:

$$\Delta r \gg a , \tag{1.180}$$

i.e., the size of the wave packet or the area where the function Ψ_n is localized is far larger than the interatomic distance. In real space, the wave packet whose wave vector is defined in the range that is smaller than the dimensions of the Brillouin zone smears itself out over a large number of elementary cells.

Suppose an electromagnetic field slowly varies within a wave packet (Figure 1.52):

$$\lambda \sim \frac{\left|\vec{B}\right|}{\left|\partial\vec{B}/\partial x\right|} \sim \frac{\left|\vec{E}\right|}{\left|\partial\vec{E}/\partial x\right|} \gg \Delta r \gg a . \tag{1.181}$$

Having formulated data as above and by using a semiclassical model, we can calculate how the wave packet moves under the influence of external electric and magnetic fields. In other words, we can find out how the wave vector $\vec{k}$, the packet center $\vec{r}$, and the zone number vary in time. Such predictions can be made based on a knowledge of the function $\varepsilon_n(\vec{k})$ without adding any information about the periodic potential of the lattice.

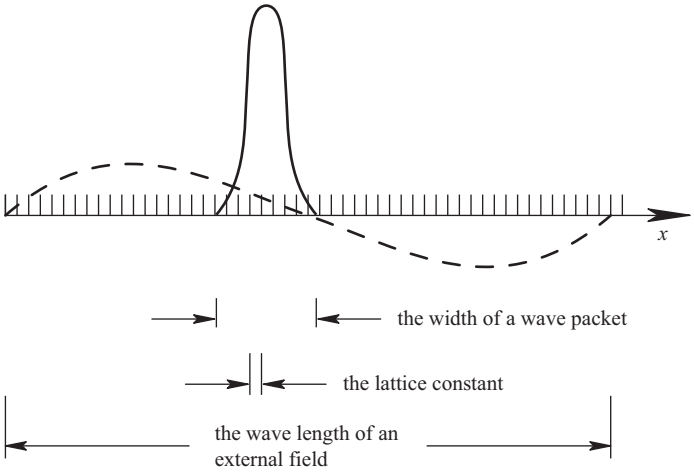


Fig. 1.52: The ratios of characteristic scales in the semiclassical model.

In the semiclassical model, the functions $\varepsilon_n(\vec{k})$ are thought to be known. Therefore, a method to calculate them is not specified. This is a delicate moment. The fact of the matter is that, to determine the function $\varepsilon_n(\vec{k})$, we should either conduct an experiment or solve the Schrödinger equation:

$$\left[-\frac{\hbar^2}{2m}\Delta + U(\vec{r})\right]\Psi_{n\vec{k}} = \varepsilon_n(\vec{k})\Psi_{n\vec{k}}\,, \tag{1.182}$$

with the potential $U(\vec{r})$ varying at distances of the order of the lattice constant.

The semiclassical model is only partly the classical limit. It deals classically with external fields but not the periodic field of a crystal.

The coordinates, the wave vector of an electron, and the zone number vary in time according to the following laws:

1. The zone number is an integral of motion. All zone numbers remain unchanged. The semiclassical model ignores interband transitions. This means that every zone has a fixed number of electrons.
2. Time changes of the coordinates and the wave vector $\vec{k}$ of an electron with the zone number n is determined by the following equations of motion:

$$\dot{\vec{r}} = \vec{V}_n(\vec{k}) = \frac{1}{\hbar}\frac{\partial\varepsilon_n(\vec{k})}{\partial\vec{k}}\,. \tag{1.183}$$

$$\hbar\dot{\vec{k}} = -|e|\cdot\left(\vec{E}(\vec{r},t) + \frac{1}{c}\left[\vec{V}_n(\vec{k})\times\vec{B}(\vec{r},t)\right]\right)\,. \tag{1.184}$$

Attention should be drawn to the fact that an electron in a crystal is subjected to forces of both the lattice and the external fields $\vec{B}$ and $\vec{E}$. In equations (1.183) and (1.184), only the external forces are responsible for the resulting electron motion.

The function $\varepsilon_n(\vec{k})$ that plays the role of the kinetic energy of the electron in a periodic field of the crystal has already taken the forces of the crystal lattice into account. That is why equations (1.183) and (1.184) do not have the form of Newton's second law.

3. The electron wave vector is determined with accuracy up to a term equal to the reciprocal lattice vector $\vec{K}$. The symbols $n, \vec{r}, \vec{k}$ and $n, \vec{r}, (\vec{k} + \vec{K})$ represent two completely equivalent ways of describing one and the same electron. At thermodynamic equilibrium, the electrons of the n-th zone with energies from ε to $(\varepsilon + \mathrm{d}\varepsilon)$ contribute to the electron density. Their contribution can be governed by the Fermi–Dirac distribution:

$$\frac{\Delta N}{V} = \nu(\varepsilon) f(\varepsilon)\,\mathrm{d}\varepsilon \equiv \int\limits_{\varepsilon<\varepsilon(\vec{k})<\varepsilon+\mathrm{d}\varepsilon} \frac{\mathrm{d}^3\vec{k}}{4\pi^3}\left[\exp\left(\frac{\varepsilon(\vec{k})-\mu}{k_\mathrm{B}T}\right)+1\right]^{-1}. \tag{1.185}$$

In the case of thermodynamic equilibrium, the energy bands being above the Fermi energy by a magnitude that is greater than $(\varepsilon_\mathrm{F}+k_\mathrm{B}T)$ prove to be empty, and the energy bands that are below $(\varepsilon_\mathrm{F}-k_\mathrm{B}T)$ are completely filled with electrons. For this reason, a theoretical description of physical properties of real metals and semiconductors must consider only a small number of the bands and electrons.

The Limits of Applicability of the Semiclassical Model

In the Schrödinger equation (1.179), the validity of the semiclassical model must be broken in the limit of zero potential $U(\vec{r})$ because an electron becomes free at this limit. Being in a uniform and constant electric field, the free electron can continuously accumulate its kinetic energy due to the electrostatic potential energy. However, the semiclassical model prohibits interband transitions and requires limiting the electron energy by its initial band boundaries. Consequently, the semiclassical model can be applied to cases where the amplitude of the interatomic periodic potential exceeds some minimum acceptable value.

The restrictions imposed on the amplitudes of slowly varying external electric and magnetic fields have the form:

$$|e|Ea \ll \frac{\left[\varepsilon_\mathrm{g}(\vec{k})\right]^2}{\varepsilon_\mathrm{F}}. \tag{1.186}$$

$$\frac{\hbar|e|B}{mc} \ll \frac{\left[\varepsilon_\mathrm{g}(\vec{k})\right]^2}{\varepsilon_\mathrm{F}}. \tag{1.187}$$

The proof of these conditions is given, for example, in the book by Ashcroft and Mermin [1].

Comments

1. The length a has the order of the lattice constant.
2. The Fermi energy ε_F must be measured from the bottom of the energy band at hand. $\varepsilon_g(\vec{k})$ is the difference between the energy $\varepsilon_n(\vec{k})$ and the nearest energy $\varepsilon_{n'}(\vec{k})$ at the same point of k-space, but in another zone:

$$\varepsilon_g(\vec{k}) = \min_{n'} \left| \varepsilon_n(\vec{k}) - \varepsilon_{n'}(\vec{k}) \right| , \quad n' = n \pm 1 .$$

3. The condition (1.186) is never violated for metals but can be broken for semiconductors and dielectrics under huge electrical fields. In the case of the latter, electrons can make interband transitions. This effect is called *electrical breakdown*.
4. The condition (1.187) can be broken simply enough. When moved, electrons cannot follow the paths obtained from the semiclassical equations. This effect is called *magnetic breakdown*.
5. It is also necessary that the frequency of the fields be small:

$$\hbar\omega \ll \varepsilon_g . \tag{1.188}$$

   Otherwise, a single photon may possess sufficient energy to cause the interband transition.
6. The wavelength of the applied fields must be restricted:

$$\lambda \gg a . \tag{1.189}$$

In doing so, the correspondence of the wave packet to the electron makes sense.

1.14 The Justification of Semiclassical Equations of Motion, the Hamiltonian Formulation and Liouville's Theorem

Let us try to prove the validity of the semiclassical equations of motion. The equation:

$$\dot{\vec{r}} = \vec{V}_n = \frac{1}{\hbar} \frac{\partial \varepsilon_n(\vec{k})}{\partial \vec{k}}$$

is equivalent to saying that the velocity of a semiclassical electron is the group velocity of the underlying wave packet. We show, that equation:

$$\hbar\dot{\vec{k}} = -|e| \left(\vec{E}(\vec{r}, t) + \frac{1}{c} \left[\vec{V}_n(\vec{k}) \times \vec{B}(\vec{r}, t) \right] \right) ,$$

at least, is not contradictory.

In the case of a time independent electric field, the above equation seems quite plausible because it provides the conservation of energy. Indeed, if the field is given

by the expression $\vec{E} = -\vec{\nabla}\varphi$, every wave packet would move so that the electron energy should remain constant. The time derivative of the energy is equal to zero:

$$\frac{\partial \varepsilon_n}{\partial \vec{k}} \cdot \dot{\vec{k}} - |e|\vec{\nabla}\varphi \cdot \dot{\vec{r}} = 0 \,. \tag{1.190}$$

Replacing $(\partial\varepsilon_n)/(\partial\vec{k})$ by $\hbar\dot{\vec{r}}$, we get:

$$\hbar\dot{\vec{r}} \cdot \dot{\vec{k}} - |e|\vec{\nabla}\varphi \cdot \dot{\vec{r}} = 0 \,,$$

or:

$$\dot{\vec{r}} \cdot \left[\hbar\dot{\vec{k}} + |e|\vec{E}\right] = 0 \,, \tag{1.191}$$

which is the same. Equation (1.191) is satisfied if:

$$\hbar\dot{\vec{k}} = -|e|\vec{E} \,. \tag{1.192}$$

As can be seen, the resulting expression is nothing but the equation (1.184) in the absence of a magnetic field. However, implementing equation (1.192) is not a necessary condition for energy preservation. This is because expression (1.191) is zero even by adding any vector perpendicular to the vector $\dot{\vec{r}}$ to (1.192). To prove that the term $-|e|[\vec{V}_n(\vec{k}) \times \vec{B}(\vec{r}, t)]/c$ is the only additional one and that the final equation must be carried out also for time dependent fields is a considerably daunting task, and it should be omitted.

The semiclassical equations of motion (1.183) and (1.184) can be written in a compact Hamiltonian form:

$$\dot{\vec{r}} = \frac{\partial H}{\partial \vec{p}} \,, \quad \dot{\vec{p}} = -\frac{\partial H}{\partial \vec{r}} \,, \tag{1.193}$$

where the Hamiltonian for an electron of the n-th zone is:

$$H(\vec{r}, \vec{p}) = \varepsilon_n \left[\frac{1}{\hbar}\left(\vec{p} + \frac{|e|}{c}\vec{A}(\vec{r}, t)\right)\right] - |e|\varphi(\vec{r}, t) \,. \tag{1.194}$$

The fields $\vec{B}$ and $\vec{E}$ can be expressed through the vector and scalar potentials:

$$\vec{B} = \operatorname{rot}\vec{A} \,, \quad \vec{E} = -\vec{\nabla}\varphi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t} \,. \tag{1.195}$$

The variable $\vec{k}$ in the equation (1.183) and (1.184) is determined as follows:

$$\hbar\vec{k} = \vec{p} + \frac{|e|}{c}\vec{A}(\vec{r}, t) \,. \tag{1.196}$$

Note that the Hamiltonian formulation contains the canonical momentum $\vec{p}$ not coinciding with the quasimomentum $\hbar\vec{k}$ of an electron in metals, and has the appearance:

$$\vec{p} = \hbar\vec{k} - \frac{|e|}{c}\vec{A}(\vec{r}, t) \,. \tag{1.197}$$

For the semiclassical equations of motion for an electron having the Hamiltonian form, Liouville's theorem holds true.

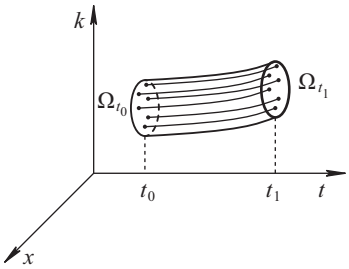


Fig. 1.53: Semiclassical trajectories in *x*, *k*-space (the one-dimensional case).

Liouville's Theorem

In a six-dimensional phase space $(\vec{r}, \vec{p})$, any areas can evolve in time only if their volumes are preserved.

The vector $\hbar\vec{k}$ differs from $\vec{p}$ only by an additional vector, which does not depend on $\vec{p}$ (1.196). Therefore, any area in the $(\vec{r}, \vec{p})$-space has the same volume as in the $(\vec{r}, \hbar\vec{k})$-space. This means the following. Imagine Ω_{t_0} to be a region in the $(\vec{r}, \vec{k})$-space at a moment t_0. Traveling outward according to the semiclassical equations of motion, its points will comprise a set Ω_{t_1} at an arbitrary moment $t_1 > t_0$. As a result, their volumes Ω_{t_0} and Ω_{t_1} will match exactly (Figure 1.53):

$$V(\Omega_{t_0}) = V(\Omega_{t_1}) \,.$$

Let us discuss what this means for the completely filled electron energy bands. The wave vectors of the electrons lie within the reciprocal space volume $\mathrm{d}^3\vec{k}$. Since the Fermi–Dirac distribution function is $f = 1$, the electron density is equal to $\mathrm{d}^3\vec{k}/(4\pi^3)$. Therefore, the number of electrons in the small volume $\mathrm{d}^3\vec{r}$ of the coordinate space amounts to:

$$\mathrm{d}N = \frac{1}{4\pi^3}\mathrm{d}^3\vec{k}\mathrm{d}^3\vec{r} \,. \tag{1.198}$$

Here, $\mathrm{d}^3\vec{k}\mathrm{d}^3\vec{r}$ is the volume element in the six-dimensional phase space $(\vec{r}, \vec{k})$ containing points corresponding to states of the electrons. By Liouville's theorem, a certain number of the points ($\mathrm{d}N = \mathrm{const}$), when driven, keep their phase volume unchanged: $\mathrm{d}^3\vec{k}\mathrm{d}^3\vec{r} = \mathrm{const}$. It follows that when the electrons move in accordance with the semiclassical equations of motion, the density of the points corresponding to states of the electrons in $(\vec{r}, \vec{k})$-space remains constant:

$$\frac{\mathrm{d}N}{\mathrm{d}^3\vec{k}\mathrm{d}^3\vec{r}} = \frac{1}{4\pi^3} = \mathrm{const} \,.$$

From the perspective of full quantum mechanical theory, this assertion emanates directly from the Pauli principle. The density of points corresponding to states of the electrons inside the completely filled band in the phase space cannot be enlarged because every level contains a maximum electron number allowed by the Pauli principle. In our case, interband transitions are forbidden, so the electrons cannot leave

their band limits. Consequently, the density of points featuring states of the electrons in the completely filled band cannot be reduced.

Thus, the quasiclassical equations of motion for an electron are consistent with the Pauli principle. The semiclassical model does not contradict to a more fundamental quantum theory, and it can be used instead of the full quantum mechanical theory.

1.15 The Lack of Contribution of Bands Completely Filled with Electrons to an Electric Current and a Flux of Heat

To begin with, we prove the validity of the following auxiliary theorem. Suppose $f(\vec{k})$ is an arbitrary function with lattice periodicity in reciprocal space: $f(\vec{k} + \vec{K}) = f(\vec{k})$. Then we have:

$$\int_{V_b} \mathrm{d}^3\vec{k} \frac{\partial}{\partial \vec{k}} f(\vec{k}) = 0 \,, \tag{1.199}$$

where the integration is carried out over the first Brillouin zone V_b.

Since V_b is a primitive cell of the reciprocal lattice, for any vector $\vec{k}'$ we have:

$$I(\vec{k}') = \int_{V_b} \mathrm{d}^3\vec{k} f(\vec{k} + \vec{k}') = \text{const} \,. \tag{1.200}$$

We now differentiate equation (1.200) over $\vec{k}'$. After that, in the integrand, we replace the differentiation over $\vec{k}'$ by the differentiation over $\vec{k}$:

$$\int_{V_b} \mathrm{d}^3\vec{k} \frac{\partial}{\partial \vec{k}} f(\vec{k} + \vec{k}') = 0 \,. \tag{1.201}$$

Calculating this expression for $\vec{k}' = 0$, we obtain:

$$\int_{V_b} \mathrm{d}^3\vec{k} \frac{\partial f(\vec{k})}{\partial \vec{k}} = 0 \,,$$

which was to be proved.

Now we will find the contribution of electrons to an electric current. If the wave vectors of the filled band electrons occupy the volume $\mathrm{d}^3\vec{k}$ in k-space, then the contribution of these electrons to the total electron concentration is $(\mathrm{d}^3\vec{k})/(4\pi^3)$. If one multiplies the above value by $(-|e|\vec{V}_n(\vec{k}))$, the electron contribution to the electric current density is given by:

$$\mathrm{d}\vec{j}_\mathrm{e} = \frac{\mathrm{d}^3\vec{k}}{4\pi^3}\left(-|e|\vec{V}_n(\vec{k})\right) = -\frac{|e|}{4\pi^3}\mathrm{d}^3\vec{k}\frac{1}{\hbar}\frac{\partial \varepsilon_n}{\partial \vec{k}} \,. \tag{1.202}$$

The contribution of all the electrons of the completely filled energy band to the current density is controlled by an integral of expression (1.202). In line with the previously

proven theorem, the integral vanishes:

$$\vec{j}_{\mathrm{e}} = -\frac{|e|}{4\pi^3\hbar}\int\limits_{V_b} \mathrm{d}^3\vec{k}\frac{\partial\varepsilon_n(\vec{k})}{\partial\vec{k}} = 0\,.$$

The function $\varepsilon(\vec{k})$ is periodic in the reciprocal lattice. The integration is over the entire Brillouin zone.

Analogously, the heat flux carried by the electrons of the completely filled band is also equal to zero:

$$\vec{j}_q = \frac{1}{4\pi^3}\int\limits_{V_b} \mathrm{d}^3\vec{k}\vec{V}_n(\vec{k})\varepsilon_n(\vec{k}) = \frac{1}{8\pi^3\hbar}\int\limits_{V_b} \mathrm{d}^3\vec{k}\frac{\partial}{\partial\vec{k}}\varepsilon_n^2(\vec{k}) = 0\,. \tag{1.203}$$

Thus, when calculating the electrical characteristics of a solid body, it is necessary to take into account only partially filled bands. This explains the emergence of a mysterious parameter of theory of metals; the number of conduction electrons. The conductivity is caused only by the electrons of the partially filled bands.

Note that the identity:

$$\int\limits_{V_b} \mathrm{d}^3\vec{k}\vec{V}_n(\vec{k}) = 0\,, \tag{1.204}$$

is true even when not all the energy levels have electrons. That the integral (1.204) vanishes due to the fact that $V_n(\vec{k}) = \partial\varepsilon_n(\vec{k})/\hbar\partial\vec{k}$, where $\varepsilon_n(\vec{k}+\vec{K}) = \varepsilon_n(\vec{k})$.

1.16 Holes

One of the most impressive results of the semiclassical model is to explain phenomena in terms of the free electron theory, by just assuming that the carriers have a positive charge.

To understand why the quasiparticles in metals and semiconductors can sometimes give such a contribution to a current as if they were positively charged, it is necessary to take into account three factors listed below [1].

1. All calculations, as used herein, are done at zero temperature, so that the Fermi–Dirac distribution function $f(\vec{k})$ reduces to a step. Then the contribution of electrons of an energy band labeled by n to the current density is determined by the formula:

$$\vec{j}_n = -\frac{|e|}{4\pi^3}\int \mathrm{d}^3\vec{k}f(\vec{k})\vec{V}_n(\vec{k}) = -\frac{|e|}{4\pi^3}\int\limits_{\substack{\text{filled}\\\text{levels}}} \mathrm{d}^3\vec{k}\vec{V}_n(\vec{k}) \neq 0\,, \tag{1.205}$$

where the integration is performed over only filled levels of the energy band. Given the identity (1.204):

$$\int\limits_{\substack{\text{filled}\\\text{levels}}} \mathrm{d}^3\vec{k}\vec{V}_n(\vec{k}) + \int\limits_{\substack{\text{unfilled}\\\text{levels}}} \mathrm{d}^3\vec{k}\vec{V}_n(\vec{k}) = 0\,, \tag{1.206}$$

expression (1.205) can be rewritten in the equivalent form:

$$\vec{j}_n = \frac{|e|}{4\pi^3} \int\limits_{\substack{\text{unfilled} \\ \text{levels}}} \mathrm{d}^3\vec{k}\,\vec{V}_n(\vec{k})\,. \tag{1.207}$$

Consequently, the current produced by electrons filling a certain set of the levels coincides exactly with what can be obtained by leaving the levels unfilled but by occupying the rest of the levels in the Brillouin zone by particles having a positive charge.

Thus, although the only real charge carriers are electrons, the current can be viewed, when necessary, as consisting of entirely fictitious particles with a positive charge, with those filling the band levels devoid of electrons. These fictitious particles are called *holes*.

Both ways of describing electron and hole processes taking place in it are impossible to apply simultaneously to one and the same band. Every particular problem requires choosing the most convenient way of describing the processes.

2. Under external electromagnetic fields, unfilled levels in the band evolve in such a way as if they were occupied by real electrons:

$$\begin{aligned} \dot{\vec{r}} &= \vec{V}_n = \frac{1}{\hbar}\frac{\partial\varepsilon_n(\vec{k})}{\partial\vec{k}}\,, \\ \hbar\dot{\vec{k}} &= -|e|\left(\vec{E}(\vec{r},t) + \frac{1}{c}\left[\vec{V}_n(\vec{k})\times\vec{B}(\vec{r},t)\right]\right)\,. \end{aligned} \tag{1.208}$$

The fact is that, under the given initial conditions, equation (1.208) describe certain paths, regardless of whether the electrons are in the levels $\varepsilon_n(\vec{k})$ or not. In this case, the resulting trajectories do not intersect. Consequently, if all the trajectories can be divided into occupied and empty ones at the initial time, they remain as such at all subsequent times.

We have arrived at a statement equivalent to the result of classical mechanics. It reads that the position and momentum of a particle completely determine its motion trajectory at any instant of time.

An occupied level as well as an empty one evolves completely according to the shape of the trajectory. The latter, in turn, depends on the form of the semiclassical equations only, but it does not depend on whether an electron resides in the trajectory or not (Figure 1.54).

3. Therefore, to find out the behavior of a hole, it suffices to establish how an electron responses to applied fields. The semiclassical equation

$$\hbar\dot{\vec{k}} = -|e|\left(\vec{E}(\vec{r},t) + \frac{1}{c}[\vec{V}_n(\vec{k})\times\vec{B}(\vec{r},t)]\right) \tag{1.209}$$

governs its motion.

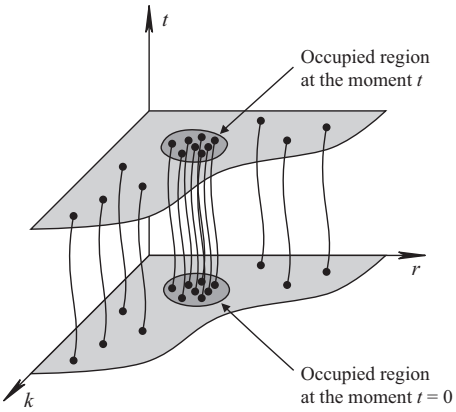


Fig. 1.54: Schematic picture of trajectories in phase space.

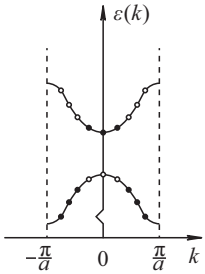


Fig. 1.55: Energy levels filled by electrons and holes in a one-dimensional crystal.

In our case, two variants are possible:

(a) The solution $\dot{\vec{k}} \uparrow\uparrow (\mathrm{d}\vec{V}/\mathrm{d}t)$ of the semiclassical equations of motion corresponds to the motion of a free particle with a negative charge.

(b) The solution $\dot{\vec{k}} \uparrow\downarrow (\mathrm{d}\vec{V}/\mathrm{d}t)$ of the semiclassical equations of motion corresponds to the motion of a positively charged particle.

The second variant can be realized only when $\vec{k}$ is the wave vector of the unfilled level (of a hole). Ultimately, this is explained by often being the unoccupied electron levels (holes) near the top of the valence band (Figure 1.55). Let us explain this statement.

Suppose that, for $\vec{k} = \vec{k}_0$, the energy of the band $\varepsilon(\vec{k})$ have a maximum. For values of $\vec{k}$ sufficiently close to $\vec{k}_0$, the function $\varepsilon(\vec{k})$ can be expanded in powers of $(\vec{k} - \vec{k}_0)$ and still keep the first terms of the expansion:

$$\varepsilon(\vec{k}) = \varepsilon(\vec{k}_0) + \sum_{i=1}^{3} \left.\frac{\partial \varepsilon}{\partial k_i}\right|_{\vec{k}=\vec{k}_0} (k - k_0)_i + \frac{1}{2} \sum_{i,j=1}^{3} \left.\frac{\partial^2 \varepsilon}{\partial k_i \partial k_j}\right|_{\vec{k}=\vec{k}_0} (k - k_0)_i (k - k_0)_j \,.$$

At the maximum point $\partial\varepsilon/\partial\vec{k}|_{\vec{k}=\vec{k}_0} = 0$, the term linear in $(\vec{k} - \vec{k}_0)$ disappears. For simplicity, we suppose that the point with the radius vector $\vec{k}_0$ possesses a high degree of symmetry, that is:

$$\left.\frac{\partial^2 \varepsilon}{\partial k_i \partial k_j}\right|_{\vec{k}=\vec{k}_0} = -2A\delta_{ij} \,.$$

Since the function $\varepsilon(\vec{k})$ has a maximum at the point with the radius vector $\vec{k}_0$, then $A > 0$. So, we have:

$$\varepsilon(\vec{k}) \approx \varepsilon(\vec{k}_0) - A(\vec{k} - \vec{k}_0)^2 \,. \tag{1.210}$$

Normally, it is necessary to determine a positive quantity m^* with the dimension of mass, assuming that:

$$A = \frac{\hbar^2}{2m^*} . \tag{1.211}$$

Consequently,

$$\varepsilon(\vec{k}) \approx \varepsilon(\vec{k}_0) - \frac{\hbar^2}{2m^*}(\vec{k} - \vec{k}_0)^2 . \tag{1.212}$$

For electron states with wave vectors close to $\vec{k}_0$, the relation holds true:

$$\vec{V}(\vec{k}) = \frac{1}{\hbar}\frac{\partial \varepsilon}{\partial \vec{k}} \approx -\frac{\hbar}{m^*}(\vec{k} - \vec{k}_0) .$$

We come up with the result:

$$\dot{\vec{V}} = -\frac{\hbar}{m^*}\dot{\vec{k}} . \tag{1.213}$$

In other words, the electron acceleration is opposite to $\dot{\vec{k}}$. Using (1.213), we give the equation of motion (1.209) the form:

$$\hbar\dot{\vec{k}} = -|e|\left(\vec{E} + \frac{1}{c}[\vec{V} \times \vec{B}]\right) = -m^*\dot{\vec{V}} . \tag{1.214}$$

In the long run, we can reach a conclusion that a negatively charged electron responds to driving fields as if it had a negative mass $-m^*$. Having changed the sign on both sides of the equation, we might as well assume that this equation describes the motion of a positively charged particle with positive mass m^*:

$$m^*\dot{\vec{V}} = |e|\left(\vec{E} + \frac{1}{c}[\vec{V} \times \vec{B}]\right) . \tag{1.215}$$

We have proceeded from the statement that, in the external fields, the hole behaves the same as the electron as if the latter were in an unoccupied level. According to the above result (1.215), the holes behave like ordinary particles with a positive charge. The quantity m^*, responsible for the dynamics of the holes near the points of maximum of the band with a high degree of symmetry, is called the *hole effective mass*.

The requirement that the unoccupied levels should lie near the point of maximum of the band with a high degree of symmetry can be significantly weakened. Consider a general situation, when:

$$\varepsilon(\vec{k}) = \varepsilon(\vec{k}_0) + \frac{1}{2}\sum_{i,j=1}^{3}\left.\frac{\partial^2 \varepsilon}{\partial k_i \partial k_j}\right|_{\vec{k}=\vec{k}_0}(k - k_0)_i(k - k_0)_j . \tag{1.216}$$

If $\vec{k}_0$ is the point of a local maximum of the function $\varepsilon(\vec{k})$, then:

$$\sum_{i,j=1}^{3}\left.\frac{\partial^2 \varepsilon}{\partial k_i \partial k_j}\right|_{\vec{k}=\vec{k}_0} s_i s_j < 0 , \quad \forall s_i, s_j \neq 0 . \tag{1.217}$$

Furthermore, we compute the group velocity of the wave packet underlying a quasiparticle with the energy (1.216):

$$V_i = \frac{1}{\hbar}\left(\frac{\partial\varepsilon}{\partial\vec{k}}\right)_i = \frac{1}{\hbar}\sum_{j=1}^{3}\left.\frac{\partial^2\varepsilon}{\partial k_i \partial k_j}\right|_{\vec{k}=\vec{k}_0}(k-k_0)_j \ .$$

As the solutions of the quasiclassical equations behave differently at different relative directions of $\dot{\vec{k}}$ and $\dot{\vec{V}}$, it can be assumed that the behavior of the quasiparticle depends on the angle between the vector $\dot{\vec{k}}$ and acceleration $\dot{\vec{V}}$. Using equation (1.217), in the vicinity of the local maximum of the function $\varepsilon(\vec{k})$, the following condition meets:

$$\dot{\vec{k}}\cdot\dot{\vec{V}} = \frac{1}{\hbar}\sum_{i,j=1}^{3}\left.\frac{\partial^2\varepsilon}{\partial k_i \partial k_j}\right|_{\vec{k}=\vec{k}_0}\dot{k}_i\dot{k}_j < 0 \ . \tag{1.218}$$

In other words, the angle between the vectors $\dot{\vec{k}}$ and $\dot{\vec{V}}$ is obtuse. Consequently, suppose the wave vector of the quasiparticle (the hole) remains close to a neighborhood of the maximum of the function $\varepsilon(\vec{k})$. This would lead to the hole being affected by external fields as if it had a positive charge.

Near the local minimum of the function $\varepsilon(\vec{k})$, similar calculations yield $\dot{\vec{k}}\cdot\dot{\vec{V}} > 0$. In other words, the same semiclassical equations of motion describing the dynamics of quasiparticles near the bottom of the conduction band results in a path that is characteristic of a negative charge (an electron).

In general, it would be reasonable to introduce the concept of the *effective mass tensor* near extrema of the function $\varepsilon(\vec{k})$:

$$\left(M^{-1}\right)_{ij} = \pm\frac{1}{\hbar^2}\left.\frac{\partial^2\varepsilon}{\partial k_i \partial k_j}\right|_{\vec{k}=\vec{k}_0} = \pm\frac{1}{\hbar}\left.\frac{\partial V_i}{\partial k_j}\right|_{\vec{k}=\vec{k}_0} \ . \tag{1.219}$$

The nature of the extremum manages the choice of the sign in this expression: if the extremum is a local maximum, then we choose the minus sign (the tensor describes holes), and if the extremum is a minimum, we choose the plus sign (electrons). In both cases, the effective mass tensor M_{ij} is positive definite:

$$\sum_{i,j=1}^{3} M_{ij}x_i x_j > 0 \ , \quad \forall x_i, x_j \neq 0 \ . \tag{1.220}$$

Given $\dot{\vec{V}} = \pm\widehat{M}^{-1}\hbar\dot{\vec{k}}$, the equation of motion takes the form:

$$\widehat{M}\dot{\vec{V}} = \mp|e|\left(\vec{E} + \frac{1}{c}[\vec{V}\times\vec{B}]\right) \ , \tag{1.221}$$

where the sign "–" is chosen for the electrons, and the sign "+" – for the holes.

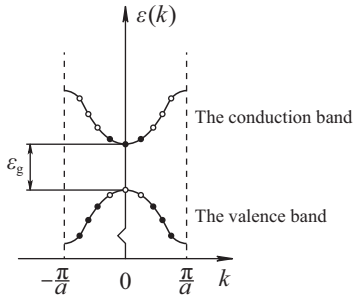


Fig. 1.56: Energy levels are filled, typically for semiconductors (a one-dimensional crystal).

Thus, we have derived a linear equation, which looks like Newton's equation, where the mass substitutes for a tensor quantity. Such equations quite accurately describe the dynamics of electrons and holes in semiconductors. This is explained by the filling of the energy levels that is inherent in semiconductors as shown in Figure 1.56.

The band gap width ε_g in *intrinsic semiconductors* is small. So, a significant part of thermally excited electrons can tunnel from the highest completely filled band (called the *valence band*) to the nearest unfilled zone (the *conduction band*).

In the valence band, the holes, that are the spaces empty of the electrons are near a local maximum of the function $\varepsilon(\vec{k})$. Therefore, they behave in external fields as positive charges. The electrons are located near the local minimum of the function $\varepsilon(\vec{k})$ and move in external fields as negative charges.

1.17 Semiclassical Motion of Electrons in a Crystal in Constant Electric and Magnetic Fields

I. The Motion of Electrons in an Electric Field

Consider the motion of electrons in a crystal in the presence of a constant electric field:

$$\vec{E} = \text{const}\,, \quad \vec{B} = 0\,. \tag{1.222}$$

In such a field the semiclassical equations of motion appear as:

$$\dot{\vec{r}} = \vec{V}_n = \frac{1}{\hbar}\frac{\partial \varepsilon_n(\vec{k})}{\partial \vec{k}}\,. \tag{1.223}$$

$$\hbar\dot{\vec{k}} = -|e|\vec{E}\,. \tag{1.224}$$

Integrating equation (1.224), we get:

$$\vec{k}(t) = \vec{k}(0)\frac{|e|\vec{E}}{\hbar}t\,. \tag{1.225}$$

As can be seen, during the time t the wave vectors of all the electrons acquire the same displacement, regardless of whether their initial states $\vec{k}(0)$ belong to a filled or unfilled band. This agrees with the statement that a completely filled energy band does not create an electric current (see the proof of the assertion). From the point of view of classical physics, the same displacement of the wave vectors of all the electrons in the completely filled energy band without creating a current configuration seems to be strange.

An understanding of why quantum theory approves the previously said comes from the proportionality of the contribution of the group velocity of an electron rather than its wave vector to the total current. The electron velocity in a crystal at the time t is given by:

$$V(\vec{k}(t)) = \frac{1}{\hbar}\frac{\partial \varepsilon_n(\vec{k}(t))}{\partial \vec{k}(t)}\,, \tag{1.226}$$

where $\vec{k}(t)$ is determined by equation (1.225). Consequently, the constant electric field causes the velocity of an electron in a metal to be a periodic and time limited function. If the electric field strength vector is codirected to the reciprocal lattice vector ($\vec{E} \uparrow\uparrow \vec{b}_i$), the function $\vec{V}$ is even oscillating! However, free electrons behave in an external electric field absolutely differently because their velocity $\vec{V}$ is proportional to $\vec{k}$ and linearly rises with time.

Now we turn our attention to a one-dimensional crystal model. Figure 1.57 follows the dependencies of the electron energy and its velocity on the wave number $\vec{k}$ in the reduced zone scheme.

Given that $\hbar\dot{k} = -|e|E = F$, where F is a force acting on the electron, we find the acceleration acquired by the electron under the force:

$$\dot{V} = \left(\frac{\partial V}{\partial k}\dot{k}\right) = \frac{\partial V}{\partial k}\frac{F}{\hbar}\,. \tag{1.227}$$

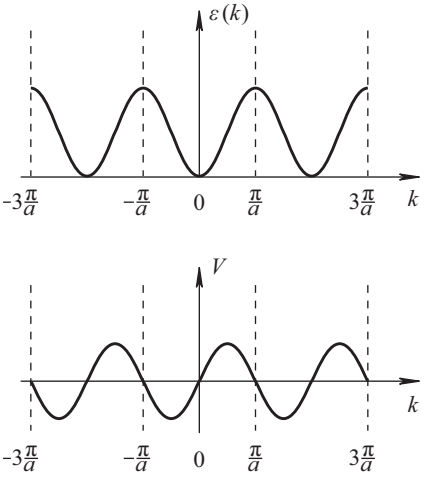


Fig. 1.57: The dependencies of the $\varepsilon(k)$ and $V(k)$ quantities on k for a one-dimensional crystal.

In classical physics, acceleration is always codirected to a force. However, we get that:

$$\begin{aligned} &\dot{V} \uparrow\uparrow F \quad \text{only when} \quad \frac{\partial V}{\partial k} > 0 \,, \\ &\dot{V} \uparrow\downarrow F \quad \text{when} \quad \frac{\partial V}{\partial k} < 0 \,. \end{aligned} \tag{1.228}$$

Of course, the case of a one-dimensional crystal talks about directions of the force and acceleration only along a straight line.

The electron behaves so unusually due to an additional force caused by the periodic potential of the lattice ions. Nevertheless, though the semiclassical model does not explicitly contain the force, it takes the force into account through the dispersion law of $\varepsilon(\vec{k})$.

At $T = 0\,\mathrm{K}$ the density of the electric current in a one-dimensional crystal is calculated from the expression:

$$j = -|e| \int V\left(k_0 - \frac{|e|\,Et}{\hbar}\right) \frac{\mathrm{d}k_0}{2\pi} \,. \tag{1.229}$$

When the energy band is completely filled, the integration is over a unit cell of the reciprocal space, from $-\pi/a$ to π/a, for example. The result is a zero. When the band is partially filled, the integration is over a part of the segment $[-\pi/a, \pi/a]$. In this case, a static electric field in the crystal would create an alternating electrical current. In reality this does not happen due to the high frequency of collisions.

With reasonable values of the electric field strength and the relaxation time, the change of the wave vector of each electron between collisions is small in comparison to the band sizes. For the electric field of the order of 10^{-2} V/cm and the relaxation time of $\tau = 10^{-14}$ sec, the magnitude of $|e|E\tau/\hbar$ has the order of $10^{-1}\mathrm{cm}^{-1}$. The band sizes are about $1/a \sim 10^{8}\ \mathrm{cm}^{-1}$.

The motion of electrons in a conductor can be treated as follows. Under the electric field, the electron moves with acceleration until it loses a part of its energy due to collisions with a lattice defect, an impurity atom, or any other scatterer. Then the electron is again accelerated by the applied field, and the process is repeated. Such a stepwise motion of the electron can be characterized by some approximate average velocity of ordered motion. The periodic motion of electrons in real metals is impossible to observe because the electrons move a little between collisions. On average, for a long period of time, their motion looks almost uniform.

At the same time, the oscillatory motion of electrons in a constant electric field, in principle, can be expected to occur in sufficiently pure crystals at low temperatures. Let us estimate the possible parameters of such motion in a three-dimensional crystal. Equations (1.223) and (1.224) allow the law of conservation of energy:

$$\varepsilon(k(t)) - |e|\,\vec{E} \cdot \vec{r}(t) = \mathrm{const} \,. \tag{1.230}$$

Since

$$\vec{k}(t) = k(0) - \frac{|e|\,Et}{\hbar} \,,$$

(1.230) implies that the motion of the electron in r-space is finite. If $\vec{E} \uparrow\uparrow \vec{b}_i$, the electron in reciprocal space oscillates. We can now estimate the oscillation period:

$$\frac{|e|\,|\vec{E}|}{\hbar} T = |\vec{b}_i| \,. \tag{1.231}$$

$$T = \frac{\hbar}{|e|\,|\vec{E}|} |\vec{b}_i| \sim \frac{\hbar}{a|e|\,|\vec{E}|} \,, \tag{1.232}$$

as $|\vec{b}_i| \sim 1/a$ (a is the lattice constant). The amplitude of these oscillations amounts to $\Delta r \sim \Delta\varepsilon/|eE|$, where $\Delta\varepsilon$ is the width of the energy band.

II. The Motion of Electrons in a Magnetic Field

Much important information about the electronic properties of metals and semiconductors can be obtained by taking measurements of their response to various disturbances under a constant magnetic field:

$$\vec{E} = 0 \,, \quad \vec{B} = \text{const} \,.$$

For definiteness, suppose that the magnetic field is directed along the axis Oz. The semiclassical equations in such a field take the form:

$$\dot{\vec{r}} = \vec{V}_n = \frac{1}{\hbar} \frac{\partial \varepsilon_n(\vec{k})}{\partial \vec{k}} \,, \tag{1.233}$$

$$\hbar \dot{\vec{k}} = -\frac{|e|}{c} \left[\vec{V}_n(\vec{k}) \times \vec{B} \right] \,. \tag{1.234}$$

Of these, it follows directly that the vector component $\vec{k}$ along the field and the energy of the quasiparticle $\varepsilon(\vec{k})$ are integrals of motion:

$$\dot{\vec{k}} \cdot \vec{B} = 0 \quad \Rightarrow \quad k_z = \text{const} \,,$$

$$\frac{\mathrm{d}\varepsilon(\vec{k}(t))}{\mathrm{d}t} = \left(\frac{\partial \varepsilon}{\partial \vec{k}} \cdot \dot{\vec{k}} \right) = -\frac{|e|\hbar}{c\hbar} \vec{V} \cdot [\vec{V} \times \vec{B}] = 0 \quad \Rightarrow \quad \varepsilon(k(t)) = \text{const} \,.$$

The two conservation laws completely govern the trajectory of electrons and holes in k-space. The electrons (holes) move along the curves defined by the intersection of constant energy surfaces $\varepsilon(\vec{k}) = \text{const}$ and planes $k_z = \text{const}$ (Figure 1.58).

The properties of the crystal are determined mainly by electrons and holes having energy close to the Fermi energy. It can be thought that the surface $\varepsilon(\vec{k}) = \text{const}$ coincides accurately with the Fermi surface. In other words, it is the shape of the Fermi surface that attention should be paid to when moving the electrons and holes in a magnetic field.

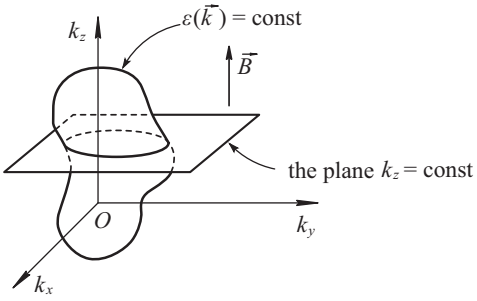


Fig. 1.58: The trajectory of an electron in k-space.

In order to establish the direction of orbital motion, we note that:

$$\vec{V} = \frac{1}{\hbar}\frac{\partial \varepsilon}{\partial \vec{k}} = \frac{1}{\hbar}\vec{\nabla}_k \varepsilon \,. \tag{1.235}$$

The gradient in turn evolves towards increasing energy. Consequently, the velocity of the quasiparticle is directed towards higher energies. In combination with the equation:

$$\hbar\dot{\vec{k}} = -\frac{|e|}{c}[\vec{V} \times \vec{B}] \,, \tag{1.236}$$

this means that if one travels in k-space along the path $\vec{k}(t)$ towards the motion of the quasiparticle, and the magnetic field is directed upwards, the right-hand side has values of $\vec{k}$ corresponding to higher energy levels.

This rule again yields the following conclusion. The filled energy electronic states near the bottom of the conduction band correspond to trajectories for negatively charged particles of classical physics. At the same time, the states near the top of the valence band and unoccupied by electrons form hole paths, which are similar to those for positively charged particles of classical physics (Figure 1.59). For definiteness, we will only talk about electrons.

In coordinate space, the electron has motion of more complex patterns. Consider the projection of its trajectory on a plane perpendicular to the vector of magnetic induction: $\vec{r}_\perp = \vec{r} - \vec{n}(\vec{n} \cdot \vec{r})$ where $\vec{n} = \vec{B}/|\vec{B}|$. The desired projection can be found by vectorially multiplying both sides of the equation (1.236) by the unit vector $\vec{n}$ directed

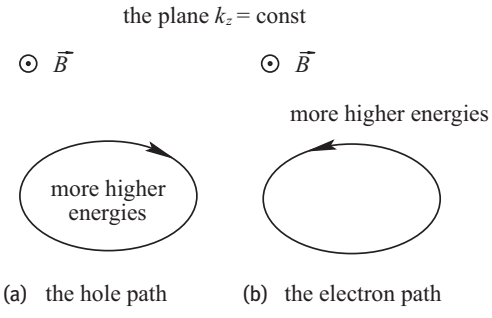


Fig. 1.59: The dependence of the direction of the quasiparticle motion in a magnetic field on the nature of the energy level filling.

along the field. This gives the following equality:

$$[\vec{n} \times \dot{\vec{k}}] = -\frac{|e|\,|\vec{B}|}{\hbar c}\left(\dot{\vec{r}} - \vec{n}(\vec{n} \cdot \dot{\vec{r}})\right) ,$$

or

$$[\vec{n} \times \dot{\vec{k}}] = -\frac{|e|\,|\vec{B}|}{\hbar c}\dot{\vec{r}}_{\perp} . \tag{1.237}$$

After integrating expression (1.237), we come to:

$$\vec{r}_{\perp}(t) - \vec{r}_{\perp}(0) = -\frac{c\hbar}{|e|\,|\vec{B}|}\left[\vec{n} \times \left(\vec{k}(t) - \vec{k}(0)\right)\right] . \tag{1.238}$$

Note that the cross product of the unit vector and its perpendicular vector equals to the latter rotated around the unit vector by an angle of 90°. Therefore, in real space the projection of the electron trajectory onto a plane perpendicular to the field is a trajectory rotated by 90° around the field direction in k-space, with the scales of the trajectories differing $(c\hbar)/(|e|\,|\vec{B}|)$ times.

In addition, attention should be drawn to the fact that, in the case of free electrons, the constant energy surfaces are the spheres $\varepsilon = (\hbar^2 k^2)/(2m)$. Their intersections with the planes $k_z = \text{const}$ yield circumferences. A circumference rotated by 90° remains a circumference. Thus, we again get the familiar result: being projected onto a plane perpendicular to the field, where the trajectory of the free electron outlines a circumference in the plane. In general semiclassical cases, the orbits are not necessarily circular; they are often even unclosed (Figure 1.60).

We cannot exactly say how the electron travels along the axis Oz. This is because the electron coordinate along the axis Oz is given by the relation:

$$z(t) = z(0) + \int_0^t V_z(\vec{k}(t'))\mathrm{d}t' , \tag{1.239}$$

where $V_z(\vec{k}(t)) = (1/\hbar)(\partial\varepsilon/\partial k_z)$. Even when $k_z = \text{const}$, the quantity V_z may be variable. Therefore, the electron's motion along the magnetic field may be uneven. This is another illustration of the fact that an electron in a crystal behaves differently to an isolated electron in a constant magnetic field.

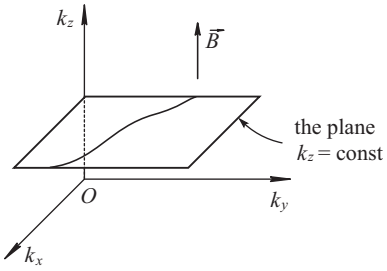


Fig. 1.60: The path in the plane $k_z = \text{const}$ (in this case, the trajectory is unclosed).

The velocity of motion along the orbit in k-space can be expressed in terms of the geometric characteristics of the band structure. Suppose an orbit with the energy ε lies in the plane k_z = const, perpendicular to the applied field. The time required for the passage of a part of the orbit between the points with the $\vec{k}_1$ and $\vec{k}_2$ radius vectors is estimated as the integral:

$$t_2 - t_1 = \int_{t_1}^{t_2} \mathrm{d}t = \int_{k_1}^{k_2} \frac{\mathrm{d}k}{|\dot{\vec{k}}|} .$$

In what follows, it is considered that $\mathrm{d}k = |\mathrm{d}\vec{k}|$. Given that:

$$|\dot{\vec{k}}| = \frac{|e|}{c\hbar} \left|[\vec{V} \times \vec{B}]\right| = \frac{|e|\,|\vec{B}|}{c\hbar} |V_\perp| = \frac{|e|\,|\vec{B}|}{c\hbar^2} \left| \left(\frac{\partial \varepsilon}{\partial \vec{k}} \right)_\perp \right| ,$$

we come up with:

$$t_2 - t_1 = \frac{\hbar^2 c}{|eB|} \int_{k_1}^{k_2} \frac{\mathrm{d}k}{\left|(\partial\varepsilon/\partial\vec{k})_\perp\right|} . \tag{1.240}$$

Here, $((\partial\varepsilon)/(\partial\vec{k}))_\perp$ is the component of the vector $(\partial\varepsilon)/(\partial\vec{k})$, perpendicular to the field or the projection of this vector onto the plane k_z = const.

Refer to Figure 1.61. Suppose $\vec{\Delta}(\vec{k})$ is a vector lying in the plane k_z = const and that it is perpendicular to the trajectory with the energy ε at the point $\vec{k}$. Let the energy $(\varepsilon + \mathrm{d}\varepsilon)$ corresponds to a trajectory lying in the same plane and this vector connects the two paths. Now, when the magnitude of $\mathrm{d}\varepsilon$ is small and positive, we will have the following:

$$\mathrm{d}\varepsilon = \left(\frac{\partial \varepsilon}{\partial \vec{k}} \cdot \vec{\Delta}(\vec{k}) \right) = \left| \left(\frac{\partial \varepsilon}{\partial \vec{k}} \right)_\perp \right| |\vec{\Delta}(\vec{k})| . \tag{1.241}$$

As can be seen from the figure, the electron's trajectory in reciprocal space is the intersection of the surfaces $\varepsilon(\vec{k})$ = const and k_z = const. The vector $(\partial\varepsilon/\partial\vec{k})$ is perpendicular to the surface $\varepsilon(\vec{k})$ = const. Consequently, the vector $(\partial\varepsilon/\partial\vec{k})$ is perpendicular to the trajectory as well but, generally speaking, does not belong to the plane k_z = const. At the same time, the vector $(\partial\varepsilon/\partial\vec{k})_\perp$ is perpendicular to the trajectory

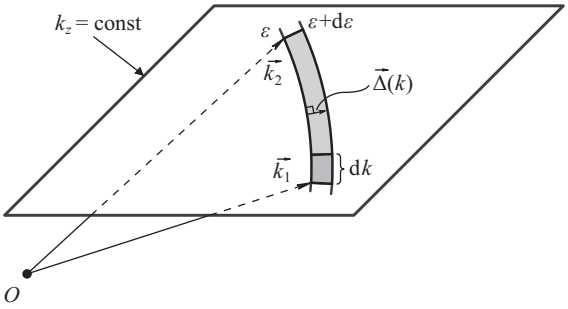


Fig. 1.61: The shaded strip in the plane k_z = const, bounded by the trajectories with the ε and $(\varepsilon + \mathrm{d}\varepsilon)$ energies.

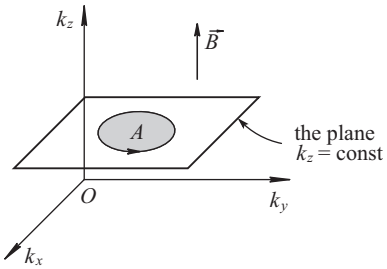


Fig. 1.62: The electron motion along a closed path in k-space.

and lies in the plane k_z = const. Moreover, under the condition $d\varepsilon > 0$ it is codirected to the vector $\vec{\Delta}(\vec{k})$. Then expression (1.240) can be written in the form:

$$t_2 - t_1 = \frac{\hbar^2 c}{|eB|} \frac{1}{d\varepsilon} \int_{k_1}^{k_2} dk |\vec{\Delta}(k)| = \frac{\hbar^2 c}{|eB|} \frac{dA_{12}}{d\varepsilon} . \tag{1.242}$$

The integral computes the area of A_{12}, a segment of the plane k_z = const. The segment is bounded between two nearest trajectories from the point with the radius vector $\vec{k}_1$ to the point with the radius vector $\vec{k}_2$. The quantity $(dA_{12})/(d\varepsilon)$ is the rate of change of the area swept out by a part of the trajectory between the points with the above radius vectors in the plane k_z = const.

When the trajectory is a simple closed curve, the $\vec{k}_1$ and $\vec{k}_2$ vectors can be chosen in such a way to obtain a closed loop. Then the value of $(t_2 - t_1)$ equals to a period T of the electron's motion along the closed orbit. If A is the area of that part of the plane k_z = const that our closed trajectory covers (Figure 1.62), formula (1.242) provides:

$$T(\varepsilon, k_z) = \left| \frac{\hbar^2 c}{eB} \frac{d}{d\varepsilon} A(\varepsilon, k_z) \right| . \tag{1.243}$$

For various values of k_z, information of the Fermi surface shape can be obtained through measuring the period of the electron motion.

1.18 General Properties of Semiconductors: the Concentration of Electrons and Holes and the Law of Mass Action

As we have previously seen, electrons of a completely filled band cannot carry a current. It is this property that determines the difference between metals and insulators. An insulator, being in the ground state, has all its bands either completely full or completely empty. A metal in the ground state contains at least a partially filled band.

Insulators can be characterized by the value of the energy gap or band gap between the top of the valence band and the bottom of the conduction band (Figure 1.63).

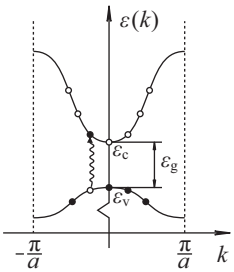


Fig. 1.63: The spectrum of electron energy in a one-dimensional crystal. The black points indicate states occupied by electrons.

Solids, being insulators at $T = 0\,\mathrm{K}$ but having energy gaps less than $\varepsilon_g \leq 2\,\mathrm{eV}$, are called *semiconductors*. These materials show noticeable conductivity when thermally excited at temperatures below the melting point.

At nonzero temperature, the smallness of band gaps causes high probability of transferring some thermally excited electrons from the valence band to the conduction band. As a result, equal small numbers of the electrons with the charge $-|e|$ and the holes with the charge $+|e|$ emerge in the conduction band and in the valence band, respectively. Consequently, the low electrical conductivity of semiconductors is provided by both electrons of the conduction band and holes (missing electrons) of the valence band.

The typical specific resistivities of semiconductors at room temperature lie in the range of 10^{-3} and $10^{9}\,\Omega \cdot \mathrm{sm}$. As a reference source, the specific resistance of metals is $\rho \approx 10^{-6}\,\Omega \cdot \mathrm{sm}$. For good insulators, ρ can reach up to $10^{22}\,\Omega \cdot \mathrm{sm}$. We can say that a semiconductor is a bad insulator at room temperature.

The conductivity coefficient σ for metals in the free electron approximation is given by:

$$\sigma = \frac{ne^2\tau}{m}\,, \tag{1.244}$$

and decreases with increasing temperature. In metals, the carrier concentration n is not temperature dependent. The entire temperature dependence is associated with the relaxation time τ. The latter diminishes as the temperature grows due to intensifying effects of electron scattering by lattice vibrations.

Unlike metals, semiconductors change not only τ but their carrier density n under varying temperature. Namely, the concentration rises steeply with increasing temperature. Moreover, the contribution of an increase in the charge carrier concentration prevails over the contribution of a decrease in the relaxation time τ. As a result, the temperature goes up, and the conductivity coefficient σ increases as well. Therefore, in contrast to metals, the main feature of semiconductors is for their resistance to fall sharply as the temperature rises.

The case study indicates that the most important issue in semiconductor physics is to calculate the charge carrier concentration. To put it differently, the density of electrons in the conduction band and of holes in the valence band must be found.

The carrier concentration in semiconductors is very low. Therefore, the behavior of electrons and holes obeys Maxwell–Boltzmann statistics. Furthermore, the electron and hole concentration values are strongly impurity dependent. That is why the early stages of development of semiconductor physics significantly delayed the accumulation of reliable information on the semiconductor materials.

Semiconductor physics of the twentieth century took great strides towards developing world civilization. High efficiency solar cells, devices of modern microelectronics, powerful compact lasers, methods of recording, storing, and reading information in modern computers, etc., use the unique properties of semiconductors.

The Number of Carriers at Thermodynamic Equilibrium

The most important characteristic of semiconductors at a temperature T is the number of electrons in the conduction band per unit volume and the number of holes in the valence band per unit volume. Our task is to learn to calculate the concentrations as functions of temperature. This is because it is these electrons and holes being close to the maximum and minimum of the conduction band and valence band, respectively, that are responsible for the conductivity of semiconductors. Near the extrema specified, the relations between the energy and the wave vector of electrons and holes are approximated with good accuracy by quadratic forms:

– For electrons:

$$\varepsilon(\vec{k}) \approx \varepsilon_c + \frac{\hbar^2}{2m_e}(k - k_c)^2 , \tag{1.245}$$

– For holes:

$$\varepsilon(\vec{k}) \approx \varepsilon_v - \frac{\hbar^2}{2m_h}(k - k_v)^2 , \tag{1.246}$$

where m_e is the mass of an electron and m_h is the mass of a hole; $k = |\vec{k}|$. Formulas (1.245) and (1.246) result in the series expansion of the dispersion law:

$$\varepsilon = \varepsilon(|\vec{k}|) . \tag{1.247}$$

Here, we are viewing a simple dispersion law (1.247) in order to not complicate further exposition.

Let us designate the density of energy levels in the conduction band as $\upsilon_c(\varepsilon)$, in the valence band as $\upsilon_v(\varepsilon)$. The level densities can be taken from the theory of an ideal gas of fermions, where the relationship between ε and k is similar to (1.245) and (1.246):

– For electrons in the conduction band ($\varepsilon > \varepsilon_c$):

$$\nu_c(\varepsilon) = \frac{m_e}{(\pi\hbar)^2}\sqrt{\frac{2m_e}{\hbar^2}(\varepsilon - \varepsilon_c)} , \tag{1.248}$$

– For holes in the valence band ($\varepsilon < \varepsilon_v$):

$$\nu_v(\varepsilon) = \frac{m_h}{(\pi\hbar)^2}\sqrt{\frac{2m_h}{\hbar^2}(\varepsilon_v - \varepsilon)}\,. \tag{1.249}$$

The concentration values, as we will see later, are strongly impurity dependent. However, there are some general relations which should be considered in the first instance.

I. The Concentration of Electrons

It is worth recollecting that the conductivity of semiconductors is governed both by electrons residing at levels in the conduction band and by holes being at levels in the valence band. Then, whatever the impurity concentration, the number of carriers available at a given temperature T in the range of energies $(\varepsilon, \varepsilon + d\varepsilon)$ is described by the following formulas:

– For electrons:

$$\frac{\Delta N}{V} = \nu_c(\varepsilon) f(\varepsilon) d\varepsilon\,, \tag{1.250}$$

– For holes:

$$\frac{\Delta N}{V} = \nu_v(\varepsilon)(1 - f(\varepsilon)) d\varepsilon\,, \tag{1.251}$$

where the Fermi–Dirac distribution function appears as:

$$f(\varepsilon) = \frac{1}{\exp\left[(\varepsilon - \mu)/k_B T\right] + 1}\,. \tag{1.252}$$

In determining the concentration of electrons and holes, the effect of impurities makes itself felt only through the chemical potential μ. For the latter, as will be shown below, the following relations hold:

$$\begin{aligned} \varepsilon_c - \mu &\gg k_B T\,, \\ \mu - \varepsilon_v &\gg k_B T\,. \end{aligned} \tag{1.253}$$

Next, we employ the following procedure: taking the conditions (1.253), we find the concentration values for electrons and holes in semiconductors. Knowing information about the possible impurity levels, we can calculate the real value of the chemical potential μ. In conclusion, we check whether μ falls in the range defined by the conditions (1.253).

The concentration of electrons in the conduction band is given by:

$$n = \int_{\varepsilon_c}^{+\infty} \nu_c(\varepsilon) f(\varepsilon) d\varepsilon\,. \tag{1.254}$$

Under the conditions (1.253) the Fermi–Dirac distribution reduces to the Maxwell–Boltzmann distribution:

$$f(\varepsilon) = \left[\exp\left(\frac{\varepsilon - \mu}{k_B T}\right) + 1\right]^{-1} \approx \exp\left(\frac{\mu - \varepsilon}{k_B T}\right)\,. \tag{1.255}$$

As can be noticed, the multiplier (1.255) in the integrand of (1.254) rapidly decreases. Therefore, the replacement of the upper limit of integration in formula (1.254) by $+\infty$ makes no discernible errors.

Entering the new variable of integration:

$$\frac{\varepsilon - \varepsilon_c}{k_B T} = u \,,$$

from (1.254) and (1.255), we find:

$$n \approx \int_{\varepsilon_c}^{\infty} \frac{m_e}{(\pi\hbar)^2} \sqrt{\frac{2m_e}{\hbar^2}(\varepsilon - \varepsilon_c)} \exp\left(\frac{\mu - \varepsilon}{k_B T}\right) d\varepsilon =$$

$$= \frac{m_e}{(\pi\hbar)^2} \sqrt{\frac{2m_e}{\hbar^2}} \exp\left(\frac{\mu - \varepsilon_c}{k_B T}\right) (k_B T)^{3/2} \underbrace{\int_0^{\infty} u^{1/2} e^{-u} du}_{\sqrt{\pi}/2} \,,$$

$$n \approx N_c \exp\left(-\frac{\varepsilon_c - \mu}{k_B T}\right) . \tag{1.256}$$

Here, the effective density of states in the conduction band is:

$$N_c = \frac{1}{4}\left(\frac{2m_e k_B T}{\pi\hbar^2}\right)^{3/2} . \tag{1.257}$$

It is necessary to emphasize that formula (1.257) is valid both for impurity and for pure semiconductors.

We estimate N_c for silicon. At $T = 300$ K and for $m_e = 0.2$ m, where m is the tabular mass of an electron, we have $N_c = 2.2 \cdot 10^{18} \text{cm}^{-3}$.

II. The Concentration of Holes

The function $f(\varepsilon)$ gives the average number of spin projected electrons occupying the energy level ε. According to the Pauli principle:

$$0 \le f(\varepsilon) \le 1 \,. \tag{1.258}$$

The function $f(\varepsilon)$ can be interpreted as the probability of occupying the energy state ε by spin projected electrons. Then, the state occupation probability for holes is:

$$f_h(\varepsilon) = 1 - f(\varepsilon) = 1 - \frac{1}{\exp\left[(\varepsilon - \mu)/k_B T\right] + 1} = \frac{1}{\exp\left[(\mu - \varepsilon)/k_B T\right] + 1} \,. \tag{1.259}$$

Under the conditions (1.253), the distribution (1.259) also reduces to the Boltzmann distribution:

$$f_{\mathrm{h}}(\varepsilon) \approx \exp\left(-\frac{\mu-\varepsilon}{k_{\mathrm{B}}T}\right) . \tag{1.260}$$

Hence, it is easy to estimate the concentration of holes in the same way as the electron density has been recently calculated:

$$p \approx \int_{-\infty}^{\varepsilon_{\mathrm{v}}} \nu_{\mathrm{v}}(\varepsilon) f_{\mathrm{h}}(\varepsilon)\mathrm{d}\varepsilon . \tag{1.261}$$

$$p \approx N_{\mathrm{v}} \exp\left(-\frac{\mu-\varepsilon_{\mathrm{v}}}{k_{\mathrm{B}}T}\right) . \tag{1.262}$$

Here N_{v} is the effective density of states in the valence band, both in impurity and in pure semiconductors. It equals to:

$$N_{\mathrm{v}} = \frac{1}{4}\left(\frac{2m_{\mathrm{h}}k_{\mathrm{B}}T}{\pi\hbar^2}\right)^{3/2} . \tag{1.263}$$

The electron and hole density values cannot be computed until the chemical potential μ is unknown. However, the product of these two concentrations do not depends on μ:

$$np = N_{\mathrm{c}}N_{\mathrm{v}} \exp\left(-\varepsilon_{\mathrm{g}}/k_{\mathrm{B}}T\right) , \tag{1.264}$$

where $\varepsilon_{\mathrm{g}} = (\varepsilon_{\mathrm{c}} - \varepsilon_{\mathrm{v}})$ is the energy gap width. The result obtained is referred to as the *law of mass action*. It means that it is sufficient to know the concentration of one charge carrier type, in order to find the concentration of another charge carrier type (at a given temperature). Methods of calculating the densities depend on how much the impurities contribute to the carrier concentration.

The universal formula (1.264) originates from the following simple reasoning. In a state of thermodynamic equilibrium, the number of electrons passing from the valence band to the conduction band per second is proportional to the transition probability of a single electron from the valence to conduction band. It can therefore be written as:

$$N_{\mathrm{trans}} = A \exp\left(-\varepsilon_{\mathrm{g}}/k_{\mathrm{B}}T\right) , \tag{1.265}$$

where A is a coefficient of proportionality. The electron can go back only if the valence band has a hole. Then the number of transitions of electrons per second from the upper conduction band to the lower valence band must be proportional to the product of the electron and hole concentrations. In a state of thermodynamic equilibrium, the numbers of direct and inverse transitions are equal, so we have:

$$np \sim \exp\left(-\varepsilon_{\mathrm{g}}/k_{\mathrm{B}}T\right) . \tag{1.266}$$

1.19 Intrinsic Semiconductors

A semiconductor crystal that is so pure that the impurities make a negligible contribution to the carrier concentration is called an *intrinsic semiconductor*. In intrinsic semiconductors, the concentrations of electrons n_i and holes p_i are ideally equal, since these particles emerge and annihilate in pairs. Using the law of mass action (1.264), we find the concentration value:

$$n_i = p_i = \sqrt{N_c N_v} \exp\left(-\varepsilon_g / 2k_B T\right) . \tag{1.267}$$

In formula (1.267), the main temperature dependence is related to the exponential factor. In the calculations, it can be considered that $N_c N_v \approx$ const in the main approximation (Figure 1.64). The index i, as labeled herein, denotes parameters of pure semiconductors.

We can find the chemical potential μ_i for intrinsic semiconductors, setting equal the expressions for the concentrations (1.267) and (1.262):

$$\mu_i = \varepsilon_v + \frac{1}{2}\varepsilon_g + \frac{k_B T}{2} \ln \frac{N_v}{N_c} = \frac{\varepsilon_v + \varepsilon_c}{2} + \frac{3}{4} k_B T \ln \frac{m_h}{m_e} . \tag{1.268}$$

At $T \to 0\,\mathrm{K}$, the chemical potential lies exactly in the middle of the band gap (Figure 1.65). Moreover, since the ratio m_h/m_e is a magnitude of the order of unity, the chemical potential μ_i does not differ much from the value $(\varepsilon_v + \varepsilon_c)/2$, even at room temperatures. As a result, we have:

– For electrons:

$$\varepsilon - \mu \geq \varepsilon_g / 2 , \tag{1.269}$$

– For holes:

$$\mu - \varepsilon \geq \varepsilon_g / 2 . \tag{1.270}$$

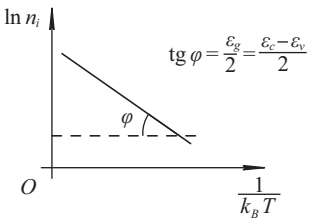


Fig. 1.64: The concentration of electrons in intrinsic semiconductors.

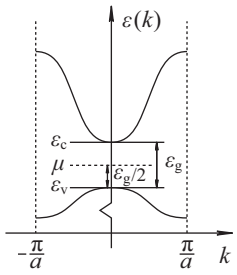


Fig. 1.65: A chemical potential energy diagram for intrinsic semiconductors (a one-dimensional crystal).

In the calculations, the assumption has been done that $|\varepsilon - \mu| \gg k_B T$ (1.253). It is equivalent to $\varepsilon_g \gg k_B T$. The latter condition is fulfilled for all semiconductors at about room temperature.

Let us carry out some numerical estimates. The concentration of charge carriers in intrinsic semiconductors at room temperature is $n_i \sim 10^6\ \mathrm{cm}^{-3}$. This value is many orders less than the concentration of electrons in metals ($n_i \sim 10^{22}\mathrm{cm}^{-3}$). Consequently, the conductivity of intrinsic semiconductors is extremely small. Experimentally, the electrical conductivity observable exceeds the intrinsic electrical conductivity of even very pure semiconductors. This is explained by the capability of even small proportions of impurities to significantly alter the charge carrier concentration and, hence, the conductivity of semiconductors.

1.20 Impurity Levels

An impurity is called a *donor* that supplies extra electrons to the conduction band. An impurity is called an *acceptor* that supplies extra holes to the valence band or captures electrons from it.

Donor impurities are atoms of a higher valence than those forming a pure substance. Acceptors, in turn, are lower valence atoms.

The most common semiconductor of the group IV elements of the periodic table is germanium (Ge). Each atom of germanium is at the center of a regular triangular pyramid whose vertices have four neighbors. Silicon and diamond are the same crystal lattice. Figure 1.66 shows a unit cell of the crystal lattice of germanium (atoms of germanium reside in vertices of a tetrahedron and in its center).

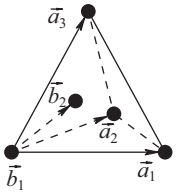


Fig. 1.66: The unit cell of the crystal lattice of germanium, where $\vec{a}_1, \vec{a}_2, \vec{a}_3$ are Bravais lattice vectors; $\vec{b}_1 = (0, 0, 0)$, $\vec{b}_2 = (1/4)(\vec{a}_1\vec{a}_2\vec{a}_3)$ are vectors characterizing the positions of germanium atoms (black spheres) in the unit cell.

For clarity, we project the picture of bonds on a plane (Figure 1.67). The double lines illustrate covalent bonds between germanium atoms. Each atom of germanium donates four valence electrons to bond its neighbors. Finding two electrons in the double line area is highly probable.

Let us discuss what happens to a semiconductor doped by a small amount of arsenic (As); a pentavalent impurity. Arsenic has five valence electrons. Four of them are adjacent to germanium atoms, according to the scheme previously proposed. The fifth electron is redundant. The extra thermally excited electrons tear away from the arsenic ions, leaving a fixed positive charge $+|e|$ in the lattice site. These provide con-

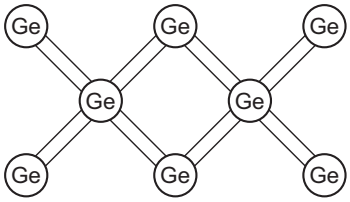


Fig. 1.67: The picture of germanium atom bonds in a plane.

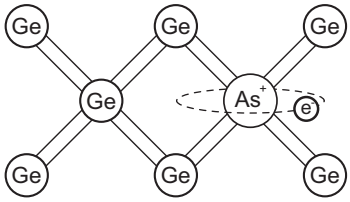


Fig. 1.68: The picture of atom bonds of germanium with an arsenic donor impurity atom.

ductivity. An arsenic atom, when ionized, is a donor because it gives its electron into the conduction band.

We estimate the energy of a loose chemical bond between the fifth electron and the As^+ ion core. The problem of seeking the wave functions and energies of the excessive (fifth) electron is similar to the problem of a hydrogen atom. The As^+ ion core plays the role of a proton; the fifth electron is like a single electron in a hydrogen atom (Figure 1.68).

If the impurity was not inside the semiconductor but in the empty space, the impurity ionization energy would be just equal to the first ionization potential of the hydrogen atom, equal to 13.6 eV. However, due to being the impurity inside the semiconductor, the ionization energy is significantly reduced. This occurs for the following reasons.

1. The effective mass of the electron in the crystal is significantly different from the free electron mass.
2. The excessive electron in the Coulomb field of the impurity atom has the potential energy $-e^2/\varepsilon r$ (in the CGS system), where ε is the dielectric permittivity of covalent crystals. The multiplier $1/\varepsilon$ ($\varepsilon > 1$) allows for the screening of the Coulomb forces due to electronic polarization of the medium.

It should be especially emphasized that microscopics validly describes the screening of the electronion interactions in terms of the dielectric permittivity ε only for the electron orbits sufficiently large compared with the interatomic distance: $R_{\text{orb}} \gg a$. In this case, the motion of the electron is slow enough because its orbital revolution frequency ω_{orb} must be much smaller than the characteristic frequency $\varepsilon_g/\hbar$ of energy transitions in the semiconductor. Only under these conditions does the electron feel the mean field of the crystal.

The conditions formulated above are well met for pentavalent impurities being added to semiconductor crystals with covalent bonds of tetravalent atoms. Later we will confirm this by numerically estimating R_{orb} and ω_{orb}.

It is known that the binding energy of an electron in a hydrogen atom is:

$$\varepsilon_b = -\frac{e^4 m}{2\hbar^2}\,. \tag{1.271}$$

In our case, e^2 substitutes for e^2/ε, and m – for m_e. Therefore, the ionization energy of the donor impurity in the semiconductor (the ionization energy differs in the sign from the binding energy) takes the following form:

$$W_d = \frac{e^4 m_e}{2\varepsilon^2 \hbar^2} = \frac{m_e}{m\varepsilon^2} 13.6\,\text{eV}\,. \tag{1.272}$$

We see that an additional factor appears here.

The Bohr radius of an electron in a hydrogen atom is:

$$a_B = \frac{\hbar^2}{me^2}\,. \tag{1.273}$$

Therefore, we can determine the electron orbital radius around the As^+ ion:

$$R_{orb} = \frac{\varepsilon\hbar^2}{m_e e^2} = \frac{\varepsilon m}{m_e} a_B\,. \tag{1.274}$$

Since the mass m_e of an electron in semiconductors is about one order of magnitude smaller than the free electron mass and the relative dielectric constant is one order greater:

$$\frac{m}{m_e} \sim 10\,, \quad \varepsilon \sim 10\,, \tag{1.275}$$

from (1.274) and (1.275) we obtain:

$$R_{orb} \approx 100 a_B \gg a_B\,. \tag{1.276}$$

Formula (1.276) justifies our calculation because of the assumption that the orbital radius of the fifth electron is much greater than the interatomic distance underlying both the semiclassical model and the macroscopic dielectric permittivity being used.

In addition, we conclude from (1.275) and (1.272) that the donor ionization energy may diminish a thousand times or more as compared to the value of 13.6 eV:

$$W_d \approx 0.013\,\text{eV}\,, \quad \varepsilon_g \approx (1 \div 2)\,\text{eV} \quad \Rightarrow \quad W_d \ll \varepsilon_g\,. \tag{1.277}$$

Therefore, the condition of slow motion along the orbit is fulfilled:

$$\omega_{orb} \sim \frac{W_d}{\hbar} \ll \frac{\varepsilon_g}{\hbar}\,. \tag{1.278}$$

For the hydrogen configuration at hand, the conduction band plays the role of a continuous spectrum. So the electron impurity binding energy is measured from the bottom of the conduction band downwards. This yields the logical conclusion that the donor impurities cause extra electronic levels to emerge, with the latter energies ε_d being less than the energies ε_c corresponding to the conduction band bottom. Moreover, the energy difference $\varepsilon_c - \varepsilon_d = W_d$ is small compared with the width of the forbidden band.

If one somehow succeeds in increasing the electron energy of the impurity by W_d, the electron will transition to the bottom of the conduction band where its wave function is not localized near the donor site. At low enough temperatures, donor impurity semiconductors have electrons as major charge carriers. These are called *n-type semiconductors* (*n* stands for negative current carriers).

At high temperatures, the transition of electrons from the valence to the conduction band is possible, so that the impurity and intrinsic conductivities are combined.

Similar reasoning is applicable to acceptor impurities, whose valence is unity less than the valence of the basic substance. For example, a boron atom (B) has only three valence electrons. It needs one electron from the bonds Ge–Ge to complete its covalent tetrahedral bonds with the basic substance atoms, forming a hole in the germanium valence band. The hole formed and the B^- ion are weakly connected with each other (Figure 1.69). The hole, when thermally excited, easily breaks the bond, taking part in the conductivity of the semiconductor. The boron atom is an acceptor because it turns into the B^- ion through capturing an electron from the valence band. Semiconductors doped with acceptor impurities are known as *p-type semiconductors*. Their main charge carriers are holes, i.e., positive charge carriers.

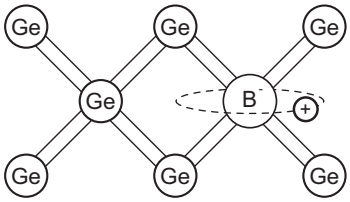


Fig. 1.69: The pattern of bonds between Ge atoms and an acceptor impurity atom of B.

The binding energy between the positive hole and the negative ion B^- can be calculated in a similar manner. At the same time, the energy diagram indicates that the hole increases its energy when passed towards the negative energy region. Therefore, unlike the previous case, the binding energy between the hole and the impurity atom turns out to be positive. This energy is small, compared to the width of the forbidden zone, and has a value of:

$$W_a = \frac{m_h e^4}{2\varepsilon\hbar^2} \,. \tag{1.279}$$

From a physical point of view, the ionization of the acceptor atom is to transition an electron from the valence band to the level ε_a lying slightly above the valence band bottom: $\varepsilon_a - \varepsilon_v = W_a > 0$.

Formally, that process can be treated as follows. Being initially in the orbit near the B^- ion, the hole changes its ground state ε_a to states of a continuous spectrum of the hydrogen like configuration. These states belong to the valence band of the semiconductor.

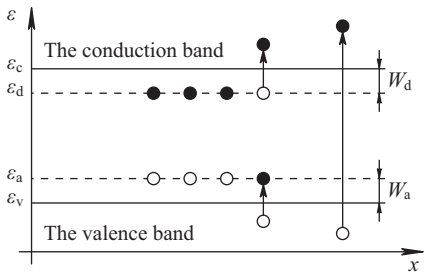


Fig. 1.70: Energy diagram: $(\varepsilon_c - \varepsilon_d)$ is the ionization energy of an electron localized on a donor impurity, $(\varepsilon_a - \varepsilon_v)$ is the binding energy of a hole localized near an acceptor impurity.

General Conclusions

The impurity energy levels always lie in the forbidden band near the top of the valence band and the conduction band bottom. As temperature grows, the transition of an electron into the conduction band from the donor level or a hole into the valence band from the acceptor level is much easier to cause than the transition of the electron from the valence band to the conduction band, as it did in the case of intrinsic semiconductors (Figure 1.70).

To invent new semiconductor devices with required electrical properties, we have to be able to govern the concentrations of electrons and holes.

1.21 Concentrations of Charge Carriers and the Chemical Potential of Impurity Semiconductors

To determine how many current carriers can be transferred from impurity levels by thermal excitation, it is necessary to know the average number of electrons at these levels.

Donor Level

For definiteness, we consider a crystal of germanium Ge (the IV group of the periodic table) doped with arsenic As (the V group). Suppose that a unit volume of the sample has N_d arsenic atoms. Suppose the impurity concentration is low, so the interaction of electrons localized on different atoms of As can be neglected.

We can find the electron density n_d by simply multiplying the donor concentration N_d by the average number of electrons $f(\varepsilon_d)$ localized on an individual impurity:

$$n_d = N_d \widetilde{f}(\varepsilon_d) \,. \tag{1.280}$$

Earlier, we used the Fermi–Dirac distribution function:

$$f(\varepsilon_d) = \frac{1}{\exp\left(\frac{\varepsilon_d - \mu}{k_B T}\right) + 1} \,. \tag{1.281}$$

However, in this case we should use the revised function $\widetilde{f}(\varepsilon_\mathrm{d})$. Let us explain the reason.

The level of each impurity atom can be either empty or occupied by one electron with arbitrary directed spin, or occupied by two electrons with opposite spins. The Coulomb repulsion of two electrons strongly localized on the impurities increases the energy level that finding two electrons on the same level is not possible. With this in mind, the average number of electrons localized on a single impurity atom is equal to:

$$\widetilde{f}(\varepsilon_\mathrm{d}) = \frac{1}{\frac{1}{2}\exp\left(\frac{\varepsilon_\mathrm{d}-\mu}{k_\mathrm{B}T}\right)+1}\,. \tag{1.282}$$

The proof of (1.282) is given, for example, in the book by Ashcroft and Mermin [1].

Therefore, from (1.280) and (1.282), we come up with an expression for the electron density:

$$n_\mathrm{d} = \frac{N_\mathrm{d}}{\frac{1}{2}\exp\left(\frac{\varepsilon_\mathrm{d}-\mu}{k_\mathrm{B}T}\right)+1}\,. \tag{1.283}$$

Acceptor Level

Analogously, the configuration that meets two holes localized near the acceptor impurity has a very large amount of energy because of the mutual Coulomb repulsion of two holes. Therefore, it is not implemented. This circumstance leads to a modification of the distribution function of the holes localized on the acceptor impurities. As a result, the concentration of holes connected with the acceptor impurities, is defined by the relation:

$$p_a = \frac{N_\mathrm{a}}{\frac{1}{2}\exp\left(\frac{\mu-\varepsilon_\mathrm{a}}{k_\mathrm{B}T}\right)+1}\,. \tag{1.284}$$

Here N_a is the number of acceptor atoms per unit volume of the sample.

Returning to the example of germanium doped with arsenic, we calculate the chemical potential μ through an equation called the electroneutrality equation. The latter replaces the previous normalization condition not suitable for this task since the total number of electrons and holes is not constant. As temperature increases, the ionization of both the impurity atoms and the atoms of the host material intensifies. As a result, the number of electrons and holes grows. We use the fact that, in any small volume of a substance, the total charge of all particles in the medium must remain zero.

The positive charge is created by 1) the holes; 2) the As^+ ions.

The negative charge is created by the electrons only. By the charge conservation law, we have the equation:

$$n = p + N_\mathrm{d}^+\,, \tag{1.285}$$

where N_d^+ is the number of the ionized impurity atoms per unit volume of the semiconductor, n is the total electron concentration, and p is the total hole concentration.

Since the concentration of donor atoms is N_d, we get:

$$N_\mathrm{d}^+ = N_\mathrm{d} - n_\mathrm{d} \,, \tag{1.286}$$

where n_d is the concentration of electrons remaining at the donor levels. For example, at $T = 0\,\mathrm{K}$ we have $N_\mathrm{d}^+ = 0$. The impurity is not ionized because $n_\mathrm{d} = N_\mathrm{d}$. Plugging (1.286) into (1.285), we obtain:

$$n = p + N_\mathrm{d} - n_\mathrm{d} \,. \tag{1.287}$$

Based on (1.282) and (1.287), we arrive at:

$$n = p + N_\mathrm{d}(1 - \widetilde{f}(\varepsilon_\mathrm{d})) = p + N_\mathrm{d}\left[1 - \left(\frac{1}{2}\exp\left(\frac{\varepsilon_\mathrm{d} - \mu}{k_\mathrm{B}T}\right) + 1\right)^{-1}\right] \,. \tag{1.288}$$

This is the condition of electrical neutrality.

Let us recall the general formulas for calculating the concentrations of electrons and holes in a semiconductor:

$$n \approx N_\mathrm{c}\exp\left(-\frac{\varepsilon_\mathrm{c} - \mu}{k_\mathrm{B}T}\right) \,, \tag{1.289}$$

$$p \approx N_\mathrm{v}\exp\left(-\frac{\mu - \varepsilon_\mathrm{v}}{k_\mathrm{B}T}\right) \,. \tag{1.290}$$

Equations (1.288)–(1.290) determine the value of μ and the concentrations of electrons and holes in semiconductors doped with a donor impurity.

Semiconductors with a Donor Impurity

Our goal is to find working formulas for calculating the concentrations of charge carriers and the chemical potential μ for semiconductors doped with a donor impurity.

The chemical potential in general depends on a number of parameters:

$$\mu = \mu(m_\mathrm{e}, m_\mathrm{h}, T, N_\mathrm{d}, \varepsilon_\mathrm{c}, \varepsilon_\mathrm{v}, \varepsilon_\mathrm{d}) \,. \tag{1.291}$$

As mentioned earlier, the total number of electrons and holes is not constant. With increasing temperature, their number rises. Therefore, the chemical potential needs to be determined from the condition of electrical neutrality. Equations (1.288)–(1.290) are impossible to solve analytically with respect to μ. Consequently, we should analyze the limiting cases.

I. Region of High Temperatures

At high temperatures, the impurity is ionized throughout: $N_\mathrm{d}^+ = N_\mathrm{d}, \widetilde{f} \approx 0$. Therefore, the relation (1.288) is simplified:

$$n = p + N_\mathrm{d} \,. \tag{1.292}$$

We use the universal equation – the law of mass action:

$$np = n_i^2 \quad \Rightarrow \quad p = n_i^2/n \,. \tag{1.293}$$

From (1.292) and (1.293), the equation

$$n^2 - nN_d - n_i^2 = 0 \tag{1.294}$$

emanates to compute n. It has two roots, but one of them is redundant because of its negative value. Only one solution remains:

$$n = \frac{N_d}{2}\left(1 + \sqrt{1 + \frac{4n_i^2}{N_d^2}}\right) . \tag{1.295}$$

1. Consider the region of ultrahigh temperatures. It meets the following condition:

$$\frac{4n_i^2}{N_d^2} \gg 1 \,. \tag{1.296}$$

From (1.293), (1.295), and (1.296) it follows that $n \approx n_i$, $p \approx n_i$ and, therefore, the chemical potential can be expressed by the formula:

$$\mu = \mu_i = \frac{\varepsilon_c + \varepsilon_v}{2} + k_B T \ln \frac{N_v}{N_c} = \varepsilon_c - \frac{\varepsilon_g}{2} + k_B T \ln \frac{N_v}{N_c} \,. \tag{1.297}$$

In other words, at ultrahigh temperatures, the concentration of electrons and holes, as well as the chemical potential are the same as in an intrinsic semiconductor. This is easy to understand. At very high temperatures, the thermal energy is sufficient to excite a large number of electrons from the valence to the conduction band. A small number of electrons of the impurity play no significant role.

2. Lower temperatures are characterized by the opposite inequality:

$$\frac{4n_i^2}{N_d^2} \ll 1 \,, \tag{1.298}$$

(1.295) and (1.298) imply:

$$n \approx N_d \,. \tag{1.299}$$

At lower temperatures, an impurity is the source of electrons. Almost all of the electrons in the impurity level are donated to the conduction band. The minority carrier concentration (of holes) and the chemical potential are of the form:

$$p \approx \frac{n_i^2}{N_d} = \frac{N_c N_v}{N_d} \exp\left(-\frac{\varepsilon_g}{k_B T}\right) . \tag{1.300}$$

$$\mu = \varepsilon_c + k_B T \ln \frac{N_d}{N_c} \,. \tag{1.301}$$

At very high temperatures, the chemical potential of (1.297) is close to a mean value of the forbidden gap. With decreasing temperature, the chemical potential is shifted from the middle of the gap to the bottom of the conduction band (1.301).

II. Region of Relatively Low Temperatures

In this case, not all impurity atoms are ionized, but their contribution is large so that $n \gg p$. The impurities are a main supplier of electrons to the conduction band. Thermal energy is not enough to promote the electrons from the valence to the conduction band, so the valence band has little holes. In the main approximation we can assume that:

$$p \approx 0 \,. \tag{1.302}$$

Then the equation (1.288) again simplifies and takes the form:

$$n \approx N_\mathrm{d} \left(1 - \widetilde{f}(\varepsilon_\mathrm{d})\right) \,. \tag{1.303}$$

Or we can rewrite it in a more detailed form as:

$$N_\mathrm{c} \exp\left(\frac{\mu - \varepsilon_\mathrm{c}}{k_\mathrm{B}T}\right) = N_\mathrm{d} \frac{1}{1 + 2\exp\left[\left(\mu - \varepsilon_\mathrm{d}\right)/k_\mathrm{B}T\right]} \,. \tag{1.304}$$

The relation (1.304) reduces to a quadratic equation with respect to the unknown variable $\exp(\mu/k_\mathrm{B}T)$. Since the exponent cannot be negative, only one root of two remains:

$$\exp\left(\frac{\mu}{k_\mathrm{B}T}\right) = \frac{1}{4}\exp\left(\frac{\varepsilon_\mathrm{d}}{k_\mathrm{B}T}\right)\left(-1 + \sqrt{1 + \frac{8N_\mathrm{d}}{N_\mathrm{c}}\exp\left(\frac{\varepsilon_\mathrm{c} - \varepsilon_\mathrm{d}}{k_\mathrm{B}T}\right)}\right) \,. \tag{1.305}$$

To simplify the analysis, we denote:

$$\frac{8N_\mathrm{d}}{N_\mathrm{c}}\exp\left(\frac{\varepsilon_\mathrm{c} - \varepsilon_\mathrm{d}}{k_\mathrm{B}T}\right) = a \,. \tag{1.306}$$

1. Consider the upper temperature limit in the low temperature region. The condition $a \ll 1$ helps to determine the limit. In expression (1.305), we expand the square root in a Taylor series and restrict ourselves to the two first terms. As a result, we have:

$$\exp\left(\frac{\mu}{k_\mathrm{B}T}\right) \approx \frac{N_\mathrm{d}}{N_\mathrm{c}}\exp\left(\frac{\varepsilon_\mathrm{c}}{k_\mathrm{B}T}\right) \,, \tag{1.307}$$

$$\mu = \varepsilon_\mathrm{c} + k_\mathrm{B}T \ln\frac{N_\mathrm{d}}{N_\mathrm{c}} \,. \tag{1.308}$$

Substituting the value of μ into expression (1.289) for the electron density, we obtain:

$$n = N_\mathrm{c}\exp\left(\frac{\mu - \varepsilon_\mathrm{c}}{k_\mathrm{B}T}\right) \approx N_\mathrm{d} \,. \tag{1.309}$$

The hole concentration can be found by using the mass action law:

$$p \approx \frac{n_\mathrm{i}^2}{N_\mathrm{d}} = \frac{N_\mathrm{c}N_\mathrm{v}}{N_\mathrm{d}}\exp\left(-\frac{\varepsilon}{k_\mathrm{B}T}\right) \ll 1 \,. \tag{1.310}$$

As expected, it was small. The calculation suggests that $p \approx 0$. Thus, even at moderate temperatures (about room temperature) the electron density coincides with the impurity concentration. The solutions obtained have been merged at the border of the high and relatively low temperature regions.

As an example, consider a silicon sample doped with a donor impurity and taken at room temperature $T = 300\,\mathrm{K}$. The typical values such as $N_\mathrm{d} \sim 10^{16}\mathrm{cm}^{-3}$, $N_\mathrm{c} \sim 2.2 \cdot 10^{18}\mathrm{cm}^{-3}$ provide $k_\mathrm{B}T \ln(N_\mathrm{d}/N_\mathrm{c}) \approx -0.14\,\mathrm{eV}$, therefore:

$$\mu \approx (\varepsilon_\mathrm{c} - 0.14)\,\mathrm{eV}\,, \tag{1.311}$$

i.e., the chemical potential lies below the conduction band bottom by a value of 0.14 eV.

Since, at temperatures well above room temperature, the silicon has $-\varepsilon_\mathrm{g}/2 \approx -0.55\,\mathrm{eV}$, from (1.297) it follows that:

$$\mu \approx (\varepsilon_\mathrm{c} - 0.55)\,\mathrm{eV}\,. \tag{1.312}$$

Thus, doping and temperature can cause a significant shift of the chemical potential.

2. Let us get a close up look at the ultralow temperature region determined by the condition $a \gg 1$. In this region we know that $-1 + \sqrt{1+a} \approx \sqrt{a}$, therefore:

$$\exp\left(\frac{\mu}{k_\mathrm{B}T}\right) = \sqrt{\frac{N_\mathrm{d}}{2N_\mathrm{c}}} \exp\left(\frac{\varepsilon_\mathrm{c} + \varepsilon_\mathrm{d}}{2k_\mathrm{B}T}\right)\,, \tag{1.313}$$

$$\mu = \frac{\varepsilon_\mathrm{c} + \varepsilon_\mathrm{d}}{2} + \frac{k_\mathrm{B}T}{2} \ln \frac{N_\mathrm{d}}{2N_\mathrm{c}}\,. \tag{1.314}$$

At $T \approx 0\,\mathrm{K}$, the chemical potential is shifted to the middle between the conduction band bottom and the level ε_d. At a temperature above absolute zero, there appears a local maximum, which is described by the second term in formula (1.314) for the chemical potential. The fact is that the second term is not a linear function of temperature, although $N_\mathrm{d} = \mathrm{const}$, $N_\mathrm{c} = N_\mathrm{c}(T)$. The local maximum of the chemical potential of (1.314) lies within the range of $((\varepsilon_\mathrm{c} + \varepsilon_\mathrm{d})/2, \varepsilon_\mathrm{c})$.

The concentrations of electrons and holes at ultralow temperatures are given by:

$$n = \sqrt{\frac{N_\mathrm{c}N_\mathrm{d}}{2}} \exp\left(-\frac{\varepsilon_\mathrm{c} - \varepsilon_\mathrm{d}}{2k_\mathrm{B}T}\right)\,. \tag{1.315}$$

$$p = n_\mathrm{i}^2/n\,. \tag{1.316}$$

These asymptotic formulas derived allow one to plot temperature dependencies of the chemical potential and the electron density for semiconductors with donor impurity (Figures 1.71 and 1.72)

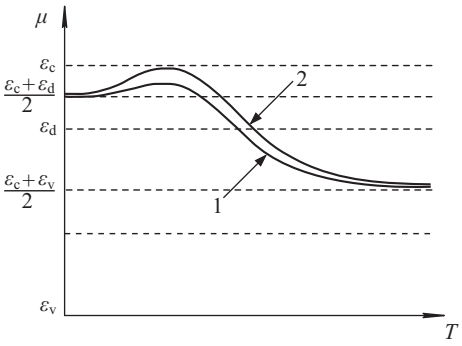


Fig. 1.71: The temperature dependence of the chemical potential for semiconductors with donor impurity. Curve 1 corresponds to the donor impurity concentration N_{d_1}, curve 2 corresponds to the impurity concentration N_{d_2}, $N_{d_2} > N_{d_1}$.

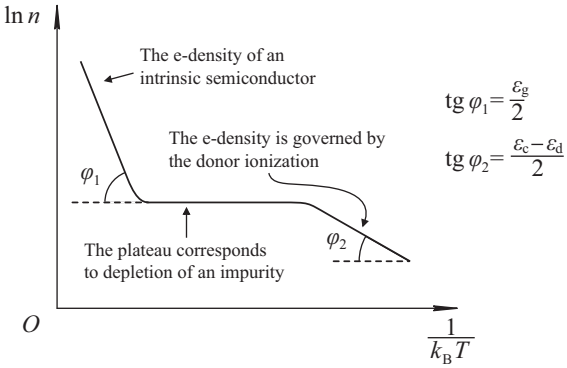


Fig. 1.72: The temperature dependence of the majority charge carrier concentration in an *n*-type semiconductor.

1.22 The Electrical Conductivity of Semiconductors

Since the quantity of electrons and holes in semiconductors is small, their contribution to the heat and thermal conductivity is also small. To a large extent, these properties are determined by the crystal lattice. Let us dwell on the electrical conductivity of semiconductors.

The electrical resistance of semiconductors, as well as metals, is due to charge carrier scattering, either by defects and lattice vibrations or by impurities. At low temperatures, it is impurities that cause the charge carrier scattering. With increasing temperature, an increasingly important role is played by lattice vibration scattering.

Let us estimate the conductivity of semiconductors in the framework of the semiclassical model. In doing so, we take into account the following claims.

1. The distributions of holes and electrons obey the Maxwell–Boltzmann law.
2. In semiconductors, the energies of electrons and holes are approximated by quadratic forms to be characteristic of free particles. In this case, it is possible to introduce the concept of mass.

For these reasons, electrons and holes in semiconductors conform to the laws of classical physics. In particular, the quasiclassical equations of motion of electrons and holes in semiconductors are reduced to Newton's equations. Next, we describe phenomenologically collisions between the charge carriers and various scatterers, as Drude did.

By taking electron relaxation into account, Newton's equation for electrons in the conduction band appears as follows:

$$m_e \dot{\vec{V}}_e = -|e|\,\vec{E} - m_e \vec{V}_e/\tau_e \,. \tag{1.317}$$

The second term on the right-hand side of (1.317) is responsible for the momentum loss associated with scattering mechanisms (τ_e is the electron relaxation time).

A similar equation can be written for holes in the valence band:

$$m_h \dot{\vec{V}}_h = |e|\,\vec{E} - m_h \vec{V}_h/\tau_h \,. \tag{1.318}$$

In a stationary state, $\dot{\vec{V}}_h = \dot{\vec{V}}_e = 0$. Therefore, from (1.317) and (1.318) we can find the drift velocities of the charge carriers:

– For electrons:

$$\vec{V}_e = -|e|\,\frac{\tau_e}{m_e}\vec{E}\,, \tag{1.319}$$

– For holes:

$$\vec{V}_h = |e|\,\frac{\tau_h}{m_h}\vec{E}\,. \tag{1.320}$$

Let us find the current density:

$$j = -|e|\,n\vec{V}_e + |e|\,p\vec{V}_h = \sigma\vec{E}\,, \tag{1.321}$$

where the specific conductivity of the medium is:

$$\sigma = e^2\left(\frac{n\tau_e}{m_e} + \frac{p\tau_h}{m_h}\right)\,. \tag{1.322}$$

The conductivity coefficient (1.322) is involved in temperature dependence through the relaxation time and the carrier concentration. At room temperatures, the relaxation time in semiconductors is temperature dependent as follows:

$$\tau \sim T^{-3/2}\,.$$

However, the temperature dependence of the concentration of charge carriers such as electrons and holes gives the greatest contribution.

Intrinsic Semiconductors

For intrinsic semiconductors $n_i = p_i$, so the electrical conductivity is of the form:

$$\sigma = n_i(T)\,(\tau_e/m_e + \tau_h/m_h)\,. \tag{1.323}$$

The multiplier $n_i(T)$ expresses the basic dependence on temperature. This dependence has an exponential nature:

$$n_i(T) \sim \exp\left(-\varepsilon_g/2k_BT\right) ,$$

so

$$\ln\sigma \sim -\frac{\varepsilon_g}{2k_BT} + \text{const}^* . \tag{1.324}$$

Here, the symbol const* stands for a value that weakly depends on the temperature. For germanium, the multiplier $\exp(-\varepsilon_g/k_BT)$ changes 10^{30} times while varying the temperature from 100 K to 700 K.

Impurity Semiconductors

For donor semiconductors $n \gg p$, therefore, the conductivity coefficient has the form:

$$\sigma \approx e^2\frac{n\tau_e}{m_e} \quad \Rightarrow \quad \ln\sigma \sim \ln n + \text{const}^* . \tag{1.325}$$

We have already discussed the dependence of the function $\ln n$ on temperature.

Conductivity provides important information on semiconductors. Knowing the electrical conductivity values, we can estimate the width of the band gap ε_g, the donor level energy ε_d, and even estimate the donor atom concentration N_d (Figure 1.72).

1.23 Rectifying Action of a *p*-*n* Junction and a Simplified Calculation of the Current Voltage Characteristics of a Diode

To start with, we model the contact of two semiconductors with different types of conduction: *p*-type and *n*-type (Figure 1.73). For this purpose, we inject a donor impurity into the right side of the crystal during its growth, and into the left side – an acceptor impurity. Such a production technology furnishes the width of the *p*-*n* junction of the order of 10^{-4} cm.

Now, the temperatures are assumed to be so high that the left side of the crystal has a lot of free holes, and their concentration value coincides with the acceptor impurity concentration ($p \approx N_a$). At the same time, many electrons reside in the right side of the crystal; their concentration is equal to the donor impurity concentration

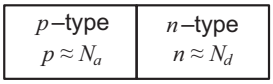


Fig. 1.73: Contact between two semiconductors of *p*-type and *n*-type. Concentrations of the carriers coincide with the concentration of impurities.

($n \approx N_d$). This is a range of temperatures close to room temperature. Besides the majority charge carriers, each region of the p-n junction has the minority charge carriers (of the opposite sign) generated by hopping the electrons from the valence to the conduction band. Such transitions produce an electron-hole pair. To be excited, each pair requires a lot of thermal energy about the gap width of ($\varepsilon_c - \varepsilon_c$). So, the quantity of the minority charge carriers are small. However, they are important for further analysis.

First, consider semiconductors of the p- and n-type separately. Then we bring them into contact and view what changes.

A p-type Semiconductor

The valence band donates electrons to an acceptor level. As a result, the valence band forms many majority charge carriers such as holes (Figure 1.74 (a)). Simultaneously, the conduction band has only a small number of electrons skipped there from the valence band. They are the minority carriers.

Although Figures 1.74 (a) and 1.74 (b) illustrate many identical charge carriers at the same energy level, this does not contradict the Pauli principle. The charge carriers differ in coordinates x, which are the quantum numbers for electrons and holes. The electrons (or holes) have the same energy but different coordinates. Consequently, they tell one from the other with different quantum numbers.

An n-type Semiconductor

There is a donor level from which electrons "jump" into the conduction band, where they become the majority charge carriers (Figure 1.74 (b)). The transitions from the valence band to the conduction band are also possible to occur. Such transitions provide a small number of minority charge carriers (holes) in the valence band for the n-type semiconductor.

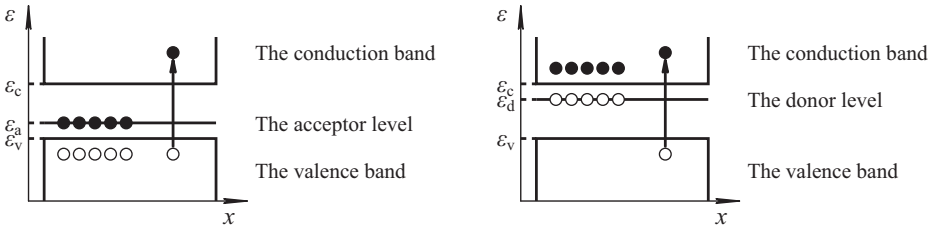


Fig. 1.74: Energy diagrams for isolated semiconductors: The symbols "○" depicts holes, "●" depicts electrons.

Now let us look what happens when *p*- and *n*-type semiconductors are brought together into contact (Figure 1.73). The left part has a lot of holes and few electrons and the right part has vice versa. All in all, the half of the *p*-type semiconductor is electrically neutral because the charge of holes is compensated by the charge of negative ions, which come off from the holes. The same takes place inside the *n*-type semiconductor: the negative electron charge is compensated by positively charged ions, disposed in the lattice nodes. The ions cannot move freely, and the electrons and holes in turn are mobile. Then the holes begin to diffuse across the junction from left to right and the electrons – from right to left (Figure 1.73). Naturally, this immediately violates the electroneutrality. When the holes leave the *p*-type region, the negative ions become "bared." Similarly, when the electrons diffuse from the *n*- to the *p*-region, the immobile positive impurity ions remain inside the *n*-region. Consequently, in the narrow near boundary area, an electric field emerges. Its direction prevents the diffusion process from spreading further (Figure 1.75). The diffusion ceases due to the electric field of the electric double layer formed by the charges of the impurity ions. The electric field of the double layer establishes a dynamic equilibrium: as many electrons (holes) go by diffusion as many return. At the thermal equilibrium, the total electric current through the *p*-*n* junction is zero.

Beyond the junction, the electroneutrality of the sample is not broken and there is no electric field. In the area of contact of the semiconductors, we can observe a jump of the electrostatic potential $\Delta\varphi$ (Figure 1.76).

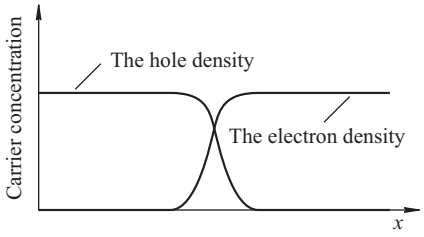


Fig. 1.75: Distribution of electrons and holes in semiconductors after being brought into contact. The intersection point of the curves corresponds to the place of the conductor contact.

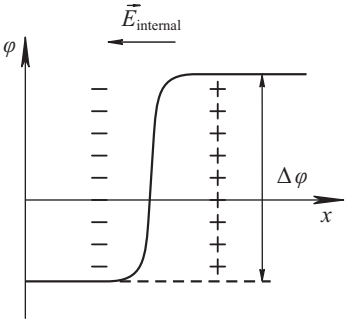


Fig. 1.76: The dependence of the electrical potential on the coordinates. The symbols "–" and "+" conventionally depict ion charges. The internal electric field is directed from right to left. It exists only in the narrow transition layer.

Important Notice

In a state of thermodynamic equilibrium, when two conductors are in contact, the chemical potential must be constant at all points of a crystal (thermodynamics requires this).

The chemical potential is aligned due to the change in the energy of electrons and holes under the internal electric field $\vec{E}_{\text{internal}}$. Let us dwell upon this statement in detail. Please compare Figures 1.77 and 1.78.

We will measure the energy of electrons from the energy ε_c, i.e., from the conduction band bottom. To find the energy of an electron, we should subtract ε_c from the current energy ε and take into account the interaction energy between the electron and the internal electrostatic field $-|e|\varphi(x)$. The potential φ, and hence the electrostatic energy, are defined with accuracy up to a constant. The electrostatic energy of the electrons in the n-region far away from the p-n junction (as $x \to +\infty$) is assumed to be equal to zero. Then the energy of the electrons at finite x has the form:

$$\varepsilon_{\text{electron}} = \varepsilon - \varepsilon_c - |e|\left[\varphi(x) - \varphi(+\infty)\right] . \tag{1.326}$$

The above formula allows calculating the total energy of the electrons in the internal electrostatic field. The bending of the conduction band is caused by the fact that the potential energy of the electrons in the p-region is greater than in the n-region.

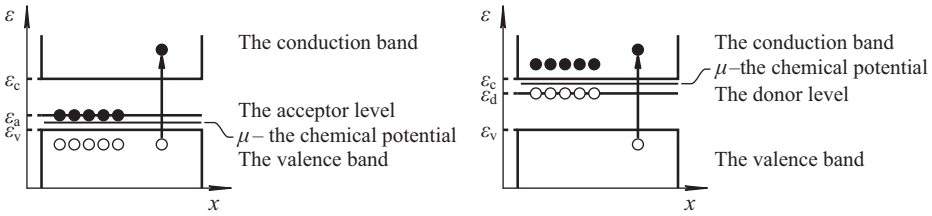


Fig. 1.77: The position of the chemical potential in the energy diagram for isolated p-type and n-type semiconductors.

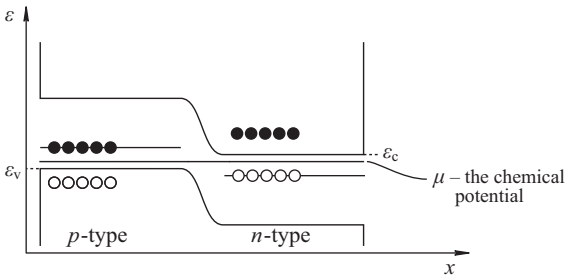


Fig. 1.78: The energy diagram and the position of the chemical potential for semiconductors after being brought into contact.

The hole energy is measured from the level ε_v in the opposite direction of the axis ε. Given that the holes have a positive charge, we get:

$$\varepsilon_{hole} = \varepsilon_v - \varepsilon + |e|\left[\varphi(x) - \varphi(-\infty)\right] . \tag{1.327}$$

We believe that the electrostatic energy of a hole in the *p*-region is zero as $x \to -\infty$. Note that we select different levels of the energy to measure. There is no contradiction in this because we describe the conduction and valence bands in different ways.

Due to the additional allowances $\pm|e|\varphi(x)$, the energy levels of the electrons and holes are bent until the chemical potentials of the *p*- and *n*-regions are equalized (Figure 1.78 and formulas (1.326), (1.327)).

Note that Figure 1.78 shows that the values of ε_c and ε_v as $x \to +\infty$ and $x \to -\infty$ remain the same as the *p*- and *n*-type semiconductors had before their contact.

Let us analyze the state of dynamic equilibrium in the *p*-*n* junction [1, 15]. We also need to focus our attention on holes. For electrons, the arguments are similar.

1. A weak hole current from the *n*- to the *p*-region exists even at thermal equilibrium without an external field. This current is called the *generation current* I_h^{gen} of holes. It is due to minority carriers.

Figure 1.79 shows that thermal fluctuations produce electron-hole pairs, either in region 1 or in region 2. These pairs are formed by transferring electrons from the valence band to the conduction band.

If the region 2 produces the electron-hole pairs, to describe the generation current, we should also take into account the diffusion of holes from the region 2 into 1. For simplicity, the electron–hole pairs are assumed to appear in the region 1 only (inside the double layer). Then, the electrostatic field immediately throws holes from the *n*-type semiconductor into the *p*-type semiconductor. Electrons travel to the region 2. Here, we are talking about only minority carriers. It is the minority carriers that are responsible for the generation current of holes. It is important to emphasize that the generation current does not depend on the external potential V, since the latter has no appreciable effect on the internal electric field around the contact region. Note also that the generation current of holes in the coordinate system at hand is negative:

$$I_h^{gen}(V) \approx I_h^{gen}(0) < 0 .$$

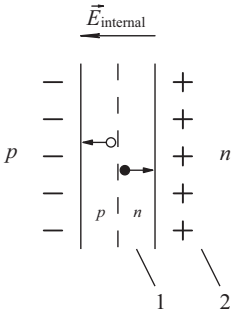


Fig. 1.79: . 1 and 2 regions of an *n*-type semiconductor are near the interface of semiconductors. The dashed line depicts the interface between the *p*- and *n*-semiconductors; the symbols "–" and "+" denote negative and positive ions, respectively.

2. Apart from the generation current, there is a *current of recombination*. It is associated with a current of holes from the p-type into the n-region (as long as we follow only the holes). The holes move from the p-region against the internal field. Therefore, they can get into the n-region only if their kinetic energy is greater than the potential barrier height. The name of the current comes from the hole-electron recombination process in the n-region.

At thermal equilibrium and under the external potential $V = 0$, the total current through the junction is zero:

$$I_{\mathrm{h}}^{\mathrm{gen}}(0) + I_{\mathrm{h}}^{\mathrm{rec}}(0) = 0 \,.$$

The generation current is negative and the recombination current is positive. Hence, we obtain:

$$I_{\mathrm{h}}^{\mathrm{rec}}(0) = -I_{\mathrm{h}}^{\mathrm{gen}}(0) = \left| I_{\mathrm{h}}^{\mathrm{gen}}(0) \right| \,.$$

Against the generation current, the recombination current is sensitive to the external potential difference V. A change in the recombination current can be determined by the Boltzmann factor:

$$I_{\mathrm{h}}^{\mathrm{rec}}(V) = I_{\mathrm{h}}^{\mathrm{rec}}(0) \exp\left(-\frac{\Delta E}{k_{\mathrm{B}} T}\right) \,. \tag{1.328}$$

Here $\Delta E = -|e|V$ is an extra potential barrier for holes. It emerges due to an external electric field. Next we choose the external potential V so that the external field diminishes the potential barrier by the magnitude $\Delta E = -|e|V > 0$ when $V > 0$. Eventually, we arrive at:

$$I_{\mathrm{h}}^{\mathrm{rec}}(V) = I_{\mathrm{h}}^{\mathrm{rec}}(0) \exp\left(\frac{|e|\, V}{k_{\mathrm{B}} T}\right) \,.$$

To explain where the formula has come from, let us recall that the holes obey the laws of classical statistical physics. Therefore, the Boltzmann multiplier determines the change in the number of holes per unit volume of a crystal under an external potential field:

$$p(V) = p(0) \exp\left(\frac{|e|\, V}{k_{\mathrm{B}} T}\right) \,.$$

Hence the recombination current is equal to:

$$I_{\mathrm{h}}^{\mathrm{rec}}(V) = |e|\, p(V) \upsilon_{\mathrm{h}} = |e|\, p(0) \upsilon_{\mathrm{h}} \exp\left(\frac{|e|\, V}{k_{\mathrm{B}} T}\right) = I_{\mathrm{h}}^{\mathrm{rec}}(0) \exp\left(\frac{|e|\, V}{k_{\mathrm{B}} T}\right) \,,$$

where υ_{h} is the velocity of the holes.

Given that $I_{\mathrm{h}}^{\mathrm{rec}}(V) = |I_{\mathrm{h}}^{\mathrm{gen}}(0)|$ we have:

$$I_{\mathrm{h}}^{\mathrm{rec}}(V) = \left| I_{\mathrm{h}}^{\mathrm{gen}}(0) \right| \exp\left(|e|\, V / k_{\mathrm{B}} T\right) \,.$$

Let us find the total hole current through this junction. Because $|I_{\mathrm{h}}^{\mathrm{gen}}(0)| = -|I_{\mathrm{h}}^{\mathrm{gen}}(0)|$, we get:

$$I_{\mathrm{h}} = I_{\mathrm{h}}^{\mathrm{gen}}(V) + I_{\mathrm{h}}^{\mathrm{rec}}(V) = -\left|I_{\mathrm{h}}^{\mathrm{gen}}(0)\right| + \left|I_{\mathrm{h}}^{\mathrm{gen}}(0)\right| \exp\left[\frac{|e|\,V}{k_{\mathrm{B}}T}\right] =$$
$$= \left|I_{\mathrm{h}}^{\mathrm{gen}}(0)\right| \left\{\exp\left[\frac{|e|\,V}{k_{\mathrm{B}}T}\right] - 1\right\} . \tag{1.329}$$

In deriving the formula, we have believed that the potential difference is $V > 0$ in the event of increasing the potential of the *p*-region and diminishing it of the *n*-region.

Let us now dwell upon the electron current through the *p*-*n* junction. The electron charge is opposite to the hole charge. Therefore, the electrons in the recombination/generation current travel in directions opposite to hole motion direction. At the same time, the electron and hole generation/recombination currents coincide in direction with each other. It is important that the potential barrier of the *p*-*n* junction diminishes by $|e|V$ as $V > 0$ for both the electrons and the holes. The electron and hole energies are measured along the axis ε in opposite directions. Therefore, the electrons and holes must overcome one and the same barrier to pass from the *n*- to the *p*-region and from the *p* to the *n*-region, respectively. Thus, the electron current through the *p*-*n* junction can be calculated by a formula close to (1.329):

$$I_{\mathrm{e}} = \left|I_{\mathrm{e}}^{\mathrm{gen}}(0)\right| \left[\exp\left(\frac{|e|\,V}{k_{\mathrm{B}}T}\right) - 1\right] . \tag{1.330}$$

Suppose the *p*-*n* junction is subjected to an external potential difference V. What changes in the energy spectrum of the electrons and holes, as well as in the chemical potential, can be observed? The charge carrier density at the interface of the semiconductors is less than in the homogeneous areas. Consequently, the *p*-*n* junction has a much higher resistance and all the voltage drop occurs across it. As a result, the junction potential difference becomes equal to $(\Delta\varphi - V)$. Accordingly, when $V > 0$, the potential barrier height lowers to the value $|e|(\Delta\varphi - V)$ for the electrons and holes, and increases to the value $|e|(\Delta\varphi + |V|)$ when $V < 0$ (Figure 1.80).

Note that, in this case, the difference between the chemical potentials of the *p*- and *n*-semiconductor portions must be equal to $|e|V$. Indeed, the chemical potential is the energy, which increases the system's energy by adding an electron. The energy required to transfer an electron from the *p*- into the *n*-region is $|e|V$. Here, we have the sign of V taken into account.

Using formulas (1.329) and (1.330), we find the total electron-hole current through the *p*-*n* junction:

$$I = I_{\mathrm{h}} + I_{\mathrm{e}} = \left(\left|I_{\mathrm{h}}^{\mathrm{gen}}(0)\right| + \left|I_{\mathrm{e}}^{\mathrm{gen}}(0)\right|\right) \left[\exp\left(\frac{|e|\,V}{k_{\mathrm{B}}T}\right) - 1\right] . \tag{1.331}$$

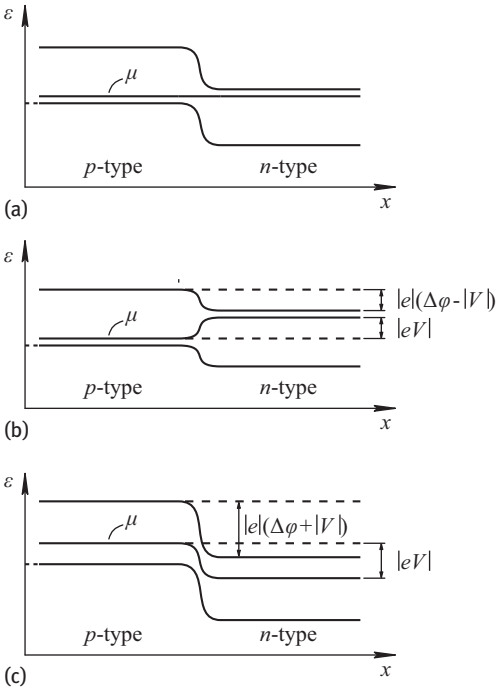


Fig. 1.80: The curvature of the edges of the energy bands and the chemical potential of the *p*-*n* junction under an external voltage: (a) The voltage lacks; (b) By applying a trigger voltage, $V > 0$; (c) By applying a locking voltage, $V < 0$.

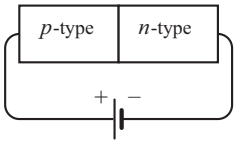


Fig. 1.81: Connecting the *p*-*n* junction to a voltage source in a forward direction $V > 0$: the battery plus connects the *p*-type region, the minus-*n*-region.

It is evident that the current flowing through the *p*-*n* junction is great when $V > 0$. In other words, the plus of a battery connects the *p*-type region; the minus of the battery connects the *n*-region (Figure 1.81). The battery reverse polarity gives $\{\exp(|e|V/k_BT) - 1\} \approx -1$; $(V < 0)$. Thus, the current through the junction is determined by the low generation currents:

$$I \approx -\left(\left|I_h^{gen}(0)\right| + \left|I_e^{gen}(0)\right|\right) .$$

Let us recall that the generation current of electrons and holes is due to minority carriers whose concentration can be easily found from the law of mass action: $pn = n_i^2$. For example, the electron concentration in an *n*-type semiconductor is known: $n \approx N_d$. Therefore, the hole concentration is $p = n_i^2/N_d \sim \exp[-\varepsilon_g/(k_BT)]$. The hole concentration in the *p*-region is $p \approx N_a$, and consequently $n = n_i^2/N_a \sim \exp[-\varepsilon_g/(k_BT)]$.

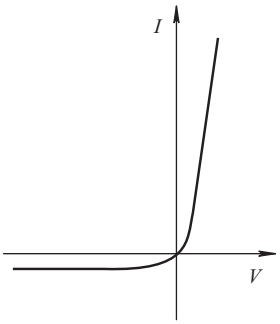


Fig. 1.82: The current voltage characteristic of the *p-n* junction.

The minority carrier concentrations up to a weakly temperature dependent multiplier are proportional to $\exp[-\varepsilon_g/(k_B T)]$. Therefore generation current can be approximated by the expression:

$$\left|I_h^{gen}(0)\right| + \left|I_e^{gen}(0)\right| = I_0 \exp\left(-\frac{\varepsilon_g}{k_B T}\right) .$$

Here, the coefficient I_0 depends so weakly on temperature that we can regard it as $I_0 = \text{const}$.

As a result, the current voltage characteristic of the *p-n* junction (of a diode) (1.331) acquires the form (Figure 1.82):

$$I = I_0 \exp\left(-\frac{\varepsilon_g}{k_B T}\right)\left[\exp\left(\frac{|e|\,V}{k_B T}\right) - 1\right] . \tag{1.332}$$

When $V > 0$, Ohm's law $I \sim V$ holds true for small values of V. For large positive values of V, Ohm's law is violated. A large direct current passes through the *p-n* junction.

When $V < 0$, the current through the junction changes only for small values of $|V|$ and then rapidly reaches saturation:

$$I = I_{\text{saturation}} = -I_0 \exp\left(-\frac{\varepsilon_g}{k_B T}\right) = \text{const} .$$

To be precise, the saturation is quite conditional. This is explained by seriously changing the curve as V diminishes. Then the breakdown voltage of the *p-n* junction diode arrives. Note that there are many Zener diodes working with critical voltage values. The current of the reverse biased *p-n* junction is thousands of times smaller than the direct current.

2 Crystal Lattice Vibrations

2.1 The Dynamics of the Crystal Lattice in the Harmonic Approximation

Earlier, we discussed ions of a crystal forming a fixed immobile lattice. In classical theory, such a model can hold only at zero temperature ($T = 0\,\mathrm{K}$). At nonzero temperature, every ion has some thermal energy and must vibrate around its equilibrium position.

In quantum theory, the static lattice model appears false even when $T = 0\,\mathrm{K}$. This is because, according to the uncertainty principle, the ions cannot be strictly localized.

The main cases of when the lattice vibrations are important to consider are listed below:

1. Ions can vibrate around their equilibrium positions. This circumstance governs all the equilibrium properties of a solid without any electron contribution. For example, the lattice dynamics manifests itself in dielectrics, where electrons are in filled bands and therefore are passive. Dielectrics conduct not only heat but also sound in the form of ion lattice vibrations. The model of the static lattice makes dielectrics sound insulators.
2. As to the mechanisms of electron-electron interactions and energy transfer in solids, the lattice dynamics are extremely important. For example, the theory of electron scattering by lattice vibrations explains what resistance and heat capacity of metals at room temperature are. In turn, superconductivity of metals at very low temperatures is due to weak attraction between electrons affected by the lattice vibrations.
3. Lattice vibrations play a role in the response of a solid to any probing radiation affecting ions (visible light, x rays and neutrons, for example). The lattice vibrations diminish the amplitude of the Bragg peaks, and create the background of scattered radiation.

Let us recall the mathematical description of a static crystal. We have previously discussed a three-dimensional crystal being made up of unit cells; each cell is a parallelepiped built on three noncoplanar vectors $\vec{a}_1, \vec{a}_2, \vec{a}_3$. To describe any crystal structure, we should build a lattice of mathematical points – the Bravais lattice. The positions of the Bravais lattice points are defined by the vectors:

$$\vec{R}^l = l_1\vec{a}_1 + l_2\vec{a}_2 + l_3\vec{a}_3\ , \tag{2.1}$$

where l_i are integers. Next, using a set of three numbers of l_i, we enumerate the elementary cells of the Bravais lattice.

https://doi.org/10.1515/9783110586183-002

For lattices with a basis, the positions of real atoms inside the unit cell are determined by the vectors $\vec{R}_\kappa$, where $\kappa = 1, 2, \ldots r$. Ultimately, the vector

$$\vec{R}_\kappa^l = \vec{R}_\kappa + \vec{R}^l \tag{2.2}$$

specifies a spatial position of an atom of the l-th unit cell indexed by κ.

The unit cell volume can be written as:

$$V_a = \vec{a}_1 \cdot [\vec{a}_2 \times \vec{a}_3] \ . \tag{2.3}$$

Any function $f(\vec{R}^l + r) = f(\vec{r})$ with the periodicity of the Bravais lattice can be expanded in a Fourier series:

$$f(\vec{r}) = \sum_{\vec{K}} f_{\vec{K}} \exp\left(\mathrm{i}\vec{K} \cdot \vec{r}\right) \ . \tag{2.4}$$

The coefficients $f_{\vec{K}}$ in the expansion (2.4) are given by:

$$f_{\vec{K}} = \frac{1}{V_a} \int\limits_{V_a} f(\vec{r}) \exp\left(-\mathrm{i}\vec{K} \cdot \vec{r}\right) \mathrm{d}^3\vec{r} \ ,$$

where the integration is over the volume V_a of the unit cell of the Bravais lattice. The vectors $\vec{K}$ appear as follows:

$$\vec{K} = n_1\vec{b}_1 + n_2\vec{b}_2 + n_3\vec{b}_3 \ , \tag{2.5}$$

where n_i are integers and $\vec{b}_1, \vec{b}_2, \vec{b}_3$ are the reciprocal lattice vectors:

$$\vec{b}_1 = \frac{2\pi}{V_a}[\vec{a}_2 \times \vec{a}_3] \ , \quad \vec{b}_2 = \frac{2\pi}{V_a}[\vec{a}_3 \times \vec{a}_1] \ , \quad \vec{b}_3 = \frac{2\pi}{V_a}[\vec{a}_1 \times \vec{a}_2] \ . \tag{2.6}$$

The basic properties of the reciprocal lattice are as follows:

1. The volume of the unit cell of the reciprocal lattice is:

$$V_b = \vec{b}_1 \cdot \left[\vec{b}_2 \times \vec{b}_3\right] = (2\pi)^3 / V_a \ . \tag{2.7}$$

2. Direct and reciprocal lattices are inverse to each other.
3. Every reciprocal lattice vector $\vec{K}$ is perpendicular to an infinite set of planes passing through nodes of the direct Bravais lattice.
4. The distance d between adjacent crystal planes with the normal $\vec{K}$ is determined by the formula:

$$d = \frac{2\pi}{\left|\vec{K}\right|} n_0 \ , \tag{2.8}$$

where n_0 is the greatest common divisor of n that specifies the vector $\vec{K}$ (2.5).

Leaving aside the artificial assumption of fixed ions located at the points with the radius vectors $\vec{R}_\kappa^l$, we use the following two weaker assumptions.

1. It is assumed that the atoms (ions) of the crystal can shift from the positions $\vec{R}_\kappa^l$ by the vectors $\vec{U}\binom{l}{\kappa}$. The position of the κ-th atom in the l-th unit cell is determined by the vector:
$$\vec{R}_\kappa^l + \vec{U}\begin{pmatrix} l \\ \kappa \end{pmatrix} . \tag{2.9}$$
The vector $\vec{R}_\kappa^l$ is time independent and characterizes the mean position of the atom. The vector $\vec{U}\binom{l}{\kappa}$ is dependent on time and determines the vibrations of a particular atom in the crystal near its equilibrium position.
2. Typical deviations of every atom of the crystal from its equilibrium position are small as compared to the distance a between neighboring atoms:
$$\left|\vec{U}\begin{pmatrix} l \\ \kappa \end{pmatrix}\right| \ll a . \tag{2.10}$$

Assumption 1 makes it possible to explain the observed crystal structure of solids, for it means that, despite the movement of ions in the solid, the Bravais lattice persists. However, we can describe, with regards to the Bravais lattice, only average rather than instantaneous positions of the ions (Figure 2.1). Note that although this assumption allows for a variety of ion motion, it does not allow their diffusion. This is because we believe that every ion oscillates with respect to a particular node. Such an assumption brings no serious restrictions except cases of the mutual exchange by the equilibrium ion positions.

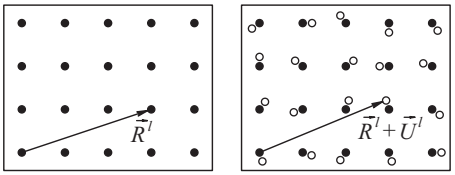


Fig. 2.1: A two-dimensional crystal with a lattice without a basis. The Bravais lattice points are given by the vector $\vec{R}^l$ and coincide with average ion positions of the crystal (a). One of the instantaneous positions of the ions (b): an ion shifts from the node with the radius vector $\vec{R}^l$ to the point with the radius vector $\vec{R}^l + \vec{U}^l$.

We are sticking with assumption 2 because it leads to a simple theory – that a harmonic approximation gives accurate quantitative results. These results are often in excellent agreement with the observed properties of a solid.

In order to write down the total kinetic energy of the lattice, suppose an atom of the unit cell labeled by κ has a mass of M_κ. Then, the kinetic energy of the lattice is:
$$T = \frac{1}{2}\sum_{l,\kappa,\alpha} M_\kappa \dot{U}_\alpha^2\begin{pmatrix} l \\ \kappa \end{pmatrix} , \tag{2.11}$$
where the index $\alpha = 1, 2, 3$ enumerates the components of the vector $\vec{U}\binom{l}{\kappa}$.

Suppose that the total potential energy Φ of a crystal depends on instantaneous atom positions. This approximation needs to be explained. The potential energy, in general, must be dependent on the coordinates of all particles comprising the crystal (that is, the ions and electrons). However, the mass of the electrons is much smaller than the mass of the ions. The electrons are so mobile that they manage to adapt to ion motion. Thus, the potential energy is supposed to be a function of only the coordinates of the ions (atoms) of the crystal:

$$\Phi = \Phi\left(\left\{\vec{R}_\kappa^l + \vec{U}\begin{pmatrix} l \\ \kappa \end{pmatrix}\right\}\right). \tag{2.12}$$

Assuming that the electrons follow the ions, we, in essence, exclude the electron-ion energy exchange. Therefore, the *approximation* in question is called *adiabatic*. Then, by means of the perturbation theory, we take the electron-lattice energy exchange into account.

With the potential energy Φ reaching a minimum as all the vectors $\vec{U}(\begin{smallmatrix} l \\ \kappa \end{smallmatrix})$ vanish, we use the standard theory of small oscillations and expand the function Φ in powers near the equilibrium position [9]:

$$\begin{aligned}\Phi\left(\left\{\vec{R}_\kappa^l + \vec{U}\begin{pmatrix} l \\ \kappa \end{pmatrix}\right\}\right) = \Phi\left(\left\{R_\kappa^l\right\}\right) + \sum_{l,\kappa,\alpha} \Phi_\alpha\begin{pmatrix} l \\ \kappa \end{pmatrix} U_\alpha\begin{pmatrix} l \\ \kappa \end{pmatrix} + \\ + \frac{1}{2}\sum_{\substack{l',\kappa',\alpha', \\ l,\kappa,\alpha}} \Phi_{\alpha\alpha'}\begin{pmatrix} l & l' \\ \kappa & \kappa' \end{pmatrix} U_\alpha\begin{pmatrix} l \\ \kappa \end{pmatrix} U_{\alpha'}\begin{pmatrix} l' \\ \kappa' \end{pmatrix}.\end{aligned} \tag{2.13}$$

Here, the constant term $\Phi(\{\vec{R}_\kappa^l\})$ is not our concern. The coefficients of the expansion terms linear in U_α must be zero (the conditions of extremum of the function Φ):

$$\Phi_\alpha\begin{pmatrix} l \\ \kappa \end{pmatrix} = \left.\frac{\partial\Phi}{\partial U_\alpha\begin{pmatrix} l \\ \kappa \end{pmatrix}}\right|_{\{\vec{U}\}=0} = 0\,. \tag{2.14}$$

Thus, the first nonvanishing expansion term is quadratic in displacements; it is the so called harmonic term. The latter relates to the matrix of constants:

$$\Phi_{\alpha\alpha'}\begin{pmatrix} l & l' \\ \kappa & \kappa' \end{pmatrix} = \left.\frac{\partial^2\Phi}{\partial U_\alpha\begin{pmatrix} l \\ \kappa \end{pmatrix}\partial U_{\alpha'}\begin{pmatrix} l' \\ \kappa' \end{pmatrix}}\right|_{\{\vec{U}\}=0}, \tag{2.15}$$

that must be defined positively because the function Φ (2.12) has a minimum at the stable equilibrium point of the crystal.

Knowing the potential Φ and kinetic T energies of the crystal, we can find the Lagrangian of lattice vibrations in the harmonic approximation:

$$L = T - \Phi \approx \frac{1}{2}\sum_{l,\kappa,\alpha} M_\kappa \dot{U}_\alpha^2\begin{pmatrix} l \\ \kappa \end{pmatrix} - \frac{1}{2}\sum_{l',\kappa',\alpha',l,\kappa,\alpha} \Phi_{\alpha\alpha'}\begin{pmatrix} l & l' \\ \kappa & \kappa' \end{pmatrix} U_\alpha\begin{pmatrix} l \\ \kappa \end{pmatrix} U_{\alpha'}\begin{pmatrix} l' \\ \kappa' \end{pmatrix}. \tag{2.16}$$

Hence, we obtain the equations of motion of the atoms of the crystal:

$$M_\kappa \ddot{U}_\alpha \begin{pmatrix} l \\ \kappa \end{pmatrix} = - \sum_{l',\kappa',\alpha'} \Phi_{\alpha\alpha'} \begin{pmatrix} l & l' \\ \kappa & \kappa' \end{pmatrix} U_{\alpha'} \begin{pmatrix} l' \\ \kappa' \end{pmatrix} . \tag{2.17}$$

Equation (2.17) can be interpreted as follows: each term of the sum in the right-hand side of (2.17) is a force acting on the κ-th atom in the l-th cell due to the displacements $U_{\alpha'}$ of the atom with the number κ' of the l' cell.

The energy of the crystal is $H = T + \Phi = \text{const}$ as the Lagrangian is explicitly time independent.

2.2 General Properties of the Force Constants

1. The definition of the matrix of the force constants (2.15) as the second derivative of the exact interaction potential yields the first property:

$$\Phi_{\alpha\alpha'} \begin{pmatrix} l & l' \\ \kappa & \kappa' \end{pmatrix} = \Phi_{\alpha'\alpha} \begin{pmatrix} l' & l \\ \kappa' & \kappa \end{pmatrix} . \tag{2.18}$$

2. The periodicity of an infinite lattice means that, if one displaces the lattice as a whole by an arbitrary translation vector $\vec{R}^m$, the displaced lattice coincides with the original one. The force constants must, therefore, not be changed:

$$\Phi_{\alpha\alpha'} \begin{pmatrix} l & l' \\ \kappa & \kappa' \end{pmatrix} = \Phi_{\alpha\alpha'} \begin{pmatrix} l+m & l'+m' \\ \kappa & \kappa' \end{pmatrix} . \tag{2.19}$$

The right-hand side of (2.19) takes into account the following:

$$\vec{R}^{l+m} = \vec{R}^l + \vec{R}^m , \quad \vec{R}^{l'+m} = \vec{R}^{l'} + \vec{R}^m . \tag{2.20}$$

The equality (2.19) holds for any vectors $\vec{R}^m$ and is possible only when satisfied by the equality:

$$\Phi_{\alpha\alpha'} \begin{pmatrix} l & l' \\ \kappa & \kappa' \end{pmatrix} = \Phi_{\alpha\alpha'} \begin{pmatrix} l-l' & \\ \kappa & \kappa' \end{pmatrix} . \tag{2.21}$$

Comment

In crystals with an inversion center (which is almost all crystals), the directions $\vec{R}^l$ and $\vec{R}^{-l} = -\vec{R}^l$ are of equal value. Hence, we conclude that the crystals with an inversion center have:

$$\Phi_{\alpha\alpha'} \begin{pmatrix} l & \\ \kappa & \kappa' \end{pmatrix} = \Phi_{\alpha\alpha'} \begin{pmatrix} -l & \\ \kappa & \kappa' \end{pmatrix} . \tag{2.22}$$

3. If a crystal was shifted as a whole by a constant vector $\vec{V}_\alpha$, then no forces acting on the atoms of the crystal will appear. The equation of motion (2.17) implies that this is possible only if the condition is met:

$$\sum_{\alpha',l',\kappa'} \Phi_{\alpha\alpha'} \begin{pmatrix} l & l' \\ \kappa & \kappa' \end{pmatrix} V_{\alpha'} = 0 \,. \tag{2.23}$$

By virtue of the arbitrariness of the vector $\vec{V}_\alpha$, the equality (2.23) is valid only if:

$$\sum_{l',\kappa'} \Phi_{\alpha\alpha'} \begin{pmatrix} l & l' \\ \kappa & \kappa' \end{pmatrix} = 0 \,. \tag{2.24}$$

According to (2.21), the matrix $\Phi_{\alpha\alpha'}$ depends only on the difference of the arguments l and l'. Therefore, in formula (2.24), we can replace the summation over l' by the summation over $\tilde{l} = l - l'$. Next, we proceed to the following form of writing:

$$\sum_{\tilde{l},\kappa'} \Phi_{\alpha\alpha'} \begin{pmatrix} & \tilde{l} & \\ \kappa & & \kappa' \end{pmatrix} = 0 \,. \tag{2.25}$$

4. If one rotates a crystal as a whole by any angle, no additional forces between the atoms of the crystal must emerge. An infinitesimal rotation by an arbitrary angle is determined by the antisymmetric matrix:

$$\omega_{\alpha\alpha'} = -\omega_{\alpha'\alpha}$$

that depends on neither l nor κ. If the atom was in the equilibrium position $R_{\alpha'}\binom{l'}{\kappa'}$ before the turn, then it would shift to the position $U_{\alpha'}\binom{l'}{\kappa'} = \sum_{\alpha''} \omega_{\alpha'\alpha''} R_{\alpha''}\binom{l'}{\kappa'}$ after the turn. In other words, the κ'-th atom of the l'-th unit cell acquires an additional displacement after the turn of the crystal as a whole:

$$U_{\alpha'} \begin{pmatrix} l' \\ \kappa' \end{pmatrix} = \sum_{\alpha''} \omega_{\alpha'\alpha''} R_{\alpha''} \begin{pmatrix} l' \\ \kappa' \end{pmatrix} \,.$$

Such displacements of the atoms must generate no forces in the equation of motion of the atoms (2.17):

$$\sum_{\alpha',l',\kappa',\alpha''} \Phi_{\alpha\alpha'} \begin{pmatrix} l - l' & \\ \kappa & \kappa' \end{pmatrix} \omega_{\alpha'\alpha''} R_{\alpha''} \begin{pmatrix} l' \\ \kappa' \end{pmatrix} = 0 \,. \tag{2.26}$$

Equation (2.26) must be true for all values of $\omega_{\alpha'\alpha''}$ that satisfy the only condition: $\omega_{\alpha'\alpha''} = -\omega_{\alpha''\alpha'}$. This is possible only when the matrix of force constants is limited by:

$$\sum_{l',\kappa'} \left[\Phi_{\alpha\alpha'} \begin{pmatrix} l - l' & \\ \kappa & \kappa' \end{pmatrix} R_{\alpha''} \begin{pmatrix} l' \\ \kappa' \end{pmatrix} - \Phi_{\alpha\alpha''} \begin{pmatrix} l - l' & \\ \kappa & \kappa' \end{pmatrix} R_{\alpha'} \begin{pmatrix} l' \\ \kappa' \end{pmatrix} \right] = 0 \,. \tag{2.27}$$

2.3 The Born–Karman Boundary Conditions – the Dynamic Matrix of a Crystal

To simplify the calculations, we suppose that a crystal has the form of a parallelepiped with sides $\vec{L}_1 = L\vec{a}_1, \vec{L}_2 = L\vec{a}_2, \vec{L}_3 = L\vec{a}_3$, where L is a large integer. Then, the crystal contains $N = L^3$ unit cells.

If the properties of the crystal inside its volume are our concern, its large sizes make the boundary conditions negligible. For the calculations, the Born–Karman boundary conditions are the most suitable. Then, according to these, each of the three main vectors $\vec{a}_i$ is satisfied by the equalities:

$$\vec{U}\begin{pmatrix} l + L_i \\ \kappa \end{pmatrix} = \vec{U}\begin{pmatrix} l \\ \kappa \end{pmatrix}, \tag{2.28}$$

where the vector $\vec{R}^l + L\vec{a}_i$ corresponds to the designation $l + L_i$.

The equation of motion of the atoms in the lattice (2.17) can be rewritten in the following form:

$$\sqrt{M_\kappa}\ddot{U}\begin{pmatrix} l \\ \kappa \end{pmatrix} = -\sum_{\alpha',l',\kappa'} (M_\kappa M_{\kappa'})^{-1/2}\,\Phi_{\alpha\alpha'}\begin{pmatrix} l & l' \\ \kappa & \kappa' \end{pmatrix}\sqrt{M_{\kappa'}}U_{\alpha'}\begin{pmatrix} l' \\ \kappa' \end{pmatrix}. \tag{2.29}$$

The solutions need to be sought in the form of a Fourier series:

$$U_\alpha\begin{pmatrix} l \\ \kappa \end{pmatrix} = \frac{1}{\sqrt{M_\kappa N}}\sum_{\vec{q}} U_\alpha\begin{pmatrix} \vec{q} \\ \kappa \end{pmatrix}\exp\left(\mathrm{i}\vec{q}\cdot\vec{R}^l\right). \tag{2.30}$$

If we add an arbitrary reciprocal lattice vector $\vec{K} = \vec{b}_1 n_1 + \vec{b}_2 n_2 + \vec{b}_3 n_3$ to the wave vector $\vec{q}$, the representation (2.30) does not change because $\vec{K}\cdot\vec{R}^l = 2\pi m$, where $m = l_1 n_1 + l_2 n_2 + l_3 n_3$ is an integer. It is always possible to ensure that the vectors $\vec{q}$ lies within a unit cell of the reciprocal lattice (the first Brillouin zone). In what follows, all the wave vectors are assumed to lie within the first Brillouin zone.

The Born–Karman (2.28) conditions impose restrictions on the possible values of the wave vector $\vec{q}$:

$$\exp\left(\mathrm{i}L\vec{q}\cdot\vec{a}_i\right) = 1, \quad (i = 1, 2, 3).$$

Hence we find the allowed values of $\vec{q}$:

$$\vec{q} = \frac{1}{L}\left(n_1\vec{b}_1 + n_2\vec{b}_2 + n_3\vec{b}_3\right), \tag{2.31}$$

where n_s are arbitrary integers ($s = 1, 2, 3$). Next, from this set, we select only physically different vectors $\vec{q}$ of the first Brillouin zone. The number of such vectors $\vec{q}$ is $N = L^3$, which is as many as the unit cells in the crystal.

By inserting the representation (2.30) into the equation of motion (2.29), we obtain:

$$\sum_{\vec{q}} \ddot{U}_\alpha \begin{pmatrix} \vec{q} \\ \kappa \end{pmatrix} \exp\left(\mathrm{i}\vec{q}\cdot\vec{R}^l\right) =$$

$$= -\sum_{\vec{q}} \sum_{\alpha',l',\kappa'} (M_\kappa M_{\kappa'})^{-1/2}\, \Phi_{\alpha\alpha'} \begin{pmatrix} l-l' \\ \kappa \quad \kappa' \end{pmatrix} \exp\left[\mathrm{i}\vec{q}\cdot\left(\vec{R}^{l'} - \vec{R}^l\right) + \mathrm{i}\vec{q}\cdot\vec{R}^l\right] U_{\alpha'} \begin{pmatrix} \vec{q} \\ \kappa' \end{pmatrix} . \tag{2.32}$$

Note that $\vec{R}^{l'} - \vec{R}^l = -\vec{R}^{l-l'}$. After replacing the summation over l' in the right-hand side of (2.32) by the summation over $\tilde{l} = l - l'$, and setting equal the coefficients of the linearly independent functions $\exp(\mathrm{i}\vec{q}\cdot\vec{R}^l)$, we have:

$$\ddot{U}_\alpha \begin{pmatrix} \vec{q} \\ \kappa \end{pmatrix} = -\sum_{\alpha',\kappa'} \varphi_{\alpha\alpha'} \begin{pmatrix} \vec{q} \\ \kappa \quad \kappa' \end{pmatrix} U_{\alpha'} \begin{pmatrix} \vec{q} \\ \kappa' \end{pmatrix} . \tag{2.33}$$

Here the *dynamic matrix of the crystal* is represented as:

$$\varphi_{\alpha\alpha'} \begin{pmatrix} \vec{q} \\ \kappa \quad \kappa' \end{pmatrix} = \sum_{\tilde{l}} (M_\kappa M_{\kappa'})^{-1/2}\, \Phi_{\alpha\alpha'} \begin{pmatrix} \tilde{l} \\ \kappa \quad \kappa' \end{pmatrix} \exp\left(-\mathrm{i}\vec{q}\cdot\vec{R}^{\tilde{l}}\right) . \tag{2.34}$$

Equation (2.34) has $\kappa = 1, 2, \ldots r$, $\alpha = 1, 2, 3$. We reduce the infinite chain of equations (2.17) to the finite system (2.33), containing $3r$ equations.

2.4 Properties of the Dynamic Matrix

To start with, let us list the basic properties of the dynamic matrix:

1.

$$\varphi_{\alpha\alpha'} \begin{pmatrix} \vec{q} \\ \kappa \quad \kappa' \end{pmatrix} = \varphi_{\alpha'\alpha} \begin{pmatrix} -\vec{q} \\ \kappa' \quad \kappa \end{pmatrix} = \varphi^*_{\alpha'\alpha} \begin{pmatrix} \vec{q} \\ \kappa \quad \kappa' \end{pmatrix} . \tag{2.35}$$

This property is a corollary of the properties (2.18) and (2.19) of the force matrix, as well as its reality condition:

$$\Phi_{\alpha\alpha'} \begin{pmatrix} l \\ \kappa \quad \kappa' \end{pmatrix} = \Phi_{\alpha'\alpha} \begin{pmatrix} -l \\ \kappa' \quad \kappa \end{pmatrix} = \Phi^*_{\alpha'\alpha} \begin{pmatrix} -l \\ \kappa' \quad \kappa \end{pmatrix} .$$

It is convenient for us to introduce a multiple index s running over $3r$ values. $s = \{\alpha, \kappa\}$, ($\kappa = 1, 2, \ldots r$; $\alpha = 1, 2, 3$). Then the dynamic matrix (2.34) can be written as:

$$\varphi_{\alpha\alpha'} \begin{pmatrix} \vec{q} \\ \kappa \quad \kappa' \end{pmatrix} \equiv \varphi_{ss'}(\vec{q}) . \tag{2.36}$$

The property (2.35), in terms of multi-index, stands for hermiticity of the dynamic matrix:

$$\varphi_{ss'}(\vec{q}) = \varphi^*_{s's}(\vec{q}) \tag{2.37}$$

and satisfies the restriction:

$$\varphi^*_{s's}(\vec{q}) = \varphi_{s's}(-\vec{q}) \ . \tag{2.38}$$

The corollary of (2.37) and (2.38) is the equality:

$$\varphi_{ss'}(\vec{q}) = \varphi_{s's}(-\vec{q}) \ . \tag{2.39}$$

Comment
In crystals with an inversion center, the following equalities are true:

$$\varphi_{ss'}(\vec{q}) = \varphi^*_{ss'}(\vec{q}) = \varphi_{s's}(\vec{q}) \ .$$

Otherwise, the dynamic matrix is real and symmetric.

2. The expanded form of $\varphi_{ss'}$ (2.34) implies that this function in reciprocal space is periodical:

$$\varphi_{ss'}(\vec{q} + \vec{K}) = \varphi_{ss'}(\vec{q}) \ . \tag{2.40}$$

2.5 The Normal Modes of Lattice Vibrations

Further analysis of the lattice dynamics equations

$$\ddot{U}_\alpha \begin{pmatrix} \vec{q} \\ \kappa \end{pmatrix} = - \sum_{\alpha', \kappa'} \varphi_{\alpha\alpha'} \begin{pmatrix} \vec{q} & \\ \kappa & \kappa' \end{pmatrix} U_\alpha \begin{pmatrix} \vec{q} \\ \kappa' \end{pmatrix} \tag{2.41}$$

leads to the more convenient and compact form:

$$\ddot{U}_s = - \sum_{s'} \varphi_{ss'} U_{s'} \ . \tag{2.42}$$

In the matrix equation (2.42), the multi-index $s = (\alpha, \kappa)$ runs over $3r$ values ($\alpha = 1, 2, 3$; $\kappa = 1, 2, \ldots r$), with the matrix $\varphi_{ss'}$ being Hermitian. That is, $\varphi^*_{ss'} = \varphi_{s's}$.

By the theorem of linear algebra, Hermitian matrix can be diagonalized by solving the eigenvalue problem:

$$\sum_{s'} \varphi_{ss'} e_{s'} = \lambda e_s \ . \tag{2.43}$$

The eigenvalues are the roots of the equation:

$$\det(\varphi - \lambda E) = 0 \ , \tag{2.44}$$

where E is a unit matrix. The left side of equation (2.44) is a $3r$-degree polynomial. Therefore, the equation (2.44) has $3r$ roots:

$$\lambda_j = \lambda_j(\vec{q}) \,,$$

where $j = 1, 2, \ldots, 3r$.

Let us analyze the properties of the eigenvalues $\lambda_j = \lambda_j(\vec{q})$.

1. Hermiticity of the matrix $\varphi_{ss'}$ guarantees that all the eigenvaluesare real: $\lambda_j^*(\vec{q}) = \lambda_j(\vec{q})$.
2. Due to the positive definiteness of the quadratic form corresponding to the potential energy of the lattice, all the numbers of $\lambda_j(q)$ are positive:

$$\lambda_j(\vec{q}) \equiv \omega_j^2(\vec{q}) > 0 \,. \tag{2.45}$$

   The restriction (2.45) is the condition for minimum potential energy of an equilibrium crystal.
3. The property (2.40) guarantees, that the functions $\omega_j(\vec{q})$ are periodic in the reciprocal space:

$$\omega_j(\vec{q} + \vec{K}) = \omega_j(\vec{q}) \,.$$

4. The eigenvalues of the matrices $\varphi(\vec{q})$ and $\varphi^{\mathrm{T}}(\vec{q})$ are equal (the symbol "T" stands for conjugation). By (2.39), we have $\varphi^{\mathrm{T}}(\vec{q}) = \varphi(-\vec{q})$. Therefore, the eigenvalues $\omega_j(\vec{q})$ must be even functions of $\vec{q}$:

$$\omega_j(\vec{q}) = \omega_j\left(-\vec{q}\right) \tag{2.46}$$

Ultimately, the evenness of the function $\omega_j(\vec{q})$ is a manifestation of invariance of the lattice dynamics equations (2.41) under the time reversal operation: $t \to -t$.

Once the eigenvalues $\lambda_j = \omega_j^2(\vec{q})$ have become known and equation (2.43) is solved, we can find $3r$ eigenvectors e_s. The vector components form a matrix $(3r \times 3r)$:

$$e_{sj} \equiv e_\alpha \left(\kappa \;\middle|\; \begin{matrix} \vec{q} \\ j \end{matrix} \right) . \tag{2.47}$$

The matrix (2.47) depends on the vector $\vec{q}$ as a parameter. Here, the index j and multiindex s, equal to $s = \{\alpha, \kappa\}$, runs over $3r$ values: $s, j = 1, 2, \ldots, 3r$.

If the matrix $\varphi_{ss'}$ is Hermitian, the eigenvectors can be chosen as orthogonal in the following sense. That is, to obey additional conditions:

$$\sum_s e_{sj}^* e_{sj'} = \delta_{jj'} \,, \quad \sum_j e_{sj}^* e_{s'j} = \delta_{ss'} \,. \tag{2.48}$$

Taking a careful look at the product of (2.48), we conclude that the matrix with the elements e_{sj} turns out to be unitary. The conditions (2.48) are the conditions of unitarity of the matrix.

We write the fields U_s of the lattice atom displacements in the form:

$$U_s = \sum_{j'} e_{sj'} Q_{j'} \ . \tag{2.49}$$

Then the dynamic equations (2.42) of the crystal acquire the form:

$$\sum_{j'} e_{sj'} \ddot{Q}_{j'} = - \sum_{s',j'} \varphi_{ss'} e_{s'j'} Q_{j'} \ . \tag{2.50}$$

Recall that the components e_{sj} are time independent as they are built by the dynamic matrix $\varphi_{ss'}$ that is also time independent. Given the identity $\sum_{s'} \varphi_{ss'} e_{s'j'} = \omega_{j'}^2(\vec{q}) e_{sj'}$, we can rewrite (2.50) in the form:

$$\sum_{j'} e_{sj'} \ddot{Q}_{j'} = - \sum_{j'} \omega_{j'}^2(\vec{q}) e_{sj'} Q_{j'} \ . \tag{2.51}$$

Taking the orthogonality conditions (2.48) into account, we multiply the equation (2.51) by e_{sj}^* and sum up the result in s. Then, we arrive at a system of equations for independent harmonic oscillators:

$$\ddot{Q}_j = -\omega_j^2(\vec{q}) Q_j \ ,$$

where $j = 1, 2, \ldots, 3r$. The problem has become trivial. Over the chosen variables, the lattice dynamics equations are easily integrated:

$$Q_j(\vec{q}, t) = c_j(\vec{q}) \exp\left(\mathrm{i}\omega_j t\right) + \tilde{c}_j(\vec{q}) \exp\left(-\mathrm{i}\omega_j t\right) \ , \tag{2.52}$$

where $c_j(\vec{q})$ and $\tilde{c}_j(\vec{q})$ are constants of integration.

After inserting (2.52) into (2.30), we obtain the general solution of the lattice dynamics equations as a superposition of traveling waves. They are called the *normal modes of the lattice*:

$$\exp\left(\pm \mathrm{i}\omega_j t + \mathrm{i}\vec{q} \cdot \vec{R}^l\right) \ .$$

The resulting formula for the displacements of lattice atoms can be written in a more convenient form for further analysis:

$$U_\alpha \begin{pmatrix} l \\ \kappa \end{pmatrix} = \frac{1}{\sqrt{M_\kappa N}} \sum_{\vec{q},j} e_\alpha \left(\kappa \,\middle|\, \begin{matrix} \vec{q} \\ j \end{matrix} \right) Q_j(\vec{q}, t) \exp\left(\mathrm{i}\vec{q} \cdot \vec{R}^l\right) \ , \tag{2.53}$$

where the whole time dependence is contained in the function $Q_j(\vec{q}, t)$.

Although the problem has been solved completely, we should clarify some properties of functions:

$$e_{sj} \equiv e_\alpha \left(\kappa \left| \begin{matrix} \vec{q} \\ j \end{matrix} \right. \right), \quad Q_j(\vec{q}, t) .$$

To do this, we write the eigenvalue equation (2.43) in more detail. Thereby, we learn more information:

$$\sum_{\alpha', \kappa'} \varphi_{\alpha\alpha'} \begin{pmatrix} \vec{q} & \\ \kappa & \kappa' \end{pmatrix} e_{\alpha'} \left(\kappa' \left| \begin{matrix} \vec{q} \\ j \end{matrix} \right. \right) = \omega_j^2(\vec{q}) e_\alpha \left(\kappa \left| \begin{matrix} \vec{q} \\ j \end{matrix} \right. \right) . \tag{2.54}$$

After the complex conjugate of (2.54), we obtain:

$$\sum_{\alpha', \kappa} \varphi_{\alpha\alpha'} \begin{pmatrix} -\vec{q} & \\ \kappa & \kappa' \end{pmatrix} e^*_{\alpha'} \left(\kappa' \left| \begin{matrix} \vec{q} \\ j \end{matrix} \right. \right) = \omega_j^2(\vec{q}) e^*_\alpha \left(\kappa \left| \begin{matrix} \vec{q} \\ j \end{matrix} \right. \right) . \tag{2.55}$$

In going over from equation (2.54) to equation (2.55), we have used the property of the dynamic matrix

$$\varphi^*_{\alpha\alpha'} \begin{pmatrix} \vec{q} & \\ \kappa & \kappa' \end{pmatrix} = \varphi_{\alpha\alpha'} \begin{pmatrix} -\vec{q} & \\ \kappa & \kappa' \end{pmatrix} ,$$

and taken into account that the eigenvalues $\omega_j^2(\vec{q})$ are real.

Recall that $\omega_j^2(\vec{q}) = \omega_j^2(-\vec{q})$ (2.46). As a result, to determine the eigenvector $e^*_\alpha(\kappa \left| \begin{smallmatrix} \vec{q} \\ j \end{smallmatrix} \right.)$, we have derived the same equation as it was for $e_\alpha(\kappa \left| \begin{smallmatrix} \vec{q} \\ j \end{smallmatrix} \right.)$ only with the replacement $\vec{q}$ by $-\vec{q}$ (this is seen from a comparison of formulas (2.54) and (2.55)). It follows that there are two possibilities:

$$e^*_\alpha \left(\kappa \left| \begin{matrix} \vec{q} \\ j \end{matrix} \right. \right) = \pm e_\alpha \left(\kappa \left| \begin{matrix} -\vec{q} \\ j \end{matrix} \right. \right) . \tag{2.56}$$

Next, we select the eigenvectors in the following way:

$$e^*_\alpha \left(\kappa \left| \begin{matrix} \vec{q} \\ j \end{matrix} \right. \right) = e_\alpha \left(\kappa \left| \begin{matrix} -\vec{q} \\ j \end{matrix} \right. \right) . \tag{2.57}$$

Now, we find a restriction to $Q_j(\vec{q}, t)$. For this, we note that the displacements of lattice atoms are real in its physical sense:

$$U^*_\alpha \begin{pmatrix} l \\ \kappa \end{pmatrix} = U_\alpha \begin{pmatrix} l \\ \kappa \end{pmatrix} .$$

Given (2.57), we come up with:

$$U_\alpha \begin{pmatrix} l \\ \kappa \end{pmatrix} = \frac{1}{\sqrt{M_\kappa N}} \sum_{\vec{q}, j} e_\alpha \left(\kappa \left| \begin{matrix} -\vec{q} \\ j' \end{matrix} \right. \right) Q^*_j(q, t) \exp\left(-\mathrm{i}\vec{q} \cdot \vec{R}^l \right) , \tag{2.58}$$

by complex conjugating of expression (2.53) for $U_\alpha\binom{l}{\kappa}$. In the equality (2.58), we replace $\vec{q}$ by $-\vec{q}$:

$$U_\alpha\begin{pmatrix} l \\ \kappa \end{pmatrix} = \frac{1}{\sqrt{M_\kappa N}} \sum_{\vec{q},j} e_\alpha\left(\kappa \,\middle|\, \begin{matrix}\vec{q} \\ j\end{matrix}\right) Q_j^*(-\vec{q}, t) \exp\left(\mathrm{i}\vec{q}\cdot\vec{R}^l\right) . \tag{2.59}$$

Expression (2.59) for $U_\alpha\binom{l}{\kappa}$ coincides with expression (2.53) for $U_\alpha\binom{l}{\kappa}$ only if:

$$Q_j^*(-\vec{q}, t) = Q_j(\vec{q}, t) . \tag{2.60}$$

A suitable choice of the integration constants in formula (2.52) easily satisfies this condition. It is enough to take

$$Q_j(\vec{q}, t) = c_j(\vec{q}) \exp(\mathrm{i}\omega_j t) + c_j^*(-\vec{q}) \exp(-\mathrm{i}\omega_j t) .$$

A further analysis requires only that the function $Q_j(\vec{q}, t)$ satisfies the restriction (2.60) and the equation of motion:

$$\ddot{Q}_j + \omega_j(\vec{q}) Q_j = 0 . \tag{2.61}$$

Substitute the displacement fields $U_\alpha\binom{l}{\kappa}$ in the form of (2.53) into the Lagrangian (2.16) and rewrite the latter in terms of Q_j, using the following four properties:

1. The orthogonality condition $\sum_s e_{sj}^* e_{sj'} = \delta_{jj'}$.
2. The property of the dynamic matrix, such as $\sum_{s'} \varphi_{ss'} e_{s'j} = \omega_j^2(\vec{q}) e_{sj}$.
3. $$\frac{1}{N}\sum_l \exp\left[\mathrm{i}(\vec{q}+\vec{q}')\vec{R}^l\right] = \begin{cases} 0 , & \text{if} \quad \vec{q} = -\vec{q}' + \vec{K} , \\ 1 , & \text{if} \quad \vec{q} = -\vec{q}' \quad \text{or} \quad \vec{q} = -\vec{q}' + \vec{K} . \end{cases}$$
   Here, $\vec{K}$ is a reciprocal lattice vector. Recall that we choose the wave vectors $\vec{q}$ and $\vec{q}'$ from the first Brillouin zone because only physically different values of $\vec{q}$ and $\vec{q}'$ are taken into account. Therefore, in calculating the Lagrangian, we exclude the case $\vec{q} = -\vec{q}' + \vec{K}$ for $\vec{K} \neq 0$.
4. The restriction associated with the reality of displacement fields is: $Q_j(-\vec{q}, t) = Q_j^*(\vec{q}, t)$.

Simple algebraic calculations lead to the following expression for the Lagrangian:

$$L = \frac{1}{2}\sum_{\vec{q},j}\left[\dot{Q}_j(\vec{q}, t)\dot{Q}_j^*(\vec{q}, t) - \omega_j^2(\vec{q}) Q_j(\vec{q}, t) Q_j^*(\vec{q}, t)\right] .$$

We introduce generalized momenta conjugate to the coordinates Q_j and Q_j^*:

$$\mathrm{P}_j(\vec{q}, t) = \frac{\partial L}{\partial \dot{Q}_j(\vec{q}, t)} = \dot{Q}_j^* , \quad \mathrm{P}_j^*(\vec{q}, t) = \frac{\partial L}{\partial \dot{Q}_j^*(\vec{q}, t)} = \dot{Q}_j . \tag{2.62}$$

Now we pass from the Lagrangian to the Hamiltonian function:

$$H = \sum_{\vec{q},j} \left[\mathrm{P}_j(\vec{q},t)\dot{Q}_j(\vec{q},t) + \mathrm{P}_j^*(\vec{q},t)\dot{Q}_j^*(\vec{q},t) \right] - L =$$
$$= \frac{1}{2} \sum_{\vec{q},j} \left\{ \mathrm{P}_j(\vec{q},t)\mathrm{P}_j^*(\vec{q},t) + \omega_j^2(\vec{q})Q_j(\vec{q},t)Q_j^*(\vec{q},t) \right\} . \tag{2.63}$$

The Hamilton equations:

$$\dot{\mathrm{P}}_j = -\frac{\partial H}{\partial Q_j} , \quad \dot{Q}_j = \frac{\partial H}{\partial \mathrm{P}_j} , \quad \dot{\mathrm{P}}_j^* = -\frac{\partial H}{\partial Q_j^*} , \quad \dot{Q}_j^* = \frac{\partial H}{\partial \mathrm{P}_j^*} , \tag{2.64}$$

reproduce the relationship between P_j and Q_j (2.62), and the lattice dynamics equations:

$$\ddot{Q}_j + \omega_j^2 Q_j = 0 .$$

Thus, the Hamiltonian approach leads to the equations of the dynamics of the crystal lattice. It should, of course, be noted that the functions of Lagrange and Hamilton are usually written through real generalized coordinates. There is no difficulty in going over to the real coordinates and momenta because the following canonical transformation helps do this:

$$\begin{aligned} Q_j(\vec{q}) &= \frac{1}{2} \left\{ X_j(\vec{q}) + X_j(-\vec{q}) + \frac{\mathrm{i}}{\omega_j(\vec{q})} \left[P_j(\vec{q}) - P_j(-\vec{q}) \right] \right\} , \\ \mathrm{P}_j(\vec{q}) &= \frac{1}{2} \left\{ P_j(\vec{q}) + P_j(-\vec{q}) - \mathrm{i}\omega_j(\vec{q}) \left[X_j(\vec{q}) - X_j(-\vec{q}) \right] \right\} . \end{aligned} \tag{2.65}$$

In the formulas (2.65) the variables X_j and P_j are real and related as follows:

$$P_j(q) = \dot{X}_j(q) . \tag{2.66}$$

Given the evenness of the function $\omega_j(\vec{q})$ and by plugging the expressions (2.65) into formula (2.63), we represent the energy of the crystal as the sum of the energies of independent real oscillators:

$$H = \frac{1}{2} \sum_{\vec{q},j} \left[P_j^2(\vec{q}) + \omega_j^2(\vec{q}) X_j^2(\vec{q}) \right] .$$

The Hamilton equations

$$\dot{X}_j = \frac{\partial H}{\partial P_j} , \quad \dot{P}_j = -\frac{\partial H}{\partial X_j}$$

reproduce the relationships (2.66) between X_j and P_j. They also presentthe lattice dynamics equations as a system of equations for real noninteracting harmonic oscillators:

$$\ddot{X}_j + \omega_j^2(\vec{q}) X_j = 0 .$$

2.6 Goldstone's Theorem – Acoustic and Optical Modes of the Normal Vibrations of a Crystal

The Lagrange function of any closed system of microparticles (electrons, ions, atoms) is invariant under transformations of a continuous translation group. That is to say, the Lagrangian form remains unchanged as the system is displaced by any constant vector. At the same time, the atoms of a crystal, being in the ground state, form a lattice whose symmetry is lower than the original symmetry of the Lagrangian of the microparticles. In other words, the physical characteristics of the equilibrium macrocrystals are invariant with respect only to a discrete translation group (there are no arbitrary translations). The values observedare described by periodic functions, reflecting the periodicity of the crystal lattice.

When the symmetry of the ground state of the system is lower than the appropriate symmetry of the Lagrangian, we can discuss this in terms of a *spontaneous symmetry breaking*. The following general theorem, proved independently by Goldstone (1961) and Bogolyubov (1963), is true.

Goldstone's Theorem

Imagine a system of a large number of particles are in the ground state. Whenever the latter's symmetry properties spontaneously break the symmetry of the micro Lagrangian of the particles under transformations of a continuous group, collective oscillations always arise, whose frequencies tend to zero as a wave vector $\vec{q}$ tends to zero as well: $\omega_j(\vec{q}) \to 0$ as $\vec{q} \to 0$.

These collective oscillations of the system are called the *Goldstone excitations, or Goldstone bosons*. The Goldstone excitations are always of such a nature, as if they tend to restore the broken symmetry of the system. The breaking of the symmetry leads to the appearance of branches (corresponding quasiparticles are called *Goldstonians*) in the Goldstone excitation spectrum. Their number is determined by the number of broken independent elements of the symmetry group of the system's Lagrangian. That is, by how many generators of the initial continuous symmetry group "disappear."

Initially, the system of microparticles forming the crystal possesses an arbitrary translational invariance. The micro Lagrangian of the particle system remains unchanged under three independent translations along the axes: Ox, Oy,and Oz. Three components of the total momentum as p_x, p_y, p_z generate these translations. As the microparticles form the crystal, the continuous translations disappear. Consequently, according to Goldstone's theorem, three crystal lattice frequencies among those $3r$ found earlier must tend to zero as the wave vector $\vec{q}$ tends to zero.

Let us verify this assertion with direct calculations. Recall that the displacement of the κ-th atom of the l-th unit cell can be represented in the form:

$$U_\alpha\begin{pmatrix} l \\ \kappa \end{pmatrix} = \frac{1}{\sqrt{M_\kappa N}} \sum_{\vec{q},j} e_\alpha\left(\kappa \,\middle|\, \begin{matrix}\vec{q}\\ j\end{matrix}\right) Q_j(\vec{q},t) \exp\left(\mathrm{i}\vec{q}\cdot\vec{R}^l\right) . \tag{2.67}$$

The polarization vectors $e_\alpha(\kappa \,|\, \begin{smallmatrix}\vec{q}\\ j\end{smallmatrix})$ are solutions to the eigenvalue problem, which can be written in detail as follows:

$$\frac{1}{M_\kappa} \sum_{l,\alpha',\kappa'} \Phi_{\alpha\alpha'}\begin{pmatrix} l & \\ \kappa & \kappa' \end{pmatrix} \frac{1}{\sqrt{M_{\kappa'}}} e_{\alpha'}\left(\kappa' \,\middle|\, \begin{matrix}\vec{q}\\ j\end{matrix}\right) \exp\left(\mathrm{i}\vec{q}\cdot\vec{R}^l\right) = \omega_j^2(\vec{q}) \frac{1}{\sqrt{M_\kappa}} e_\alpha\left(\kappa \,\middle|\, \begin{matrix}\vec{q}\\ j\end{matrix}\right) . \tag{2.68}$$

We argue that the three frequencies $\omega_j(\vec{q})$ vanish for $\vec{q} = 0$. Let us prove this statement.

Suppose that $\vec{q} = 0$ and the values $(1/\sqrt{M_\kappa}) e_\alpha(\kappa \,|\, \begin{smallmatrix}0\\ j\end{smallmatrix})$ do not depend on the parameter κ for all values of the index α:

$$\frac{1}{\sqrt{M_\kappa}} e_\alpha\left(\kappa \,\middle|\, \begin{matrix}\vec{q}=0\\ j\end{matrix}\right) = V_\alpha(j) . \tag{2.69}$$

Then equation (2.68) can be rewritten as:

$$\frac{1}{M_\kappa} \sum_{\alpha'} V_{\alpha'}(j) \sum_{l,\kappa'} \Phi_{\alpha\alpha'}\begin{pmatrix} l & \\ \kappa & \kappa' \end{pmatrix} = \omega_j^2(0) V_\alpha(j), \alpha = 1,2,3 . \tag{2.70}$$

Due to the property of the force matrix (2.25), we have:

$$\sum_{l,\kappa'} \Phi_{\alpha\alpha'}\begin{pmatrix} l & \\ \kappa & \kappa' \end{pmatrix} = 0 .$$

Then from (2.70) we get:

$$\omega_j^2(0) V_\alpha(j) = 0 . \tag{2.71}$$

The number of the independent vectors $V_\alpha(j)$ in three-dimensional space cannot be greater than three. Therefore, the equality (2.71) implies that only three frequencies from $\omega_j(0)$ can vanish. Thus, we are convinced of the validity of Goldstone's theorem.

According to (2.53), if combinations of the constants (2.69) do not depend on the parameter κ, all the atoms in unit cells of the crystal vibrate in the same phase, with the same amplitude. The parameter κ enumerates the atoms within the unit cell. Such oscillations are characteristic of the displacements of an elastic medium through which sound propagates. Therefore, three modes of lattice vibrations with frequencies vanishing for $\vec{q} = 0$ are referred to as *acoustic modes*.

Since $\omega_j^2(-\vec{q}) = \omega_j^2(\vec{q})$, it can be assumed that the frequencies of the acoustic modes for small $\vec{q}$ appear as:

$$\omega_j^2(q) \approx \sum_{\gamma,\delta=1}^{3} \alpha_{j\gamma\delta} q_\gamma q_\delta , \quad j = 1,2,3 .$$

A detailed calculation confirms this assumption. In an isotropic medium, the acoustic modes must have the form:

$$\omega_j^2 \approx c_j^2 \vec{q}^2 \,, \quad j = 1, 2, 3 \,,$$

where c_j are the velocities of sound waves. The velocities of two transverse sound waves with different polarizations are the same, $c_1 = c_2$ and differ from the longitudinal sound wave velocity, $c_3 \neq c_1, c_2$.

The rest $(3r - 3)$ of the oscillations whose frequencies do not vanish as $\vec{q} \to 0$ are called *optical lattice vibrations*. The name is sometimes misleading. Let us explain where it comes from. The fact is that, if we have two atoms in the unit cell, the atoms oscillate in optical modes of vibrations towards or away from each other in opposite phase. If the atoms bear a charge (they are ions), their vibrations correspond to alternating electric dipole moments capable of emitting electromagnetic waves in the optical range. This circumstance was the reason for all the $(3r - 3)$ lattice vibrations becoming known as optical.

If the unit cell has one atom $(r = 1)$, there are only three acoustic modes of the lattice vibrations. It is not worth forgetting that the optical modes can be lost by taking a simple lattice with identical atoms for theoretical calculations.

Counting the Number of States

In most cases, we have $3r$ functions of $\omega_j(\vec{q})$. The wave vector $\vec{q}$ runs over quasicontinuous values within the first Brillouin zone. The first Brillouin zone has as many permitted values of $\vec{q}$ as the crystal has unit cells, i.e., N. Therefore, the quantity $\omega_j(\vec{q})$ takes $3rN$ values.

The density of the wave vectors in reciprocal space can be found by dividing the number N of the permitted values of $\vec{q}$ in the first Brillouin zone by the zone volume V_b:

$$\frac{N}{V_b} = \frac{NV_a}{(2\pi)^3} = \frac{V}{(2\pi)^3} \,. \tag{2.72}$$

By transforming formula (2.72), we have taken into account that if V_a is the volume of the unit cell of the direct lattice, then $V_b = (2\pi)^3/V_a$, $NV_a = V$ is the volume of the crystal.

In reciprocal space, the first Brillouin zone can be split into small volumes $\mathrm{d}^3\vec{q}$. According to (2.72), the volume contains $(V/(2\pi)^3)\mathrm{d}^3\vec{q}$ permitted wave vectors. We have obtained a multiplier that allows replacing the summation in the vectors $\vec{q}$ by integration:

$$\sum_{\vec{q}} \ldots = \frac{V}{(2\pi)^3} \int \mathrm{d}^3\vec{q} \ldots \,. \tag{2.73}$$

2.7 Lattice Vibrations Using an Example of a Linear Chain of Atoms

Consider a set of atoms with mass M. Let them be arranged along a straight line Ox at points spaced apart by a distance of a (see. Figure 2.2). On the axis Ox, $R^n = na$, where n are integers, are sites of a one-dimensional Bravais lattice. In this case, the unit cell contains one atom, and the atomic equilibrium positions coincide with the Bravais lattice sites.

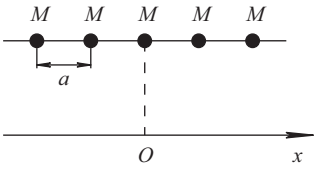


Fig. 2.2: A monatomic linear chain.

Let u_n be the displacement of the atom along the Ox axis, measured from its equilibrium position. The function $u_n(t)$ describes the vibrations of the atom near the point na.

The potential energy of atoms in the chain in the harmonic approximation has the form:

$$U = \frac{1}{2}\sum_{n,n'} \Phi(an, an')u_n u_{n'} \,, \tag{2.74}$$

where $\Phi(an', an) = (\partial^2\Phi/\partial u_n \partial u_{n'})|_{\{u_s\}=0}$ is the force matrix and $\Phi(\{na + u_n\})$ is the precise value of the potential energy of atom-atom interaction in the chain. We discuss the basic properties of the force matrix next.

The matrix $\Phi(an', an)$ is positive definite. A necessary (but not sufficient) condition for positive definiteness of the matrix is that all its diagonal elements must be positive:

$$\Phi(an, an) > 0 \,. \tag{2.75}$$

From the definition of the force matrix, it follows that:

$$\Phi(an, an') = \Phi(an', an) \,. \tag{2.76}$$

Any endless chain (we handle an infinite chain) must remain invariant under translations by the Bravais lattice vectors $R^m = ma$, hence it all comes to:

$$\Phi(an, an') = \Phi(a(n - n')) \,. \tag{2.77}$$

The consequence of the properties (2.76) and (2.77) is that the chain directions are equivalent:

$$\Phi(an) = \Phi(-an) \,. \tag{2.78}$$

That is, a one-dimensional monatomic chain always has a center of inversion.

Under arbitrary translations of the endless chain as a whole, no forces must emerge:

$$\sum_{n'} \Phi(a(n-n')) = 0 \,. \tag{2.79}$$

Let us make an additional assumption that the atom labeled by the number n interacts only with its nearest neighbors on the left and right. With this assumption, the matrix $\Phi(an', an) = \Phi(a(n-n'))$ accepts only the following index values n': $n' = n-1, n, n+1$.

For $n' = n$, we have $\Phi(0) > 0$ (see equations (2.75) to (2.77)). For further analysis, it is convenient to denote:

$$\Phi(0) = 2\alpha > 0 \,.$$

For $n' = n-1$, $n' = n+1$, from (2.78) we find that $\Phi(a) = \Phi(-a)$.

Next, we write down the condition (2.79) in the *nearest neighbor approximation*:

$$\Phi(0) + \Phi(a) + \Phi(-a) = 0 \,.$$

This gives $2\alpha + 2\Phi(a) = 0$, because $\Phi(0) = 2\alpha$, $\Phi(a) = \Phi(-a)$. Consequently, $\Phi(a) = -\alpha$.

So we have ascertained that:

$$\Phi(0) = 2\alpha > 0, \Phi(-a) = \Phi(a) = -\alpha < 0 \,. \tag{2.80}$$

The equations of motion of the atoms in the chain in the nearest neighbor approximation have the form:

$$M\ddot{u}_n = -\alpha\,[2u_n - u_{n+1} - u_{n-1}] = -\partial U/\partial u_n = F_n \,, \tag{2.81}$$

where F_n the force that acts on the n-th atom from its surroundings. When $u_n = u_{n+1} = u_{n-1}$, the force F_n vanishes, as it should.

The potential energy U in the nearest neighbor approximation is written as:

$$U = \frac{1}{2}\sum_n \alpha(u_n - u_{n-1})^2 \,. \tag{2.82}$$

We rewrite equation (2.81) in a different way:

$$M\ddot{u}_n = -\alpha(u_n - u_{n+1}) - \alpha(u_n - u_{n-1}) \,. \tag{2.83}$$

The equations of motion of material points with mass M, coupled with ideal weightless springs with the stiffness of α, have the exact same form. The equations do not include the equilibrium spring length a because the forces arise only when the springs stretch.

Now we estimate the parameter α. The main forces to stabilize the crystal structure are electrostatic forces. Typically, these forces are due to the electrons of unfilled

atomic shells, whose number is small. We believe that the effective charges of atoms in the chain are equal in magnitude to the electron charge $|e|$:

$$\Phi \sim \frac{e^2}{4\pi\varepsilon_0 |a + u_n|} \quad \text{(in the International System of Units)} ,$$
$$\Phi \sim \frac{e^2}{|a + u_n|} \quad \text{(in the CGS)}$$

Consequently,

$$\alpha = \left.\frac{\partial^2 \Phi}{\partial u_n^2}\right|_{u_n=0} \sim \frac{e^2}{4\pi\varepsilon_0 a^3} \quad \text{(in the SI)} , \qquad \alpha \sim \frac{e^2}{a^3} \quad \text{(in the CGS)} .$$

In the SI, we have:

$$a \sim 10^{-10}\,\text{m} , \quad e \sim 10^{-19}\,\text{C} , \quad 1/4\pi\varepsilon_0 \sim 10^{10}\,\text{m/F} ,$$

Thus,

$$\alpha \sim \frac{10^{-38}}{10^{-30}} 10^{10} \approx 100\,\frac{\text{kg}}{\text{sec}^2} = 100\,\frac{\text{N}}{\text{m}} .$$

The solutions of the equations of motion of atoms in the chain (2.81) need to be sought in the following manner:

$$u_n = \text{Re}\,(\tilde{u}_n) , \quad \tilde{u}_n = A \exp(\text{i}\varphi_n) , \quad \varphi_n = -\omega t + qna + \beta ,$$

where ω, A, β are real parameters. Then, we arrive at:

$$\ddot{\tilde{u}}_n = -\omega^2 A \exp(\text{i}\varphi_n) , \quad \tilde{u}_{n+1} = A \exp(\text{i}\varphi_n + \text{i}qa) , \quad \tilde{u}_{n-1} = A \exp(\text{i}\varphi_n - \text{i}qa) .$$

Plugging these expressions into the equations of motion, we find:

$$-M\omega^2 A \exp(\text{i}\varphi_n) = -\alpha\,[2 - \exp(-\text{i}qa) - \exp(\text{i}qa)]\,A \exp(\text{i}\varphi_n) =$$
$$= -2\alpha(1 - \cos qa) A \exp(\text{i}\varphi_n) .$$

After reducing by the factor $A \exp(\text{i}\varphi_n)$, we obtain the dispersion relation (Figure 2.3):

$$\omega^2 = \frac{2\alpha}{M}(1 - \cos qa) = \frac{4\alpha}{M} \sin^2 \frac{qa}{2} . \tag{2.84}$$

Hence it follows that:

$$\omega(q) = 2\sqrt{\frac{\alpha}{M}} \left|\sin \frac{qa}{2}\right| . \tag{2.85}$$

The solution describing the displacements of atoms in the chain is determined by the real part of the function $\tilde{u}_n$:

$$u_n(t) = A \cos[qna + \beta - \omega t] . \tag{2.86}$$

For $0 < |q| < \pi/a$, the solution (2.86) is a traveling plane wave.

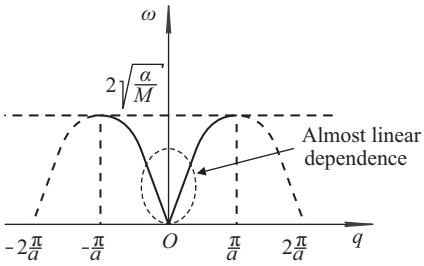


Fig. 2.3: A dispersion curve for a monoatomic linear chain in the nearest neighbor approximation.

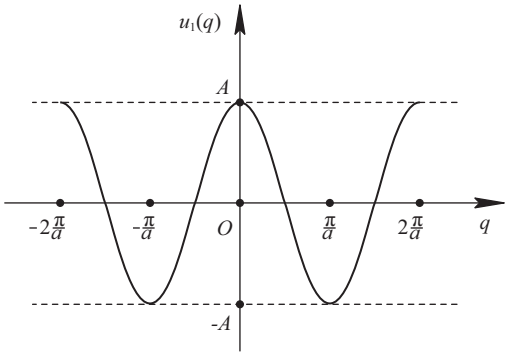


Fig. 2.4: The dependence of the displacement of the atom with the number n = 1 on the wave number q for $\beta = 0$, $t = 0$.

As can be seen, all the possible vibrations can be described by searching q through the interval $-\pi/a < q \leq \pi/a$ corresponding to the first Brillouin zone. The values of q, which lie outside the first Brillouin zone, lead to the already familiar repeated motions (see Figure 2.4).

Suppose that the number of atoms N in the chain are finite but great. If effects taking place at the ends of the chain are not our concern, we can use the Born–Karman boundary conditions $u_{n+N} = u_n$. The latter corresponds to closure of the chain to form a ring. These require the following:

$$\exp(\mathrm{i}qNa) = 1 \,.$$

Hence, in turn, it implies that the quantity q must have the form:

$$q = \frac{2\pi}{a}\frac{n}{N} \,, \tag{2.87}$$

where n is an arbitrary integer.

The first Brillouin zone contains the points q whose values of n run over the interval:

$$-\frac{N}{2} < n \leq \frac{N}{2} \quad \text{(The total number of pieces is } N) \,. \tag{2.88}$$

For large values of N, the allowed wave numbers are arranged quasicontinuously:

$$\Delta q = \frac{2\pi}{aN} \to \infty \quad \text{as} \quad N \to \infty \,.$$

When $q = 0$, the solution (2.86) describes the static displacement of all the atoms by one magnitude:

$$u_n(t) = A\cos\beta\ . \tag{2.89}$$

When $q = \pi/a$, we obtain a standing wave. The neighbor atoms in the chain vibrate in opposite phase:

$$u_n(t) = (-1)^n A\cos[\beta - \omega t]\ , \quad \omega = 2\sqrt{\alpha/M}\ . \tag{2.90}$$

When $|qa| \ll 1$, the oscillation frequency of the atoms in the chain is given by:

$$\omega \approx \sqrt{\frac{\alpha}{M}}a\,|q|\ .$$

Such a dependence of ω on q is typical for sound waves in the event of coinciding their phase $V_{\text{ph}} = \omega/|q|$ and group $V_{\text{group}} = \partial\omega/\partial|q|$ velocities:

$$V_{\text{ph}} = V_{\text{group}} \equiv s = \sqrt{\frac{\alpha}{M}}a\ . \tag{2.91}$$

We estimate the quantity s. Earlier, we already evaluated the parameter α: $\alpha \sim 100\text{N/m}$. The atoms of the middle part of the periodic table have $M \sim 10^{-25}\text{kg}$. The interatomic distances in the crystals are $a \sim 10^{-10}\text{m}$. Therefore, $s \sim \sqrt{10^{27}}10^{-10} \approx 10^3\text{m/ sec}$. We have obtained a typical velocity of sound in the crystal.

Near the Brillouin zone boundary, the dependence of the frequency ω on the wave number q is nonlinear. Therefore, the group and phase velocities are not equal: $V_{\text{group}} < V_{\text{ph}}$. In general, when $\omega = \omega(|\vec{q}|)$, the relation holds:

$$V_{\text{group}} = V_{\text{ph}} + |\vec{q}|\,\partial V_{\text{ph}}/\partial\,|\vec{q}|\ .$$

Media, for which $V_{\text{group}} < V_{\text{ph}}$, are called the media with normal dispersion.

An arbitrary motion of the chain can be represented as a superposition of N independent waves (2.86)–(2.88). Therefore, we have found a complete solution.

2.8 A Diatomic Chain: A One-Dimensional Lattice with Basis

We will now deal with a one-dimensional chain with two kinds of atoms in the unit cell, with the distance between them being equal to a. The period of the unit cell is $2a$. Along the axis Ox, the position of the Bravais lattice points are defined by the formula: $R^n = 2an$ (Figure 2.5). The positions of the atoms in the unit cell are characterized by the numbers R_κ ($\kappa = 1, 2$): $R_1 = 0$, $R_2 = a$. In this case, the coordinates of the equilibrium atoms can be written as $R_\kappa^n = 2an + R_\kappa$, where n are integers.

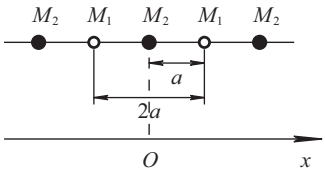


Fig. 2.5: Schematic representation of a diatomic linear chain.

Analysis of the diatomic chain is similar to the analysis of the single-atom chain. The equations of motion of the atoms in the chain have a similar form:

$$M_n \ddot{u}_n = -\alpha \left[2u_n - u_{n+1} - u_{n-1}\right] . \tag{2.92}$$

Now, the points with even and odd numbers of n identify the atoms with masses M_2 and $M_1 (M_2 > M_1)$, respectively. For simplicity, the constant α is assumed to be independent of the number of the atom.

The solutions of (2.92) should be sought in the following form:

$$\begin{cases} u_{2n+1} = A_1 \exp\{-\mathrm{i}\omega t + \mathrm{i}q(2n+1)a + \mathrm{i}\beta\} \\ u_{2n} = A_2 \exp\{-\mathrm{i}\omega t + \mathrm{i}q2na + \mathrm{i}\beta\} . \end{cases} \tag{2.93}$$

After substituting (2.93) into (2.92) and reducing by the common factor, we get:

$$\begin{cases} [M_1\omega^2 - 2\alpha]A_1 + 2\alpha A_2 \cos qa = 0 \\ 2\alpha A_1 \cos qa + [M_2\omega^2 - 2\alpha]A_2 = 0 . \end{cases} \tag{2.94}$$

The system (2.94) has a nontrivial solution only if its determinant vanishes. The vanishing of the determinant dictates the frequencies of the normal modes:

$$\omega_{\pm}^2 = \alpha\left(\frac{1}{M_1} + \frac{1}{M_2}\right) \pm \alpha\sqrt{\left(\frac{1}{M_1} + \frac{1}{M_2}\right)^2 - \frac{4}{M_1 M_2}\sin^2 qa} . \tag{2.95}$$

The two dependencies of ω on q bear the name of two branches of the dispersion relation. The lower branch $\omega_-(q)$ has the same nature as the only branch found for the monatomic one-dimensional lattice. This branch is called *acoustic*, because its dispersion relation for small wave numbers has the form: $\omega \approx sq$, which is characteristic of sound waves. The second branch $\omega_+(q)$ is called the *optical branch*. Let us first analyze what constitutes each type of the oscillation when $q \approx 0$ $(q \neq 0)$.

For the acoustic branches we have:

$$\omega_-^2 \approx 0 , \quad A_1 \approx A_2 .$$

Within the elementary cell, different atoms vibrate in phase (Figure 2.6).

Fig. 2.6: Motion of neighboring lattice atoms for the acoustic branch of the spectrum when $q \approx 0$.

For the optical branch:

$$\omega_+^2 = 2\alpha\left(\frac{1}{M_1} + \frac{1}{M_2}\right) \neq 0\,, \quad -\frac{M_1}{M_2}A_1 \approx A_2\,.$$

Within the elementary cell, different atoms oscillate in opposite phase, and their oscillation amplitudes are inversely proportional to their masses (Figure 2.7).

Fig. 2.7: Motion of neighboring lattice atoms for the optical branch of the spectrum when $q \approx 0$.

For the diatomic chain, the first Brillouin zone is twice less than in the case of a chain of identical atoms:

$$-\frac{\pi}{2a} < q \leq \frac{\pi}{2a}\,. \tag{2.96}$$

When $q = \pi/2a$ at the boundary of the Brillouin zone for the optical branch of the spectrum, we have:

$$\omega_+^2 = \frac{2\alpha}{M_1}\,, \quad A_1 \neq 0\,, \quad A_2 = 0\,.$$

Within each unit cell, only light atoms vibrate (Figure 2.8)

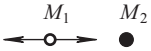


Fig. 2.8: Oscillations of the light atoms for the optical branch of the spectrum when $q = \pi/2a$.

For the acoustic branch of the spectrum:

$$\omega_- = \frac{2\alpha}{M_2}\,, \quad A_1 = 0, A_2 \neq 0\,,$$

In this case, in each unit cell, the light atoms are immobile and only the heavy atoms oscillate (Figure 2.9) but slower than the light atoms of the optical branch: $\omega_-^2 < \omega_+^2$.

Fig. 2.9: Oscillations of the heavy atoms for the optical branch of the spectrum when $q = \pi/2a$.

Since the frequencies $\omega_-(q)$ and $\omega_+(q)$ are not equal the gap emerges at the boundary of the Brillouin zone (Figure 2.10). Moreover, at the edges of the Brillouin zone we have $\partial\omega_\pm/\partial q|_{q=\pm\pi/2a} = 0$.

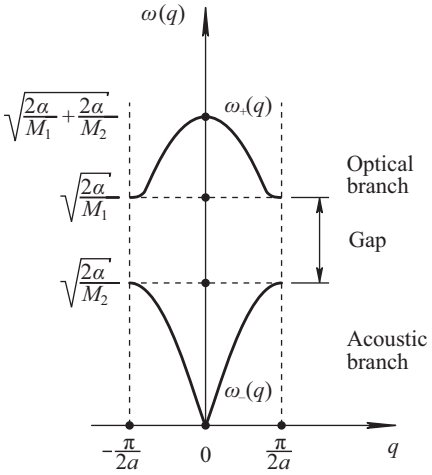


Fig. 2.10: The dispersion relation for a diatomic linear chain. The lower branch is acoustic and the upper is optical.

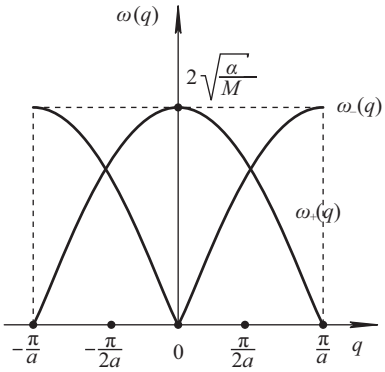


Fig. 2.11: Dependencies of $\omega_-(q)$, $\omega_+(q)$, in the limit of $M_2 \to M_1$.

In the limit $M_2 \to M_1$, the gap disappears and we obtain the dependencies depicted in Figure 2.11.

When $M_2 \to M_1$, the Brillouin zone of the chain of atoms expands twice:

$$-\frac{\pi}{a} < q \leq \frac{\pi}{a}\,. \tag{2.97}$$

This is true because, when $M_2 \to M_1$, the unit cell in the direct space shrinks in half (compare Figure 2.2 and Figure 2.5). In the new Brillouin zone (2.97), the upper branch of the spectrum becomes equivalent to the lower, as:

$$\omega_\pm^2 = \frac{2\alpha}{M_1} \pm \alpha\sqrt{\frac{4}{M_1^2}}\sqrt{1-\sin^2 qa} = \frac{2\alpha}{M_1}(1 \pm \cos qa)\,.$$

Then:

$$\omega_-^2 = \frac{2\alpha}{M_1}(1-\cos qa)\,, \quad \omega_+^2 = \frac{2\alpha}{M_1}(1+\cos qa)\,.$$

Consequently, we come up with: $\omega_-(q \pm \pi/a) = \omega_+(q)$ (see Figure 2.11). In other words, when $M_2 = M_1$, the branch $\omega_+(q)$ is no longer new because it is a product of the shift of the branch $\omega_-(q)$ to the next Brillouin zone.

The passage to the limit $M_2 \to M_1$ is especially illustrative if one depicts the optical and acoustic branches of the original spectrum in the expanded zone scheme (Figures 2.12 and 2.13).

In the repeated zone scheme, the oscillation spectrum of a diatomic lattice has the form as displayed in Figure 2.14.

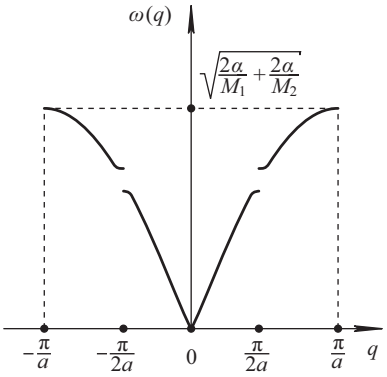


Fig. 2.12: The appearance of the gap in the vibration spectrum of a diatomic chain (the extended zone scheme).

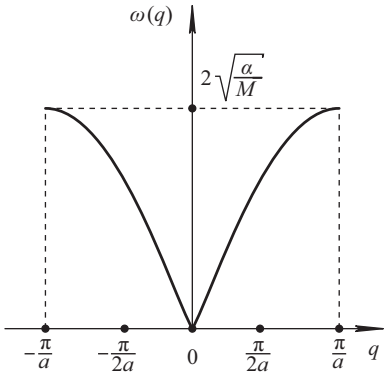


Fig. 2.13: The change in the spectrum of a diatomic chain when $M_2 \to M_1$.

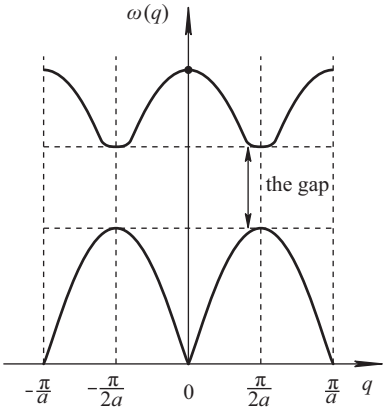


Fig. 2.14: The frequency spectrum of a diatomic chain in the repeated zone scheme.

Thus, the alternate variation of the atom masses in the chain leads to the appearance of new Brillouin zone boundaries at the points $q = \pm\pi/2a$. The spectrum of the atom vibrations shows these boundaries as forbidden frequency ranges (the gaps are opened).

2.9 Quantum Theory of the Harmonic Crystal

To begin with, let us write the Hamiltonian of a crystal:

$$H = \frac{1}{2}\sum_{\vec{q},j}\left(P_j^2(\vec{q}) + \omega_j^2(\vec{q})X_j^2(\vec{q})\right) . \tag{2.98}$$

Furthermore, we replace the dynamic variables in (2.98) by Hermitian operators that satisfy the canonical commutation relations:

$$\left[\hat{p}_i(\vec{q})\,,\ \hat{p}_j(\vec{q}')\right] = \left[\hat{x}_i(\vec{q})\,,\ \hat{x}_j(\vec{q}')\right] = 0\,, \quad \left[\hat{x}_s(\vec{q})\,,\ \hat{p}_j(\vec{q}')\right] = \mathrm{i}\hbar\delta_{sj}\delta_{\vec{q}\vec{q}'}\,. \tag{2.99}$$

With the operators $\hat{x}_s, \hat{p}_j$ being Hermitian, their eigenvalues correspond to observed values.

However, the annihilation and creation operators $\hat{a}_j, \hat{a}_i^+$ not being Hermitian is convenient against the operators $\hat{x}_s, \hat{p}_j$, as it allows us to run further analysis despite them not corresponding to the observed values:

$$\hat{x}_j(\vec{q}) = \sqrt{\frac{\hbar}{2\omega_j(\vec{q})}}\left(\hat{a}_j(\vec{q}) + \hat{a}_j^+(\vec{q})\right)\,, \quad \hat{p}_j(\vec{q}) = -\mathrm{i}\sqrt{\frac{\hbar\omega_j(\vec{q})}{2}}\left(\hat{a}_j(\vec{q}) - \hat{a}_j^+(\vec{q})\right)\,. \tag{2.100}$$

It is easy to verify that the previously entered variables P_i, Q_i (2.65) correspond to the operators:

$$\begin{aligned}
\widehat{Q}_j(\vec{q}) &= \frac{1}{2}\left\{\hat{x}_j(\vec{q}) + \hat{x}_j(-\vec{q}) + \frac{\mathrm{i}}{\omega_j(\vec{q})}\left(\hat{p}_j(\vec{q}) - \hat{p}_j(-\vec{q})\right)\right\} = \sqrt{\frac{\hbar}{2\omega_j(\vec{q})}}\left(\hat{a}_j(\vec{q}) + \hat{a}_j^+(-\vec{q})\right)\,. \\
\widehat{\mathrm{P}}_j(\vec{q}) &= \frac{1}{2}\left\{\hat{p}_j(\vec{q}) + \hat{p}_j(-\vec{q}) - \mathrm{i}\omega_j(\vec{q})\left(\hat{x}_j(\vec{q}) - \hat{x}_j(-\vec{q})\right)\right\} = -\mathrm{i}\sqrt{\frac{\hbar\omega_j(\vec{q})}{2}}\left(\hat{a}_j(\vec{q}) - \hat{a}_j^+(-\vec{q})\right)\,.
\end{aligned} \tag{2.101}$$

The commutation relations for the operators $\hat{a}_s, \hat{a}_j^+$ follow from the commutation relations (2.99):

$$\begin{aligned}
\left[\hat{a}_s(\vec{q}), \hat{a}_j(\vec{q}')\right] &= \left[\hat{a}_s^+(\vec{q}), \hat{a}_j^+(\vec{q}')\right] = 0\,, \\
\left[\hat{a}_s(\vec{q})\,,\ \hat{a}_j^+(\vec{q}')\right] &= \delta_{sj}\delta_{\vec{q}\vec{q}'}\,.
\end{aligned} \tag{2.102}$$

The expression for the Hamilton operator, in terms of creation and annihilation operators that corresponds to the classical Hamiltonian (2.98), can be written as:

$$\widehat{H} = \frac{1}{2}\sum_{\vec{q},j}\hbar\omega_j(\vec{q})\left[\hat{n}_j + \frac{\hat{I}}{2}\right]\,, \tag{2.103}$$

where the Hermitian operator $\hat{n}_j(\vec{q}) = \hat{a}_j^+(\vec{q})\hat{a}_j(\vec{q})$ has the eigenvalues $n_j(\vec{q}) = 0, 1, 2, 3, \ldots$ and the eigenvectors:

$$|\{n_j(\vec{q})\}\rangle = \prod_{j=1}^{3r}|n_j(\vec{q})\rangle\,,$$

where r is the number of atoms of different sorts in the unit cell.

The number $n_j(\vec{q})$ characterizes the degree of excitation of the normal mode with the wave vector $\vec{q}$ of the j-th branch of the lattice vibration spectrum. Such terminology is inconvenient for describing processes in a crystal, where the normal modes redistribute energy among themselves, or their energy quanta feed other subsystems of the crystal, such as electrons. Therefore, the quantum theory introduces the concept of a

phonon in the same way as the electromagnetic field theory introduces the concept of a *photon*. Rather than saying that the normal mode of the j-th branch with the wave vector $\vec{q}$ is in the $n_j(\vec{q})$-th excited state, it is convenient to believe that the crystal contains $n_j(\vec{q})$ quasiparticles – phonons with the energy $\hbar\omega_j(\vec{q})$. Compare the photon and phonon concepts.

A photon describes quanta of an electromagnetic field; in particular, light. The speed of a photon coincides with the speed of light. In contrast, a phonon describes vibrations of a crystal (sound) in a certain frequency range. Phonons move at the speed of sound.

Photons can exist in a vacuum. Phonons, in turn, are responsible for collective motion of real particles (atoms in a crystal); hence they reside only in the crystal (in the matter). Both phonons and photons are bosons, i.e., any number of phonons may be in one quantum mechanical state. Moreover, both photons and phonons have spin equal to unity. At the same time, a phonon has three projections of the angular momentum, and a photon has only two.

2.10 The Debye Interpolation Theory of the Heat Capacity of a Crystal

Debye was based on the assumption of exciting the intrinsic normal lattice vibrations (phonons) by heating the crystal. The acoustic phonons possessing the minimum excitation energy ($\omega(\vec{q}) \to 0$ for $\vec{q} \to 0$) govern the thermal properties of the crystal. Debye regarded a crystal as a vessel containing an ideal gas of acoustic phonons. He derived an approximate formula for the heat capacity of the crystal over the following criteria, which is listed below:

1. Out of all the branches of the vibrational spectrum of the crystal, Debye took only three acoustic ones into account. For the latter, he used one and the same linear dispersion law $\omega = s|\vec{q}|$ (Figure 2.15).

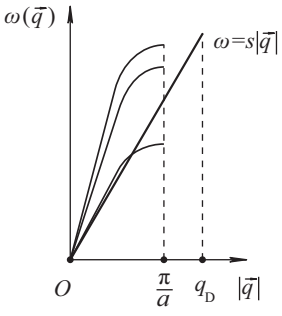


Fig. 2.15: Three acoustic branches obeying one and the same linear dispersion law.

2. Debye replaced the true Brillouin zone by a sphere of radius q_D; the volumes of the sphere and the true Brillouin zone coincide:

$$\frac{4}{3}\pi q_D^3 = V_b \,. \tag{2.104}$$

Since $V_b = (2\pi)^3/V_a$, where V_a is the volume of the unit cell of the direct lattice, an expression for q_D can be written as:

$$q_D = \sqrt[3]{6\pi^2/V_a} \,. \tag{2.105}$$

To evaluate q_D, note that $V_a \sim a^3$, where a is the interatomic distance. Hence, we find:

$$q_D \sim \frac{\pi}{a} \sim k_F \,, \tag{2.106}$$

i.e., q_D has the same order of magnitude as the Fermi wave vector for electrons in a crystal.

The maximum possible value of the wave number q_D correlates with the minimum wavelength λ_D:

$$q_D = \frac{2\pi}{\lambda_D} \sim \frac{\pi}{a} \rightarrow \lambda_D \sim 2a \,. \tag{2.107}$$

The condition (2.107) has a certain geometric meaning. The wavelength of the lattice vibrations cannot be less than $\lambda_D = 2a$. This is explained by the fact that the motion can be observed only in those places where the atoms are (Figure 2.16). All the wavelengths of the lattice vibrations must be greater than λ_D and, therefore, all the wave vectors must be less than $q_D \sim 2\pi/\lambda_D \sim \pi/a$. The frequency $\omega_D = sq_D$ is called the *Debye frequency*.

The density of the allowed wave vectors $\vec{q}$ in reciprocal space is $V/(2\pi)^3$.

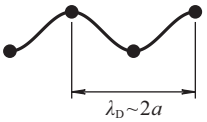


Fig. 2.16: The minimum wavelength of the atom vibrations in a crystal.

The volume $d^3\vec{q}$ contains $(V/(2\pi)^3)d^3\vec{q}$ allowed wave vectors. As we look into the three branches of the energy spectrum of the crystal, the number of the allowed quantum states in the volume element $d^3\vec{q}$ of the reciprocal space is given by:

$$3\frac{V}{(2\pi)^3}d^3\vec{q} \,. \tag{2.108}$$

The average number of thermally equilibrium phonons with the energy $\varepsilon(\vec{q}) = \hbar s|\vec{q}|$ is determined by the Bose–Einstein distribution:

$$f = \frac{1}{\exp\left(\hbar s\,|\vec{q}|\,/k_B T\right) - 1} \,. \tag{2.109}$$

This is an appropriate place to note an important detail: phonons are not particles, but quasiparticles. They can appear and disappear. At given parameters of T and V, the number of phonons $\widetilde{N}$ is not constant, but it derives from the thermodynamic conditions of minimum free energy. Therefore, the chemical potential μ is equal to zero for the phonon system:

$$\mu = \left(\frac{\partial F}{\partial \widetilde{N}}\right)_{T,V} = 0 \,. \tag{2.110}$$

That is why we put that $\mu = 0$ in formula (2.109).

The total number of phonons in the volume element $\mathrm{d}^3\vec{q}$ centered at a radius vector $\vec{q}$ at temperature T can be estimated by the formula:

$$\frac{3Vf(\vec{q})}{(2\pi)^3}\mathrm{d}^3\vec{q} \,, \tag{2.111}$$

where $\mathrm{d}^3\vec{q} = 4\pi q^2 \mathrm{d}q$; their energy is equal to:

$$\frac{3Vf(\vec{q})\hbar s\,|\vec{q}|}{(2\pi)^3}\mathrm{d}^3\vec{q} \,. \tag{2.112}$$

Given the above, we can write the following expression for the total energy of the crystal:

$$\varepsilon = \frac{3V\hbar s}{2\pi^2}\int\limits_0^{q_\mathrm{D}} \mathrm{d}q \frac{q^3}{\exp\left(\hbar s q/k_\mathrm{B}T\right) - 1} \,. \tag{2.113}$$

Note, that:

$$V_a = \frac{V}{N} = \frac{1}{n} \,. \tag{2.114}$$

where n is the number of unit cells per unit volume. Or, in other words, n is the number of mathematical Bravais lattice points per unit volume. N is the total number of unit cells in the crystal.

From (2.105) and (2.114), we get:

$$q_\mathrm{D}^3 = \frac{6\pi^2}{V_a} = 6\pi^2 n \,. \tag{2.115}$$

Now, it is convenient to determine the Debye temperature T_D:

$$k_\mathrm{B}T_\mathrm{D} = \hbar\omega_\mathrm{D} \equiv \hbar s q_\mathrm{D} \,. \tag{2.116}$$

Hence, we can make the estimate:

$$T_\mathrm{D} = \frac{\hbar s q_\mathrm{D}}{k_\mathrm{B}} \sim \frac{\hbar s \pi}{k_\mathrm{B} a} = \frac{10^{-34}\cdot 10^3}{10^{-23}\cdot 10^{-10}} \approx 10^2 \mathrm{K} \,.$$

Let us change the variables $\hbar s q/k_\mathrm{B}T = x$ in formula (2.113) and calculate the specific heat of the crystal at constant volume:

$$c_\mathrm{V} = \frac{1}{V}\left(\frac{\partial \varepsilon}{\partial T}\right)_V = 9nk_\mathrm{B}\left(\frac{T}{T_\mathrm{D}}\right)^3 \int\limits_0^{T_\mathrm{D}/T} \mathrm{d}x \frac{x^4 \exp x}{(\exp x - 1)^2} \,. \tag{2.117}$$

Now we can examine the limiting cases.

If $T \ll T_D$, then the upper limit in the integral (2.117) can be replaced by ∞:

$$\int_0^\infty dx \frac{x^4 \exp x}{(\exp x - 1)^2} = \frac{4\pi^4}{15} .$$

Consequently, we come to

$$c_V \approx \frac{12}{5}\pi^4 n k_B \left(\frac{T}{T_D}\right)^3 . \tag{2.118}$$

If $T \gg T_D$ then:

$$\int_0^{T_D/T} dx \frac{x^4 \exp x}{(\exp x - 1)^2} \approx \int_0^{T_D/T} dx x^2 = \frac{1}{3}\left(\frac{T_D}{T}\right)^3 ,$$

and we get:

$$c_V = 3 n k_B = \text{const} . \tag{2.119}$$

We have arrived at the *Dulong–Petit law* of classical physics.

It is worth emphasizing that only the quantum theory can explain the law $c_V \sim T^3$, observed experimentally in dielectrics and metals at low temperatures (Figure 2.17). Classical physics for any temperature yields the heat capacity of (2.119).

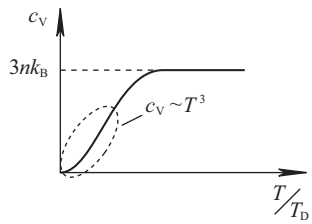


Fig. 2.17: Temperature dependence of the specific heat of the crystal.

2.11 The Role of the Anharmonic Terms in the Energy of a Crystal

There are many physical phenomena that are entirely caused by the highest terms of the expansion of the potential energy of a crystal in powers of the ionic displacements. These terms, always neglect the harmonic approximation. One of such properties is the *effect of thermal expansion*.

In the harmonic approximation, the thermal vibrations of atoms near the equilibrium positions are spherically symmetrical: each atom is as close to one of its neighbors as it is far away from it. As temperature grows, the amplitude of the vibrations

increase as well. However, the average distance between the atoms of the crystal does not change. It follows that the harmonic approximation cannot explain the thermal expansion of bodies.

When the expansion of the elastic energy of the crystal takes the highest terms into account, the vibrations of each atom are essentially amplitude dependent and are of such a nature that the atom is easier to remove from a neighbor than to approach it. Due to the asymmetry of the vibrations, the average distances between the atoms become larger as the crystal heats up. That is why bodies increase their volume upon heating.

Anharmonic interactions are an important cause of finite *thermal conductivity of dielectrics*. In other words, a perfect harmonic crystal would possess infinite conductivity.

Let us explain this statement. In the harmonic approximation, the heat transfer should be described by wave packets, each of which represents a superposition of normal modes of the lattice vibrations of one (of the j-th) branch of the spectrum with close wave vectors:

$$\left|\vec{q} - \vec{q}'\right| \leq \Delta q \ . \tag{2.120}$$

The size Δr of the wave packet in the coordinate space satisfies the constraint

$$L \gg \Delta r \sim \frac{1}{\Delta q} \gg a \ , \tag{2.121}$$

where L is the size of the crystal and a is the interatomic distance.

With the group velocity of the wave packet being $\vec{V} = \partial\omega_j(\vec{q})/\partial\vec{q} = \text{const}$, a perfect harmonic crystal has infinite thermal conductivity. Earlier, we noted a similar situation for the electrical conductivity of metals.

The finite thermal conductivity of real crystals appears for several reasons.

1. Various imperfections of a crystal lattice, impurities, and free electrons (in the case of metals) act as scattering centers for phonons and prevent heat flux. Phonons collide with the surface of the sample, which also limits the heat flux.
2. A special role is played by phonon interactions, such as processes of scattering, merger or decay of phonons. The phonon interactions can be described by the anharmonic terms in the crystal's Hamiltonian. Therefore, they are always there and they cannot be eliminated.

For a theoretical description of the phonon interactions, the expansion of the potential energy of the lattice Φ in powers of the ionic displacements $\vec{U}$ keeps necessarily the fourth-order terms together with cubic ones [2]. This is due to the following reasons:

(a) Hamiltonian that contains only third-order terms in $\vec{U}$ have no ground state: by choosing the values of $\vec{U}$ in an appropriate way, we may set the potential energy as arbitrarily large in absolute magnitude (as desired) and negative.

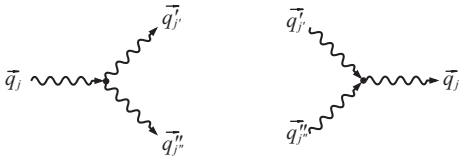


Fig. 2.18: Processes corresponding to the third-order anharmonic terms: one phonon splits into two, two phonons merge into one.

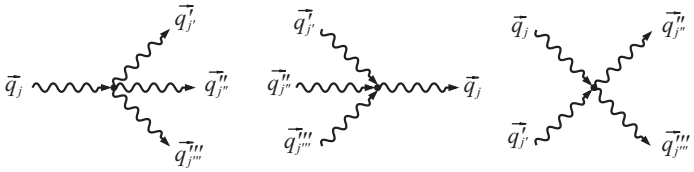


Fig. 2.19: Processes corresponding to the fourth-order anharmonic terms: one phonon decays into three, three phonons merge into one, phonon scattering.

(b) The laws of conservation of energy and momentum must be fulfilled when phonons collide. These laws impose such severe restrictions on the scattering processes associated with the third-order terms that even if $|\vec{U}|$ is small (consequently, the fourth-order terms in the potential energy of the crystal are small too), the contributions of the third- and fourth-order terms to the thermal conductivity are comparable in magnitude (Figures 2.18 and 2.19).

The anharmonic terms of higher order (5th, 6th, etc.) also lead to the phonon scattering and reactions between them. However, under the assumption of smallness of the crystal vibrations, the third- and fourth-order terms are still the most important in the expansion in powers of $\vec{U}$.

It is important to point out that the law of conservation of the total quasimomentum of the phonons holds only up to the term $\hbar\vec{K}$. For example, for the decay of a phonon (Figure 2.18), in a general case, we have:

$$\vec{q} = \vec{q}' + \vec{q}'' + \vec{K}\,,$$

where $\vec{K}$ is an arbitrary reciprocal lattice vector.

The processes of the phonon interactions with the nonzero reciprocal lattice vector $\vec{K}$ are called *Umklapp processes*. It is these, rather than the processes with $\vec{K} = \vec{0}$, that govern the phonon momentum relaxation and finite thermal conductivity of the crystal. In this case, the crystal acquires the quasimomentum $\hbar\vec{K} \neq \vec{0}$ as a whole. With decreasing temperature T, the number of phonons capable of participating in the Umklapp processes diminishes exponentially: $\sim \exp(-\theta//T)$, with θ being a positive parameter.

2.12 Electron-Phonon Interaction

To begin with, we write the potential energy of an electron in a periodic field of fixed ions as:

$$U^{\text{reg}}(\vec{r}) = \sum_{l,\kappa} \widetilde{V}_\kappa \left(\vec{r} - \vec{R}_\kappa^l\right) , \quad U^{\text{reg}}(\vec{r} + \vec{R}^l) = U^{\text{reg}}(\vec{r}) , \tag{2.122}$$

where $\widetilde{V}_\kappa(\vec{r} - \vec{R}_\kappa^l)$ is the pair interaction potential between the κ-th ion of the l-th cell and an electron situated at the point with the radius vector $\vec{r}$. The vector $\vec{R}_\kappa^l = \vec{R}^l + \vec{R}_\kappa$ specifies the position of the ion in the crystal.

When the ions are displaced outwards from their equilibrium positions, their position vectors are described by the formula:

$$\vec{R}_\kappa^l \to \vec{R}_\kappa^l + \vec{U}\begin{pmatrix} l \\ \kappa \end{pmatrix} .$$

In addition, the interaction energy between the electron and the crystal lattice takes the following form:

$$U(\vec{r}) = \sum_{l,\kappa} \widetilde{V}_\kappa \left(\vec{r} - \vec{R}_\kappa^l - \vec{U}\begin{pmatrix} l \\ \kappa \end{pmatrix}\right) . \tag{2.123}$$

The displacement fields $\vec{U}\binom{l}{\kappa}$ are assumed to be small, i.e., they meet the condition (2.10). Then, we expand the energy (2.123) into a Taylor series, restricting ourselves to only the first terms [11]:

$$U(\vec{r}) = \sum_{l,\kappa} \widetilde{V}_\kappa \left(\vec{r} - \vec{R}_\kappa^l - \vec{U}\begin{pmatrix} l \\ \kappa \end{pmatrix}\right) \approx U^{\text{reg}} - \sum_{l,\kappa,\alpha} \left[\frac{\partial}{\partial r_\alpha} \widetilde{V}_\kappa \left(\vec{r} - \vec{R}_\kappa^l\right) U_\alpha \begin{pmatrix} l \\ \kappa \end{pmatrix}\right] . \tag{2.124}$$

From here we get the interaction Hamiltonian between the electrons and the lattice vibrations:

$$\begin{aligned} H_{\text{int}} &= -\sum_{l,\kappa,\alpha} \left[\frac{\partial}{\partial r_\alpha} \widetilde{V}_\kappa \left(\vec{r} - \vec{R}_\kappa^l\right) U_\alpha \begin{pmatrix} l \\ \kappa \end{pmatrix}\right] = \\ &= \sum_{\vec{q},j} Q_j(\vec{q}) \sum_{l,\kappa,\alpha} \frac{1}{\sqrt{NM_\kappa}} \left[\frac{\partial}{\partial r_\alpha} \widetilde{V}_\kappa(\vec{r} - R_\kappa^l)\right] e_\alpha \left(\kappa \,\middle|\, \begin{matrix} \vec{q} \\ j \end{matrix}\right) \exp\left(\mathrm{i}\vec{q} \cdot \vec{R}^l\right) . \end{aligned} \tag{2.125}$$

Here, we have used the expansion (2.67).

To go from the Hamiltonian of classical physics (2.125) to the Hamiltonian operator $\widehat{H}_{\text{int}}$ of the quantum theory, we replace the dynamic variable $Q_j(\vec{q})$ in the relation (2.125) by the operator $\widehat{Q}_j(\vec{q})$. As shown previously, the operator $\widehat{Q}_j(\vec{q})$ is expressed in terms of the creation and annihilation of phonons:

$$\widehat{Q}_j(\vec{q}) = \sqrt{\frac{\hbar}{2\omega_j(\vec{q})}} \left[\widehat{a}_j(\vec{q}) + \widehat{a}_j^+(-\vec{q})\right] . \tag{2.126}$$

As a result, we have:

$$\begin{aligned} \widehat{H}_{\text{int}} = &-\sum_{\vec{q},j} \sqrt{\frac{\hbar}{2\omega_j(\vec{q})}} \left[\widehat{a}_j(\vec{q}) + \widehat{a}_j^+(-\vec{q})\right] \cdot \\ &\cdot \sum_{l,\kappa,\alpha} \frac{1}{\sqrt{NM_\kappa}} \left[\frac{\partial}{\partial r_\alpha} \widetilde{V}_\kappa(\vec{r} - R_\kappa^l)\right] e_\alpha \left(\kappa \,\middle|\, \begin{matrix} \vec{q} \\ j \end{matrix}\right) \exp\left(\mathrm{i}\vec{q} \cdot \vec{R}^l\right) . \end{aligned} \tag{2.127}$$

The perturbation theory allows calculating various observables through the matrix elements of the operator $\widehat{H}_{\text{int}}$ on the Bloch functions of electrons. Recall that, in the periodic field of ions, the electron Bloch functions have the form:

$$\Psi_{s\vec{k}} = \exp\left(\mathrm{i}\vec{k}\cdot\vec{r}\right) u_{s\vec{k}}(\vec{r}) \,, \quad u_s(\vec{r}+\vec{R}^l) = u_s(\vec{r}) \,, \tag{2.128}$$

where s is the energy band number and $\vec{k}$ is the electron wave vector.

Let the electron Bloch functions be orthonormalized in the crystal volume V:

$$\int\limits_V \mathrm{d}^3\vec{r}\Psi^*_{s'\vec{k}'}(\vec{r})\Psi_{s\vec{k}}(\vec{r}) = \delta_{ss'}\delta_{\vec{k}\vec{k}'} \,. \tag{2.129}$$

To find the matrix element $\langle s'\vec{k}'|\widehat{H}_{\text{int}}|s\vec{k}\rangle$, it is sufficient to calculate the matrix element $\langle s'\vec{k}'|\partial\widetilde{V}_\kappa(\vec{r}-R^l_\kappa)/\partial r_\alpha|s\vec{k}\rangle$. To calculate the latter, we replace the integration over the coordinates $\vec{r}$ of the electron by the integration over the variable $\vec{r}\,' = \vec{r} - \vec{R}^l$:

$$\left\langle s'\vec{k}' \left| \frac{\partial}{\partial r_\alpha}\widetilde{V}_\kappa(\vec{r}-\overbrace{\vec{R}^l_\kappa}^{\vec{R}^l+\vec{R}_\kappa}) \right| s\vec{k} \right\rangle =$$
$$\left[\int \mathrm{d}^3\vec{r}\,' \frac{\partial}{\partial r'_\alpha}\widetilde{V}_\kappa(\vec{r}-\vec{R}_\kappa)u^*_{s'\vec{k}'}(\vec{r}\,')u_{s\vec{k}}(\vec{r}\,')\exp\left\{-\mathrm{i}\vec{r}\,'\cdot\left(\vec{k}'-\vec{k}\right)\right\}\right]\exp\left\{\mathrm{i}\vec{R}^l\cdot(\vec{k}-\vec{k}')\right\} \,. \tag{2.130}$$

Using formula (2.130) and the identity:

$$\sum_l \exp\left[\mathrm{i}\vec{R}^l\cdot(\vec{q}-\vec{k}'+\vec{k})\right] = N\delta_{\vec{k}'-\vec{k},\vec{q}+\vec{K}} \,, \tag{2.131}$$

where $\vec{K}$ is an arbitrary reciprocal lattice vector and N is the number of unit cells in the crystal, we may reduce the matrix element of the interaction operator between the electrons and the lattice vibrations to the following form:

$$\left\langle s'\vec{k}'\right| \widehat{H}_{\text{int}} \left|s\vec{k}\right\rangle = \sum_{\vec{q},j}\left[\widehat{a}_j(\vec{q}) + \widehat{a}^+_j(-\vec{q})\right] g_{s'\vec{k}',s\vec{k}}\begin{pmatrix}\vec{q}\\ j\end{pmatrix} \,, \tag{2.132}$$

where

$$g_{s'\vec{k}',s\vec{k}}\begin{pmatrix}\vec{q}\\ j\end{pmatrix} = -\sqrt{\frac{\hbar N}{2\omega_j(\vec{q})}}\sum_{\kappa,\alpha}\left\langle s'\vec{k}'\left|\frac{\partial}{\partial r_\alpha}\widetilde{V}_\kappa(\vec{r}-\vec{R}_\kappa)\right|s\vec{k}\right\rangle\frac{1}{\sqrt{M_\kappa}}e_\alpha\left(\kappa\,\middle|\,\begin{matrix}\vec{q}\\ j\end{matrix}\right)\sum_{\vec{K}}\delta_{\vec{k}'-\vec{k},\vec{q}+\vec{K}}$$
$$g^*_{s'\vec{k}',s\vec{k}}\begin{pmatrix}\vec{q}\\ j\end{pmatrix} = g_{s\vec{k},s'\vec{k}'}\begin{pmatrix}-\vec{q}\\ j\end{pmatrix} \tag{2.133}$$

From formula (2.133), an important statement follows. In the crystal, under the electron-phonon interaction the quasimomentum conservation law holds true only up to the quantity $\hbar\vec{K}$ related to some reciprocal lattice vector $\vec{K}$:

$$\hbar(\vec{k}'-\vec{k}) = \hbar(\vec{q}+\vec{K}) \,. \tag{2.134}$$

Recall that the first Brillouin zone physically limits different values of the phonon wave vector $\vec{q}$. According to the relation (2.134), under the electron-phonon interactions, the processes with $\vec{K} \neq 0$, when the quasimomentum $\hbar(\vec{k}'-\vec{k})$ transfers the

phonon wave vector from the first Brillouin zone in the neighboring ones, cannot be excluded. From the standpoint of classical physics, there are certainly "normal" electron-phonon interaction processes when $\vec{K} = \vec{0}$.

For further analysis, the parameter $g_{s'\vec{k}',s\vec{k}}\left(\begin{smallmatrix}\vec{q}\\ j\end{smallmatrix}\right)$ in formula (2.132) needs to be estimated. In doing so, we assume that each unit cell of the crystal contains only one ion, then:

$$\kappa = 1\,, \quad \vec{R}_\kappa = 0\,, \quad M_\kappa = M\,, \quad \widetilde{V}_\kappa\left(\vec{r} - \vec{R}_\kappa^{\,l}\right) = \widetilde{V}\left(\vec{r} - \vec{R}^{\,l}\right)\,.$$

It should be recalled that the potential energy $\widetilde{V}(r)$ of the Coulomb interaction between the electron and the ion is screened so that it is of the order e^2/a within a sphere of radius a and vanishes outside this sphere. The parameter a is of the order of the interatomic distance in the crystal. It means that:

$$\int \widetilde{V}(\vec{r})\mathrm{d}^3\vec{r} \sim \frac{e^2}{a}a^3 = e^2a^2\,. \tag{2.135}$$

Given (2.135), we replace the potential $\widetilde{V}(\vec{r})$ by the effective point potential:

$$\widetilde{V}(\vec{r}) = e^2a^2\delta(\vec{r})\,. \tag{2.136}$$

When estimating, we describe the electrons by plane waves, i.e., we put forward that $u_{s\vec{k}} = 1\sqrt{V}$, where V is the volume of the crystal. In addition, for simplicity, we assume that the electron-phonon interactions happen within one energy band: $s = s'$.

Under the above conditions, the formulas (2.133) and (2.135) yield the following:

$$g_{s'\vec{k}',s\vec{k}}\left(\begin{smallmatrix}\vec{q}\\ j\end{smallmatrix}\right) \sim \mathrm{i}\sqrt{\frac{\hbar N}{2\omega_j(\vec{q})}}\frac{\left(\vec{e}\cdot\tilde{\vec{q}}\right)}{V\sqrt{M}}e^2a^2\delta_{\vec{k}'-\vec{k},\tilde{\vec{q}}}\,, \tag{2.137}$$

where $\vec{e} \equiv \vec{e}\left(\kappa\middle|\begin{smallmatrix}\vec{q}\\ j\end{smallmatrix}\right)$ is the polarization vector of the phonon, $\tilde{\vec{q}} = \vec{q} + \vec{K}$. When $\vec{K} = 0$, the value of $|(\vec{e}\cdot\vec{q})|$ is maximal when $\vec{e}\,|\,|\vec{q}$, i.e., the interactions of the electrons with longitudinal lattice vibrations are the most substantial.

The scattering of electrons by lattice vibrations can be regarded as a process of interaction of two gases – electron and phonon. The Hamiltonian (2.127) describes the electron-phonon interactions, which are characterized by Feynman diagrams presented in Figure 2.20.

If an electric field accelerates electrons, the electrons, in turn, generate phonons (Figure 2.20 (a)). This stands for heating the conductor by the electric current. Therefore, the above theory makes it possible to theoretically describe the rise in temperature of the conductor as the current flows through the latter.

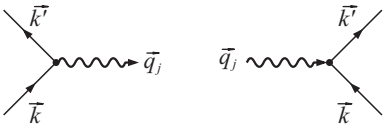


Fig. 2.20: Typical electron-phonon interactions. The solid line corresponds to an electron and the wavy line to a phonon.

3 Superconductivity

3.1 The Basic Physical Properties of Superconductors

In 1911, Kamerlingh Onnes, a Dutch physicist, studying the temperature dependencies of the resistance of mercury discovered that the resistance to an electric current suddenly dropped at a temperature of about 4 K. Shortly after, the same properties were found in some other metals. The new phenomenon was called "superconductivity," and the respective metals came to be named "superconductors." We will now take a closer look at the phenomenon of superconductivity.

I. The temperature at which the resistance disappears is called the critical temperature. This temperature is very different for different superconductors. To date, among pure metals, niobium has the greatest critical temperature: $T_c = 9.25$ K. The lowest one was found for wolfram: $T_c = 0.0154$ K. Superconductivity is not a rare phenomenon in nature. About twenty pure metals at low temperatures become superconductors: Ti, Zr, Hf, V, Nb, Ta, Mo, W, Tc, Re, Ru, Os, Ir, Zn, Cd, Hg, Al, Ga, In, Tl, Sn, Pb, La, Th, Pa, U. Even some semiconductors, such as Si, Ge, can be transformed into a superconducting state under certain conditions (such as when under high pressure, or when used as a thin film sample).

There are thousands of alloys exhibiting superconducting properties. For alloys the critical temperature turns out to be even higher than for pure metals. The record holder among the alloys is a Nb_3Ge compound with a critical temperature of $T_c = 23$ K.

The observation of superconductivity in metals and alloys is possible only by cooling them with expensive and "capricious" liquid helium. Until 1987, no superconducting materials with $T_c > 23$ K were synthesized. This inhibited practical applications of superconductors.

One, and only one, replacement of copper wires by superconducting ones increases power generation by 30%. Superconductors make it possible to manufacture high efficient electric motors and generators. For example, superconducting coils have long since created huge magnetic fields with induction of $B \sim 100$ T (CI) or $\sim 10^8$ Gauss (CGS). This is important, for instance, to solve problems of controlled thermonuclear fusion to develop new types of high speed magnetic levitation transport.

Superconducting magnets are of great importance to produce strong magnetic fields because simple coils made of aluminum and copper wire, in this case, consume enormous amount of energy. Almost all of this energy is released in the form of heat, and heat dissipation requires a cumbersome and expensive water cooling system. Water pumps experience vibration problems. A second difficulty is a strong attraction between the solenoid turns as an electric current passes through them. The copper solenoid coil begins to flow like a liquid. As a result, magnetic fields with the induction of a 10 Tesla cannot be achieved by means of the coils of copper wire. A coil made of a superconductor would eliminate the first difficulty. This is because the energy re-

https://doi.org/10.1515/9783110586183-003

quired only for maintaining the temperature of liquid helium is 1000 times less than the energy a conventional electromagnet consumes. In addition, the entire installation becomes compact.

A superconducting coil can be short circuited, then it becomes a permanent magnet. According to Lenz's rule, any change in external magnetic fields is compensated by the induced superconducting currents in the material. Therefore, superconducting magnets possess not only a great magnitude but also an extremely constant magnetic field.

Superconductors are the basis for a new generation of computers, which are compact, are high speed, and have low heat dissipation. One of the main problems arising when designing compact chips is heat dissipation. A liquid or gas can cool off only the surface but are unable to provide heat removal from the volume, which leads to a rise in temperature and, consequently, to destruction of the chips. The minimum area of contact between two superconductors, which serves as a logical element of a computer, is 10^{-9} cm^2, its speed performance is 10^{-10} sec, and the switching heat dissipation is negligible.

Today, there are devices called SQUIDs (Superconducting Quantum Interference Device), which already work on the basis of superconducting materials. These can detect extremely weak magnetic fields with strength in the order of 10^{-14} (CGS), which are generated, for example, by the heart or the cortex of the human brain. Methods of magnetodiagnostics supply much more information than electro-, cardio-, and encephalography. Moreover, they can be carried out in a noncontact manner. For example, they make it possible to take cardiograms from the heart of a child while still in the womb.

Until 1987, various theories of superconductivity thoroughly explained the available experimental data and left little hope for creating superconductors with $T_c > 30$ K. Therefore, the scientific world was extremely shocked when, in 1986, Bednorz and Muller (an IBM US company in Switzerland) declared the discovery of high temperature superconductivity in La-Ba-Cu-O metal-oxide ceramics at temperatures of liquid nitrogen which were cheap and available. Recall that nitrogen comprises 80% of the earth's atmosphere and its condensation temperature is 77 K. Early initial experiments greatly overcame the nitrogen barrier. The critical temperature of the first superconducting ceramics turned out to be about 90 K. For this discovery, Bednorz and Muller were awarded the Nobel Prize in 1987.

It is disappointing that ceramics are not stable; they crack, chemically decompose, and undergo structural transitions. The mechanism of their high temperature superconductivity has not still been elucidated. Therefore, we will talk about the high temperature superconductivity of ceramics separately, but now, let us return to the discussion of well-studied superconducting metals and alloys at temperatures of about 1–10 K.

II. Investigation of properties of metals and alloys has shown that superconductivity can be destroyed not only by increasing the temperature, but also by applying a suffi-

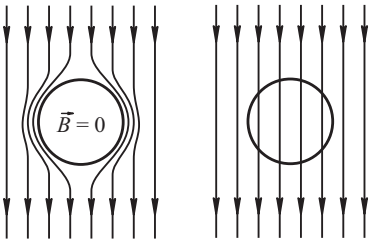


Fig. 3.1: The Meissner effect: (a) a superconductor; (b) a normal metal.

ciently strong magnetic field H. The critical field value H_c, at which superconductivity disappears, decreases with increasing temperature.

One of the main properties of superconductors is the so called the Meissner effect. *The Meissner effect* can be treated as the expulsion of a magnetic field with strength less than H_c from a superconductor during its transition to the superconducting state. In simpler words, the average macroscopic magnetic induction field in the superconductor is $\vec{B} = 0$ (Figure 3.1). This statement needs to be clarified and detailed.

When placed in an external magnetic field, a superconducting sample produces a persistent surface current. The latter creates its own field, fully compensating the external field inside the superconductor. It is the true magnetic field $\vec{B}$ (magnetic induction); it determines forces acting on the electric charge in the electromagnetic field:

$$\vec{F} = q\vec{E} + \frac{q}{c}\left[\vec{v} \times \vec{B}\right] \quad \text{(CGS)} \tag{3.1}$$

Here, q is the amount of charge, $\vec{v}$ is its velocity, and $\vec{E}$ and $\vec{B}$ are the electric field strength and magnetic flux density, respectively. The magnetic field strength $\vec{H}$ is not a real physical field, but is of an artificial mathematical construction, which relates the true field by means of the formula:

$$\vec{B} = \vec{H} + 4\pi\vec{M}\,, \tag{3.2}$$

where $\vec{M}$ is the magnetization of the sample per unit volume.

From this point of view, the condition $\vec{B} = \vec{H} + 4\pi\vec{M} = 0$ allows one formally to regard the superconductor as a perfect diamagnetic with a negative definite tensor of the magnetic susceptibility:

$$\chi_{sp} = \frac{\partial M_s}{\partial H_p} = -\frac{1}{4\pi}\delta_{sp}\,. \tag{3.3}$$

More detailed studies have found that the magnetic field is zero only inside the bulk of the sample. In a thin surface layer, the field diminishes gradually from a predetermined surface value to zero. The thickness of this layer called the *penetration depth* is typically of the order of $\delta = (10^{-5} \div 10^{-6})\,\text{cm}$ at $T = 0\,\text{K}$. The penetration depth depends on the temperature so that $\delta(T) \to \infty$ as $T \to T_c$.

The expulsion of the magnetic field towards the outside of the superconductor, in fact, is not always complete. There are two types of superconductors, both of which differ in how they penetrate an external magnetic field into the material.

The penetration of the field into the superconductor depends, generally speaking, on what geometry the sample has. However, the field behaves clearly differently in the simplest sample form (a long, thin cylinder with an axis parallel to the applied magnetic field) and in the bulk samples.

Type-I Superconductors

A magnetic field never penetrates inwards superconductors when it is less than the critical field $H_c(T)$. When $H \geq H_c(T)$, the entire sample jumps from the superconducting state to the normal state. Empirically, it has been found that the dependence $H_c(T)$ is described well by the formula (Figure 3.2):

$$H_c(T) = H_c(0)\left[1 - (T/T_c)^2\right] . \tag{3.4}$$

The accepted way of penetrating a magnetic field into a *type-I superconductor* is often described as dependence of macroscopic (diamagnetic) magnetization M on the strength of the applied field or by the dependence $B = B(H)$ (Figure 3.3).

In type-I superconductors, an interface between the superconducting and normal states is associated with positive surface energy. The penetration of the magnetic field inwards a type-I superconductor requires energy consumption to create the interface and the field inside the superconductor.

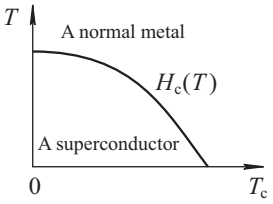


Fig. 3.2: Phase boundary of the plane *H-T* between superconducting and normal states of a type-I superconductor.

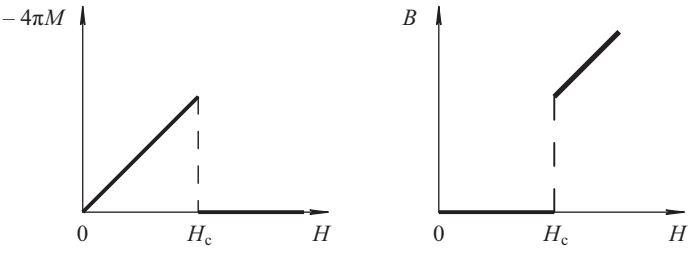


Fig. 3.3: The dependence of magnetization (magnetic induction) on the strength of an external magnetic field for type-I superconductors.

Type-II Superconductors

In contrast to type-I superconductors, a magnetic field gradually penetrates *type-II superconductors*. When less than the lowest critical field $H_{c1}(T)$, the magnetic field never penetrates at all into the samples as type-I superconductors do. Within the interval $H_{c1}(T) < H < H_{c2}(T)$, the magnetic field forms thin filaments inside the bulk of the sample. Inside the filaments, the superconducting state is destroyed; beyond the filaments, it is maintained. According to Lenz's rule, eddy superconducting currents arise around the filaments. Their magnetic fields tend to shield the external magnetic field inside each filament. This state of the superconductor is known as a vortex state. The number of filaments (*Abrikosov vortices*) within the superconducting material increases as the external magnetic field grows.

The external magnetic field in the form of filaments penetrates partially into the type-II superconductor due to negative energy of the interface between the superconducting and normal phases. As a result, the penetration of the magnetic field into the superconductor proves to be energetically advantageous in a certain field interval. We will dwell later on the microscopic reasons for the difference in the surface energies of different superconductors.

The theory of type-II superconductors was worked out by Ginzburg, Landau, Abrikosov and Gor'kov. Abrikosov predicted theoretically that the vortex filaments in the bulk of the superconductors form a regular and, often, a triangular lattice. The lattice spacing between the vortices is larger than the vortex core size and can reach 10^{-5}cm. The images of the vortex structures were captured with an electron microscope. As can be seen from Figure 3.4, the end face of a superconductor powdered with a contrasting powder of a ferromagnetic demonstrates a periodic lattice with triangular cells through thickening the powder at the outlet points of the vortices.

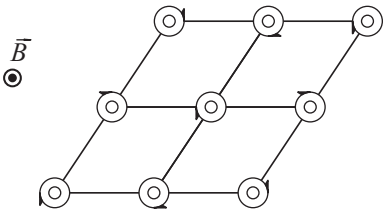


Fig. 3.4: Schematic representation of a lattice of Abrikosov vortices. Arrows indicate the directions of the eddy currents.

When $H > H_{c2}(T)$, the superconducting state in type-II superconductors is destroyed completely. Dependencies $M = M(T)$ and $B = B(H)$ for type-II superconductors are shown in Figure 3.5.

In any superconducting samples, superconductivity is destroyed when the magnetic field produced by a current flowing through the sample exceeds a critical surface value. The critical current value can reach 100 A in a wire with a diameter of 1 mm. Type-II superconductors possess the highest critical fields and critical currents.

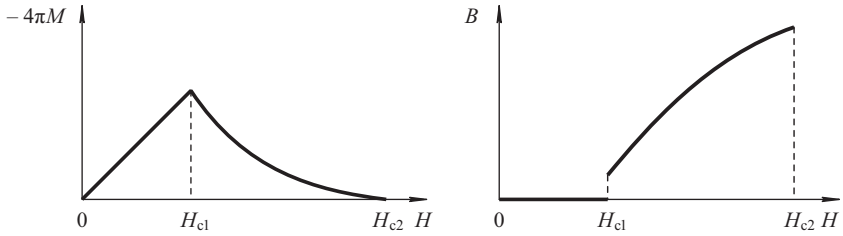


Fig. 3.5: The dependence of magnetization (magnetic induction) on the strength of an external magnetic field for type-II superconductors.

III. With ideal conductivity, superconductors in contrast to metals are poor heat conductors and have even lower heat capacity. Let us recall that, at low temperatures, the temperature dependence of the heat capacity of a normal metal has the form $AT+BT^3$, where the linear term is due to electrons, and cubic one describes lattice vibrations. The behavior of the heat capacity of the metal varies considerably below the superconducting transition temperature. Detailed studies show that the temperature linear term describing the electron contribution to the heat capacity is replaced by another. At very low temperatures, the new term descends much faster. Its low temperature behavior is determined by the multiplier $\exp(-\Delta/k_B T)$. This means that, in superconductors, the excited states of the electron subsystem are separated from the ground state by an energy gap Δ, as in semiconductors. The gap in magnitude is very small and amounts to $\Delta \approx k_B T_c$, where T_c is the transition temperature ($T_c \sim 10\,\mathrm{K}$).

IV. Superconductors with different isotopic compositions are found to exhibit the following dependencies of T_c and H_c on the mass of the crystal lattice ions (the *isotopic effect*):

$$T_c \sim M^{-1/2}\,, H_c \sim M^{-1/2}\,.$$

The mass of the ions becomes essential as soon as the lattice vibrations begin. Consequently, the superconductivity phenomenon is not purely electronic. Its appearance involves a lattice as well.

3.2 The Qualitative Features of the Microscopic Theory

An electric current in normal (i.e., not superconducting) metals is the transition of electrons under an electric field to higher energy levels in the conduction band. This is always possible because the electron energy levels in the conduction band are arranged densely. Current-carrying electrons can lower their energy by scattering on impurities, phonons and lattice defects and come back to energy levels that are not current carrying. From the energy point of view, this is the reason for the resistance of metals. The structure of the electron energy spectrum in a normal metal, and thus its

electrical resistance, can be explained by the fact that the electrons in the metal are Fermi particles obeying the Pauli exclusion principle.

Particles with integer spin, bosons, behave differently: as $T \to 0\,\mathrm{K}$, all Bose particles can condense in the same lowest energy level. When n bosons are in the same quantum state in particular, their velocities are equal. As a result, the probability of falling into this state increases for other bosons. The collective synchronous boson motion to be destroyed requires energy. At low temperatures, thermal energy is sometimes not enough to transfer the condensate of the synchronously moving bosons into other energy states.

The superfluidity of liquid helium isotopes with an atomic weight of four can serve as an example of the Bose condensate motion without energy loss. The liquid ^{4}He when $T < 2.19\,\mathrm{K}$ flows through fine capillaries without friction (Kapitza P.L., 1940).

If, in metals, one manages to combine electrons in pairs to have integer spin, such a system would exhibit superfluidity. In other words, we can observe superfluidity with the charged Bose particles as electron vapor or superconductivity.

In a fixed lattice, the Coulomb repulsion between electrons is screened at distances of the order of interatomic. Nevertheless, it is still repulsion, not attraction. So the electron pairing, at first glance, seems to be absurd. At the same time, the critical temperature dependence of a superconductor on mass of the crystal lattice ions suggests that deformations and vibrations of the lattice contribute to the attraction between electrons in a metal.

When moved through the crystal lattice, an electron with a wave vector $\vec{k}_1$ attracts positively charged ions. Thus, it causes deformations and vibrations of the lattice with the wavelength of the order of the interatomic distance. Such lattice vibrations correspond to acoustic phonons with a maximum frequency ω_D, and thus, with the energy $\hbar\omega_\mathrm{D}$.

As a result, the lattice ions are displaced, and an excess positive charge arises around the electron №1. Such a positively charged cloud attracts another electron №2 with a wave vector $\vec{k}_2$ turned around. This electron tends to reduce the excess positive charge. Because of the inertia, some ions cannot separate from each other. As a consequence, the electron №2 may be a relatively large distance away from the electron №1 (at a distance up to 10^{-5}cm). In the language of quantum mechanics, the process described stands for the generation of virtual phonon (or phonons) by the electron №1 and for the absorption them by the electron №2. The virtual particles, in contrast to conventional particles, exist only in intermediate states, have a short duration, and serve as interaction carriers. If such an electron-electron interaction, mediated by phonons, blocks the Coulomb repulsion, the resulting interaction between the electrons is attraction. The entire collective energy can further decrease as a result of condensation of bound states of the electron pairs.

Next, by perturbation theory, we calculate the change in energy of two free electrons. In doing so, we show that the energy of the system of two electrons, as a result of the exchange of virtual phonons, indeed goes down if the initial energies of the

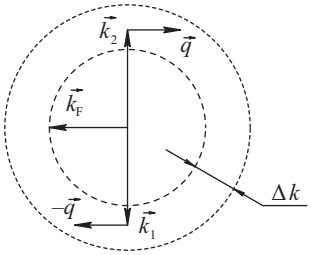


Fig. 3.6: The wave vectors $\vec{k}_1$ and $\vec{k}_2$ of electrons interacting with phonons are antiparallel and lie in a narrow spherical layer with a thickness of $\Delta k \approx \omega_D/V_F$ near the Fermi sphere in the reciprocal space. The wave vectors of virtual phonons interacting with the electrons are equal to $\pm\vec{q}$.

electrons №1 and №2 lie in the narrow layer near the Fermi surface:

$$\hbar\omega_D \approx \Delta\varepsilon = \frac{\partial\varepsilon}{\partial k}\Delta k = \hbar V_F \Delta k \,,$$

and the wave vectors $\vec{k}_1$ and $\vec{k}_2$ and the electron spins are antiparallel (Figure 3.6).

We can estimate the range of the forces of attraction between the electrons. The attraction forces appear due to the electron-phonon exchange. According to the uncertainty relation, the exchange takes place for the time $\tau \sim \hbar/\Delta\varepsilon \approx 1/\omega_D$. During this time, the electrons move apart at a distance of $r_0 \approx \tau 2V_F$. Therefore, the radius of action of the attraction forces between the electrons is given by:

$$r_0 \approx \frac{2V_F}{\omega_D} \sim \frac{V_F}{s}a \approx \frac{10^5\ \text{m/s}}{10^3\ \text{m/s}}a \approx 10^2 a \gg a \,,$$

where a is the interatomic distance. The phonon attraction turns out to be long range.

Recall that the Coulomb repulsion is short range, i.e., it manifests itself within the distances of the order of interatomic spacing. However, estimates show that the constants of the two interactions are the same order of magnitude:

$$\left|\frac{g_{rep}}{g_{attr}}\right| \sim 1 \,.$$

This statement will be justified below.

Hence, an important conclusion follows: the phonon attraction may prevail in some metals, and in the other: the Coulomb repulsion. It is this fact that explains why some metals pass into the superconducting state, while others do not.

The phonon attraction of electrons was discovered by Bardin and Fröhlich in 1950. Scientists of that time believed that the theory of superconductivity had been built already. However, it took another twenty years before the theory was developed. Let us clarify the main reason for the difficulties faced.

As mentioned earlier, the superfluidity in the electron system, i.e., the superconductivity, could be caused by the electron pairing. However, quantum mechanics denies the bound states if the interactions are not strong enough.

A measure of the binding energy Δ of electrons in a pair can be a critical temperature T_c:

$$\Delta \sim k_B T_c \,.$$

Since $T_c \approx (1 \div 10\,\mathrm{K})$, the ratio of the binding energy Δ and the kinetic electron energy ε_F turns out to be an extremely small magnitude:

$$\frac{\Delta}{\varepsilon_F} \sim 10^{-4}\,.$$

Therefore, the electron pairing appears to be impossible.

A one-dimensional case, when the particles move along a straight line, is the exception. Here, any attraction forms a bound state. However, a one-dimensional model, at first glance, has nothing to do with this problem.

In 1956, Cooper resolved the paradox. He noticed that it came, not to the interaction of isolated particles, but quasiparticles interacting with each other as the Fermi sphere is filled. This results in a one-dimensional problem instead of a three-dimensional one. From a formal point of view, the replacement is that the three-dimensional integrals over the wave vectors should be calculated as one-dimensional by the rule:

$$\int \frac{\mathrm{d}^3\vec{k}}{(2\pi)^3} \ldots = \frac{1}{2}\nu\,(\varepsilon_F) \int \mathrm{d}\varepsilon \ldots\,,$$

where $\nu(\varepsilon_F)$ is the density of energy levels on the Fermi surface. The multiplier $1/2$ appears because of the summation over the one quasiparticle states. In this case, the particle's spin projection is assigned as the density of states, determined previously as the form:

$$\nu(\varepsilon) = \frac{km}{(\pi\hbar)^2} = \frac{m}{(\pi\hbar)^2}\sqrt{\frac{2m\varepsilon}{\hbar^2}}\,,$$

and takes both the spin projections into account.

Thus, the filled Fermi sphere makes the Pauli principle radically change the problem at hand. As a consequence, the electrons fall into the bound state, whatever the weakness of their attraction. Breaking a coupled pair of the electrons requires energy. At low temperatures, the thermal motion energy is not enough to destroy the bound states. As a result, the ground state of a superconductor is not like one of a normal metal. The former lies below the energy ground state of a normal metal and is separated by an energy gap.

3.3 The Second Order Correction to the Energy of a Two Electron System Due to Electron-Phonon Interaction

Consider a gas of free electrons. For simplicity, the quasiparticle states, unperturbed by lattice vibrations, are assumed to be described by a free fermion model in a potential box. Let us calculate the change in energy of any two electrons as a result of their interaction with the lattice ion vibrations. Suppose the electron-phonon interaction is weak. We can find the correction to the energy of the two electrons from the general

formula of perturbation theory [4, 6]:

$$\Delta\varepsilon = \langle\tilde{n}|\,\widehat{H}_{\text{int}}\,|\tilde{n}\rangle + \sum_{\substack{\tilde{m}\\ \tilde{m}\neq\tilde{n}}} \frac{\left|\langle\tilde{n}|\,\widehat{H}_{\text{int}}\,|\tilde{m}\rangle\right|^2}{E^{(0)}_{\tilde{n}} - E^{(0)}_{\tilde{m}}}\,. \tag{3.5}$$

Here, $\widehat{H}_{\text{int}}$ is the interaction Hamiltonian of electrons №1 and №2 with lattice vibrations, and $|\tilde{n}\rangle$ is the unperturbed states of the two electrons without lattice vibrations (phonons):

$$|\tilde{n}\rangle = \left|\vec{k}_1\right\rangle\left|\vec{k}_2\right\rangle|0\rangle_\Phi \equiv |\vec{k}_1, \vec{k}_2, 0\rangle\,, \quad E^{(0)}_{\tilde{n}} = \varepsilon_0(\vec{k}_1) + \varepsilon_0(\vec{k}_2)\,. \tag{3.6}$$

The ket-vector $|0\rangle_\Phi$ corresponds to the vacuum state of the phonon subsystem of crystals. The ket-vectors $|\vec{k}_i\rangle$ characterize the states of free electrons with wave vectors $\vec{k}_i$ and energies $\varepsilon_0(\vec{k}_i)$ $(i = 1, 2)$.

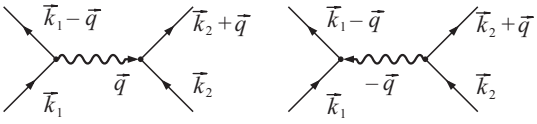


Fig. 3.7: Virtual processes where two electrons are exchanged by a phonon.

In formula (3.5), the summation over the intermediate states $|\tilde{m}\rangle$ of the whole system takes into account the processes of virtual phonon exchange between these electrons. For simplicity, attention should be drawn only to two main virtual processes (Figure 3.7). They correspond to the intermediate states $|\tilde{m}\rangle$ of the whole system with a virtual phonon:

$$|\tilde{m}\rangle = \begin{cases} \left|\vec{k}_1 - \vec{q}\right\rangle\left|\vec{k}_2\right\rangle|1\rangle_\Phi \\ \left|\vec{k}_1\right\rangle\left|\vec{k}_2 + \vec{q}\right\rangle|1\rangle_\Phi \end{cases} \equiv \begin{cases} \left|\vec{k}_1 - \vec{q}\,, \vec{k}_2, 1\right\rangle \\ \left|\vec{k}_1\,, \vec{k}_2 + \vec{q}, 1\right\rangle \end{cases}, \tag{3.7}$$

$$E_{\tilde{m}} = \begin{cases} \varepsilon_0(\vec{k}_1 - \vec{q}) + \varepsilon_0(\vec{k}_2) + \hbar\omega(\vec{q}) \\ \varepsilon_0(\vec{k}_1) + \varepsilon_0(\vec{k}_2 + \vec{q}) + \hbar\omega(-\vec{q})\,. \end{cases} \tag{3.8}$$

Here, $\hbar\omega(\pm\vec{q})$ are virtual phonon energies. The calculation looks only at the main interaction between the electrons with longitudinal phonons. In what follows, the polarization index $j = 3$ of these phonons is omitted.

Given an explicit form of the interaction operator $\widehat{H}_{\text{int}}$ (2.127) and the identity:

$$\langle 0|\,\hat{a}\,|0\rangle_\Phi = \langle 0|\,\hat{a}^+\,|0\rangle_\Phi = 0\,,$$

it is easy for us to verify that the first order correction to the energies of the two electrons is zero: $\langle\tilde{n}|\widehat{H}_{\text{int}}|\tilde{n}\rangle = 0$. As a result, formula (3.5) acquires the form:

$$\Delta\varepsilon = \frac{\left|\left\langle\vec{k}_1, \vec{k}_2, 0\left|\widehat{H}_{\text{int}}\right|\vec{k}_1 - \vec{q}, \vec{k}_2, 1\right\rangle\right|^2}{\varepsilon_0(\vec{k}_1) - \varepsilon_0(\vec{k}_1 - \vec{q}) - \hbar\omega(\vec{q})} + \frac{\left|\left\langle\vec{k}_1, \vec{k}_2, 0\left|\widehat{H}_{\text{int}}\right|\vec{k}_1, \vec{k}_2 + \vec{q}, 1\right\rangle\right|^2}{\varepsilon_0(\vec{k}_2) - \varepsilon_0(\vec{k}_2 + \vec{q}) - \hbar\omega(-\vec{q})}\,. \tag{3.9}$$

Let us transform expression (3.9) using the following claims.

1. The function $\omega(\vec{q})$ is even:

$$\omega(\vec{q}) = \omega\left(-\vec{q}\right) .$$

2. The squared modulus of the matrix elements $|\langle\widetilde{n}|\widehat{H}_{\text{int}}|\widetilde{m}\rangle|^2$ are easy to calculate using formula (2.132) for matrix elements of the interaction operator of the electron Bloch functions and by the formulas:

$$\langle 0|\ \widehat{a}\ |1\rangle_{\Phi} = 1 , \quad \langle 0|\ \widehat{a}^{+}\ |1\rangle_{\Phi} = 0 .$$

After doing simple calculations, we come up with:

$$\left|\left\langle \vec{k}_1, \vec{k}_2, 0\left|\widehat{H}_{\text{int}}\right|\vec{k}_1 - \vec{q}, \vec{k}_2, 1\right\rangle\right|^2 = \left|\left\langle \vec{k}_1, \vec{k}_2, 0\left|\widehat{H}_{\text{int}}\right|\vec{k}_1, \vec{k}_2 + \vec{q}, 1\right\rangle\right|^2 = \left|g_{\vec{k}_1, \vec{k}_1 - \vec{q}}\right|^2 .$$

For free electrons in a potential box, the index s of the function $g_{s\vec{k}_1, s\vec{k}_1 - \vec{q}}$ should be omitted. This is because it corresponds to the number of the energy band, and the bands appear only when we take the periodic potential of the crystal into account.

3. By virtue of the law of conservation of energy, we have:

$$\varepsilon_0(\vec{k}_1) + \varepsilon_0(\vec{k}_2) = \varepsilon_0(\vec{k}'_1) + \varepsilon_0(\vec{k}'_2) ,$$

or

$$\varepsilon_0(\vec{k}_1) - \varepsilon_0(\vec{k}'_1) = -\left[\varepsilon_0(\vec{k}_2) - \varepsilon_0(\vec{k}'_2)\right] , \tag{3.10}$$

where $\vec{k}'_i$ $(i = 1, 2)$ are the wave vectors of electrons in the final state.

According to the law of conservation of momentum, for the free electrons (see Figure 3.7), we get:

$$\vec{k}'_1 = \vec{k}_1 - \vec{q} , \quad \vec{k}'_2 = \vec{k}_2 + \vec{q} . \tag{3.11}$$

Therefore, the formulas (3.10), (3.11) yield:

$$\varepsilon_0(\vec{k}_1) - \varepsilon_0(\vec{k}_1 - \vec{q}) = -\left[\varepsilon_0(\vec{k}_2) - \varepsilon_0(\vec{k}_2 + \vec{q})\right] . \tag{3.12}$$

The above claims simplify the correction form to the two electron energy (3.9) that is due to the electron-phonon interaction:

$$\Delta\varepsilon = -\frac{2|g_{\vec{k}_1, \vec{k}_1 - \vec{q}}|^2 \hbar\omega(\vec{q})}{(\hbar\omega(\vec{q}))^2 - \left[\varepsilon_0(\vec{k}_1) - \varepsilon_0(\vec{k}_1 - \vec{q})\right]^2} . \tag{3.13}$$

At low electron energies, this is:

$$(\hbar\omega)^2 \gg \left[\varepsilon_0(\vec{k}_1) - \varepsilon_0(\vec{k}_1 - \vec{q})\right]^2$$

when the lattice manages to follow the ion motion, we have:

$$\Delta\varepsilon = -\frac{2|g_{\vec{k}_1, \vec{k}_1 - \vec{q}}|^2}{\hbar\omega(\vec{q})} < 0 . \tag{3.14}$$

The estimation of the matrix element (2.137) gives rise to:

$$|g_{\vec{k}_1,\vec{k}_1-\vec{q}}|^2 \approx \frac{\hbar N q^2 e^4 a^4}{2\omega V^2 M} .$$

Using the formulas:

$$n = \frac{N}{V} , \quad na^3 \sim 1 , \quad \omega^2 = s^2 q^2 , \quad s = \sqrt{\frac{\alpha}{M}} a , \quad \alpha \sim \frac{e^2}{a^3} ,$$

we arrive at:

$$\Delta\varepsilon = -\frac{2|g_{\vec{k}_1,\vec{k}_1-\vec{q}}|^2}{\hbar\omega} \approx -\frac{n\hbar}{\omega V}\frac{q^2 e^4 a^4}{M\hbar\omega} \approx -\frac{q^2 e^4 a}{VM\omega^2} \approx -\frac{1}{V}\frac{e^4 a}{M\left(e^2/aM\right)} \approx -\frac{e^2 a^2}{V} . \quad (3.15)$$

The electron-phonon interaction is long range, extending over the whole crystal volume. Therefore, we can treat the correction to the two electron energy as an average of certain effective potential energy of their interaction U_{eff} over the crystal volume:

$$\Delta\varepsilon \approx \frac{1}{V}\int U_{\text{eff}} \mathrm{d}^3\vec{r} = -\frac{a^2 e^2}{V} . \quad (3.16)$$

Formula (3.16) implies that the electron-electron interaction transmitted by phonons is equivalent to a point interaction:

$$U_{\text{eff}} = -e^2 a^2 \delta(\vec{r}) , \quad (3.17)$$

with its sign corresponding to attraction.

With the effective interaction being independent on the rotation angle of the momenta $\hbar\vec{k}_1$ and $\hbar\vec{k}_2$ of the electrons, the electron pair has the orbital angular momentum equal to $l = 0$.

The interaction found is equivalent to attraction of particles as their coordinates coincide. In this case, the coordinate two electron wave function $|k_1\rangle|k_2\rangle$ turns to be symmetrical with respect to the permutation of the quantum numbers $\vec{k}_1$ and $\vec{k}_2$ of the two electrons. However, electrons are the Fermi particles. This means that their total wave function $|k_1\rangle|k_2\rangle\chi_{12}$ must be antisymmetric with respect to the permutation of all quantum numbers of electrons, including spin ones. It follows that the calculation proposed is fair only when the spin part of the total two electron wave function is antisymmetric with respect to the permutation of their spin projections:

$$\chi_{12} = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle_1|\downarrow\rangle_2 - |\downarrow\rangle_1|\uparrow\rangle_2\right) . \quad (3.18)$$

The spin function χ_{12} (3.18) corresponds to the singlet state of the electron system with total spin $S = 0$.

Previously, we have shown that the screened Coulomb repulsion is equivalent to the effective point interaction $e^2 a^2 \delta(\vec{r})$. Now we have just proved that the effective phonon attraction is the same order of magnitude: $-e^2 a^2 \delta(\vec{r})$.

Let us give some estimates which are useful for further analysis. In metals, the Coulomb energy of two electron interactions is about their kinetic energy:

$$\frac{e^2}{a} \sim \frac{p_\mathrm{F}^2}{m} ,$$

where a is the interatomic distance. According to the uncertainty relation, $p_\mathrm{F} a \sim \hbar$. This brings us to the chain of equalities:

$$e^2 a^2 \approx \frac{p_\mathrm{F}^2}{m} a^3 \approx \frac{\hbar^3}{m p_\mathrm{F}} = \frac{\hbar^2}{m k_\mathrm{F}} . \tag{3.19}$$

3.4 Cooper Pairs

In 1956, Cooper put forward an assumption that, at low temperatures, the ground state of a normal metal with the Fermi sphere filled up with electrons becomes energetically unfavorable, and therefore unstable. A superconducting phase where electrons form bound states, later called *Cooper pairs*, has less energy.

The Cooper pair formation is caused by the effective electron-electron attraction involving virtual phonon exchange. The pairing origin can be understood through a simple model problem that illustrates the mutual electron interaction by means of a two-particle potential $U(\vec{r}_1, \vec{r}_2)$ [4, 10]. The essential point here is the fact that the model takes into account the interaction of only two electrons. The rest of them form a ground state of a normal metal. In other words, the other electrons fill up all the energy levels inside the Fermi sphere. It appears that the Fermi sphere, when filled completely, substantially affects the binding of electrons in pairs.

We write the Schrödinger equation for two quasiparticles:

$$\left[\widehat{H}_0(\vec{r}_1) + \widehat{H}_0(\vec{r}_2) + U(\vec{r}_1, \vec{r}_2)\right] \Psi(\vec{r}_1, \vec{r}_2) = E\Psi(\vec{r}_1, \vec{r}_2) . \tag{3.20}$$

Here, $\widehat{H}_0(\vec{r})$ is the free quasiparticle Hamiltonian:

$$\widehat{H}_0(\vec{r})\Psi_{\vec{k}}(\vec{r}) = \varepsilon_0(\vec{k})\Psi_{\vec{k}}(\vec{r}) . \tag{3.21}$$

For simplicity, we consider an isotropic case when the problem is (3.21) where eigenfunctions and eigenvalues have the form:

$$\Psi_{\vec{k}}(r) = \frac{\exp\left(\mathrm{i}\vec{k}\cdot\vec{r}\right)}{\sqrt{V}} , \quad \varepsilon_0(\vec{k}) = \frac{\hbar^2 k^2}{2m} \geq \varepsilon_\mathrm{F} .$$

The system is thought to be in a volume V and is thought to satisfy the Born–Karman boundary conditions.

Suppose an electron pair of the system has the least energy. Then its total wave function needs to be sought in the following form:

$$\Psi(\vec{r}_1 s_1, \vec{r}_2 s_2) = \sum_{|\vec{k}'|>k_\mathrm{F}} c_{\vec{k}'} \exp\left(\mathrm{i}\vec{k}'\cdot\vec{r}\right) \exp\left(-\mathrm{i}\vec{k}'\cdot\vec{r}_2\right)\chi_{12} , \tag{3.22}$$

where $c_{\vec{k}'} = c_{-\vec{k}'}$, $\chi_{12} = 1/\sqrt{2}(|\uparrow\rangle_1|\downarrow\rangle_2 - |\downarrow\rangle_1|\uparrow\rangle_2)$. It is important that all the coefficients $c_{\vec{k}'}$ are different from zero only when $|\vec{k}'| > k_F$. We assume that the electrons with opposite spins and wave vectors are "more ready" to pair. The total wave function (3.22) is antisymmetric with respect to the permutation of electrons.

We substitute formula (3.22) into formula (3.20). Then, we multiply the result of the left-hand side by the function $\exp(-i\vec{k}\cdot\vec{r}_1)\exp(i\vec{k}\cdot\vec{r}_2)$ and integrate it over $\vec{r}_1$ and $\vec{r}_2$ considering the orthogonality conditions:

$$\iint\limits_V d^3\vec{r}_1 d^3\vec{r}\exp\left\{-i(\vec{k}-\vec{k}')\cdot(\vec{r}_1-\vec{r}_2)\right\} = V^2\delta_{\vec{k},\vec{k}'}\,.$$

Finally, we get:

$$\left[E - 2\varepsilon_0(\vec{k})\right]c_{\vec{k}} = \sum_{|\vec{k}'|>k_F} U_{\vec{k}\vec{k}'}c_{\vec{k}'}\,, \tag{3.23}$$

where

$$U_{\vec{k}\vec{k}'} = \frac{1}{V^2}\iint\limits_V d^3\vec{r}_1 d^3\vec{r}_2\exp\left\{-i(\vec{k}-\vec{k}')\cdot(\vec{r}_1-\vec{r}_2)\right\}U(\vec{r}_1,\vec{r}_2)\,.$$

In accordance with the qualitative picture just previously set forth, we put:

$$U_{\vec{k}\vec{k}'} = \frac{1}{V}\begin{cases} -g\,, & \text{for} \quad \varepsilon_F \le \varepsilon_0(\vec{k}) \le \varepsilon_F + \hbar\omega_D\,, \quad \varepsilon_F \le \varepsilon_0(\vec{k}') \le \varepsilon_F + \hbar\omega_D\,, \\ 0\,, & \text{in the rest of cases}\,. \end{cases} \tag{3.24}$$

Formula (3.24) takes into account the fact that the attraction between the electrons occurs only when their energies fall into the interval $[\varepsilon_F, \varepsilon_F + \hbar\omega_D]$. At the same time, the parameter is $g \sim \hbar^2/k_F m \sim e^2a^2$, i.e., it is close to our previous estimates.

The potential (3.24) is rather a simple model of existing effective attractions between electrons. Therefore, the outcomes, depending on the potential type, should not be treated too seriously. Fortunately, the theory yields a number of relations not dependent on the phenomenological parameter g. These results are in good agreement with experimental data for a wide range of superconductors.

Given an explicit form of the matrix element $U_{\vec{k}\vec{k}'}$ (3.24) we solve equation (3.23) for $c_{\vec{k}'}$:

$$c_{\vec{k}} = \frac{gI}{V\left[2\varepsilon_0(\vec{k}) - E\right]}\,, \tag{3.25}$$

where

$$I = \sum_{\varepsilon_F \le \varepsilon_0(\vec{k}') \le \varepsilon_F + \hbar\omega_D} c_{\vec{k}'}\,. \tag{3.26}$$

Now we find a two electron bound state, corresponding to the eigenvalue E that is less than the two free electron energy $2\varepsilon_F$. Suppose E is designated as $2\varepsilon_F - 2\Delta$. Then we insert formula (3.25) into (3.26). After canceling the common multiplier I, we come to an equation for calculating Δ:

$$1 = \sum_{\varepsilon_F \le \varepsilon_0(\vec{k}') \le \varepsilon_F + \hbar\omega_D} \frac{g}{2V[\varepsilon_0(\vec{k}') - \varepsilon_F + \Delta]}\,. \tag{3.27}$$

We simplify the equation (3.27) by replacing the summation (or integration) in three-dimensional k-space through integration over the electron energy ε within the interval $[\varepsilon_F, \varepsilon_F + \hbar\omega_D]$:

$$\frac{1}{V}\sum_{\vec{k}} \ldots = \frac{1}{(2\pi)^3}\int \mathrm{d}^3\vec{k}\ldots = \frac{1}{2}\nu(\varepsilon_F)\int \mathrm{d}\varepsilon \ldots . \tag{3.28}$$

This is a mathematical method that reduces the three-dimensional problem in the k-space to the equivalent one-dimensional problem on the axis ε. Formula (3.28) involves $\nu(\varepsilon_F)$ as the density of the energy levels at the Fermi sphere. Let us once again draw our attention to the multiplier 1/2 in (3.28). The latter appears because of the summation over the allowed states of a quasiparticle, whose spin projection is given. As previously determined, the density of states has taken both spin projections into account. Transformed by the rule (3.28), the equality (3.27) becomes:

$$1 = \frac{g\nu(\varepsilon_F)}{4}\int_{\varepsilon_F}^{\varepsilon_F+\hbar\omega_D} \frac{\mathrm{d}\varepsilon}{\varepsilon - \varepsilon_F + \Delta} = \frac{g\nu(\varepsilon_F)}{4}\ln\frac{\hbar\omega_D + \Delta}{\Delta} \approx \frac{g\nu(\varepsilon_F)}{4}\ln\frac{\hbar\omega_D}{\Delta} . \tag{3.29}$$

In the right-hand side of (3.29) we have used the approximation $0 < \Delta \ll \hbar\omega_D$ to be justified later.

From equation (3.29), we find the binding energy of the electrons:

$$\Delta = \hbar\omega_D \exp\left[-\frac{4}{\nu(\varepsilon_F)g}\right] > 0 . \tag{3.30}$$

So, in a pair of quasiparticles, the electrons have a finite binding energy. At $T = 0$ K all the pairs of electrons must form a Bose condensate.

Note that the binding energy (3.30) is not an analytic function of the phonon attraction constant g (it has an essential singularity when $g = 0$). Therefore, Cooper's result cannot be derived by perturbation theory, i.e., it cannot be expanded in a series in powers of g. It should be especially emphasized that we have used the perturbation theory earlier, only to make an estimate the parameter g.

Now we give numerical estimates for the binding energy Δ. Since, for free electrons:

$$g \sim \frac{\hbar^2}{k_F m} , \quad \nu(\varepsilon_F) = \frac{m}{(\pi\hbar)^2}k_F ,$$

we get $g\nu \sim 1/\pi^2 < 1$. There are arguments in favor of the inequality $0 < g\nu < 1$, which turns out to hold true in general cases. Therefore, we have $4/g\nu(\varepsilon_F) > 1$. The exponent in (3.30) contains this number with a negative sign. As a result, the estimate $0 < \Delta \ll \hbar\omega_D$ proved to be justified. Since $\Delta \sim k_B T_c$, $\hbar\omega_D \sim k_B T_D$, we come up with the result:

$$\frac{\Delta}{\hbar\omega_D} = \frac{T_c}{T_D} \ll 1 .$$

We have shown that the superconducting transition temperature is much below than the Debye temperature. More accurate estimates yield the following outcome:

$$\frac{T_c}{T_D} = \exp\left[-\frac{4}{g\nu(\varepsilon_F)}\right] \sim 10^{-2} .$$

We evaluate the radius of the resulting bound state – the size of a Cooper pair. By analogy with the estimates for hydrogen atom, we think that the uncertainty of the kinetic energy of an electron in the pair is of the order of the binding energy Δ:

$$\Delta \sim k_B T_c \sim \delta\left(\frac{p^2}{2m}\right) = \frac{p_F \delta p}{m} = V_F \delta p .$$

Hence we arrive at $\delta p \sim k_B T / V_F$. The uncertainty relation $\xi_0 \delta p \sim \hbar$ for the position and momentum of an electron leads us to an estimate of the characteristic size of a Cooper pair:

$$\xi_0 \sim \frac{\hbar V_F}{k_B T_c} \sim \frac{\varepsilon_F}{k_F} \frac{1}{k_B T_c} .$$

At $T_c \approx 10\,\mathrm{K}$, we have $k_B T_c \approx 10^{-4}\varepsilon_F$. Recall that $k_F \sim 10^{-8}\mathrm{cm}$ at any temperature. Ultimately, we obtain a numerical estimate of the size of a Cooper pair:

$$\xi_0 \sim 10^{-4}\mathrm{cm} \sim 10^4 a \gg a .$$

Thus, the size of a Cooper pair amounts to tens of thousands of interatomic distances (tens of nanometers). The pairs are very loose formations. The question arises of how the bound pairs can be placed in metals without interfering with each other if the distance between electrons in a normal metal, in fact, is of the order of a.

The Cooper pairs cannot be interpreted as a classical gas of particles because the distance between them is much greater than their size. The pairs penetrate each other. At the same time, they are free as gas particles and almost never interact between themselves. Such a situation is possible only in quantum mechanics. To point out this feature of the Cooper pairs, the strict theory talks not about pairs having a certain size, but about a pair electron correlation spread over a certain distance – namely, the "*correlation length*." It is in the order of:

$$\xi_0 \sim \frac{\hbar V_F}{\Delta} .$$

The above reasoning covers a single Cooper pair of electrons in the presence of conventional, nonsuperconducting electrons obeying the Fermi distribution. The Bardeen–Cooper–Schrieffer (BCS) theory takes a considerable step forward: it allows the building of the ground state of a superconductor in which all the electrons (not only two) form bound pairs.

3.5 The Bardeen–Cooper–Schrieffer Theory (Qualitative Results)

Proceeding from the idea that electrons with opposite momenta can form coupled, overlapping pairs, we choose the wave function of the entire collective of the electrons in the BCS theory as a single wave function:

$$\Psi_{\mathrm{BCS}}(\vec{r}_1 s_1, \ldots, \vec{r}_N s_N) = \widehat{A}\left\{\varphi(\vec{r}_1 s_1, \vec{r}_2 s_2)\cdot\varphi(\vec{r}_3 s_3, \vec{r}_4 s_4)\cdot\ldots\cdot\varphi(\vec{r}_{N-1} s_{N-1}, \vec{r}_N s_N)\right\} . \tag{3.31}$$

The action of the antisymmetrization operator $\widehat{A}$ on a function reduces to adding $(N! - 1)$ functions obtained through all possible permutations of the arguments $\vec{r}_1 s_1, \vec{r}_2 s_2, \ldots \vec{r}_N s_N$. Besides, these functions are multiplied by the number +1 or −1, depending on if a given permutation is a result of an even or odd number of paired permutations. The wave function describing a single pair takes the form:

$$\varphi(\vec{r}_1 s_1, \vec{r}_2 s_2) = \sum_{\vec{k}} c_{\vec{k}} \exp\left[\mathrm{i}\vec{k}\cdot(\vec{r}_1 - \vec{r}_2)\right]\chi_{12} . \tag{3.32}$$

where $c_{\vec{k}} = c_{-\vec{k}}$. The total spin of the pair is assumed to be zero, because the spin part χ_{12} of the function (3.32) is antisymmetric:

$$\chi_{12} = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle_1 |\downarrow\rangle_2 - |\downarrow\rangle_1 |\uparrow\rangle_2\right) .$$

At the same time, the coordinate part in (3.32) is symmetric and, moreover, it corresponds to the zero angular momentum of the pair.

Comment

A question can arise about the pairing of quasiparticles with spin 1/2 in the states with $l \neq 0$ and $S = 1$. How is it possible? In general, such pairings are possible. However, they lead to magnetic properties which are not observed in conventional metals and alloys, and which have been studied before, prior to discovering high temperature metal ceramics. The triplet pairing with $l = 1$ and $S = 1$ is observed in liquid ^{3}He. It can be emphasized that the case in hand is the helium isotope with an atomic mass of three. In contrast to the ^{4}He, ^{3}He atoms are not bosons but fermions. At very low temperatures, ^{3}He atoms form a Bose condensate of pairs with spin $S = 1$, which turns out to be superfluid. However, it is an electrically neutral liquid.

Among metals, the pairing with nonzero orbital angular momentum and parallel spins is possible in the so called “heavy”-fermion systems. However, to date, this issue has not been adequately investigated. The difficulty is that pairing with the nonzero orbital angular momentum stands for the nonlocal electron-electron interaction. Therefore, only a new nonphonon mechanism of electron attraction can be considered (the phonon mechanism corresponds to the local interaction).

The structure of the wave function chosen in the BCS theory implies that:

a) The wave function contains all electrons as Cooper pairs.
b) Electron pairs existing in a superconductor correlate with each other.

It can be pointed out that the wave function of a single pair cannot be changed without completely destroying the whole superconducting state. The wave function (3.31) for the ground state of a superconductor assumes the following: if this state really exists, i.e., its energy is lower than the energy of the ground state of a normal metal, the superconductivity cannot be easily destroyed.

As to the form of the Hamiltonian of the electron system in a metal, there are two rough assumptions underlying the BCS theory:

1. Effects related to the band structure are not taken into account. The free electron approximation is used.
2. Effective electron pair attraction having a complicated form is replaced by a model potential with the following matrix elements:

$$U_{\vec{k}\vec{k}'} = \frac{1}{V}\begin{cases} -g\,, & \text{when} \quad \left|\varepsilon(\vec{k}) - \varepsilon_{\mathrm{F}}\right| < \hbar\omega_{\mathrm{D}}\,, \quad \left|\varepsilon(\vec{k}') - \varepsilon_{\mathrm{F}}\right| < \hbar\omega_{\mathrm{D}}\,, \\ 0\,, & \text{in the rest of cases}\,, \end{cases}$$

where $g \sim \hbar^2/(k_{\mathrm{B}}m) \sim e^2a^2$. Such a potential takes into account the main features of the electron-phonon attraction and allows all the calculations to be carried out completely.

General Conclusions

Suppose the wave function Ψ_{BCS} is a testing one to evaluate the energy of the ground state of an electron collective by means of the variation principle. It is claimed that, for arbitrarily small $g > 0$ (the electron-electron attraction is arbitrarily small), it corresponds to lower system energy as compared to functions chosen in the free electron approximation.

At $T = 0$ K, the total energy F_S per unit volume of the BCS ground state is below the total energy F_N per unit volume of the ground state in a normal metal. The electrons in the energy band width $\sim\Delta(0)$ (their number is $\sim\nu(\varepsilon_{\mathrm{F}})\Delta(0)$) are coupled in pairs and the pairing energy is $\sim(-\Delta)$. This process diminishes the energy per unit volume in a metal by magnitude of the order of $\nu(\varepsilon_{\mathrm{F}})\Delta^2(0)$. The exact calculation within the BCS theory yields:

$$F_S - F_N = -\frac{1}{4}\nu(\varepsilon_{\mathrm{F}})\Delta^2(0)\,,$$

where ε_{F} is the Fermi energy of a normal metal.

$$\Delta(0) = 2\hbar\omega_{\mathrm{D}}\exp\left(-\frac{4}{g\nu(\varepsilon_{\mathrm{F}})}\right)$$

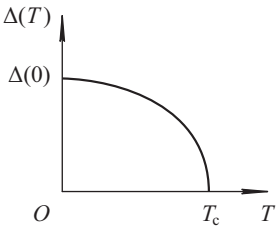


Fig. 3.8: $\Delta(T)$-dependence.

is the binding energy of two electrons. It differs from the findings of Cooper (3.30) only by a factor equal to 2. The additional factor takes the tuning of the energy spectrum and the ground state of the entire electron collective into account.

In the ground state of a superconductor, all states are filled pairwise. If the state with the wave vector $\vec{k}$ and spin-downwards is occupied, the state with the wave vector $(-\vec{k})$ and spin-upwards is occupied as well. Being broken, the pair gives birth to two unpaired electrons – quasiparticles with the same energy:

$$E(\vec{k}) = \sqrt{(\varepsilon_0(\vec{k}) - \varepsilon_F)^2 + \Delta^2} \, .$$

The energy of each quasiparticle is $E(\vec{k}) \approx \varepsilon_0(\vec{k}) - \varepsilon_F$ only when $|\varepsilon_0(\vec{k}) - \varepsilon_F| \gg \Delta$.

The minimum energy required for breaking the pair and transferring the two electrons into the normal state is given by:

$$\min\left\{E(\vec{k}) + E(\vec{k}')\right\} = 2\Delta(0) \, .$$

The magnitude of the energy gap is affected by increasing the temperature $\Delta = \Delta(T)$ (Figure 3.8).

A more accurate mathematical calculation was carried out by Bogolyubov and Gor'kov. Bogolyubov developed a special method of canonical transformations. Gor'kov proposed a new method of Green's functions to calculate properties of superconductors.

It is natural to assume that the temperature, at which the energy spectrum gap $\Delta(T)$ of superconductors disappears, corresponds to the temperature of the superconducting transition. Carrying out a particular calculation, Bardeen, Cooper and Schrieffer get [11]:

$$k_B T_c = 1.14\,\hbar\omega_D \exp\left(-\frac{4}{g\nu(\varepsilon_F)}\right) . \tag{3.33}$$

Since $\omega_D \sim 1/\sqrt{M}$, this explains the isotope effect:

$$T_c \sim 1/\sqrt{M} \, ,$$

where M is mass of the crystal lattice ions.

Note the universal relation:

$$\frac{\Delta(0)}{k_B T_c} = 1.76 \tag{3.34}$$

that holds up to 10% for most superconducting metals and alloys.

The BCS theory gives the following expression for the specific heat of a superconductor at low temperatures. It does not depend on the parameter g:

$$\frac{c_s}{\gamma T_c} = 1.34 \left(\frac{\Delta(0)}{k_B T} \right)^{\frac{3}{2}} \exp\left(-\frac{\Delta(0)}{k_B T} \right) , \tag{3.35}$$

Here, γ is the coefficient of the linear term in the temperature dependence of the heat capacity of a metal in the normal state: $c_n \approx \gamma T$. The result obtained is in strongly supports experimental data. The temperature dependence of the specific heat of a superconductor is of an exponential nature, and is ascertained by the energy gap width $\Delta(0)$.

The Existence of a Critical Current in Superconductors

In normal metals, the electrons scattering by impurities, defects, and vibrating lattice ions transfer their kinetic energy to the crystal, leading to the conductor heating up. This lost energy is compensated by an external source whose removal makes the current decay.

In superconductors, the current does not disappear, even without an external energy source until it exceeds a certain critical value. Let us estimate the critical current in a superconductor.

If the mean quasimomentum of a Cooper pair moving directionally is $\hbar\vec{K}_0$, the wave vectors of the pair of electrons are equal to $(\vec{k} + \vec{K}_0/2)$ and $(-\vec{k} + \vec{K}_0/2)$. When scattered by lattice irregularities, due to energetic profitability, the electron with the momentum $\hbar(\vec{k} + \vec{K}_0/2)$ passes to a state with the momentum $\hbar(-\vec{k} + \vec{K}_0/2)$ to diminish its kinetic energy by the magnitude:

$$\frac{\hbar^2}{2m}\left(\vec{k} + \vec{K}_0/2\right)^2 - \frac{\hbar^2}{2m}\left(-\vec{k} + \vec{K}_0/2\right)^2 = \frac{\hbar^2}{m}\vec{k} \cdot \vec{K}_0 . \tag{3.36}$$

It is clear that the energy gain is possible only when the angle between the vectors $\vec{k}$ and $\vec{K}$ are acute.

At the same time, the change in the state of one electron out of the pair is equivalent to breaking the pair, which requires the energy expenditure $\sim 2\Delta(0)$. Consequently, destroying the pair is possible only if the following energy relation is fulfilled:

$$\frac{\hbar^2}{m}\vec{k} \cdot \vec{K}_0 \geq 2\Delta(T) . \tag{3.37}$$

Using (3.37), we can estimate the maximum possible total momentum of the pair. Setting $k \sim k_F$, we find the critical momentum of the pair as a whole:

$$\hbar K_0 \sim \frac{2m\Delta(T)}{\hbar k_F} .$$

Hence, it follows that the estimate of the critical current density of the superconductor is given by:

$$j_c = -|e|\left(\frac{\hbar K_0}{2m}\right)n_s = -n_s\frac{|e|\,\Delta(T)}{\hbar k_F} \sim (10^6 \div 10^7)\frac{A}{cm^2}\,, \tag{3.38}$$

where n_s is the concentration of superconducting electrons.

When $j < j_c$, the pair breakings and the scattering of the electrons forming the pairs by the irregularities are energetically impossible. The scattering does not occur because of "mutual support" of the electrons in Cooper pairs, such that the Cooper pairs "flow around" the defects without energy loss. Since the crystal does not receive the superconducting current energy, switching off an external source of electromotive force does not entail the decay of the superconducting current.

With increasing temperature, phonons emerge and their energies are sufficient enough to break the pairs. The breakings of the Cooper pairs lead to a decrease in the superconducting electron concentration n_s. In addition, the energy gap decreases with increasing temperature. According to formula (3.38), the factors listed above are the reason for a decrease in the critical current j_c.

In general, at $T < T_c$, a unit volume of the superconductor has n_s normal electrons – quasiparticles with energy $E(\vec{k})$ and $n_s/2$ Cooper pairs (n_s superconducting electrons). The total number of electrons per unit volume remains unchanged: $n_s + n_n = \text{const}$. The total current in the superconductor is the sum of superconducting and normal currents. The superconducting current is carried by $n_s/2$ Cooper pairs with the charge $-2|e|$. The normal current is associated with the density of ordinary electrons. In contrast to the superconducting current, the normal current decays when switching off an external source of emf. The coexistence of the normal and superconducting components of the total current in the superconductor is interdependent and inseparable.

In 1972, Bardeen, Cooper and Schrieffer were awarded the Nobel Prize for the explanation of the phenomenon of superconductivity.

3.6 The Ginzburg–Landau Theory – The London Penetration Depth

In 1958, Lev Gor'kov generalized the microscopic BCS theory. However, the equations derived by him are complicated and rarely used for solving practical problems.

In 1950, long before the formulation of the BCS theory, Ginzburg and Landau came up with a simple, semiclassical theory, which perfectly reproduces the main results of the microscopic theory of superconductivity. As Gor'kov showed in 1959, the Ginzburg–Landau equations are an exact limit of consistent microscopic theory

under two conditions:

$$|T_c - T| \ll T_c \,; \tag{3.39}$$

$$\delta(T) \gg \xi_0 \,, \tag{3.40}$$

where $\delta(T)$ is the depth of penetration of an external magnetic field in a superconductor and ξ_0 is the correlation length at $T = 0$ K:

$$\xi_0 \sim \frac{\hbar V_F}{\Delta(0)} \,.$$

The Ginzburg–Landau theory assumes that the condensate of Cooper pairs moves as a whole and is described by a single stationary wave function $\Psi(\vec{r})$ even under a static external magnetic field. To calculate $\Psi(\vec{r})$, they proposed a relatively simple set of differential equations involving temperature as a parameter.

One of the theory's assumptions is of particular interest: the assumption that the current density in a superconductor under an external magnetic field is defined by the usual quantum-mechanical formula for a particle with the charge $-2|e|$ and mass $2m$:

$$\vec{j} = -\frac{|e|}{2m}\left[\Psi^*\left(-i\hbar\vec{\nabla} + \frac{2|e|}{c}\vec{A}\right)\Psi + \Psi\left(i\hbar\vec{\nabla} + \frac{2|e|}{c}\vec{A}\right)\Psi^*\right] \,, \tag{3.41}$$

where $\vec{A}$ is the vector potential to be introduced for solving one of the Maxwell equations.

Consider the case when the modulus of the wave function $\Psi(\vec{r})$ remains constant, but only the phase changes:

$$\Psi(\vec{r}) \approx |\Psi| \exp\left[i\varphi(\vec{r})\right] \,, \tag{3.42}$$

where $|\Psi| = \text{const}$. Then an expression for the current density in the superconductor is simplified and takes the form:

$$\vec{j} = -\left[\frac{2e^2}{mc}\vec{A} + \frac{|e|\hbar}{m}\vec{\nabla}\varphi\right] n_s \,, \tag{3.43}$$

where $n_s = |\Psi|^2$ is the density of superconducting electrons (with proper normalization of the wave function).

Gor'kov demonstrated that $|\Psi(\vec{r})| \sim \Delta(\vec{r})$ in terms of the microscopic theory. It follows that, generally, the energy gap in the spectrum of the superconductor is a function of coordinates. Within the Ginzburg–Landau equations, the density of superconducting electrons is thought to be linearly temperature dependent at $T < T_c$: $n_s = |\Psi|^2 \sim (T - T_c)$. In this case, the temperature dependencies of the energy gap and the correlation length have the form:

$$\Delta(T) \sim (T_c - T)^{\frac{1}{2}} \,, \quad \xi(T) = \hbar V_F/\Delta(T) \sim (T_c - T)^{-\frac{1}{2}} \,.$$

The quantity $\Delta(T)$ serves as the order parameter. Its temperature dependence shown in Figure 3.8 is typical of phase transitions of the second kind.

According to Maxwell's equation, in the stationary case the magnetic field and the superconductor current are related as follows:

$$\operatorname{rot}\vec{B} = \frac{4\pi}{c}\vec{j} \quad \text{(the CGS system)}$$

Let us take the curl of both sides of this equation. Given formula (3.43), with the identities

$$\operatorname{rot}(\operatorname{rot}\vec{B}) \equiv -\Delta\vec{B} + \vec{\nabla}\operatorname{div}B\,, \quad \operatorname{rot}\vec{\nabla}\varphi \equiv 0\,,$$

and the equalities

$$\operatorname{rot}\vec{A} = \vec{B}\,, \quad \operatorname{div}\vec{B} = 0\,,$$

we come to the closed equation:

$$\Delta\vec{B} = \delta^{-2}\vec{B}\,, \quad \delta = \sqrt{\frac{2mc^2}{4\pi n_s(2e)^2}}\,. \tag{3.44}$$

Equation (3.44) allows us to theoretically describe the picture of penetration of a weak magnetic field deep into a flat boundary superconductor.

Suppose a superconductor is in an external magnetic field $\vec{B} = (0, 0, B_0)$ and is occupying a half space $x \geq 0$ (Figure 3.9). The field inside the superconductor is described by the solution of the boundary problem:

$$\frac{\mathrm{d}^2B_z}{\mathrm{d}x^2} = \delta^{-2}B_z\,, \quad B_z(x=0) = B_0 \tag{3.45}$$

that has the form:

$$B_z(x) = B_0\exp(-x/\delta)\,. \tag{3.46}$$

This implies that the applied magnetic field penetrates into the superconductor only at the distance:

$$\delta = \sqrt{\frac{2mc^2}{4\pi n_s(T)(2e)^2}}\,. \tag{3.47}$$

At $T = 0$ K, we have $\delta(0) \sim (10^{-5} - 10^{-6})$ cm.

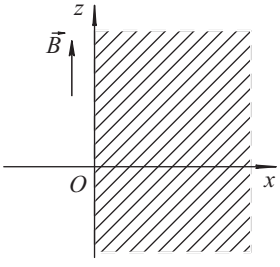


Fig. 3.9: Schematic representation of a superconductor in a magnetic field.

In the Ginzburg–Landau theory, the density of superconducting electrons is equal to $n_s(T) \sim (T_c - T)$, so the penetration depth amounts to $\delta(T) \sim (T_c - T)^{-1/2}$. Since $\delta(T) \to \infty$ as $T \to T_c$, then one of the conditions of applicability of the Ginzburg–Landau theory $\delta(T) \gg \xi_0$ holds for all superconductors near the phase transition temperature.

In 1935, the London brothers were the first to enter the quantity δ, without delving into the microscopic causes of superconductivity. Therefore, the parameter δ is called the *London penetration depth*.

3.7 Quantization of a Magnetic Flux

From the equation:

$$\vec{j} = -\left[\frac{2e^2}{mc}\vec{A} + \frac{|e|\hbar}{m}\vec{\nabla}\varphi\right] n_s \tag{3.48}$$

we can get another interesting consequence. Consider a ring shaped superconductor placed in a constant magnetic field directed along the axis of the ring (Figure 3.10).

Let us integrate expression (3.48) for the current density over an arbitrary closed contour C lying inside the ring. Since appreciable currents can flow only near the surface of the sample, then $\int_C \vec{j} \cdot d\vec{l} = 0$. This is equivalent to:

$$\frac{2e^2}{mc}\int_C \vec{A} \cdot d\vec{l} + \frac{|e|\hbar}{m}\int_C \underbrace{\vec{\nabla}\varphi \cdot d\vec{l}}_{d\varphi} = 0\,. \tag{3.49}$$

According to Stokes' theorem, we have:

$$\int_C \vec{A} \cdot d\vec{l} = \int_S \underbrace{\operatorname{rot}\vec{A}}_{\vec{B}} \cdot d\vec{S} = \int_{S_0} \vec{B} \cdot d\vec{S} = \Phi\,. \tag{3.50}$$

Because $\vec{B} = 0$ inside the superconductor, in formula (3.50), Φ is the magnetic flux through the hole in the ring; S is a surface based on the contour C; S_0 is a part of the surface S, covering the hole in the ring.

Now the wave function of the Cooper pairs, having the form

$$\Psi(\vec{r}) = \sqrt{n_s}\exp(i\varphi)$$

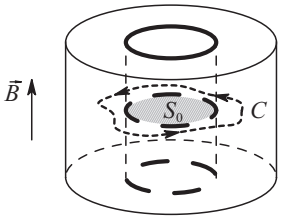


Fig. 3.10: A superconductor in the form of the ring being in an external constant magnetic field; a surface S_0 covering the hole in the ring.

should be clearly defined. In doing so its phase, when traversing along the contour C, must be incremented by $2\pi n$, where n is an integer. Emanating from this, we have:

$$\oint_C \mathrm{d}\varphi = 2\pi n \,. \tag{3.51}$$

According to (3.49)–(3.51), the magnetic flux through the hole in the ring must be quantized:

$$\Phi = -\frac{nhc}{2\,|e|} = -n\Phi_0 \,, \quad \Phi_0 = \frac{hc}{2\,|e|} = 2.07 \cdot 10^{-7}\ \mathrm{Gs/cm^2} \,. \tag{3.52}$$

The quantity Φ_0 is referred to as a *fluxon* or a *quantum of magnetic flux*.

Macroscopic quantization of the magnetic flux can be observed experimentally and serves as a convincing argument in favor of the Ginzburg–Landau theory.

3.8 The Microscopic Nature of Two Types of Superconductors – Vortex Lattices and Superconducting Magnets

Let us discuss the coexistence of normal and superconducting phases in a critical external magnetic field $H_c(T)$. The superconductor itself is assumed to be nonmagnetic. That is, $\vec{B} = \vec{H}$ in the CGS system.

Electron states in the superconductor are correlated at distances of the order $\xi(T) = hV_F/\Delta(T)$. That is why the parameter Δ cannot change sharply from the value Δ_0 in a superconductor to the value $\Delta = 0$ in a normal metal. There is a transition region of size $\sim\xi(T)$ where Δ is close to zero (Figure 3.11). This region has no magnetic field, so it is already in the normal phase.

In the normal phase, a unit volume of the substance must have the energy $H_c^2/8\pi$. In the absence of a magnetic field, the boundary layer of a thickness ξ possesses an excess energy:

$$\frac{H_c^2}{8\pi}\xi S \,,$$

where S is area of the boundary.

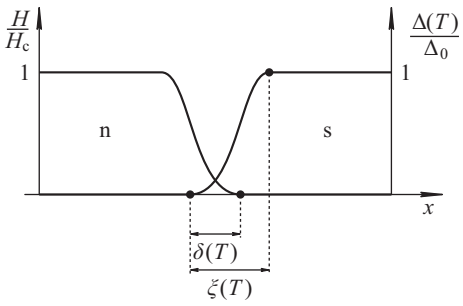


Fig. 3.11: Boundary between normal and superconducting phases of a conductor placed in a critical external magnetic field.

On the other hand, a magnetic field nevertheless penetrates into the superconductor to a depth of δ due to the Meissner effect. This reduces the boundary layer energy of the superconductor by the magnitude

$$\frac{H_c^2}{8\pi}\delta S\,.$$

As a result, the total energy per unit surface separating the normal and superconducting phases has the form:

$$\sigma_{\text{surf}} \sim \frac{H_c^2}{8\pi}(\xi - \delta)\,. \tag{3.53}$$

If $\xi > \delta$, then $\sigma_{\text{surf}} > 0$ and we have a type-I superconductor. If $\xi < \delta$, the surface energy is negative; it stands for a type-II superconductor. From a microscopic point of view, the existence of two types of superconductors is due to the difference in the δ and ξ parameters.

It should be noted that $\xi(T) \sim \delta(T) \sim (T_c - T)^{-1/2}$ as $T \to T_c$. This means that the ratio of the characteristic parameters near T_c does not depend on temperature:

$$\frac{\delta(T)}{\xi(T)} \sim O(1)\,.$$

Thus, temperature is not the cause of dividing superconductors into type-I and type-II.

A negative surface energy is energetically favorable for the magnetic field to penetrate deep into the superconductor, not abruptly but gradually due to the increase in the total number of magnetic flux filaments in the sample. Each filament carries a magnetic flux quantum $\Phi_0 = ch/(2|e|)$ (a fluxon of the same magnitude as that penetrates into the hole in the superconducting ring). The filament is surrounded by superconducting currents tending to shield the magnetic field inside the filament.

Ginzburg–Landau's approach proved to be perfect for describing type-II superconductors with vortex filaments. With the Ginzburg–Landau equations, Alexei Abrikosov, a Russian physicist, has succeeded in building the theory of type-II superconductors.

Near the center of the vortex filament in the plane perpendicular to it, the dependence of the wave function of a Bose condensate of Cooper pairs on the spatial coordinates is of the form:

$$\Psi \sim \rho \exp(\mathrm{i}\theta)\,, \tag{3.54}$$

where $\rho = \sqrt{x^2 + y^2}$ is the distance to the filament center, $\theta = \arg(x + \mathrm{i}y)$ (Figure 3.12).

It follows that the wave function vanishes in the points located in the vortex filament. The velocity of the vortex motion of superconducting pairs around the filament decreases in inverse proportion to the distance to the center of the filament:

$$V \sim \left|\frac{\Psi^*(-\mathrm{i}\hbar\vec{\nabla}\Psi)}{2m\,|\Psi|^2}\right| \sim \frac{\hbar}{2m\rho}\,. \tag{3.55}$$

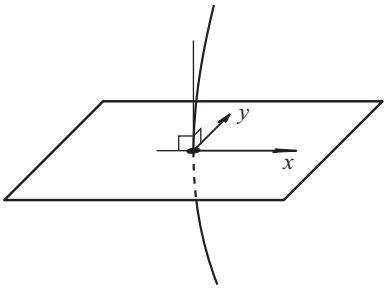


Fig. 3.12: A vortex filament and a plane perpendicular to it.

It can be seen that the velocity V_s tends towards infinity as $\rho \to 0$. Remarkably, in this case the superconducting current $\vec{j} \approx -(\hbar|e|/m)|\Psi|^2\vec{\nabla}\varphi$ is finite because $|\Psi|^2 \sim \rho^2$ and $|\vec{\nabla}\varphi| \sim 1/\rho$ as $\rho \to 0$.

The wave function of the condensate of Cooper pairs is restored to its equilibrium value at distances $\rho > \xi$.

Each filament tends to push off the other filament. Abrikosov showed that, in an isotropic crystal, the mutual repulsion of the filaments is minimal in a triangular lattice of filaments. Under the influence of the anisotropy of the crystal, it is possible, for example, to reconstruct the triangular vortex lattice in the square.

It is important that the balance between the δ and ξ parameters can always be changed by adding nonmagnetic impurities. Any type-I superconductor can be turned into a type-II superconductor by increasing the concentration of impurities. It is interesting, and important, that type-II superconductors have the highest critical fields and temperatures. It is worth noting that recently discovered high temperature superconductors, whose conduction mechanism is not yet clear, are also type-II superconductors. Thus, beyond a shadow of a doubt, type-II superconductors are one of the greatest discoveries of the second half of the twentieth century, and the leading role in the development of the theory was played by Abrikosov.

It should be kept in mind that, in practical use, the critical current flowing through type-II superconductors is limited by the vortex lattice motion. This is because when interacting with currents appearing around Abrikosov filaments, an external current loses its energy. However, there is a way to overcome this difficulty. In order to do so, vortex pinning centers need to be produced inside the superconductor, for example, by irregularities of the crystal. These centers form a potential relief for the vortex lattice and make it difficult for the latter to move. Nevertheless, at a finite temperature, there are always thermal fluctuation transitions of the vortex lattice, whereby so called *creep* arises. This phenomenon involves fluctuation jumps of the vortex lattice. Although the creep occurs at any temperature, fortunately the resistance caused by the creep is so low that the current through the superconductor does not change significantly for about a hundred years.

The creep creates another problem. If a superconductor has vortex bundle jumping, it heats up in this segment and can convert to the normal state, which, in turn, entails further heat release to destroy the superconducting wire or magnet. To prevent and avoid this, superconducting cables are stranded. Copper and superconducting wires alternate. Copper (normal metal) plays a role of an insulator. A superconductor has low thermal conductivity, but copper conducts heat away from the superconductor when the latter is suddenly heated. In addition, in the case when the superconducting current ceases to flow, the normal conductor connected in parallel acts as a bypass.

Note that electrons with opposite spins in a magnetic field have the energy difference $2\mu_B|B|$, where μ_B is the Bohr magneton. The magnetic field tends to rotate electron spins so as to make their directions and the magnetic field coincide. The destruction of a Cooper pair occurs when the difference in energies of electrons in the pair in an external magnetic field is of the order of the binding energy: $2\mu_B|B| \sim \Delta$, i.e., when $|B| \sim \Delta/2\mu_B$.

For the same reason, doping of magnetic impurities reduces superconductivity, and ferromagnetic ordering completely suppresses it. The coexistence of antiferromagnetic ordering and superconductivity is a debatable issue today. It should also be pointed out that, against ferromagnetic conductors, antiferromagnetic metals are extremely rare.

Abrikosov and Gor'kov showed that in contrast to conventional nonmagnetic impurities, an increase in the concentration of magnetic impurities gives rise to lowering of T_c and binding energy of some part of Cooper pairs. As a result, at a limited temperature range there is so called *gapless superconductivity*. Pairs with low binding energy prove to be broken, which eliminates the gap but, nevertheless, the remaining of Cooper pairs continue carrying the superconducting current.

There is another reason for the destruction of Cooper pairs by a magnetic field: momenta of electrons in a Cooper pair are oppositely oriented, so the electrons are affected by the oppositely directed Lorentz forces. This causes the Larmor twisting of the pairs (Figure 3.13).

Next we estimate the Larmor radius R of an electron in an external magnetic field within classical physics:

$$\frac{mV_F^2}{R} = \frac{V_F}{c}|eB| \quad \Rightarrow \quad R = \frac{mcV_F}{|eB|}.$$

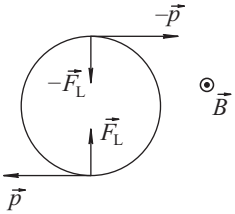


Fig. 3.13: Larmor twisting of a Cooper pair in a magnetic field with induction $\vec{B}$: $\vec{F}_L$ and $-\vec{F}_L$ are the Lorentz forces, $\vec{p}$ and $-\vec{p}$ are electron momenta.

In the calculation we have taken into account that the electron velocity is of the order of V_F.

A Cooper pair can exist as long as the Larmor radius exceeds its size:

$$R = \frac{mcV_F}{|eB|} > \xi \,. \tag{3.56}$$

The Larmor radius can be extended by increasing the effective mass of the charge carriers, which in turn causes the critical magnetic field value to grow.

In 2003, Abrikosov and Ginsburg became the Nobel Prize winners for the development of theory of superconductors.

3.9 Possible Physical Mechanisms of High Temperature Conductivity

The BCS theory contains prerequisites for the explanation of physical mechanisms of high temperature conductivity. This theory yields the formula

$$k_B T_c = 1.14\,\hbar\omega_D \exp\left(-\frac{4}{g\nu(\varepsilon_F)}\right) , \tag{3.57}$$

which indicates the possible directions for improving the critical temperature. Let us discuss some of them [4, 10].

I. In the BCS theory, the parameter $g\nu(\varepsilon_F)/4 < 1$. In other words, the electron-phonon interactions are assumed to be weak. The generalization of the theory to the case of strong electron-phonon coupling with $g\nu(\varepsilon_F)/4 \approx 1$ has been carried out by Eliashberg. It has been turned that the approximation has taken only thermal phonons with frequencies $\hbar\omega \sim k_B T_c$ into account to describe the electron-electron attraction, which is a rough estimate. It has been found that the electron-electron interaction is affected by virtual phonons with all frequencies, with the main role being played by phonons with frequencies lying near extrema of the function $\omega(\vec{q})$. The fact of the matter is the regions lying near the extrema correspond to the maximum phonon density (the *Van Hove singularities* in the phonon spectrum are important).

The theory of strong coupling for the critical superconducting transition temperature brings the formula to:

$$k_B T_c = \hbar\widetilde{\omega} \exp\left(-1/g(\widetilde{\omega})\right) \,. \tag{3.58}$$

It differs from the formula derived in the BCS theory, in that:

a) $\widetilde{\omega}$ is not the Debye frequency but a characteristic frequency. Its calculation requires knowledge of the entire vibrational spectrum.
b) The coupling constant $g(\widetilde{\omega})$ depends strongly on frequency: $g(\widetilde{\omega}) \sim \widetilde{\omega}^{-2}$. This means that a high temperature T_c corresponds to low frequencies.

Eliashberg's theory provides two conclusions:

1. In the normal state, strong electron-phonon coupling corresponds to great resistance. Consequently, superconductors with high temperatures T_c in the normal state must conduct an electric current poorly.
2. There is a probability of finding superconductors with high T_c among "soft"-phonon spectrum structures whose $\widetilde{\omega}$ are low.

However, although decreasing $\widetilde{\omega}$ corresponds to increasing the exponential factor in (3.58), at the same time, it leads to a reduction of the coefficient $\hbar\widetilde{\omega}$ before the exponent. Besides, the lattices with low $\widetilde{\omega}$ are structurally unstable. Ultimately, the Eliashberg theory gives little hope of improvement of T_c above 30 K.

II. Let us discuss the possibility of increasing T_c due to the pre-exponential factor. Since $\omega_D \sim 1/\sqrt{M}$, high T_c can be obtained by reducing the mass of the lattice atoms.

Solid hydrogen, if it were metal, could be a unique opportunity in this regard. Unfortunately, under normal conditions, hydrogen is a dielectric crystal of H_2 molecules. According to theoretical calculations, molecular hydrogen must pass to the atomic phase under high pressure. The phase bears similarities to the phase for alkali metals. However, the transition pressure is extremely high and amounts to 2.5Mbar. The question arises of whether this phase remains in a metastable state after the removal of the pressure.

At first, these difficulties were overcome by producing *hydrides*, i.e., metal-hydrogen compounds to generate high frequency branches in the phonon spectrum. These frequencies had to correspond to proton oscillations. Getting metastable solid solutions with a high concentration of hydrogen became possible by means of special methods. However, the substances obtained had $T_c < 10$ K.

Abrikosov suggested another idea which would give the same advantages as solid hydrogen did. Imagine a substance containing an equal number of electrons and holes. Suppose the mass of the hole was much greater than the mass of the electrons, and that the potential energy of the former exceeds the kinetic one. The holes can then form a periodic structure, i.e., "a lattice in a lattice." The substance resembles solid hydrogen, but protons are replaced by heavy holes. In this medium, phonons propagate with the Debye frequency ten times greater than the Debye frequency $\hbar\widetilde{\omega}_D$ for solid hydrogen. To produce high T_c, it is also necessary that such substances should have a high charge carrier density but the dielectric constant is small. Unfortunately, at the present time, such substances are not synthesized, and it is unclear whether it is possible to create them.

III. Are there other more powerful interaction mechanisms which can lead to electron pairing and superconductivity phenomenon other than the interaction through a vibrating crystal lattice? In principle, there are.

The overall scheme of interaction between electrons and a transmission medium A can be represented as follows:

$$e_1 + A \rightarrow e_1' + A^*, e_2 + A^* \rightarrow e_2' + A \,, \tag{3.59}$$

where e_i corresponds to the electron ($i = 1, 2$); A is the ground state and A^* is an excited state of the transmission system. Such "a double response" results in returning the system A to its original state, and the electrons e_1 and e_2 change their momenta (that is, as if they scatter by each other). It can be shown that such an interaction is necessarily the attraction. If the interaction is not too great, it yields the following formula for calculating the T_c:

$$k_B T_c \sim \Delta E \exp(-1/\lambda) \,, \tag{3.60}$$

where ΔE is the energy difference for the states A and A^*, and λ depends on the interaction between the electrons and the system A. In fact, the above mechanism bears similarity to the phonon attraction.

In this case, the most promising mechanism is thought to be the so called exciton mechanism. An *exciton* in a molecular crystal can be regarded as electronic excitation of one of the molecules propagating through the lattice due to the presence of translational symmetry (the Frenkel exciton). In a semiconductor, an exciton is a bound state of an electron and a hole that resembles a hydrogen atom. The bound state is free to migrate through the lattice (the Wannier–Mott exciton). The radius of such a quasi-atomic formation can be as much as tens of times larger than the lattice constant. It should be pointed out that the Wannier–Mott excitons, being bound states of an electron and a hole, may be at different lattice sites. At the same time, the Frenkel excitons can be interpreted as the Wannier–Mott limit: an electron and a hole reside at one and the same site.

The exciton energy $\Delta E = k_B \theta_{\text{exiton}}$ corresponds to a temperature of about $\theta_{\text{exiton}} \sim 10^4$ K. Formula (3.60) gives an estimation of T_c even for $\lambda = 1/4$: $T_c \sim 300$ K.

However, the exciton mechanism is extremely difficult to implement in practice. The fact is that the excitons exist in molecular crystals, semiconductors or dielectrics due to these being poor conductors of electric current under normal conditions. Superconductivity, at least before the discovery of high temperature superconductors, was observed experimentally in materials with metallic conductivity. Therefore, the electron-electron attraction through the exciton exchange was expected to be brought about in a quantum system consisting of two subsystems – a metal and a dielectric (semiconductor) one. In 1964, Ginzburg came up with an idea of producing heterogeneous structures of alternating metal and dielectric films. Due to the quantum tunneling effect, electrons of the metal would go inside the dielectric to be exchanged by excitons. This would lead to an effective attraction between the electrons and, ul-

timately, to the phenomenon of high temperature superconductivity in the metallic system.

The limiting case of the heterogeneous structures is quasi-two-dimensional superconductors with one single molecule thick conductive layer, separated by relatively thick nonconductive space. Generally, superconductivity in both two-dimensional and one-dimensional lattices is impossible to exist due to its destruction by thermal fluctuations. However, it turns out that even a slight interaction between the conductive layers or chains suppresses the fluctuations and restores superconductivity. In addition, low dimensionality of the system stabilizes the exciton states, making them more stable under a broader range of temperatures and external electric fields. Despite not reaching high T_c in "sandwiches," layered and filamental structures, these materials and the mechanism itself should be still deemed as promising.

3.10 High Temperature Superconductors

High temperature superconductivity was found where it seemed absurd to look for it – in metal oxide ceramics, most of which are known to be good insulators at room temperatures [8, 10, 12].

Let us discuss the properties of compounds of the type $La_{2-x}(Ba, Sr)_xCuO_4$ with $T_c \sim 40\,K$. It is these that were the first materials where high temperature superconductivity was discovered. The optimum composition corresponds to the order of $x \approx 0.2$ of Ba or Sr impurities. Here, x indicates the proportion of Ba (or Sr) atoms, which replaced La in some lattice sites. All ions in such compounds form a crystal lattice through ionic-covalent bonds rather than metallic one. Taking into account the valence state, a chemical formula of $La_{2-x}Ba_xCuO_4$ should be written in the form $La^{3+}_{2-x}Ba^{2+}_x Cu^{3+}_x Cu^{2+}_{1-x} O^{2-}_4$. It is worth pointing out that the copper ions appear in two different valence states. There is an approach that links high temperature superconductivity with this fact.

The unit cell of the compound is a right prism with a square base. Such a crystal system is called a *tetragonal crystal system*. Its structure is shown in Figure 3.14. The structure can be represented as flat layers of oxygen octahedra, intergrown together through common oxygen ions at the position I (Figure 3.15). Neighboring layers of the octahedra are separated by two atomic planes containing La atoms. In the La planes, La is partially substituted for Ba atoms. Furthermore, each flat layer of the oxygen octahedra is shifted relative to the adjacent plane layer. The vertices of the pyramids of the upper layer face the pits formed by the pyramids of the lower layer (Figure 3.15).

It turns out that it is the layers of oxygen octahedra with Cu-O atoms which are responsible for the superconducting properties of the structure. The conductivity of the layer is explained by the overlap of the wave functions of the Cu and O atoms. The Cu and O ions are coupled via strong Coulomb interaction.

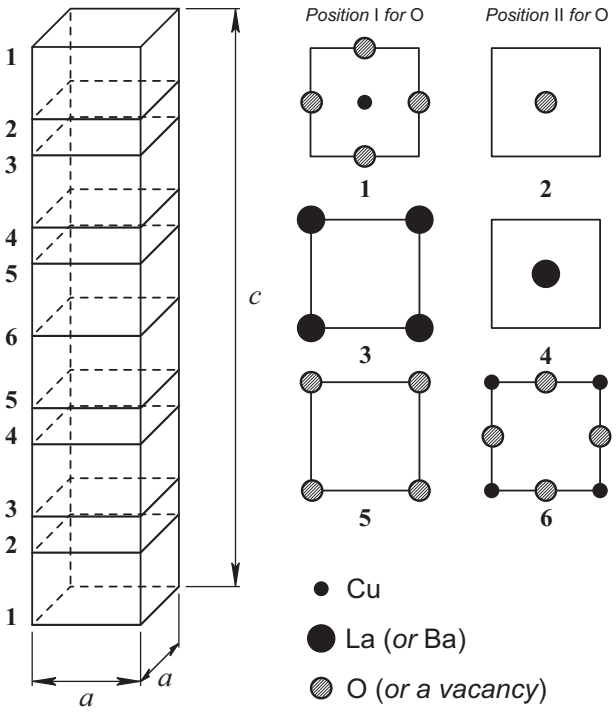


Fig. 3.14: The unit cell of $La_{2-x}Ba_xCuO_4$ is a rectangular parallelepiped with a height of $c = 13.25$ Å and a square base with side $a = 3.78$ Å. On the right cross sections of the parallelepiped and the arrangement of atoms. On the left from the parallelepiped – the number of the cross sections.

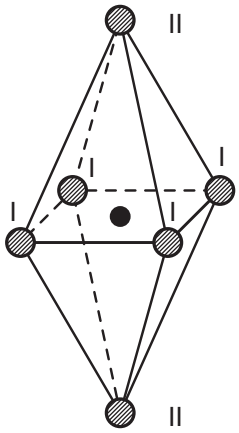


Fig. 3.15: Each Cu atom is the center of an octahedron elongated along the axis c; two oxygen atoms at position II and four oxygen atoms at position I share its vertices.

Let us focus on the layered, quasi-two-dimensional and anisotropic nature of the structure: the conductive layers are separated by nonconductive planes where atoms of La are partially and randomly replaced by Ba atoms.

The band structure calculation in the case of $x = 0$ shows that the La_2CuO_4 compound has an exactly half filled conduction band. The Fermi surface is almost like a square cross section pipe with an axis parallel to the axis c (in the extended zone scheme). Thus, the electronic structure of the crystal is highly anisotropic. Such an electronic structure makes the metal semiconductor phase transition energetically favorable (the *Peierls transition*). A soft mode $\omega(\vec{q}) \approx 0$ with $\vec{q} \neq 0$ appears in the spectrum of lattice vibrations. As a result, the *lattice symmetry* reduces to *rhombic* and the unit cell converts into a cuboid with unequal edge lengths. The changes give rise to the appearance of a gap in the conduction band. The lower edge of the gap coincides with the Fermi surface of electrons. The phase transition forms two zones: one is filled with electrons, and the other, separated by the gap, is completely free. The crystal becomes a semiconductor. A phase transition of the type of a metal/an insulator in La_2CuO_4 is indeed observed. However, even partly replacing trivalent La by divalent Ba lowers the Fermi level and violates conditions for realizing the transition. Then the lattice stabilizes in the tetragonal metallic phase. It occurs due to the decrease of the total number of electrons. The stabilization takes place even when $x = 0.07$ and eliminates the soft mode and, thus, weakens the impact of the electrons on the lattice vibrations. But there are physical reasons for the electron-phonon interaction to still be strong enough to be taken into account when discussing the possible mechanisms of electron pairing in high temperature superconductors. Let us list the main reasons leading to a large electron-phonon coupling constant.

First, a layered crystal structure of high temperature *cuprates* has a sufficiently high density of electron states at the Fermi surface, despite a very small number of electrons per unit cell. Secondly, the major role in electron pairing is assumed to be played by intense (because of small mass) oscillations of the oxygen octahedra centered on the copper atoms. Besides, strong hybridization of the wave functions of the Cu and O atoms in the Cu-O plane allows even the electrons bonded to a light oxygen atom to participate in the electron-phonon interaction.

Furthermore, sufficiently weak screening directed perpendicularly to the Cu-O planes results in a significant fraction of the ionic bond in the cuprates and strong electron-phonon effects on these planes. This is because the potential energy of the Cu and O atoms changes as the O ions in the position II and even La and Ba ions vibrate.

Next we proceed to discussing other superconducting ceramics, such as $ABa_2Cu_3O_{7-x}$, where A can be any element, beginning with yttrium and ending with heavier, rare earth trivalent elements:

$$A = Y, La, Nd, Sm, Eu, Gd, Ho, Er, Lu\ .$$

For all of these compounds T_c is 90–98 K, which is higher than that of compounds considered. Their lattice is orthorhombic but close to tetragonal. For $A = Y$, we have $c = 11.70\,\text{Å}$, $a = 3.83\,\text{Å}$, $b = 3.89\,\text{Å}$ (the unit cell is a cuboid with uneven edge lengths). These compounds are also layered anisotropic materials with flat sequences of oxygen octahedra surrounding copper atoms. Conductive Cu-O layers in such com-

pounds are more separated from each other (not two but three nonconductive atomic planes). Perhaps it is this difference in structure that explains the almost threefold increase in their T_c.

These compounds possess a small isotopic effect: $T_c \sim 1\sqrt{M}$, where M is the mass of the ^{16}O and ^{18}O oxygen isotopes.

The measurement of the value of magnetic flux quanta passing through the hole in a superconducting ring shows that the superconducting current in the ceramics is carried by Cooper pairs with a charge $-2|e|$. It has been found that the effective mass of electrons in ceramics, as measured by the London penetration depth, is unusually large; about a hundred electron masses.

At the same time, estimation of the Debye temperature yields normal values of $T_D \sim (20–400)\,K$. It can be concluded that the BCS theory with weak electron-phonon coupling cannot explain such a high T_c. The standard approach to strong electron-phonon coupling is based on the idea of isotropic (singlet) electron pairing and on the Eliashberg equations. This way also does not explain much of the superconducting properties, including the *d*-type symmetry of the energy gap. There must be something else that defines, together with the electron-phonon interaction, a complete mechanism of high temperature superconductivity.

To date, there are numerous schemes of electron pairing. Various systems of quasiparticle electron excitations are most often discussed. Only instead of phonons and excitons leading to electron-electron attraction and electron pairing, other carriers of the interaction can be considered. This may be: a) antiferromagnetic spin fluctuations (in the Cu-O planes, spins 1/2 of the copper atoms are antiferromagnetic ordered state), b) specifics of the band structure, so called "*nesting*", etc. *Nesting* is the coincidence of individual segments of a circuit bounding the cross section of the cylindrical Fermi surface of cuprates after being displaced by a certain wave vector. So Gor'kov put forward a hypothesis of forming Bose particles – bipolarons in the crystal even in the normal phase at $T > T_c$. Bipolarons are bound states of two heavy Fermi quasiparticles (polarons).

A *polaron* is an electron placed in a potential well formed by electron induced polarization of the crystal lattice. A *bipolaron* is a pair of two electrons surrounding the lattice ions polarized by the pair.

Local bipolarons of small spatial extension may appear due to strong electron-phonon interaction. Superfluidity of charged Bose particle bipolarons is expected to occur at low temperatures. This means superconductivity. Such superconductivity differs principally from the BCS superconductivity, but namely the bipolaron destruction temperature is much higher than their condensation temperature. That is to say, against Cooper pairs, bipolarons "live" at temperatures above T_c as well. In this case, a bipolaron gas can be treated as a nearly ideal Bose gas. Its condensation temperature ($m^* = 100\,m$, where m is the mass of a free electron):

$$T_0 = \frac{3.31\hbar^2}{m^* k_B}\left(\frac{n_s}{2}\right)^{2/3} \sim 100\,K$$

must be identified with a superconducting transition temperature. As can be seen, the Gor'kov point of view provides, at least, the correct order of magnitude of T_c.

The copper-oxygen planes of cuprate compounds build up a strongly correlated two-dimensional electron system. In this system, the main role is played by a single electron band, and Coulomb repulsion of electrons at one and the same site is dominant. In cuprates, kinetic and potential energies of interaction between particles have the same order of magnitude. Therefore, there is no small parameter for consistent theoretical description of such systems. The absence of a smallness parameter limits the possibility to use the simplest approximations of strongly or weakly interacting electrons to describe even the dielectric phase. Such a phase is observed in the lack of substitutional impurities (the terminology commonly used is "*in the absence of doping*"). It is interesting and important that the superconductivity mechanisms governed by direct electron correlations can be realized in cuprates. Most theoretical studies in this direction deal with the Hubbard model. The latter takes into account both the electron transfer from a site to a site in a crystal lattice and strong Coulomb repulsion of two electrons with opposite spins at a single site. It is the Hubbard model and its related models have helped to propose the two most "exotic" ideas concerning the nature of high temperature superconductivity.

According to the first of them, in cuprates, an electron having a charge and spin ceases to be a well-defined elementary excitation. It is assumed that the electronic system of cuprates, both in the normal and in the superconducting states, is theoretically described in terms of new weakly interacting quasiparticles – "holons" and "spinons." Holons are spinless collective excitations of cuprates; they carry a charge. Spinons are noncharged Fermi quasiparticles carrying spin. A weakly nonideal gas of spinons and holons is called the *Luttinger liquid*. Simultaneous Bose condensation of holons and spinons (analog of Cooper pairs) corresponds to the superconducting state.

The second idea constitutes a hypothesis proposed by Robert Betts Laughlin. His statement says that collective excitations in high temperature systems are neither bosons nor fermions and are described by special quasiparticles, "anions," that obey fractional statistics.

Both ideas are quite radical and constantly undergo modification. Their legitimacy is the subject of lively debates due to the absence of any well-established findings and conclusions in the research.

The Hubbard model offers insight into other possible occurrences of superconductivity in a strongly correlated two-dimensional electron-electron repulsion system. It has been found that although the interaction between two electrons of the copper ion is great and repulsive, the interaction between electrons of the copper and electrons of the nearest neighboring oxygen ions turns out to be attractive and can lead to the superconducting state with anisotropic *d*-electron pairing. The emergence of the anisotropic energy gap is in qualitative agreement with experimental data.

Interestingly, if one leaves aside the complicating factors associated with the specifics of high temperature samples studied, their magnetic properties are fully described in terms of the Ginzburg–Landau theory for type-II superconductors.

Currently there is no common point of view about the nature of high temperature superconductivity. This is explained by the fact that the oxide compounds are complex: they are composed of a huge number of the lattice atoms whose location significantly affects the physical properties of the compounds. Moreover, the metal oxides are unsuitable for producing single crystals because their melting point lies in the temperature range (~1000 °C), where chemical degradation of the compounds begins to take place. There are difficulties in controlling the degree of doping and the homogeneity of the samples, etc.

The discovery of high temperature superconducting compounds has helped to provide a powerful impetus for the development of fundamental research in condensed matter physics. As a result of these investigations, not only does new knowledge originate but novel and unique superconductors with high critical parameters appear. As long as the high temperature materials are technologically difficult to apply, their use is very limited but some progress is already evident.

4 Quantum Coherent Optics: Interaction of Radiation with Matter

4.1 Maxwell's Equations and Natural Oscillations of an Electromagnetic Field in a Closed Cavity

In classical physics, electromagnetic phenomena in matter are described by Maxwell's equations. The electromagnetism equations are written differently in the CGS and SI systems. Earlier, we have used the CGS system when considering electromagnetic phenomena. In what follows, we present Maxwell's equations as applied to optical phenomena in the SI system of units. This is because all the recommended literature widely uses the SI system. The conversion from the CGS to the SI system does not meet any difficulties. We also recall the form of the electromagnetism laws in the CGS and the SI systems of units.

We begin with Maxwell's equations in terms of true physical fields $\vec{B}$ and $\vec{E}$. The latter are average values over physically small periods of time and macroscopic medium volumes of microscopic fields. We call the $\vec{B}$ and $\vec{E}$ fields true because they define force $\vec{F}$ as acting on a charge q from the direction of the electromagnetic field in the medium.

SI		**CGS**	
Lorentz Force $\vec{F}$			
$\vec{F} = q\vec{E} + q\left[\vec{V} \times \vec{B}\right]$	(4.1a)	$\vec{F} = q\vec{E} + \frac{q}{c}\left[\vec{V} \times \vec{B}\right]$	(4.1a')
Maxwell's Equations			
$\mathrm{rot}\,\vec{E} = -\frac{\partial \vec{B}}{\partial t}$	(4.2a)	$\mathrm{rot}\,\vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t}$	(4.2a')
$\mathrm{rot}\,\vec{B} = \mu_0\left\{\vec{j}_{\mathrm{ext}} + \vec{j}_{\mathrm{pol}} + \vec{j}_{\mathrm{mol}} + \varepsilon_0\frac{\partial \vec{E}}{\partial t}\right\}$	(4.2b)	$\mathrm{rot}\,\vec{B} = \frac{4\pi}{c}\left\{\vec{j}_{\mathrm{ext}} + \vec{j}_{\mathrm{pol}} + \vec{j}_{\mathrm{mol}}\right\} + \frac{1}{c}\frac{\partial \vec{E}}{\partial t}$	(4.2b')
$\mathrm{div}\,\vec{E} = \frac{1}{\varepsilon_0}\left\{\rho_{\mathrm{ext}} + \rho_{\mathrm{pol}}\right\}$	(4.2c)	$\mathrm{div}\,\vec{E} = 4\pi\left\{\rho_{\mathrm{ext}} + \rho_{\mathrm{pol}}\right\}$	(4.2c')
$\mathrm{div}\,\vec{B} = 0$	(4.2d)	$\mathrm{div}\,\vec{B} = 0$	(4.2d')

Here, ε_0 and μ_0 are the electric and magnetic constants, respectively; $\varepsilon_0\mu_0 = 1/c^2$, where c is the speed of light in a vacuum.

Medium properties are characterized by the electric polarization vector $\vec{P}$ (dipole moment per unit volume of medium) and the magnetization vector $\vec{M}$ (magnetic moment per unit volume of medium). To simplify Maxwell's equations, two auxiliary vector fields $\vec{H}$ and $\vec{D}$ are defined. They are related to the $\vec{M}$ and $\vec{P}$ vectors, and the $\vec{B}$ and $\vec{E}$ physical fields, as follows.

https://doi.org/10.1515/9783110586183-004

SI		CGS	
Electric Displacement Vector $\vec{D}$			
$\vec{D} = \varepsilon_0 \vec{E} + \vec{P}$	(4.3a)	$\vec{D} = \vec{E} + 4\pi\vec{P}$	(4.3a')
Magnetic Field Strength $\vec{H}$			
$\vec{H} = \frac{1}{\mu_0}\vec{B} - \vec{M}$	(4.4a)	$\vec{H} = \vec{B} - 4\pi\vec{M}$	(4.4a')

$\vec{M} = 0$ in vacuum and nonmagnetic media. As a result, the $\vec{B}$ and $\vec{H}$ fields coincide in the CGS, system but differ by the factor μ_0 in the SI system.

SI		CGS	
$\vec{H} = \frac{1}{\mu_0}\vec{B}$	(4.5a)	$\vec{H} = \vec{B}$	(4.5a')

Therefore, we can say that the intensity of the magnetic field $\vec{H}$ characterizes the magnetic field without considering the influence of a medium.

With the $\vec{P}$ and $\vec{M}$ vectors, the polarization charge density ρ_{pol}, the density of polarization currents $\vec{j}_{\text{pol}}$, and the density of the so called molecular currents $\vec{j}_{\text{mol}}$ can be calculated in both of the systems of units:

SI		CGS	
$\rho_{\text{pol}} = -\operatorname{div}\vec{P}$	(4.6a)	$\rho_{\text{pol}} = -\operatorname{div}\vec{P}$	(4.6a')
$\vec{j}_{\text{pol}} = \frac{\partial \vec{P}}{\partial t}$	(4.6b)	$\vec{j}_{\text{pol}} = \frac{\partial \vec{P}}{\partial t}$	(4.6b')
$\vec{j}_{\text{mol}} = \operatorname{rot}\vec{M}$	(4.6c)	$\vec{j}_{\text{mol}} = c \cdot \operatorname{rot}\vec{M}$	(4.6c')

Then the system of Maxwell's equations for the true fields $\vec{E}$ and $\vec{B}$ and (4.2) can be rewritten in another form:

SI		CGS	
$\operatorname{rot}\vec{E} = -\frac{\partial \vec{B}}{\partial t}$	(4.7a)	$\operatorname{rot}\vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t}$	(4.7a')
$\operatorname{rot}\vec{H} = \vec{j}_{\text{ext}} + \frac{\partial \vec{D}}{\partial t}$	(4.7b)	$\operatorname{rot}\vec{H} = \frac{4\pi}{c}\vec{j}_{\text{ext}} + \frac{1}{c}\frac{\partial \vec{D}}{\partial t}$	(4.7b')
$\operatorname{div}\vec{D} = \rho_{\text{ext}}$	(4.7c)	$\operatorname{div}\vec{D} = 4\pi\rho_{\text{ext}}$	(4.7c')
$\operatorname{div}\vec{B} = 0$	(4.7d)	$\operatorname{div}\vec{B} = 0.$	(4.7d')

As a consequence, these transformations of the Maxwell equations in the SI system eliminate the fundamental constants ε_0 and μ_0. Now the Maxwell equations (in both

systems) include only the external charge density ρ_{ext} and the density of external currents $\vec{j}_{\text{ext}}$.

The system of equations (4.6) and (4.7) is common but is not closed. It must be supplemented by the so called constitutive equations as expressing $\vec{P}$ and $\vec{M}$ via their inducing fields $\vec{E}$ and $\vec{B}$. When $\vec{E}$ and $\vec{B}$ slowly change in time and space and are relatively small, the constitutive equations are given by:

$$P_i = \alpha_{ij} E_j \,, \quad M_i = \chi_{ij} H_j \,, \tag{4.8}$$

where α_{ij} is the tensor of the dielectric susceptibility of a medium, χ_{ij} is the tensor of the magnetic susceptibility of a medium; $i, j = 1, 2, 3$. The indices repeated twice imply summation.

Equation (4.8) allows one to eliminate $\vec{P}$ and $\vec{M}$ from equations (4.3a) and (4.4a):

$$D_i = \varepsilon_{ij} E_j \,, \quad B_i = \mu_{ij} H_j \,, \tag{4.9}$$

Where $\varepsilon_{ij} = \varepsilon_0 \delta_{ij} + \alpha_{ij}$ is the tensor of the dielectric permeability of a medium and $\mu_{ij} = \mu_0(\delta_{ij} + \chi_{ij})$ is the tensor of the magnetic permeability of a medium.

Generally, $\vec{D}$ and $\vec{B}$ are more complex functions of $\vec{E}$ and $\vec{H}$. The particular form of the constitutive equations to be established is beyond the frame of electrodynamics. The issue refers to the microscopic quantum theory that takes the atomic molecular structure of matter into account.

To begin with, we have a look at Maxwell's equations for charge and current free space (in a vacuum): $\rho_{\text{ext}} = 0, \vec{j}_{\text{ext}} = 0$, $\vec{M} = \vec{P} = 0$ ($\chi_{ij} = \alpha_{ij} = 0$). In this case, Maxwell's equation has the following closed form:

$$\operatorname{rot} \vec{E} = -\mu_0 \frac{\partial \vec{H}}{\partial t} \,, \tag{4.10a}$$

$$\operatorname{rot} \vec{H} = \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \,, \tag{4.10b}$$

$$\operatorname{div} \vec{E} = 0 \,, \tag{4.10c}$$

$$\operatorname{div} \vec{H} = 0 \,. \tag{4.10d}$$

In a vacuum:

$$\vec{B} = \mu_0 \vec{H} \,. \tag{4.11}$$

Acting by the curl operator on both sides of equation (4.10a) and using equation (4.10b), we find:

$$\operatorname{rot} \operatorname{rot} \vec{E} = -\mu_0 \frac{\partial \operatorname{rot} \vec{H}}{\partial t} = -\mu_0 \varepsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} \,. \tag{4.12}$$

Given that the speed of light in a vacuum is $c = 1/\sqrt{\varepsilon_0 \mu_0}$, and taking into account the equation (4.10c) and the identity:

$$\operatorname{rot} \operatorname{rot} \vec{E} = \vec{\nabla} \operatorname{div} \vec{E} - \Delta \vec{E} \,,$$

we arrive at a closed wave equation for the vector $\vec{E}$:

$$\Delta\vec{E} = \frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2}\,. \tag{4.13}$$

It is easy to verify that the field $\vec{H}$ satisfies the same equation:

$$\Delta\vec{H} = \frac{1}{c^2}\frac{\partial^2\vec{H}}{\partial t^2}\,. \tag{4.14}$$

Possible solutions of the wave equations (4.13) and (4.14) are running flat monochromatic waves:

$$\vec{E}(\vec{r},t) = \vec{E}_0\exp\left[\mathrm{i}\vec{k}\cdot\vec{r} - \mathrm{i}\omega t\right] + \text{c.c.}\,, \quad \vec{H}(\vec{r},t) = \vec{H}_0\exp\left[\mathrm{i}\vec{k}\cdot\vec{r} - \mathrm{i}\omega t\right] + \text{c.c.}\,, \tag{4.15}$$

where $\vec{E}_0$, $\vec{H}_0$ are constant complex vectors. Here and below, the symbol "c.c." denotes a "complex conjugate" term. The wave vector $\vec{k}$ can be arbitrarily oriented in space, and its modulus is related with the frequency ω: $\omega = kc$.

If one inserts (4.15) into the Maxwell equations (4.10), we obtain:

$$\left[\vec{k}\times\vec{E}_0\right] = \omega\mu_0\vec{H}_0\,, \quad \left[\vec{k}\times\vec{H}_0\right] = -\omega\varepsilon_0\vec{E}\,, \quad \vec{k}\cdot\vec{E}_0 = \vec{k}\cdot\vec{H}_0 = 0\,. \tag{4.16}$$

From (4.16), it is easy to establish that the vectors $\vec{E}_0$, $\vec{H}_0$, $\vec{k}$ form a right-handed triple of orthogonal vectors (Figure 4.1), and modules of the vectors $\vec{E}_0$ and $\vec{H}_0$ are related by:

$$H_0 = \sqrt{\frac{\varepsilon_0}{\mu_0}}E_0\,. \tag{4.17}$$

The position of the vector $\vec{E}$ in a plane perpendicular to the vector $\vec{k}$ can be characterized by two independent polarization directions, such as two real orthogonal unit vectors $\vec{e}_{\vec{k}\lambda}$ ($\lambda = 1, 2$):

$$\vec{e}_{\vec{k}\lambda}\cdot\vec{e}_{\vec{k}\lambda'} = \delta_{\lambda\lambda'}\,. \tag{4.18}$$

For further analysis, the vectors $\vec{e}_{\vec{k}1}$, $\vec{e}_{\vec{k}2}$, $\vec{n}_{\vec{k}} = \vec{k}/|\vec{k}|$ are convenient to choose as they should form a right-handed triple (Figure 4.2). This allows the following equalities to hold true:

$$\left[\vec{e}_{\vec{k}1}\times\vec{e}_{\vec{k}2}\right] = \vec{n}_{\vec{k}}\,, \quad \vec{n}_{\vec{k}}\cdot\vec{e}_{\vec{k}\lambda} = 0\,. \tag{4.19}$$

Given these remarks, we can write equation (4.15) for the linearly polarized flat monochromatic waves $\vec{E}$ and $\vec{H}$ in a particular form:

$$\begin{aligned}
\vec{E}_{\vec{k}\lambda} &= \mathrm{i}\vec{e}_{\vec{k}\lambda}\sqrt{\frac{\hbar\omega_k}{2V\varepsilon_0}}\left\{a_{\vec{k}\lambda}\exp\left[\mathrm{i}\vec{k}\cdot\vec{r} - i\omega_k t\right] - \text{c.c.}\right\}\,,\\
\vec{H}_{\vec{k}\lambda} &= \mathrm{i}\left[\vec{n}_{\vec{k}}\times\vec{e}_{\vec{k}\lambda}\right]\sqrt{\frac{\hbar\omega_k}{2V\mu_0}}\left\{a_{\vec{k}\lambda}\exp\left[\mathrm{i}\vec{k}\cdot\vec{r} - i\omega_k t\right] - \text{c.c.}\right\}\,.
\end{aligned} \tag{4.20}$$

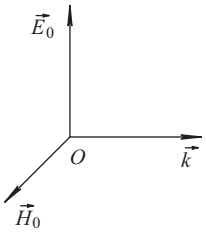


Fig. 4.1: Right-handed triple of orthogonal vectors $\vec{E}_0$, $\vec{H}_0$, $\vec{k}$.

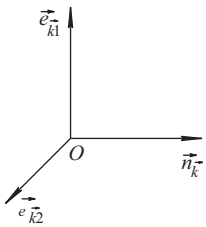


Fig. 4.2: Right-handed orthogonal vectors $\vec{e}_{\vec{k}1}$, $\vec{e}_{\vec{k}2}$, $\vec{n}_{\vec{k}}$.

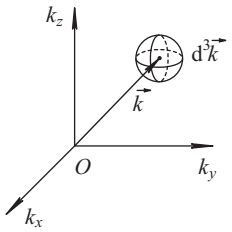


Fig. 4.3: Volume element $\mathrm{d}^3\vec{k}$ in reciprocal space.

Here, $a_{\vec{k}\lambda}$ are arbitrary complex numbers (dimensionless amplitudes) and V is the volume of the region containing the electromagnetic field; $\omega_k = kc$.

A cavity contains the field if it is large enough. The properties of the electromagnetic field inside the cavity does not depend on its size, shape, and nature of the walls. To significantly simplify the calculations, we choose the cavity in the form of a cube:

$$0 \le x, y, z \le L \,, \tag{4.21}$$

where L is the length of the cube edge, and takes the Born–Karman boundary conditions:

$$\begin{gathered}\vec{E}(x + L, y, z) = \vec{E}(x, y, z) \,, \quad \vec{E}(x, y + L, z) = \vec{E}(x, y, z) \,, \\ \vec{E}(x, y, z + L) = \vec{E}(x, y, z) \,.\end{gathered} \tag{4.22}$$

These boundary conditions isolate discrete values of the wave vector components:

$$k_i = \frac{2\pi}{L} n_i \,, \tag{4.23}$$

where n_i are arbitrary integers ($i = x, y, z$). The set of numbers (n_x, n_y, n_z) defines the mode of the electromagnetic field.

For large L, the allowed values of the vector $\vec{k}$ are quasicontinuously distributed and form a lattice in reciprocal space. Each lattice site occupies the volume $(2\pi)^3/L$. The small volume $\mathrm{d}^3\vec{k}$ centered at the point with the wave vector $\vec{k}$ (Figure 4.3) contains

$$\mathrm{d}^3\vec{k} \colon \frac{(2\pi)^3}{L^3} = \frac{V\mathrm{d}^3\vec{k}}{(2\pi)^3} \tag{4.24}$$

allowed wave vectors ($V = L^3$ is the cavity volume where the electromagnetic field is).

Whatever $\vec{k}$ is, there are two independent electromagnetic waves differing in polarization directions, $\vec{e}_{\vec{k}\lambda}$ ($\lambda = 1, 2$). Therefore, the total number of electromagnetic modes with wave vectors lying near a point with the radius vector $\vec{k}$ within the volume $\mathrm{d}^3\vec{k}$ (Figure 4.3) is given by:

$$\mathrm{d}n = \frac{2V\mathrm{d}^3\vec{k}}{(2\pi)^3} \,. \tag{4.25}$$

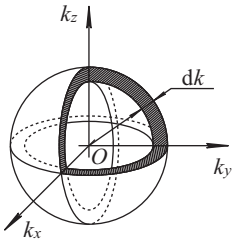


Fig. 4.4: Spherical shell capturing allowed wave vectors.

To calculate the number of the electromagnetic modes with wave vectors whose lengths are ranged from k to $(k + \mathrm{d}k)$, we should assume that all such vectors are within a thin spherical shell:

$$\mathrm{d}^3\vec{k} = 4\pi k^2 \mathrm{d}k \,. \tag{4.26}$$

Here, $4\pi k^2$ is the area of the sphere and $\mathrm{d}k$ is the thickness of the spherical shell (Figure 4.4). Substituting formula (4.26) into (4.25), we obtain the number of electromagnetic modes whose wave vectors fall within the range $(k, k+\mathrm{d}k)$, i.e., in the thin spherical shell:

$$\mathrm{d}n = \frac{2V\mathrm{d}^3\vec{k}}{(2\pi)^3} = \frac{8\pi V k^2 \mathrm{d}k}{(2\pi)^3} \,. \tag{4.27}$$

The number of modes with frequencies in the interval $(\nu, \nu + \mathrm{d}\nu)$ is our concern in many ways. The wave number k and frequency υ are related as follows:

$$k = \frac{2\pi\nu}{c} \,. \tag{4.28}$$

After inserting (4.28) into (4.27), we can calculate the number of modes per unit volume of the cavity in the frequency interval $(\nu, \nu + \mathrm{d}\nu)$:

$$\frac{\mathrm{d}n}{V} \equiv D(\nu)\,\mathrm{d}\nu = \frac{8\pi\nu^2 \mathrm{d}\nu}{c^3} \,. \tag{4.29}$$

The quantity $D(\nu)$ is called the *density of modes*, or sometimes, the *density of states* or the number of states per unit frequency interval.

The electromagnetic modes $\vec{E}_{\vec{k}\lambda}$ and $\vec{H}_{\vec{k}\lambda}$ (4.25) form a complete set. Therefore any field inside the cavity can be represented as a superposition of the electromagnetic modes $\vec{E}_{\vec{k}\lambda}$ and $\vec{H}_{\vec{k}\lambda}$:

$$\begin{aligned} \vec{E}(\vec{r}, t) &= \mathrm{i}\sum_{\vec{k},\lambda} \vec{e}_{\vec{k}\lambda} \sqrt{\frac{\hbar\omega_k}{2V\varepsilon_0}} \left\{ a_{\vec{k}\lambda}(t) \exp\left(\mathrm{i}\vec{k}\cdot\vec{r}\right) - \mathrm{c.c.} \right\} \,, \\ \vec{H}(\vec{r}, t) &= \mathrm{i}\sum_{\vec{k},\lambda} \left[\vec{n}_{\vec{k}} \times \vec{e}_{\vec{k}\lambda}\right] \sqrt{\frac{\hbar\omega_k}{2V\mu_0}} \left\{ a_{\vec{k}\lambda}(t) \exp\left(\mathrm{i}\vec{k}\cdot\vec{r}\right) - \mathrm{c.c.} \right\} \,, \end{aligned} \tag{4.30}$$

where $a_{\vec{k}\lambda}(t) = a_{\vec{k}\lambda} \exp(-\mathrm{i}\omega_k t)$.

Using the representation (4.30), we can compute the energy of electromagnetic waves in the cavity:

$$W = \frac{1}{2}\int\limits_V \mathrm{d}^3\vec{r}\left[\varepsilon_0\vec{E}^2(\vec{r},t) + \mu_0\vec{H}^2(\vec{r},t)\right] . \tag{4.31}$$

After substituting the series (4.30) into (4.31), the integrand of formula (4.31) contains terms of two types. The former are the products of the terms with wave vectors $\vec{k} \neq \mp\vec{k}'$. They include the multipliers $\exp[\mathrm{i}(\vec{k} \pm \vec{k}')\cdot\vec{r}]$ vanishing when integrated over the volume. A typical integral of this type has the form:

$$\int\limits_0^L \exp\left[\frac{2\pi}{L}\left(n \pm n'\right)x\right]\mathrm{d}x = 0 \tag{4.32}$$

when $n \pm n' \neq 0$. The latter contains the products of the terms with wave vectors $\vec{k} = \mp\vec{k}'$. The exponential factors drop out, so their integration over $\vec{r}$ yields the volume of the cavity. Note also that:

$$\left([\vec{n}_{\vec{k}} \times \vec{e}_{\vec{k}\lambda}]\cdot[\vec{n}_{\pm\vec{k}} \times \vec{e}_{\pm\vec{k}\lambda'}]\right) = \pm\left(\vec{e}_{\vec{k}\lambda}\cdot\vec{e}_{\pm\vec{k}\lambda'}\right) . \tag{4.33}$$

As a result, the electromagnetic field energy in the cavity can be expanded as follows:

$$W = \sum_{\vec{k},\lambda} \hbar\omega_k a^*_{\vec{k}\lambda}(t)a_{\vec{k}\lambda}(t) . \tag{4.34}$$

Next, we proceed from the mode variables $a_{\vec{k}\lambda}$ and $a^*_{\vec{k}\lambda}$ to the real variables $Q_{\vec{k}\lambda}$ and $P_{\vec{k}\lambda}$:

$$Q_{\vec{k}\lambda}(t) = \sqrt{\frac{\hbar}{2\omega_k}}\left[a^*_{\vec{k}\lambda}(t) + a_{\vec{k}\lambda}(t)\right] . \tag{4.35}$$

$$P_{\vec{k}\lambda}(t) = \dot{Q}_{\vec{k}\lambda} = -\mathrm{i}\sqrt{\frac{\hbar\omega_k}{2}}\left[a_{\vec{k}\lambda}(t) - a^*_{\vec{k}\lambda}(t)\right] . \tag{4.36}$$

Using formulas (4.35) and (4.36), we represent the electromagnetic field energy (4.34) in the form of Hamiltonian's functions for a set of independent real harmonic oscillators:

$$W \equiv H = \sum_{\vec{k},\lambda} H_{\vec{k}\lambda} = \sum_{\vec{k},\lambda}\frac{1}{2}\left[P^2_{\vec{k}\lambda} + \omega_k^2 Q^2_{\vec{k}\lambda}\right] . \tag{4.37}$$

The first of Hamilton's equations has the form

$$\dot{Q}_{\vec{k}\lambda} = \frac{\partial H}{\partial P_{\vec{k}\lambda}} = P_{\vec{k}\lambda} \tag{4.38}$$

and is consistent with the equality (4.36) that, thus, is a corollary of the equation of motion. This is achieved by appropriately choosing the coefficients in (4.35), and (4.36). The second Hamiltonian equation:

$$\dot{P}_{\vec{k}\lambda} = -\frac{\partial H}{\partial Q_{\vec{k}\lambda}} = -\omega_k Q_{\vec{k}\lambda} , \tag{4.39}$$

taking (4.38) into account, acquires the form:

$$\ddot{Q}_{\vec{k}\lambda} + \omega_k^2 Q_{\vec{k}\lambda} = 0\,. \tag{4.40}$$

Thus, the equations of the electromagnetic field in the cavity boil down to a set of independent equations for harmonic oscillators. Each mode of the electromagnetic oscillation corresponds to an oscillator.

4.2 Quantization of a Free Electromagnetic Field

In the previous section, we have succeeded in representing the electromagnetic field energy in a cavity as the sum of the energies of independent harmonic oscillators. Therefore, the quantization of the electromagnetic field may be carried out in the same way as in the quantum mechanics course for a classical harmonic oscillator. For this, we replace the dynamic variables $Q_{\vec{k}\lambda}$ and $P_{\vec{k}\lambda}$ by the Hermitian operators:

$$Q_{\vec{k}\lambda} \to \hat{q}_{\vec{k}\lambda}\,, \quad P_{\vec{k}\lambda} \to \hat{p}_{\vec{k}\lambda}\,, \tag{4.41}$$

that satisfy the commutation relations, typical for the operators of coordinates and momenta:

$$[\hat{q}_{\vec{k}\lambda}, \hat{p}_{\vec{k}'\lambda'}] = \mathrm{i}\hbar\delta_{\vec{k}\vec{k}'}\delta_{\lambda\lambda'}\,, \quad [\hat{q}_{\vec{k}\lambda}, \hat{q}_{\vec{k}'\lambda'}] = [\hat{p}_{\vec{k}\lambda}, \hat{p}_{\vec{k}'\lambda'}] = 0\,. \tag{4.42}$$

As is the case for the harmonic oscillator, it is useful to introduce the annihilation $\hat{a}_{\vec{k}\lambda}$ and creation $\hat{a}^+_{\vec{k}\lambda}$ operators to quantize a free electromagnetic field:

$$\hat{a}_{\vec{k}\lambda} = (2\hbar\omega_k)^{-\frac{1}{2}}\left(\omega_k\hat{q}_{\vec{k}\lambda} + \mathrm{i}\hat{p}_{\vec{k}\lambda}\right)\,, \quad \hat{a}^+_{\vec{k}\lambda} = (2\hbar\omega_k)^{-\frac{1}{2}}\left(\omega_k\hat{q}_{\vec{k}\lambda} - \mathrm{i}\hat{p}_{\vec{k}\lambda}\right)\,. \tag{4.43}$$

From formulas (4.43) we can express the $\hat{q}_{\vec{k}\lambda}$ and $\hat{p}_{\vec{k}\lambda}$ operators through $\hat{a}_{\vec{k}\lambda}$ and $\hat{a}^+_{\vec{k}\lambda}$:

$$\hat{q}_{\vec{k}\lambda} = \sqrt{\frac{\hbar}{2\omega_k}}\left(\hat{a}_{\vec{k}\lambda} + \hat{a}^+_{\vec{k}\lambda}\right)\,, \quad \hat{p}_{\vec{k}\lambda} = -\mathrm{i}\sqrt{\frac{\hbar\omega_k}{2}}\left(\hat{a}_{\vec{k}\lambda} - \hat{a}^+_{\vec{k}\lambda}\right)\,. \tag{4.44}$$

The operators $\hat{a}_{\vec{k}\lambda}$ and $\hat{a}^+_{\vec{k}\lambda}$ are not Hermitian so they do not correspond to the values observed, in contrast to $\hat{q}_{\vec{k}\lambda}$ and $\hat{p}_{\vec{k}\lambda}$.

The commutation relations for the operators $\hat{a}_{\vec{k}\lambda}$ and $\hat{a}^+_{\vec{k}\lambda}$ are derived from the corresponding relations for $\hat{q}_{\vec{k}\lambda}$ and $\hat{p}_{\vec{k}\lambda}$:

$$\left[\hat{a}_{\vec{k}\lambda}, \hat{a}_{\vec{k}'\lambda'}\right] = \left[\hat{a}^+_{\vec{k}\lambda}, \hat{a}^+_{\vec{k}'\lambda'}\right] = 0\,, \quad \left[\hat{a}_{\vec{k}\lambda}, \hat{a}^+_{\vec{k}'\lambda'}\right] = \delta_{\vec{k}\vec{k}'}\delta_{\lambda\lambda'}\,. \tag{4.45}$$

Substituting formula (4.44) into (4.37), we find the Hamiltonian operator describing the quantized electromagnetic field in the cavity:

$$\begin{aligned}\widehat{H} &= \sum_{\vec{k},\lambda}\widehat{H}_{\vec{k}\lambda} = \sum_{\vec{k},\lambda}\frac{1}{2}\left(\hat{p}^2_{\vec{k}\lambda} + \omega_k\hat{q}^2_{\vec{k}\lambda}\right) = \sum_{\vec{k},\lambda}\frac{\hbar\omega_k}{2}\left(\hat{a}^+_{\vec{k}\lambda}\hat{a}_{\vec{k}\lambda} + \hat{a}_{\vec{k}\lambda}\hat{a}^+_{\vec{k}\lambda}\right) = \\ &= \sum_{\vec{k}\lambda}\hbar\omega_k\left(\hat{a}^+_{\vec{k}\lambda}\hat{a}_{\vec{k}\lambda} + \frac{1}{2}\right) = \sum_{\vec{k}\lambda}\hbar\omega_k\left(\hat{n}_{\vec{k}\lambda} + \frac{1}{2}\right)\,.\end{aligned} \tag{4.46}$$

Here we have used the last of the commutation relations (4.45).

The Hermitian operator $\hat{n}_{\vec{k}\lambda} = \hat{a}^{+}_{\vec{k}\lambda}\hat{a}_{\vec{k}\lambda}$ has eigenvectors:

$$\left|\{n_{\vec{k}\lambda}\}\right\rangle = \left|n_{\vec{k}_1\lambda_1}\right\rangle \left|n_{\vec{k}_2\lambda_2}\right\rangle \ldots \left|n_{\vec{k}_l\lambda_l}\right\rangle \ldots \ . \tag{4.47}$$

These are constructed in the form of a direct product of the eigenvectors of the individual harmonic oscillators and satisfy the equations:

$$\hat{n}_{\vec{k}_l\lambda_l} \left|\{n_{\vec{k}\lambda}\}\right\rangle = n_{\vec{k}_l\lambda_l} \left|\{n_{\vec{k}\lambda}\}\right\rangle \ , \tag{4.48}$$

where $n_{\vec{k}\lambda} = 0, 1, 2 \ldots .$

The operator $\hat{n}_{\vec{k}\lambda}$ may be interpreted as an operator of the number of quantum fluctuations of the electromagnetic field in the cavity. They are called photons. Let us clarify the origin of the name. The states $|\{n_{\vec{k}\lambda}\}\rangle$ are eigenvectors of the Hamiltonian operator

$$\hat{H}_{\vec{k}\lambda} \left|\{n_{\vec{k}\lambda}\}\right\rangle = E_n \left|\{n_{\vec{k}\lambda}\}\right\rangle \ . \tag{4.49}$$

They meet the energy spectrum of the electromagnetic field of the following form:

$$E_n = \sum_{\vec{k},\lambda} \hbar\omega_k \left[n_{\vec{k}\lambda} + \frac{1}{2} \right] \ . \tag{4.50}$$

Therefore, it can be said that the cavity has $n_{\vec{k}_1\lambda_1}$ quasiparticles (*photons*) with the energy $\hbar\omega_{k_1}$ of the first mode of the electromagnetic field fluctuations, $n_{\vec{k}_2\lambda_2}$ – of the second mode, $n_{\vec{k}_l\lambda_l}$ – of the l-th, etc. On the one hand, each number $n_{\vec{k}\lambda}$ gives the number of photons, and on the other hand, it characterizes the degree of excitation of the electromagnetic mode with the wave vector $\vec{k}$ and polarization λ.

When the states $|\{n_{\vec{k}\lambda}\}\rangle$ are orthonormal:

$$\left\langle n_{\vec{k}_l\lambda_l} \middle| n_{\vec{k}_m\lambda_m} \right\rangle = \delta_{\vec{k}_l\vec{k}_m}\delta_{\lambda_l\lambda_m} \ ,$$

the following relations are fulfilled:

$$\begin{aligned} \hat{a}^{+}_{\vec{k}_l\lambda_l} \left|\ldots, n_{\vec{k}_l\lambda_l}, \ldots\right\rangle &= \left(n_{\vec{k}_l\lambda_l} + 1\right)^{\frac{1}{2}} \left|\ldots, n_{\vec{k}_l\lambda_l} + 1, \ldots\right\rangle \ , \\ \hat{a}_{\vec{k}_l\lambda_l} \left|\ldots, n_{\vec{k}_l\lambda_l}, \ldots\right\rangle &= \left(n_{\vec{k}_l\lambda_l}\right)^{\frac{1}{2}} \left|\ldots, n_{\vec{k}_l\lambda_l} - 1, \ldots\right\rangle \ . \end{aligned} \tag{4.51}$$

The latter allow the $\hat{a}_{\vec{k}\lambda}$ and $\hat{a}^{+}_{\vec{k}\lambda}$ operators to be called as creation and annihilation operators of quasiparticles, *photons*.

Any state of the electromagnetic field in the cavity can be represented as a superposition of the eigenstates $|\{n_{\vec{k}\lambda}\}\rangle$:

$$\begin{aligned} |\Psi\rangle = &\sum_{n_{\vec{k}_1\lambda_1}} \sum_{n_{\vec{k}_2\lambda_2}} \ldots \sum_{n_{\vec{k}_l\lambda_l}} \ldots C_{n_{\vec{k}_1\lambda_1}, n_{\vec{k}_2\lambda_2}, \ldots, n_{\vec{k}_l\lambda_l}, \ldots} \left|n_{\vec{k}_1\lambda_1}, n_{\vec{k}_2\lambda_2}, \ldots, n_{\vec{k}_l\lambda_l}, \ldots\right\rangle \equiv \\ \equiv &\sum_{\{n_{\vec{k}\lambda}\}} C_{\{n_{\vec{k}\lambda}\}} \left|\{n_{\vec{k}\lambda}\}\right\rangle \ . \end{aligned} \tag{4.52}$$

It is obvious that many photons can be in one energy state. This means that the photons are bosons. A more detailed analysis shows that the photon has spin of $s = 1$ (in units of $\hbar$). However, the projection of the photon spin on the direction of its motion (on the direction of the vector $\vec{k}$) takes only two values: ± 1 as the phonon has three such projections: $0, \pm 1$, with the transversality condition for electromagnetic waves excluding the zero value of the projection of the photon spin.

In classical theory, the states of a photon with the projections ± 1 on the direction of its motion correspond to two circular polarized monochromatic electromagnetic waves. Within the quantum theory, circular waves may also be quantized as we have done the same with linear polarized waves. For each value of the wave vector $\vec{k}$, there are only two independent waves with alternate circular polarizations.

Suppose that we are in a certain point of space and have a look from the end of the vector $\vec{k}$ at the plane perpendicular to it. Then, in one case, the strength vector of the electromagnetic field $\vec{E}_{\vec{k}\mu}$ rotates clockwise without changing its length. In another case, the vector rotates counterclockwise. These rotations correspond to the two projections of the photon spin on the direction of the vector $\vec{k}$. They are called the *photon helicity*: $\mu = \pm 1$. The value $\mu = -1$ corresponds to rotation of the vector $\vec{E}_{\vec{k}\mu}$ in a clockwise direction, and the value $\mu = +1$ – to counterclockwise rotation. Now we will explain this in detail.

Like any other particle, a photon can have a certain angular momentum. The division of the angular momentum into orbital (associated with motion of a photon in space) and spin (intrinsic angular momentum of a photon at rest) parts is, generally speaking, physically meaningless. The speed of a photon is always different from zero. Therefore there is no frame of reference where a photon would be at rest. Nevertheless, to facilitate calculations, it is useful to formally represent the $\hat{j}_\alpha$ components of the angular momentum operator for a single photon as a sum of the "orbital" and "spin" momentum operator components:

$$\hat{j}_\alpha = \hat{l}_\alpha + \hat{s}_\alpha \,, \tag{4.53}$$

where $\alpha = 1, 2, 3$.

Formula (4.53) is not difficult to derive. For this, it is sufficient to remember that a) a free photon is theoretically delineated in terms of only one vector potential $\vec{A}(\vec{r}, t)$, b) the angular momentum operator for the photon coincides, up to a factor i, with the infinitesimal transformation operator for the field $\vec{A}(\vec{r}, t)$ under spatial rotations of the coordinate system. In any inertial reference system, a photon moves at the speed of light. Therefore, there is always a preferred direction in space for it to move, coinciding with the direction of the wave vector $\vec{k}$. As such, the transformation laws for the vector potential of the photon are easier to discuss in k-space. Since the fields $\vec{A}(\vec{r})$ and $\vec{A}(\vec{k})$ are connected through a Fourier integral:

$$\vec{A}(\vec{r}) = \int \mathrm{d}^3\vec{k}\, \vec{A}(\vec{k}) \exp\left(\mathrm{i}\vec{k}\cdot\vec{r}\right) \,,$$

the rotation of axes of the Cartesian coordinate system

$$\vec{r} \to \vec{r}' = D^T(\theta)\,\vec{r}\,, \quad D^T D = DD^T = I\,, \quad \det D = 1$$

is equivalent to rotation of the k–space:

$$\vec{k} \to \vec{k}' = D(\theta)\,\vec{k}\,.$$

Here, D is an orthogonal matrix for rotation. The symbol "T" denotes transposition.

Under an infinitesimal rotation of the k-space, the vector potential $\vec{A}(\vec{k})$ can be transformed by the rule:

$$\vec{A}(\vec{k}) \to \vec{A}'\left(\vec{k} + \delta\vec{k}\right) = \vec{A}(\vec{k}) + \left[\delta\vec{\theta} \times \vec{A}(\vec{k})\right]\,, \tag{4.54}$$

where $\delta\vec{k}$ is a change of $\vec{k}$ rotated by the angle $\delta\theta$:

$$\delta\vec{k} = \left[\delta\vec{\theta} \times \vec{k}\right]\,. \tag{4.55}$$

The vector $\delta\vec{\theta}$ simultaneously defines both a rotation angle and an axis, around that the k-space revolves. According to (4.54), the vector $\vec{A}(\vec{k})$ rotated by the angle $|\delta\vec{\theta}|$ is transferred at the point with the radius vector $(\vec{k} + \delta\vec{k})$.

Using (4.54) and (4.55), we can calculate the change of the vector potential at the point with the radius vector $\vec{k}$:

$$\begin{aligned}
\delta\vec{A}(\vec{k}) &= \vec{A}'(\vec{k}) - \vec{A}(\vec{k}) = \vec{A}\left(\vec{k} - \delta\vec{k}\right) + \left[\delta\vec{\theta} \times \vec{A}(\vec{k})\right] - \vec{A}(\vec{k}) = \\
&= -\delta\vec{k} \cdot \vec{\nabla}_{\vec{k}}\vec{A}(\vec{k}) + \left[\delta\vec{\theta} \times \vec{A}(\vec{k})\right] = \underbrace{-\left[\delta\vec{\theta} \times \vec{k}\right] \cdot \vec{\nabla}_{\vec{k}}}_{-\left(\delta\vec{\theta} \cdot \left[\vec{k} \times \vec{\nabla}_{\vec{k}}\right]\right)}\vec{A}(\vec{k}) + \left[\delta\vec{\theta} \times \vec{A}(\vec{k})\right] \equiv \\
&\equiv -\mathrm{i}\delta\theta_\alpha\left[\left(\hat{l}_\alpha + \hat{s}_\alpha\right)\vec{A}(\vec{k})\right] \equiv -\mathrm{i}\delta\theta_\alpha\hat{j}_\alpha\vec{A}(\vec{k})\,.
\end{aligned} \tag{4.56}$$

In what follows, the indices α (repeated here) mean summation.

In the long run, we have substantiated the representation (4.53) for the photon angular momentum:

$$\hat{j}_\alpha = \hat{l}_\alpha + \hat{s}_\alpha\,.$$

The operators:

$$\hat{l}_\alpha = -\mathrm{i}\left[\vec{k} \times \vec{\nabla}_{\vec{k}}\right]_\alpha\,, \tag{4.57}$$

coincide with the components of the conventional quantum mechanical operator for the orbital momentum of a particle in the momentum representation (in units of $\hbar$).

The operators $\hat{s}_\alpha$ may be treated as the photon spin operator components. The action of the operator $\hat{s}_\alpha$ on the vector $\vec{A}$ is given by:

$$\left(\hat{s}_\alpha\vec{A}\right)_\beta \equiv \left(\hat{s}_\alpha\right)_{\beta\gamma}A_\gamma \equiv -\mathrm{i}\varepsilon_{\alpha\beta\gamma}A_\gamma\,, \tag{4.58}$$

where $\alpha, \beta, \gamma = 1, 2, 3$.

In a more detailed component wise form, the $\hat{l}_\alpha$ and $\hat{s}_\alpha$ operators can be written as:

$$\hat{l}_\alpha = -\mathrm{i}\varepsilon_{\alpha\beta\gamma} k_\beta \frac{\partial}{\partial k_\gamma} , \quad (\hat{s}_\alpha)_{\beta\gamma} = -\mathrm{i}\varepsilon_{\alpha\beta\gamma} .$$

It is easy to check that the $\hat{l}_\alpha$ and $\hat{s}_\alpha$ operators commute with each other, are Hermitian, and satisfy commutation relations, typical of the angular momentum operators:

$$\left[\hat{l}_\alpha, \hat{l}_\beta\right] = \mathrm{i}\varepsilon_{\alpha\beta\gamma}\hat{l}_\gamma , \quad \left[\hat{s}_\alpha, \hat{s}_\beta\right] = \mathrm{i}\varepsilon_{\alpha\beta\gamma}\hat{s}_\gamma .$$

In the formula

$$\hat{j}_\alpha \vec{A} = \left(\hat{l}_\alpha + \hat{s}_\alpha\right)\vec{A} ,$$

the operator $\hat{s}_\alpha$ acts only on the vector index of the field $\vec{A}(\vec{k})$, transforming various components of the vector $\vec{A}$ through each other and does not affect coordinates of the vector $\vec{k}$. This is the formal basis for the introduction of the photon spin operator. The operator $\hat{l}_\alpha$, in contrast, acts on the vector field $\vec{A}(\vec{k})$ as on a function of $\vec{k}$.

Since the state of a photon is characterized by the three-dimensional vector $\vec{A}$, the number of different functions transformed through each other (degeneracy multiplicity) is: $2s + 1 = 3$. Consequently, the photon corresponds to spin $s = 1$.

The peculiarity of the problem is that the vector potential $\vec{A}(\vec{k})$ obeys the gauge condition:

$$\operatorname{div}\vec{A}(\vec{r}) = 0 \quad \text{or} \quad \vec{k}\cdot\vec{A}(\vec{k}) = 0 . \tag{4.59}$$

The latter leads to further relationships between the components of the vector $\vec{A}$ and does not allow one to distinguish the orbital angular momentum of the photon from the spin one.

Consider the state of a photon with a certain wave vector $\vec{k}$ (with momentum $\hbar\vec{k}$). In such a situation, the parameters of the photon are not invariant with respect to all transformations of the three-dimensional rotations group. Here, we can talk only about the axial symmetry of the photon around the axis defined by the vector $\vec{k}$. In particular, the state of a photon with a certain wave vector has no certain angular momentum. This is because the $\vec{k}$ and $\hat{j}_\alpha$ operators do not commute with each other. However, the photon may have a certain value of the angular momentum projection on the direction of $\vec{k}$. Indeed, using (4.57), we have the following:

$$k_\alpha \hat{l}_\alpha = 0 .$$

Then, from (4.53), we find:

$$k_\alpha \hat{j}_\alpha = k_\alpha \hat{s}_\alpha . \tag{4.60}$$

Since the $\hat{\vec{k}}$ and $\hat{s}_\alpha$ operators commute, then, according to (4.60), the $\vec{k}$, $k_\alpha\hat{j}_\alpha$ operators observed commute as well. Thus, the equality (4.60) stands for coincidence of the angular momentum and spin projections of the photon on its motion direction. Also it means that they can be measured simultaneously with the photon wave vector.

In virtue of the limitation of (4.59), only two components of the vector $\vec{A}$, lying in a plane perpendicular to the vector $\vec{k}$, are transformed through themselves under the action of the operators $\hat{s}_\alpha$. Therefore, the spin projection of the photon on the direction of its motion may have only two values $\mu = \pm 1$; the value $\mu = 0$ is impossible. It should be noted that, in the above formalism, the photon states with the spin projections $\mu = \pm 1$ correspond to two types of the circular polarized electromagnetic wave matched to the photon. As mentioned earlier, these states are called spiral and the quantity is called *μ-helicity*. The main property of helicity is its invariance under space rotations, and under Lorentz transformations that do not change the direction of the vector $\vec{k}$. The helicity changes sign when spatially reflected.

4.3 Zero point Energy

To start with, we discuss the state of an electromagnetic field without photons:

$$n_{\vec{k}_1\lambda_1} = n_{\vec{k}_2\lambda_2} = \ldots = 0\,. \tag{4.61}$$

Such a state is called a vacuum. The vacuum state corresponds to a ket-vector $|0\rangle$, satisfying the condition

$$\hat{a}_{\vec{k}\lambda}\,|0\rangle = 0 \tag{4.62}$$

for any $\hat{a}_{\vec{k}\lambda}$.

The vacuum state has an amazing property: despite the absence of photons, its total energy is not zero:

$$\widehat{H}\,|0\rangle = \frac{1}{2}\sum_{\vec{k},\lambda}\hbar\omega_k\,|0\rangle \equiv \varepsilon_0\,|0\rangle\ . \tag{4.63}$$

The quantity $\varepsilon_0 = \sum_{\vec{k}\lambda}\hbar\omega_k/2$ is referred to as the zero point energy. Its existence owes to the uncertainty relations for the position and momentum of every harmonic oscillator: $\Delta p_{\vec{k}\lambda}\Delta q_{\vec{k}\lambda} \sim \hbar/2$. Even in their ground state, systems undergo fluctuations of the electromagnetic field, and they correspond to the energy of the oscillators:

$$\sum_{\vec{k},\lambda}\frac{1}{2}\left(\Delta p^2_{\vec{k}\lambda} + \omega_k^2\Delta q^2_{\vec{k}\lambda}\right)_{\min} \approx \frac{1}{2}\sum_{\vec{k},\lambda}\left(\Delta p^2_{\vec{k}\lambda} + \frac{\omega_k^2\hbar^2}{4\Delta p^2_{\vec{k}\lambda}}\right)_{\min} = \varepsilon_0\,.$$

Since the energy of one quantum of the field is equal to $\hbar\omega_k = h\nu$, and the volume $V = 1\,\mathrm{m}^3$ within the frequency ranges from ν to $(\nu + \mathrm{d}\nu)$, have

$$D(\nu)\,\mathrm{d}\nu = \frac{8\pi\nu^2\,\mathrm{d}\nu}{\mathrm{c}^3}$$

allowed field states, the total zero point energy per unit volume of a cavity can be written in the form of an integral:

$$\varepsilon_0 = \frac{1}{2}\int_0^\infty h\nu D(\nu)\mathrm{d}\nu = \frac{4\pi h}{c^3}\int_0^\infty \nu^3\mathrm{d}\nu = \left.\frac{\pi h}{c^3}\nu^4\right|_0^\infty \to \infty\,. \tag{4.64}$$

The integral (4.64) diverges because the permitted frequencies do not have the upper limit (this is the difference between photons and phonons).

In quantum theory, the above difficulty can usually be avoided as follows. There is no experiment where instruments would register a response proportional to the zero point energy. In practice, the response is always proportional to a change in total energy of an electromagnetic field relative to the zero point energy:

$$E' = E - \varepsilon_0 = \sum_{\vec{k},\lambda} \hbar\omega_k n_{\vec{k}\lambda} \,, \tag{4.65}$$

and this value is finite. In other words, the difficulties can be overcome by counting off the energy system from the zero point energy level. We will come across this as we go along.

4.4 Amplitude and Phase Operators for Single-Mode Quantum States of a Radiation Field

There is no reason to believe that the electromagnetic field of real light beams corresponds to any one of the quantum states $|\{n_{\vec{k}\lambda}\}\rangle$. Our immediate goal is to find a superposition of the states $|\{n_{\vec{k}\lambda}\}\rangle$ that would correspond to an electromagnetic wave of classical physics.

We restrict ourselves to considering a single mode of the electromagnetic field with a wave vector $\vec{k}$. Suppose the mode had certain polarization λ. Taking the above into account, we omit the indices $\vec{k}$, λ of all the variables that describe the mode and write the field vectors $\vec{E}$, $\vec{H}$ as scalars.

In the classical theory of electromagnetic waves, the complex amplitude of the wave a is represented as the product of its module and the phase factor:

$$a = |a| \exp(\mathrm{i}\varphi) \,. \tag{4.66}$$

This is followed by proceeding to the real observed fields E and H.

Recall that the expressions for the electromagnetic field operators $\widehat{E}$ and $\widehat{H}$ formally become classical after replacing: $\widehat{a} \to a$, $\widehat{a}^+ \to a^*$. To comply with classical physics, the amplitude and phase operators should be separated from the creation and annihilation photon operators, i.e., we should seek an analogue to operation (4.66). In quantum theory, there is no exact way to carry out this procedure. Therefore, the determination of the amplitude and phase quantum mechanical operators suffers from a large degree of arbitrariness. The main reasons lie in the fact that the quantum mechanical amplitude and phase must have the same values in the corresponding limit, as in classical physics. Also, they must be related to Hermitian operators to be (at least in principle) observed quantities.

As working formulas, we take the ratios:

$$\widehat{a} = \sqrt{\widehat{n}+1}\exp(\mathrm{i}\widehat{\varphi}) \,, \quad \widehat{a}^+ = \exp(-\mathrm{i}\widehat{\varphi})\sqrt{\widehat{n}+1} \,. \tag{4.67}$$

Formula (4.67) implies that the Hermitian operator $\sqrt{\hat{n}+1}$ is an amplitude operator. It should be emphasized that the expressions $\exp(\mathrm{i}\hat{\varphi})$ and $\exp(-\mathrm{i}\hat{\varphi})$ do not possess the properties of exponents of the operators $\mathrm{i}\hat{\varphi}$ and $-\mathrm{i}\hat{\varphi}$. They are designated so only because of their modification into the phase factors of classical physics within the appropriate limit.

According to (4.67), the ratios below are the definitions of the $\exp(\mathrm{i}\hat{\varphi})$ and $\exp(-\mathrm{i}\hat{\varphi})$ operators:

$$\exp\left(\mathrm{i}\hat{\varphi}\right) = (\hat{n}+1)^{-1/2}\hat{a}\,, \quad \exp\left(-\mathrm{i}\hat{\varphi}\right) = \hat{a}^{+}(\hat{n}+1)^{-1/2}\,. \tag{4.68}$$

Using formula (4.68), we can derive the properties of these operators from the known properties of creation and annihilation operators. For example using the identity:

$$\hat{a}\hat{a}^{+} = \hat{n}+1\,,$$

following from the formulas:

$$\left[\hat{a}, \hat{a}^{+}\right] = 1\,, \quad \hat{n} = \hat{a}^{+}\hat{a}\,,$$

we find a relationship that confirms, at first sight, the validity of the designations $\exp(\mathrm{i}\hat{\varphi})$ and $\exp(-\mathrm{i}\hat{\varphi})$:

$$\exp\left(\mathrm{i}\hat{\varphi}\right)\cdot\exp\left(-\mathrm{i}\hat{\varphi}\right) = (\hat{n}+1)^{-1/2}\hat{a}\cdot\hat{a}^{+}(\hat{n}+1)^{-1/2} = (\hat{n}+1)^{-1/2}(\hat{n}+1)(\hat{n}+1)^{-1/2} = 1\,.$$

At the same time, it is easy to check that:

$$\exp\left(-\mathrm{i}\hat{\varphi}\right)\cdot\exp\left(\mathrm{i}\hat{\varphi}\right) \neq 1\,.$$

The formulas

$$\hat{n}\,|n\rangle = \hat{a}^{+}\hat{a}\,|n\rangle = n\,|n\rangle\,, \quad \hat{a}\,|n\rangle = n^{1/2}\,|n-1\rangle\,, \quad \hat{a}^{+}\,|n\rangle = (n+1)^{1/2}\,|n+1\rangle \tag{4.69}$$

allow one to easily calculate the actions of the operators $\exp(\mathrm{i}\hat{\varphi})$ and $\exp(-\mathrm{i}\hat{\varphi})$ on the state $|n\rangle$:

$$\begin{aligned}
\exp\left(\mathrm{i}\hat{\varphi}\right)|n\rangle &= (\hat{n}+1)^{-1/2}(\hat{a}\,|n\rangle) = \left(\hat{n}+1\right)^{-1/2} n^{1/2}\,|n-1\rangle = \\
&= n^{1/2}(\hat{n}+1)^{-1/2}\,|n-1\rangle = \begin{cases} |n-1\rangle\,, & n \neq 0 \\ 0\,, & n = 0\,; \end{cases} \\
\exp\left(-\mathrm{i}\hat{\varphi}\right)|n\rangle &= \hat{a}^{+}\left(\hat{n}+1\right)^{-1/2}|n\rangle = (n+1)^{-1/2}(\hat{a}^{+}\,|n\rangle) = \\
&= (n+1)^{-1/2}(n+1)^{1/2}\,|n+1\rangle = |n+1\rangle\,.
\end{aligned} \tag{4.70}$$

As a result, only the following matrix elements of the operators $\exp(\mathrm{i}\hat{\varphi})$ and $\exp(-\mathrm{i}\hat{\varphi})$ do not vanish:

$$\langle n-1|\exp\left(\mathrm{i}\hat{\varphi}\right)|n\rangle = 1\,, \quad \langle n+1|\exp\left(-\mathrm{i}\hat{\varphi}\right)|n\rangle = 1\,. \tag{4.71}$$

According to (4.71), these operators do not satisfy the relations of the type

$$\langle i|\,\widehat{Q}\,|j\rangle = \langle j|\,\widehat{Q}\,|i\rangle^* \ . \tag{4.72}$$

Therefore they are not Hermitian operators and do not depict the observed properties of the electromagnetic field. However, they can be used to build a new pair of the operators:

$$\begin{aligned} \cos\widehat{\varphi} &= \{\exp(\mathrm{i}\widehat{\varphi}) + \exp(-\mathrm{i}\widehat{\varphi})\}/2 \ , \\ \sin\widehat{\varphi} &= \{\exp(\mathrm{i}\widehat{\varphi}) - \exp(-\mathrm{i}\widehat{\varphi})\}/2\mathrm{i} \ . \end{aligned} \tag{4.73}$$

The nonzero matrix elements of the operators (4.73) acquire the form:

$$\begin{aligned} \langle n-1|\cos\widehat{\varphi}\,|n\rangle &= \langle n|\cos\widehat{\varphi}\,|n-1\rangle = 1/2 \ , \\ \langle n-1|\sin\widehat{\varphi}\,|n\rangle &= -\langle n|\sin\widehat{\varphi}\,|n-1\rangle = 1/2\mathrm{i} \ . \end{aligned} \tag{4.74}$$

They meet the condition (4.72). Thus, the operators (4.73) prove to be Hermitian. Therefore, they will be taken as quantum mechanical operators describing the observed properties of the electromagnetic field phase.

Because

$$[\hat{n}, \hat{a}^+] = \hat{a}^+ \ , \quad [\hat{n}, \hat{a}] = -\hat{a} \ ,$$

the following commutation relation holds:

$$\begin{aligned} [\hat{n}, \exp(\mathrm{i}\widehat{\varphi})] \equiv [\hat{n}, (\hat{n}+1)^{-1/2}\hat{a}] &= (\hat{n}+1)^{-1/2}\underbrace{[\hat{n}, \hat{a}]}_{-\hat{a}} + \underbrace{[\hat{n}, (\hat{n}+1)^{-1/2}]}_{0}\hat{a} = \\ &= -(\hat{n}+1)^{-1/2}\hat{a} = -\exp(\mathrm{i}\widehat{\varphi}) \ . \end{aligned}$$

Similarly we can argue that:

$$[\hat{n}, \exp(-\mathrm{i}\widehat{\varphi})] = \exp(-\mathrm{i}\widehat{\varphi}) \ .$$

Hence, in turn, we arrive at:

$$[\hat{n}, \cos\widehat{\varphi}] = -\mathrm{i}\sin\widehat{\varphi} \ , \quad [\hat{n}, \sin\widehat{\varphi}] = \mathrm{i}\cos\widehat{\varphi} \ . \tag{4.75}$$

The commutation relations (4.75) show that the particle number and phase operators do not commute with each other. Consequently, the radiation field states, being simultaneously eigenstates of the operators, are impossible to find. Therefore, both the amplitude of the electromagnetic wave associated with $\hat{n}$ and the phase associated with $\cos\widehat{\varphi}$ and $\sin\widehat{\varphi}$ cannot be accurately computed.

The results of measurements of the amplitude and phase are subject to the uncertainty relations that follow from (4.75):

$$\Delta n \Delta\cos\varphi \geq \left|\langle\sin\varphi\rangle\right|/2 \ , \quad \Delta n \Delta\sin\varphi \geq \left|\langle\cos\varphi\rangle\right|/2 \ . \tag{4.76}$$

In what follows, the symbols $\langle f\rangle$ and Δf stand for the average of the observed quantity f and its variance in the quantum mechanical state $|\alpha\rangle$:

$$\langle f\rangle = \frac{\langle\alpha|\,\widehat{f}\,|\,\alpha\rangle}{\langle\alpha|\alpha\rangle} \ , \quad \Delta f = \sqrt{\langle f^2\rangle - \langle f\rangle^2} \ .$$

Thus, the attempt to separate the amplitude and phase operators from the creation and annihilation photon operators leads to the phase and amplitude operators not commuting among themselves. In doing so, the resulting uncertainty relations are typical of a quantized electromagnetic field.

4.5 Coherent Photon States: Their Properties and Relationship with Classical Electromagnetic Waves

Let us introduce the state vectors of a quantized radiation field to describe electromagnetic waves of classical physics in the limit of large amplitudes. These states, designated as $|\alpha\rangle$, are called coherent states of the radiation field. The coherent states $|\alpha\rangle$ are important, not only because they, of all the quantum states, most correctly detail classical electromagnetic waves, but also because a laser discussed further below generates coherent state radiation.

Neither the amplitude nor the phase of the electromagnetic wave in the coherent state $|\alpha\rangle$ are precisely defined. However, both these quantities have the least mean square deviations that correspond to the signs of equality in (4.76). In the mechanics of microparticles, the analogous quantum state that most fully corresponds to a mass point in classical physics is a wave packet.

Next, we proceed to discussing the properties of a set of the coherent states $\{|\alpha\rangle\}$. We define each state as a linear superposition of the eigenstates $\{|n\rangle\}$ of the particle number operator:

$$|\alpha\rangle = \exp\left\{-\frac{|\alpha|^2}{2}\right\} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle \ . \tag{4.77}$$

Here, α is an arbitrary complex number.

It is easy to verify that the states $\{|\alpha\rangle\}$ are normalized:

$$\langle\alpha\,|\alpha\rangle = \exp\left(-|\alpha|^2\right) \underbrace{\sum_{n=0}^{\infty} [(\alpha^*\alpha)^n/n!]}_{\exp\left(|\alpha|^2\right)} = 1 \ . \tag{4.78}$$

At the same time, different coherent states $|\alpha\rangle$ and $|\beta\rangle$ are nonorthogonal:

$$\langle\alpha\,|\beta\rangle = \exp\left\{-\frac{|\alpha|^2}{2} - \frac{|\beta|^2}{2}\right\} \sum_{n=0}^{\infty} \frac{(\alpha^*\beta)^n}{n!} = \exp\left(-\frac{|\alpha|^2}{2} - \frac{|\beta|^2}{2} + \alpha^*\beta\right) .$$

This implies that

$$\left|\langle\alpha|\,\beta\rangle\right|^2 = \exp\left(-|\alpha-\beta|^2\right) \neq 0 \ . \tag{4.79}$$

The complex number α is parameterized by two arbitrary real numbers. As a result, the coherent states $\{|\alpha\rangle\}$ form a double continuum. Their number is much larger than the number of the states $\{|n\rangle\}$. Various states $\{|\alpha\rangle\}$ make up an overfilled set. That

is why they are not orthogonal. According to (4.79), different states $|\alpha\rangle$ and $|\beta\rangle$ become nearly orthogonal when the magnitude of $|\alpha - \beta|$ is much greater than unity.

It is important that the ket-vectors of $|\alpha\rangle$ are eigenvectors of the annihilation operator:

$$\widehat{a}\,|\alpha\rangle = \exp\left(-|\alpha|^2/2\right)\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}\sqrt{n}\,|n-1\rangle = \updownarrow n \to n+1 \updownarrow =$$
$$= \alpha\exp\left(-|\alpha|^2/2\right)\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}\,|n\rangle = \alpha\,|\alpha\rangle\ . \tag{4.80}$$

By (4.80), the complex number α is an eigenvalue of the operator $\widehat{a}$. Note that the ket-vector of $|\alpha\rangle$ is not an eigenvector of the creation operator because the expression $\widehat{a}^{+}|\alpha\rangle$ cannot be converted so as to obtain an expression of the form $\lambda|\alpha\rangle$.

At the same time, since:

$$\widehat{a}\,|\alpha\rangle = \alpha\,|\alpha\rangle\ , \tag{4.81}$$

we have:

$$\langle\alpha|\,\widehat{a}^{+} = \langle\alpha|\,\alpha^{*}\ . \tag{4.82}$$

In other words, the bra vector of $\langle\alpha|$ is an eigenvector of the creation operator $\widehat{a}^{+}$ with its eigenvalue α^{*}.

An equivalent way of entering the coherent states is as follows. The relation (4.81) is taken as the definition of a coherent state. Then the expansion (4.77) is a consequence of the new definition.

Although the coherent states are not orthogonal, they can be used as basic functions with the completeness relation:

$$\frac{1}{\pi}\int \mathrm{d}^2\alpha\,|\alpha\rangle\,\langle\alpha| = \sum_{n=0}^{\infty}|n\rangle\,\langle n| = \widehat{I}\,, \tag{4.83}$$

where $\mathrm{d}^2\alpha = \mathrm{d}(\mathrm{Re}\alpha)\mathrm{d}(\mathrm{Im}\alpha)$. To prove the relations (4.83), we should express the complex number α through the real amplitude $|\alpha|$ and phase θ:

$$\alpha = |\alpha|\exp(\mathrm{i}\theta)\,, \tag{4.84}$$

and use the identity:

$$\int \mathrm{d}^2\alpha(\alpha^{*})^n\alpha^m\exp\left(-|\alpha|^2\right) = \int \mathrm{d}\,|\alpha|\cdot|\alpha|^{m+n+1}\exp\left(-|\alpha|^2\right)\int_0^{2\pi}\mathrm{d}\theta\exp[\mathrm{i}(m-n)] = \pi n!\delta_{nm} \tag{4.85}$$

Assume $|f\rangle$ is a state of the electromagnetic field of the form:

$$|\,f\rangle \equiv f\left(\widehat{a}^{+}\right)|\,0\rangle\ ,$$

where $f(x)$ is any function that can be expanded in a Taylor series in powers of x. Let us argue that the ket-vector of $|f\rangle$ can be represented as a superposition of coherent states.

We have the chain of equalities:

$$|f\rangle = \hat{I}\,|f\rangle = \frac{1}{\pi}\int \mathrm{d}^2\alpha\,|\alpha\rangle\,\langle\alpha|\,f(\hat{a}^+)\,|0\rangle = \updownarrow\text{taking into account that } \langle\alpha|\,\hat{a}^+ = \langle\alpha|\,\alpha^*\updownarrow =$$
$$= \frac{1}{\pi}\int \mathrm{d}^2\alpha\,|\alpha\rangle\,\langle\alpha|\,f(\alpha^*)\,|0\rangle = \frac{1}{\pi}\int \mathrm{d}^2\alpha\,|\alpha\rangle\,f(\alpha^*)\underbrace{\langle\alpha|0\rangle}_{\exp\left(-|\alpha|^2/2\right)} =$$
$$= \frac{1}{\pi}\,|\alpha\rangle\int \mathrm{d}^2\alpha\cdot f(\alpha^*)\exp\left(-|\alpha|^2/2\right).$$

As can be seen, the arbitrary vector $|f\rangle$ can be expanded by the vectors of $\{|\alpha\rangle\}$:

$$|f\rangle = \frac{1}{\pi}\int \mathrm{d}^2\alpha f(\alpha^*)\exp\left(-|\alpha|^2/2\right)|\alpha\rangle\ . \tag{4.86}$$

We may reverse the expansion (4.86) and find the function $f(\alpha^*)$ in the known vector of $|f\rangle$. To be sure of this, it should be noted that:

$$\pi^{-1}\int \mathrm{d}^2\alpha\exp\left[\beta^*\alpha - |\alpha|^2\right](\alpha^*)^n = (\beta^*)^n\ ,$$

and so the identity is valid:

$$\pi^{-1}\int \mathrm{d}^2\alpha\exp\left[\beta^*\alpha - |\alpha|^2\right]f(\alpha^*) = f(\beta^*)\ . \tag{4.87}$$

For arbitrary functions, $f(x)$ is represented as a power series in the variable x.

The relations (4.86) and (4.87) easily result in:

$$\langle\beta|f\rangle = \exp\left(-|\beta|^2/2\right)f(\beta^*)\ ,$$

where $\langle\beta|$ is the vector of a coherent state.

Thus, there is a one-to-one correspondence between the vector of the state $|f\rangle$ and the functions $f(a^*)$. The latter play the role of the expansion coefficients of the vector of $|f\rangle$ in the basis of the coherent states $\{|\alpha\rangle\}$.

The expansion of the operators of observed quantities in the coherent states is brought about in a similar way, and it also turns out to be unambiguous.

Now, we calculate the average number of photons in the coherent state $|\alpha\rangle$.

$$\langle n\rangle \equiv \langle\alpha|\,\hat{n}\,|\,\alpha\rangle = \exp\left(-|\alpha|^2\right)\sum_{n=0}^{\infty}\frac{(\alpha^*\alpha)^n}{n!}n = \updownarrow n\to n+1\updownarrow =$$
$$= \exp\left(-|\alpha|^2\right)|\alpha|^2\underbrace{\sum_{n=0}^{\infty}\frac{(\alpha^*\alpha)^n}{n!}}_{\exp\left(|\alpha|^2\right)} = |\alpha|^2\ . \tag{4.88}$$

According to (4.88), the average number of photons $\langle n\rangle$ is greater with the larger $|\alpha|$.

Analogously, we find:

$$\left\langle n^2 \right\rangle \equiv \langle \alpha | \, \hat{n}^2 \, | \alpha \rangle = \exp\left(-|\alpha|^2\right) \sum_{n=0}^{\infty} \frac{(\alpha^* \alpha)^n}{n!} n^2 =$$

$$= \exp\left(-|\alpha|^2\right) \sum_{n=0}^{\infty} \frac{\alpha^{2n}}{n!} \{n(n-1) + n\} = |\alpha|^4 + |\alpha|^2 \ . \tag{4.89}$$

Using formulas (4.88) and (4.89), we compute the variance Δn:

$$\Delta n = \sqrt{\langle n^2 \rangle - \langle n \rangle^2} = |\alpha| \ . \tag{4.90}$$

Hence, the relative uncertainty of the number of photons in the coherent state

$$\Delta n / \langle n \rangle = |\alpha|^{-1} = \langle n \rangle^{-1/2} \tag{4.91}$$

is much less than unity for $|\alpha| \gg 1$. The relative uncertainty of the number of photons in the coherent state decreases with an increasing number of photons that form this state.

It can be illustrated that, with the average number of photons in the coherent state being large, the following approximate equalities hold [13]:

$$\langle \cos\varphi \rangle \approx \cos\theta \ , \quad \langle \sin\varphi \rangle \approx \sin\theta \ ,$$

where $|\alpha| \gg 1$, $\alpha = |\alpha| \exp(\mathrm{i}\theta)$.

For a large average number of photons, the coherent state $|\alpha\rangle$ has a minimum product of uncertainties allowed by quantum mechanics:

$$\Delta n \Delta \cos\varphi = (\sin\theta)/2 \ , \quad \Delta n \Delta \sin\varphi = (\cos\theta)/2 \ . \tag{4.92}$$

Compare formulas (4.76) and (4.92).

For $|\alpha| \gg 1$ the relative uncertainties of the number of photons and phase

$$\frac{\Delta n}{\langle n \rangle} \approx \frac{1}{|\alpha|} \ , \quad \frac{\Delta \cos\varphi}{\langle \cos\varphi \rangle} \approx \frac{\mathrm{tg}\theta}{2\,|\alpha|} \tag{4.93}$$

are proportional to $|\alpha|^{-1} = \langle n \rangle^{-1/2}$. The increase in the average number of photons in the coherent state enhances accuracy in determining the amplitude and phase of the electromagnetic wave. The assertion made above can be more clearly evident if one calculates the average value of the electric field strength in the coherent state:

$$\langle E \rangle \equiv \langle \alpha | \, \hat{E} \, | \alpha \rangle =$$

$$= \mathrm{i} \left(\frac{\hbar\omega}{2\varepsilon_0 V} \right)^{1/2} \left\{ \langle \alpha | \, \hat{a} \, | \alpha \rangle \exp[-\mathrm{i}\omega t + \mathrm{i}\vec{k} \cdot \vec{r}] - \langle \alpha | \, \hat{a}^+ \, | \alpha \rangle \exp[\mathrm{i}\omega t - \mathrm{i}\vec{k} \cdot \vec{r}] \right\} =$$

$$= \mathrm{i} \left(\frac{\hbar\omega}{2\varepsilon_0 V} \right)^{1/2} \{ \alpha \exp[-\mathrm{i}\omega t + \mathrm{i}\vec{k} \cdot \vec{r}] - \alpha^* \exp[\mathrm{i}\omega t - \mathrm{i}\vec{k} \cdot \vec{r}] \} =$$

$$= -2 \left(\frac{\hbar\omega}{2\varepsilon_0 V} \right)^{1/2} |\alpha| \sin\left(\vec{k} \cdot \vec{r} - \omega t + \theta \right) \ . \tag{4.94}$$

In calculating this, we have taken into account that:

$$\langle\alpha|\,\hat{a}\,|\alpha\rangle = \langle\alpha|\,(\hat{a}\,|\alpha\rangle) = \langle\alpha|\,\alpha|\,\alpha\rangle = \alpha\overbrace{\langle\alpha|\alpha\rangle}^{1} = \alpha\,, \quad \langle\alpha|\,\hat{a}^{+}\,|\alpha\rangle = \langle\alpha|\,\hat{a}\,|\alpha\rangle^{*} = \alpha^{*}\,.$$

According to (4.94), the average value of the electric field strength $\langle E\rangle$ in the coherent state corresponds to a classical monochromatic traveling wave with the amplitude:

$$|E_0| = 2\left(\frac{\hbar\omega}{2\varepsilon_0 V}\right)^{1/2}|\alpha|\,,$$

and the phase $(\theta+\pi)$:

$$\langle E\rangle = 2\left(\frac{\hbar\omega}{2\varepsilon_0 V}\right)^{1/2}|\alpha|\sin\left(\vec{k}\cdot\vec{r}-\omega t+\theta+\pi\right)\,. \tag{4.95}$$

Similarly, we calculate the average square of the electric field strength:

$$\left\langle E^2\right\rangle = \frac{\hbar\omega}{2\varepsilon_0 V}\left\{1+4\left|\alpha^2\right|\sin^2\left(\vec{k}\cdot\vec{r}-\omega t+\theta\right)\right\}\,. \tag{4.96}$$

As a result, the dispersion of the average value of the electric field strength is constant:

$$\Delta E \equiv \sqrt{\left\langle E^2\right\rangle - \langle E\rangle^2} = \left(\frac{\hbar\omega}{2\varepsilon_0 V}\right)^{1/2} = \text{const}\,.$$

The relative uncertainty of the average value of the electric field strength $\Delta E/\langle E\rangle$ is proportional to $|\alpha|^{-1}$ and, therefore, it is small for: $|\alpha| = \langle n\rangle^{1/2} \gg 1$.

General Conclusions

1. In the coherent state, the amplitude and phase of the electromagnetic wave are not exactly determined, but their uncertainties are finite, and the product of the uncertainties is minimal.
2. The coherent state $|\alpha\rangle$ is a type of quantum mechanical state of the electromagnetic field, directly related with a classical electromagnetic wave. Quantum mechanical uncertainty effects become unimportant when the average number of photons in the coherent state is much greater than unity.
3. A laser running in the states, far exceeding a certain threshold value, generates coherent state radiation with a very high average number of photons.

4.6 Equilibrium Thermal Radiation and Its Properties

Electromagnetic radiation is the only type of radiation that can be in equilibrium with matter. Equilibrium electromagnetic radiation can be observed in a closed cavity of

volume V at a constant temperature T of the cavity walls. For thermal equilibrium to be established in the radiation, even a small amount of matter in the cavity or material cavity walls is essential. This is because photons are thought to be noninteracting with each other. In this case, the mechanism of establishing equilibrium between the radiation and the wall substance is absorption and emission of photons (electromagnetic waves) by atoms of the cavity walls. The substance of the walls and the electromagnetic field inside the cavity exchange among themselves by photons in such a way that the dynamic equilibrium establishes at a given temperature. The mount of electromagnetic field energy that is absorbed by the walls returns back into the cavity due to the reflection of previous photons or emissions of new photons by atoms of the matter.

It is possible for *thermal radiation* to exist at any temperature. All bodies above absolute zero temperature emit thermal radiation.

The properties of equilibrium electromagnetic radiation are listed below:

1. uniformity (coordinate independent radiation)
2. isotropy (direction independent radiation)
3. nonpolarizability
4. radiation characteristics do not depend on the material the cavity walls are made of, and are defined only by temperature

As far back as 1913, Lorenz showed that the failure of, at least, one of these properties leads to the possibility of constructing a perpetual motion machine of the second kind.

Photons are bosons as particles with integer spin. Consequently, the average number of thermally equilibrium photons with the energy $\hbar\omega = h\nu$ in an ideal gas is determined by the Bose–Einstein distribution:

$$f(\nu) = \frac{1}{\exp[(h\nu - \mu)/k_{\mathrm{B}}T] - 1}\,, \tag{4.97}$$

where μ is the chemical potential.

Here, it is worth noting the following circumstance. In some respects, photons are called quasiparticles because they do not behave like normal particles. In particular, the chemical potential μ of real particles is regulated by the condition of conservation of the total number of particles in the system. In the case of electromagnetic radiation, it makes no sense to talk about conservation of the total number of photons; their number is not constant and may vary. When interacting with atoms of cavity walls, photons appear and disappear. The closed radiation cavity of volume V and at a given temperature T contains the average number of photons N driven by the laws of thermodynamics. The condition of thermodynamic equilibrium corresponds to the minimum free energy F of the electromagnetic field:

$$\left.\frac{\partial F}{\partial N}\right|_{T,V=\mathrm{const}} = 0\,.$$

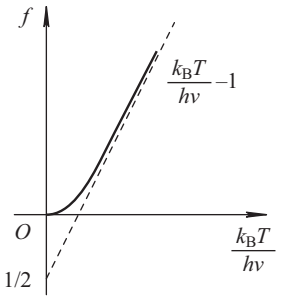


Fig. 4.5: The dependence of the Bose–Einstein photon distribution on the parameter $k_B T/h\nu$.

On the other hand, according to the laws of thermodynamics, the chemical potential of photons is:

$$\mu = \left.\frac{\partial F}{\partial N}\right|_{T,V=\text{const}} .$$

Thus, the thermal equilibrium of photons meets the vanishing of the chemical potential μ. As for an ideal Bose gas of free particles, the chemical potential is not equal to zero ($\mu \neq 0$).

Eventually, the Bose–Einstein distribution for the photon system takes the form (Figure 4.5):

$$f(\nu) = \frac{1}{\exp(h\nu/k_B T) - 1} . \tag{4.98}$$

The unit volume of the electromagnetic field in the frequency range $(\nu, \nu + \mathrm{d}\nu)$ comprises

$$D(\nu)\mathrm{d}\nu = 8\pi\nu^2 \mathrm{d}\nu/c^3 \tag{4.99}$$

allowed states for photons.

Since $f(\nu)$ is the average number of photons with energy $\varepsilon = h\nu$ at a given temperature T, the total number of the photons per unit volume at the temperature T in the frequency range of the electromagnetic radiation from ν to $(\nu + \mathrm{d}\nu)$ can be calculated as:

$$\frac{\mathrm{d}N}{V} = f(\nu)8\pi\nu^2 \mathrm{d}\nu/c^3 . \tag{4.100}$$

To derive the relation (4.100), we have multiplied expression (4.98) by (4.99).

The total number of photons in the cavity can be obtained by integrating the relation (4.100) over all frequencies:

$$\frac{N}{V} = \frac{8\pi}{c^3}\int_0^\infty \frac{\nu^2 \mathrm{d}\nu}{\exp(h\nu/k_B T) - 1} = \left(\frac{k_B T}{\hbar c}\right)^3 \underbrace{\frac{1}{\pi^2}\int_0^\infty \frac{x^2 \mathrm{d}x}{\exp(x) - 1}}_{0.244} .$$

Thus, the total number of photons in the cavity of volume V is determined by the relation:

$$N = 0.244V\,(k_B T/\hbar c)^3 .$$

If one multiplies the equality (4.100) by a single photon's energy $h\nu$, we obtain the energy of the thermal radiation per unit volume of the cavity in the frequency range $(\nu, \nu + \mathrm{d}\nu)$:

$$\rho(\nu)\mathrm{d}\nu = \frac{8\pi h\nu^3 \mathrm{d}\nu}{c^3\,[\exp(h\nu/k_\mathrm{B}T) - 1]}\,. \tag{4.101}$$

The function $\rho(\nu)$ is called the *spectral density* of the *energy of thermal electromagnetic radiation*.

Suppose a small body be inside the cavity with equilibrium electromagnetic radiation. This would not change the velocity of propagation of electromagnetic waves outside the body. The body would reflect and absorb incident photons and would also radiate away new ones. All in all, the body comes into equilibrium with the radiation. Due to isotropy of thermal radiation, each element of the volume radiates away uniformly in all directions.

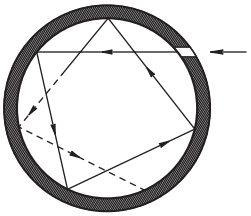


Fig. 4.6: Model of a black body.

Moreover, as an additional idealization, we assume that *the body is absolutely black*. That is, it completely absorbs all the incident photons (Figure 4.6). Under dynamic equilibrium conditions in the system, the black body must emit exactly the same amount of photons as absorbed. Consider a small area $\mathrm{d}S$ in the surface of the black body with the outer normal $\vec{n}$. We calculate the number of photons emitted by the area back into the cavity in a direction that forms an angle θ with the normal for a time $\mathrm{d}t$. The emission of photons towards the cavity rather than inside the material is our concern. Therefore, the angle θ ranges within $0 \le \theta \le \pi/2$.

Since the photons move outside the body at the speed c, they will find themselves within the inclined cylinder depicted in Figure 4.7 in the time $\mathrm{d}t$. The volume of the cylinder is $V_\mathrm{cyl} = \mathrm{d}Sh = \mathrm{d}Sc\mathrm{d}t\cos\theta$.

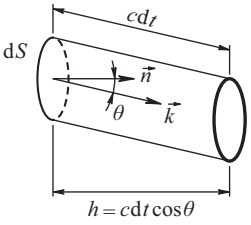


Fig. 4.7: Area $\mathrm{d}S$ in the surface of an absolute black body emitted photons: $\vec{n}$ is the outer normal to the area, $\vec{k}$ is the wave vector of a photon, θ is an angle between the vectors $\vec{n}$ and $\vec{k}$, $h = c\mathrm{d}t\cos\theta$ is the high of the cylinder.

The density of the photons within the frequency range $(\nu, \nu+\mathrm{d}\nu)$ is given by (4.100). Therefore, the area emits

$$\mathrm{d}\widetilde{N} = V_{\mathrm{cyl}}\frac{\mathrm{d}N}{V} = \mathrm{d}S c\mathrm{d}t\cos\theta\left[8\pi\nu^2/c^3\right]f(\nu)\mathrm{d}\nu$$

photons in the specified direction for the time $\mathrm{d}t$.

The number of photons emitted per unit area of the body surface per unit time along the direction corresponding to the angle θ is equal to:

$$\frac{\mathrm{d}\widetilde{N}}{\mathrm{d}t\mathrm{d}S} = \cos\theta[8\pi\nu^2/c^3]f(\nu)\mathrm{d}\nu\ . \tag{4.102}$$

The energy radiated in this direction can be obtained by multiplying formula (4.102) by the energy of one photon $h\nu$:

$$[8\pi/c^3]\cos\theta h\nu^3 f(\nu)\mathrm{d}\nu\ . \tag{4.103}$$

To calculate the energy radiated by the area in all directions, we should average expression (4.103) over all possible values of the angle θ. For this we note that the photon wave vector $\vec{k}$ corresponding to the angle θ lies within a narrow solid angle $\mathrm{d}\Omega = \sin\theta\mathrm{d}\theta\mathrm{d}\varphi$, where the angle φ controls the position of the projection of the photon wave vector $\vec{k}$ in the area $\mathrm{d}S$ (Figure 4.8). The angles θ and φ vary in the following interval:

$$0 \le \varphi \le 2\pi,\ 0 \le \theta \le \pi/2\ .$$

The full solid angle that surrounds the area $\mathrm{d}S$ amounts to 4π. The probability of dropping the photon wave vector into the angle $\mathrm{d}\Omega$ is equal to $\mathrm{d}\Omega/4\pi$. Hence we can get an average value of $\cos\theta$:

$$\langle\cos\theta\rangle = \int\cos\theta\frac{\mathrm{d}\Omega}{4\pi} = \frac{1}{4\pi}\int_0^{2\pi}\mathrm{d}\varphi\int_0^{\pi/2}\cos\theta\sin\theta\mathrm{d}\theta = \frac{1}{4}\ . \tag{4.104}$$

After averaging expression (4.103) over the possible values of the angle θ, we derive the final formula for the energy emitted per unit area of a black body for 1 sec in the frequency range $(\nu, \nu + \mathrm{d}\nu)$:

$$I(\nu)\mathrm{d}\nu = \frac{2\pi h\nu^3\mathrm{d}\nu}{c^2\left[\exp(h\nu/k_{\mathrm{B}}T) - 1\right]}\ . \tag{4.105}$$

The function $I(\nu)$ is referred to as the *spectral intensity of thermal radiation*. Formula (4.105) is the famous Planck formula.

Classical physics faced insurmountable difficulties in trying to theoretically explain the experimentally obtained curve of the spectral intensity of the thermal radiation. The graph of the function $I(\nu)$ (Figure 4.9) highlights two features: $I(\nu) \to 0$ when $\nu \to 0$ and $I(\nu) \to 0$ when $\nu \to \infty$.

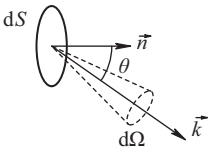


Fig. 4.8: Photon wave vector $\vec{k}$ lies within the solid angle dΩ.

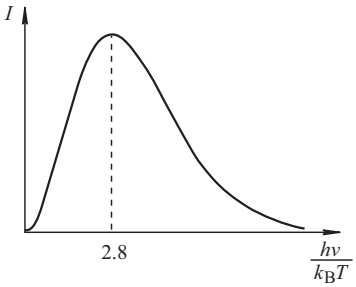


Fig. 4.9: Spectral intensity of thermal radiation I as a function of the parameter $h\nu/k_B T$.

The reason for $I(\nu)$ as $\nu \to 0$ declining is simple enough and can be understood without resorting to quantum theory. The radiation wavelength and its frequency are linked by the ratio $\lambda = c/\nu$. When $\nu \to 0$, the length of the electromagnetic wave tends to infinity ($\lambda \to \infty$). So, radiation is hard to "push" into a volume which has a size smaller than the wavelength. However, classical physics gave no explanations regarding the experimental fact that $I(\nu) \to 0$ as $\nu \to \infty$. The theory imposed no restrictions to radiation at high frequencies (short wavelengths) and resulted in a catastrophically false outcome: $I(\nu) \to \infty$ as $\nu \to \infty$. This paradox was called the "ultraviolet catastrophe."

According to quantum theory, the number of very high energy photons in radiation at any given temperature is low (4.98). Ultimately, it is the reason that $I(\nu) \to 0$ as $\nu \to \infty$.

The frequency $\nu_{\max}$, when the function $I(\nu)$ has a maximum, and temperature are related as:

$$h\nu_{\max}/k_B T = 2.82\,. \tag{4.106}$$

Expression (4.106) is one of the ways of writing *Wien's displacement law* (1893).

The total energy radiated by a black body for 1 sec within the entire frequency range can be estimated through the following integral:

$$I_0 = \int_0^\infty I(\nu)\mathrm{d}\nu = \frac{2\pi\,(k_B T)^4}{c^2 h^3} \underbrace{\int_0^\infty \frac{x^3\mathrm{d}x}{\exp(x)-1}}_{\pi^4/15} = \frac{\pi^2 k_B^4}{60\hbar^3 c^2} T^4 \equiv \sigma T^4\,. \tag{4.107}$$

The quantity I_0 is called the *total intensity of emission* of a black body. The proportionality of I_0 to the fourth power of temperature provides the *Stefan–Boltzmann law* for radiation (1879).

The theoretically calculated value of the constant σ is equal to:

$$\sigma \equiv \frac{\pi^2 k_B^4}{60\hbar^3 c^2} = 5.67 \cdot 10^{-5} \frac{\mathrm{erg}}{\mathrm{sm}^2 \cdot \mathrm{sec} \cdot \mathrm{K}^4}\,,$$

and in good agreement with experimental data.

4.7 The Einstein Coefficients: Spontaneous and Induced Energy Transitions of an Atomic System Under an Electromagnetic Field

Speculating about the Planck formula, Einstein came to a curious conclusion. To do so, he needed no formalism of quantum mechanics. Let us reproduce a logical sequence of his mental reasoning.

Assume that the cavity contains a gas consisting of identical atoms, with each atom being in only two quantum mechanical states with energies ε_1 and ε_2 ($\varepsilon_1 < \varepsilon_2$). Imagine $N_1(N_2)$ is the total number of the atoms per unit volume with energy ε_1 (ε_2). In thermal equilibrium, the atoms are distributed over the energy levels ε_1 and ε_2, in accordance with the Boltzmann formula:

$$\frac{N_2}{N_1} = \exp\left[-\frac{\varepsilon_2 - \varepsilon_1}{k_B T}\right] . \tag{4.108}$$

If $\varepsilon_1 < \varepsilon_2$, then $N_1 > N_2$.

Simultaneously, in thermal equilibrium the total number of atoms transitioning from the energy level two to level one for 1 sec must be equal to the number of atoms traveling backwards (Figure 4.10).

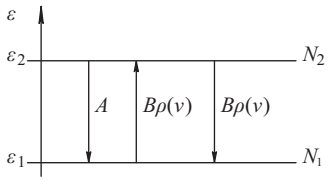


Fig. 4.10: Scheme of atomic energy levels and three ways to change the state of an atom.

The aforementioned follows from the fact that, in equilibrium, the numbers N_1 and N_2 are unchangeable over time. The specified requirements do not contradict the inequality $N_1 > N_2$ only if the probability of down transition of an individual atom for 1 sec ($2 \rightarrow 1$) is greater than the probability of its up transition for 1 sec ($1 \rightarrow 2$). Hereinafter, for the sake of simplicity, the probability of the transition of an atom for 1 sec will be called the transition rate.

Einstein guessed that the formulas for the atomic transition rates have the following algebraic structure:

$$W_{2\rightarrow 1} = B\rho(\nu) + A , \quad W_{1\rightarrow 2} = B\rho(\nu) . \tag{4.109}$$

Such a form of writing, as expected, yields the transition rate from the second level to the first one. It is found to be greater than the reverse transition rate by a positive constant A. From a physical point of view, Einstein revealed two types of *energy transitions* of atoms: *spontaneous* and *induced* by an electromagnetic field.

If an atom is in the excited state 2, even in the absence of an electromagnetic field in the cavity, there is a finite probability A of its spontaneous transition for 1 sec to the ground state 1. Vacuum fluctuations make the atom go from the excited state to its ground state, emitting a photon with the energy $h\nu = \varepsilon_2 - \varepsilon_1$.

Electromagnetic thermal radiation with the spectral density $\rho(\nu)$ in the cavity, if any, causes additional field induced atomic energy transitions; both downwards and upwards. Moreover, according to Einstein, the transition rates are equal and proportional to the field characteristic $\rho(\nu)$ with the proportionality coefficient B. Thus, the field induced transition rates are the same, and equal to $B\rho(\nu)$.

At this point it is worth emphasizing that the Einstein coefficients A and B are defined so that they do depend only on the properties of atomic states rather than the function $\rho(\nu)$.

Let us try to reproduce the Planck formula using the relations (4.109). In thermal equilibrium, the population of the levels N_1 and N_2 is constant. In this case, the upward transition rate of all of N_1 atoms $N_1 W_{1\to 2} = N_1 B\rho(\nu)$ must be equal to the downward transition rate of all of N_2 atoms: $N_2 W_{2\to 1} = N_2[A + B\rho(\nu)]$. Hence, we arrive at the equality:

$$N_1 B\rho(\nu) = N_2[A + B\rho(\nu)] \,. \tag{4.110}$$

Since $h\nu = \varepsilon_2 - \varepsilon_1$, the relation (4.108) can be rewritten as:

$$\frac{N_2}{N_1} = \exp\left[-\frac{h\nu}{k_B T}\right] \,. \tag{4.111}$$

Formulas (4.110) and (4.111) imply that:

$$\rho(\nu) = \frac{A/B}{\exp(h\nu/k_B T) - 1} \,. \tag{4.112}$$

This result should be consistent with Planck's formula (4.101) for the spectral density of the thermal energy of electromagnetic radiation. Expressions (4.112) and (4.101) coincide at all temperatures T if the ratio of the Einstein coefficients A and B satisfies the condition:

$$\frac{A}{B} = \frac{8\pi h\nu^3}{c^3} \,. \tag{4.113}$$

The chain of Einstein's reasoning is valuable because it discloses the existence of two types of energy transitions in a photon system and atoms of matter. In thermal equilibrium, these transitions are balanced, but what happens when the balance is disturbed? To illustrate this, we consider a beam of electromagnetic waves passing through the material inside the cavity. We discuss two possible cases.

1. Suppose all the atoms of matter are initially in their ground state. After passing a beam of electromagnetic waves through the material, a part of the induced atoms rises up from the ground state to an excited one. Other induced transitions may help these matter atoms go back to the ground state, and thereby return the energy quanta to the electromagnetic wave. In addition, the atoms of matter may pass spontaneously

from the excited state to the ground one. The spontaneous transitions have nothing to do with the beam of the external electromagnetic waves. Therefore, the resulting photons are scattered in a random direction. As a consequence, the electromagnetic wave beam leaves the cavity, losing a part of its energy. Such a situation is not abnormal; the matter absorbs the electromagnetic wave energy. In truth, the chosen event name is not very good, because the atoms of matter, in fact, dissipate the electromagnetic wave energy rather than absorb it.

2. Let us present a more interesting case. Imagine a beam of electromagnetic radiation is targeted in pre-excited matter atoms. We choose the frequency ν of the external electromagnetic radiation so that the condition $h\nu = \varepsilon_2 - \varepsilon_1$ should be fulfilled. Then, even weak external radiation causes induced energy transitions of the matter atoms. The atoms pass into the ground state and enrich the electromagnetic wave by energy quanta $h\nu = \varepsilon_2 - \varepsilon_1$. Since the induced transition rate $B\rho(\nu)$ is proportional to the electromagnetic field energy in the material, more and more atoms of the material will transfer their energy to the beam. As a result, the matter releases a powerful electromagnetic pulse, even if the input external radiation were initially weak. This was the general idea of a laser that belonged to Einstein.

Here, it is worth pointing out that the Einstein coefficient A is easy to measure. Let us look at doing it.

Suppose all the atoms of the substance to be initially excited (the state 2) in the absence of an external electromagnetic field is ($\rho(\nu) = 0$). Since the matter atoms can only spontaneously transition from state 2 to state 1, the mean population of the excited atoms should decay according to the law:

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = -N_2 A \,. \tag{4.114}$$

The solution of the differential equation (4.114) appears as:

$$N_2(t) = N_2(0)\exp(-t/t_{\text{spont}}) \,, \tag{4.115}$$

where $t_{\text{spont}} = 1/A$; $N_2(0)$ is the mean population of the atoms at the initial time.

Thus, the Einstein coefficient A is equal to the inverse lifetime t_{spont} of an excited atom with respect to the spontaneous transition. Since each transition is accompanied by spontaneous emission of a photon with the frequency $\nu = (\varepsilon_2 - \varepsilon_1)/h$, the change in the number N_2 may be traced. Consequently, the coefficient A can be measured. The main idea of the method is to put a large number of atoms into the excited state followed by determination of the time dependence of the spontaneous emission intensity:

$$I(t) = \text{const}\, N_2(t) \,. \tag{4.116}$$

From formulas (4.115) and (4.116) we get:

$$\ln I = -At + \text{const} \,. \tag{4.117}$$

The coefficient A can be calculated from the angular coefficient of the linear dependence of $\ln I$ on t: $\mathrm{tg}\varphi = A$ (Figure 4.11).

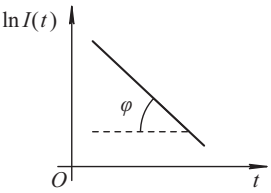


Fig. 4.11: The dependence of the natural logarithm of the spontaneous emission intensity ln *I* on time.

4.8 Interaction Between a Quantized Electromagnetic Field and a Two Level Atom – the Electric Dipole Approximation

To clarify the operating conditions for a quantum generator of radiation (laser), we should be familiar with the foundations of quantum theory that describe the photon-matter interaction.

Consider the model problem on the interaction of a quantized electromagnetic field with a single atom. For simplicity, the Hamiltonian $\widehat{H}_{\text{atom}}$ is assumed to have only two eigenvectors $|i\rangle$ $(i = 1, 2)$. The latter correspond with two quantum mechanical states of the atom; the ground state with the energy ε_1 and the excited state with the energy ε_2 $(\varepsilon_2 > \varepsilon_1)$:

$$\widehat{H}_{\text{atom}} |i\rangle = \varepsilon_i |i\rangle \quad (i = 1, 2) \ .$$

The state vectors of the atom are assumed to be orthonormalized: $\langle i|j\rangle = \delta_{ij}$.

The use of the completeness condition for eigenvectors of the atom Hamiltonian

$$\sum_{i=1,2} |i\rangle \langle i| = |1\rangle \langle 1| + |2\rangle \langle 2| = \widehat{I}$$

leads to the identity:

$$\widehat{H}_{\text{atom}} = \widehat{I} \cdot \widehat{H}_{\text{atom}} \cdot \widehat{I} = \sum_{i=1,2} |i\rangle \langle i| \widehat{H}_{\text{atom}} \sum_{j=1,2} |j\rangle \langle j| = \sum_{i,j=1,2} |i\rangle \underbrace{\langle i| \widehat{H}_{\text{atom}} |j\rangle}_{\varepsilon_i \delta_{ij}} \langle j| =$$

$$= \sum_{i=1,2} \varepsilon_i |i\rangle \langle i| \ .$$

Consider the action of the operator $|i\rangle\langle j|$ on a certain state $|l\rangle$ of the atom. In virtue of the orthonormality of eigenvectors we have:

$$| \ i\rangle \langle j|l\rangle = | \ i\rangle \delta_{jl} \ . \tag{4.118}$$

Consequently, the application of the operator $|i\rangle\langle j|$ to the atom state $|l\rangle$ gives the state $|i\rangle$ providing that the initial state is $|j\rangle$ and zero in all the rest of the cases. In other words, the operator $|i\rangle\langle j|$ generates the state $|i\rangle$ and destroys the state $|j\rangle$. By this ideology, the operator $|i\rangle\langle j|$ is usually denoted similarly to creation and annihilation operators of photons. Let us designate the creation and annihilation operators of the atomic state $|i\rangle$ as $\widehat{b}_i^+$ and $\widehat{b}_i$, and introduce the multiplicative combination $\widehat{b}_i^+ \widehat{b}_j \equiv |i\rangle\langle j|$. Equation (4.118) can be written as:

$$|i\rangle \langle j|l\rangle \equiv \widehat{b}_i^+ \widehat{b}_j |l\rangle = |i\rangle \delta_{ji} \ . \tag{4.119}$$

Note that the commutator of the bilinear combinations $\hat{b}_i^+ \hat{b}_j$ gives the same operators. This is easily demonstrated by the chain of equalities:

$$\left[\hat{b}_i^+ \hat{b}_j \, , \, \hat{b}_r^+ \hat{b}_s\right] = |i\rangle \langle j|r\rangle \langle s| - |r\rangle \langle s|i\rangle \langle j| = |i\rangle \delta_{jr} \langle s| - |r\rangle \delta_{si} \langle j| = \hat{b}_i^+ \hat{b}_s \delta_{jr} - \hat{b}_r^+ \hat{b}_j \delta_{si} \, .$$

The atom-photon interaction in the visible spectrum gives birth to no new electrons and does not annihilate the old ones. In this case, only the electron transition from one state to another is possible. Emanating from the foregoing, any atomic operators can always be written in terms of combinations of $\hat{b}_i^+ \hat{b}_j$. In particular,

$$\hat{H}_{\text{atom}} = \sum_{i=1,2} \varepsilon_i |i\rangle \langle i| = \sum_{i=1,2} \varepsilon_i \hat{b}_i^+ \hat{b}_i = \sum_{i=1,2} \varepsilon_i \hat{n}_i \, . \tag{4.120}$$

According to (4.119), the eigenvalues of the operator $\hat{n}_i = \hat{b}_i^+ \hat{b}_i = |i\rangle\langle j|$ are equal to 0, 1. This provides a reason for treating the operator $\hat{n}_i$ as the operator of the number of quasiparticles (fermions). We are now at the stage where we can theoretically describe the energy states of a system of isolated atoms in terms of an ideal fermion gas.

The transformation of the initial Hamiltonian to the form (4.120) is called second quantization of a Hamiltonian. The name reflects the fact that, to obtain such a representation, the usual procedure first needs to be carried out for determining the quantum mechanical stationary states of the system and its energy levels. This procedure can be regarded as first quantization of electron motion in an atom. The second quantization writes down the Hamiltonian through known stationary states. Ultimately, the quasiparticle creation and annihilation operators $\hat{b}_i^+ \hat{b}_j$ appear.

The atom Hamiltonian in the form of the second quantization is suitable for calculating the interaction of an atom with any physical system, such as a radiation field. Recall that the Hamiltonian of a radiation field has the following form:

$$\hat{H}_{\text{rad}} = \sum_{\lambda, \vec{k}} \hbar \omega_k \hat{a}_{\vec{k}\lambda}^+ \hat{a}_{\vec{k}\lambda} \, . \tag{4.121}$$

Here, we have omitted the zero point energy of vacuum fluctuations as the measure level for the photon field energy.

So, the second quantization delineates both the photons and the atomic states in terms of creation and annihilation operators of some quasiparticles. With the quasiparticles being different, the groups of the operators $\{\hat{a}_{\vec{k}\lambda'}^+, \hat{a}_{\vec{k}\lambda}\}$ and $\{\hat{b}_i^+, \hat{b}_j\}$ commute with each other.

The full Hamiltonian that describes the system, "an atom + radiation," should be:

$$\hat{H} = \hat{H}_{\text{atom}} + \hat{H}_{\text{rad}} + \hat{H}_{\text{int}} \, .$$

Before writing down the atom-electromagnetic field interaction Hamiltonian $\hat{H}_{\text{int}}$, we turn first to classical physics. Consider an atom consisting of a nucleus with a positive charge $Z|e|$ surrounded by Z-electrons, each of which has a charge $-|e|$. Typical values of the radii of atomic electron orbits are determined by the magnitude of the

Bohr radius: $a_B = 4\pi\varepsilon_0\hbar^2/(me^2) \approx 5\cdot10^{-11}$ m (in the SI system), where m is the mass of an electron. If the frequency of the external electromagnetic wave incident on an atom does not exceed 10^{18} Hz, its wavelength is much larger than a_B. Therefore, the spatial variation of the electric and magnetic fields within the atom can be neglected. Also, the main interaction between the atomic electrons and electromagnetic radiation reduces to the electron-electric field radiation interaction. The impact of the magnetic field of the wave on the electrons is weak enough. It amounts to $O(v/c)$ of the electric field interaction, where v is typical velocity of an electron in an atom and c is speed of light. Thus, the influence of the magnetic field of the electromagnetic wave on the atom in the main approximation can be ignored.

Here is a shortcut derivation of the expression for the atom-radiation interaction energy within the electric dipole approximation. Suppose the atom's nucleus, with the charge $Z|e|$, be located at the origin of the Cartesian coordinate system and that electrons have radius vectors $\vec{r}_i$ ($i = 1, 2, \ldots, Z$). According to the laws of electrostatics, the potential energy of interaction between electrons and an external electric field can be calculated as the work done by an external force acting to pull each of the Z negative charge $-|e|$ (originally located at the point of the nucleus) towards their actual positions along the radius vectors $\vec{r}$ (Figure 4.12). In the calculations, the Coulomb interactions between the charges are, of course, left aside because only the potential energy of the system of charges in the external field $\vec{E}(\vec{r}, t)$ needs to be determined:

$$H_{\text{int}} = \sum_{i=1}^{Z} \left(\vec{r}_i \cdot \vec{F}\right) = |e| \sum_{i=1}^{Z} \left(\vec{r}_i \cdot \vec{E}(0, t)\right) .$$

Note that here we have used Newton's third law. For the i-th electron to move, the external force $\vec{F} = |e|\vec{E}$ is necessary to apply. This is because the force $-|e|\vec{E}$ of the electromagnetic field already acts on the electron. Calculating the energy H_{int}, we have taken into account the fact that the electric field $\vec{E}$ weakly changes within the atom. Therefore, the field can be thought of, with good accuracy, as being calculated at a single point – at the point that corresponds to the radius vector $\vec{r} = 0$ of the atomic nucleus. The quantity $\vec{P} = -|e| \sum_{i=1}^{Z} \vec{r}_i$ is called the dipole moment of an atom. For further analysis, we separate the coordinate dependence and write down the atom's

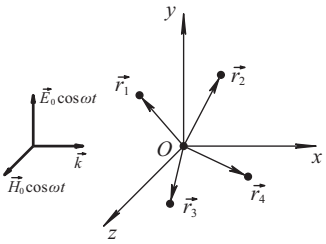


Fig. 4.12: The coordinate system used to describe an atom and an electromagnetic wave.

dipole moment in the form:

$$\vec{P} = -|e|\,\vec{D}\,, \quad \vec{D} = \sum_{i=1}^{Z} \vec{r}_i\,.$$

Thus, classical physics gives the following Hamiltonian of the interaction between an atom and an electromagnetic field:

$$H_{\text{int}} = -\left(\vec{P}\cdot\vec{E}\right) \equiv |e|\left(\vec{D}\cdot\vec{E}\right)\,. \tag{4.122}$$

Proceeding to quantum theory, we should replace the dynamic variables $\vec{D}$ and $\vec{E}$ by the operators $\widehat{\vec{D}}$ and $\widehat{\vec{E}}$. In doing so, we express these operators in terms of the creation and annihilation operators $\{\hat{a}^{+}_{\vec{k}'\lambda'}, \hat{a}_{\vec{k}\lambda}\}$ and $\{\hat{b}^{+}_{i}, \hat{b}_{j}\}$. According to the results obtained previously, we have the following for the electric field operator $\widehat{\vec{E}}$:

$$\widehat{\vec{E}}(0,t) = \mathrm{i}\sum_{\vec{k},\lambda}\sqrt{\frac{\hbar\omega_k}{2\varepsilon_0 V}}\,\vec{e}_{\vec{k}\lambda}\left(\hat{a}_{\vec{k}\lambda} - \hat{a}^{+}_{\vec{k}\lambda}\right)\,.$$

The summation runs over all the wave vectors by taking two possible polarizations of each mode: $\lambda = 1, 2$.

By applying the completeness condition for the atomic states twice, we obtain the following representation of the operator $\widehat{\vec{D}}$:

$$\widehat{\vec{D}} = \hat{I}\cdot\widehat{\vec{D}}\cdot\hat{I} = \sum_{i=1,2}|i\rangle\langle i|\,\widehat{\vec{D}}\sum_{j=1,2}|j\rangle\langle j| = \sum_{i,j=1,2}\vec{D}_{ij}\,|i\rangle\langle j| = \sum_{i,j=1,2}\vec{D}_{ij}\hat{b}^{+}_{i}\hat{b}_{j}\,,$$

where $\vec{D}_{ij} = \langle i|\widehat{\vec{D}}|j\rangle$.

Note that the quantity $\vec{D} = \sum_{i=1}^{Z}\vec{r}_i$ changes its sign under spatial inversion (when changing $\vec{r}_i \to -\vec{r}_i$, where $i = 1, 2, \ldots, Z$). Suppose each of the quantum states of the atom possessed certain evenness. In other words, suppose it either changes its sign or does not, under the inversion of all the electron coordinates, with the state evenness $|1\rangle$ and $|2\rangle$ being opposed. This would result in the equalities $\langle 1|\widehat{\vec{D}}|1\rangle \equiv \vec{D}_{11} = 0$, $\langle 2|\widehat{\vec{D}}|2\rangle \equiv \vec{D}_{22} = 0$ being valid. In this case, the off diagonal matrix elements of the operator $\widehat{\vec{D}}$ do not have to be necessarily zero:

$$\begin{aligned}\langle 1|\sum_{k=1}^{Z}\vec{r}_k\,|2\rangle &= \langle 2|\sum_{k=1}^{Z}\vec{r}_k\,|1\rangle^{*} = \\ &= \int \mathrm{d}^3\vec{r}_1\mathrm{d}^3\vec{r}_2\ldots\mathrm{d}^3\vec{r}_Z\Psi_1^{*}(\vec{r}_1,\vec{r}_2,\ldots\vec{r}_z)\sum_{k=1}^{Z}\vec{r}_k\Psi_2(\vec{r}_1,\vec{r}_2,\ldots\vec{r}_z) = \\ &= D_{12} = D_{21}^{*}\,.\end{aligned}$$

Here, $\Psi_i(\vec{r}_1,\vec{r}_2,\ldots,\vec{r}_Z) = \langle\vec{r}_1,\vec{r}_2,\ldots,\vec{r}_Z|i\rangle$ $(i = 1, 2)$ are the wave functions of the atomic states in the coordinate representation. The integration is performed over the coordinates of all the electrons in the atom.

Furthermore, to simplify the analysis, we input $\vec{D}_{12} = \vec{D}_{21}$. This is true for the real wave functions Ψ_1, Ψ_2, which imposes no limits on the generality of the discussion.

Once we have made the specified simplifications, the interaction operator for the two level atom and the electromagnetic field takes the form:

$$\widehat{H}_{\text{int}} = \mathrm{i}\,|e| \sum_{\vec{k},\lambda} \sqrt{\frac{\hbar\omega_k}{2\varepsilon_0 V}} \left(\vec{e}_{\vec{k}\lambda} \cdot \vec{D}_{12}\right) \left(\widehat{a}_{\vec{k}\lambda} - \widehat{a}^{+}_{\vec{k}\lambda}\right) \left(\widehat{b}^{+}_1 \widehat{b}_2 + \widehat{b}^{+}_2 \widehat{b}_1\right) . \tag{4.123}$$

Eventually, we find the complete Hamiltonian of the "atom + radiation" system:

$$\widehat{H} = \widehat{H}_{\text{atom}} + \widehat{H}_{\text{rad}} + \widehat{H}_{\text{int}} \equiv \widehat{H}_0 + \widehat{H}_{\text{int}} ,$$

where $\widehat{H}_0 = \widehat{H}_{\text{atom}} + \widehat{H}_{\text{rad}}$ is the Hamiltonian of an isolated atom and a free electromagnetic field, respectively.

Comments

1. An alternative description of two level atom states can be effected by introducing an energyspin vector operator $\vec{\widehat{R}}$ (*pseudospin*). It is defined by three component $\widehat{R}_1, \widehat{R}_2, \widehat{R}_3$:

$$\widehat{R}_1 = |2\rangle\langle 1| + |1\rangle\langle 2| , \quad \widehat{R}_2 = \mathrm{i}\,(|2\rangle\langle 1| - |1\rangle\langle 2|) , \quad \widehat{R}_3 = |1\rangle\langle 1| - |2\rangle\langle 2| .$$

It is easy to verify that these operators satisfy the algebraic and commutation relations characteristic of the angular momentum operator for spin 1/2:

$$[\widehat{R}_s, \widehat{R}_p] = 2\mathrm{i}\varepsilon_{spq}\widehat{R}_q , \quad \widehat{R}_1^2 = \widehat{R}_2^2 = \widehat{R}_3^2 = \widehat{I} .$$

Here, ε_{spq} is the unit completely antisymmetric Levi–Civita pseudotensor and $\widehat{I} = |1\rangle\langle 1| + |2\rangle\langle 2|$ is the unit operator.

In terms of pseudospin, we have:

$$\widehat{H}_{\text{atom}} = \sum_{i=1,2} \varepsilon_i |i\rangle\langle i| = \frac{\varepsilon_1 + \varepsilon_2}{2}\widehat{I} + \frac{\varepsilon_1 - \varepsilon_2}{2}\widehat{R}_3 ,$$

$$\widehat{H}_{\text{int}} = |e| \left(\vec{D}_{12} \cdot \vec{\widehat{E}}\right) (|2\rangle\langle 1| + |1\rangle\langle 2|) = |e| \left(\vec{D}_{12} \cdot \vec{\widehat{E}}\right) \widehat{R}_1 .$$

It follows that, from a mathematical point of view, the interaction of a two level atom with a radiation field and the behavior of a particle with spin 1/2 in an external magnetic field can be described in an equivalent way. It should be emphasized that, in this case, the physical meaning of the operator $\vec{\widehat{R}}$ is quite different. The vector operator $\vec{\widehat{R}}$ acts in the energy space of the atom rather than in the ordinary coordinate space, or in the space of spin states.

2. When analyzed, the atom-electromagnetic field interactions can be set forth in terms of classical physics, only when the resulting energy change of the system is large, as compared with the energy of each emitted or absorbed photon. Providing the amount of photons is huge at high intensities and low frequencies of the field, the discrete nature of the energy exchange between the atom and the radiation may be left aside. For example, a static electromagnetic field or a radio wave field can be outlined as a given external classical field.

4.9 The Rates of Spontaneous and Induced Atomic Transitions When Electromagnetic Waves Travel through a Medium as Well as Under Thermal Radiation Conditions

Now we illustrate how the quantum interaction Hamiltonian can be applied by calculating the rate of absorption and emission of photons by an atom that may make transitions between the states $|1\rangle$ and $\langle 2|$.

The Schrödinger equation that determines the vector states $|\chi\rangle$ of the "atom + radiation" system has the form:

$$i\hbar\frac{\partial}{\partial t}|\chi\rangle = (\widehat{H}_0 + \widehat{H}_{\text{int}})|\chi\rangle \ . \tag{4.124}$$

With no interaction between the atom and the electromagnetic field, the equation (4.124) has the stationary states acting as the eigenvectors of the operator $\widehat{H}_0 = \widehat{H}_{\text{atom}} + \widehat{H}_{\text{rad}}$:

$$\widehat{H}_0|i, n_{\vec{k}\lambda}\rangle = (\varepsilon_i + \hbar\omega_k n_{\vec{k}\lambda})|i, n_{\vec{k}\lambda}\rangle \ , \quad |i, n_{\vec{k}\lambda}\rangle = |i\rangle|n_{\vec{k}\lambda}\rangle \ .$$

Since the state vectors are $|i\rangle$ $(i = 1, 2)$ and $\{|n_{\vec{k}\lambda}\rangle\}$ for the isolated atom and the radiation field, respectively, we have chosen orthonormalized

$$\langle i|j\rangle = \delta_{ij} \ , \quad \left\langle n_{\vec{k},\lambda}|n_{\vec{k}'\lambda'}\right\rangle = \delta_{\vec{k}\vec{k}',\lambda\lambda'} \ ,$$

the vectors $\{|i, n_{\vec{k}\lambda}\rangle\}$ also prove to be orthonormalized:

$$\langle i, n_{\vec{k}\lambda}|j, n_{\vec{k}'\lambda'}\rangle = \delta_{\vec{k}\vec{k}',\lambda\lambda'}\delta_{ij} \ .$$

Furthermore, for simplicity, we confine ourselves to the consideration of the interaction of the atom with only a single mode of the electromagnetic field. The quantum numbers $\vec{k}$ and λ (the wave vector of the electromagnetic wave and its polarization) characterize this mode. Suppose $n_{\vec{k}\lambda}$ is a certain number of photons in the mode. Moreover, suppose the electromagnetic wave frequency ν is close to the frequency $(\varepsilon_2 - \varepsilon_1)/h$ of the atomic transition between the two states. Such transitions play a major role in energy benefits. For the atom-electromagnetic wave interaction, there are two ways of solving the equation (4.124), due to the energy closeness ($h\nu \approx \varepsilon_2 - \varepsilon_1$).

Variant 1:

$$|\chi(t)\rangle = c_1(t)\,|1, n_{\vec{k}\lambda}\rangle + c_2(t)\,|2, n_{\vec{k}\lambda} - 1\rangle \ , \tag{4.125}$$

where

$$c_1(0) = 1, c_2(0) = 0 \ .$$

The approximation (4.125) is suitable when the "atom + radiation" system, originally ($t = 0$) in the state $|1, n_{\vec{k},\lambda}\rangle$, undergoes an induced transition into the state $|2, n_{\vec{k},\lambda} - 1\rangle$ as $t > 0$. The superposition (4.125) contains ket-vectors of $|1, n_{\vec{k}\lambda}\rangle$ and $|2, n_{\vec{k}\lambda} - 1\rangle$ with their weights $c_i(t)$ ($i = 1, 2$). They allow one to find the probability of detecting the atom and the radiation mode in the appropriate quantum states.

Variant 2:

$$|\chi(t)\rangle = c_1(t)\,|1, n_{\vec{k}\lambda} + 1\rangle + c_2(t)\,|2, n_{\vec{k}\lambda}\rangle \ . \tag{4.126}$$

The approximation (4.126) suggests the initial condition

$$c_2(t = 0) = 1, c_1(t = 0) = 0 \ .$$

Formula (4.126) is used when the atom is initially ($t = 0$) in the excited state $|2\rangle$, and the electromagnetic field contains $n_{\vec{k}\lambda}$ photons. Following on, the atom passes into the ground state $|1\rangle$ as $t > 0$, and the electromagnetic field acquires an additional photon.

The ket-vector $|\chi(t)\rangle$ is assumed to be normalized in both cases:

$$\langle\chi(t)|\chi(t)\rangle = |c_2(t)|^2 + |c_1(t)|^2 = 1 \ . \tag{4.127}$$

Formula (4.127) has limitations for the functions $c_i(t)$. Further calculations make it possible to trace these limitations.

As we describe the interaction of the atom with a single mode of the electromagnetic field, the operator $\widehat{H}_{\text{int}}$ has the form:

$$\widehat{H}_{\text{int}} = \mathrm{i}\,|e|\sqrt{\frac{\hbar\omega_k}{2\varepsilon_0 V}}\left(\vec{D}_{12}\cdot\vec{e}_{\vec{k}\lambda}\right)\left(\widehat{a}_{\vec{k}\lambda} - \widehat{a}^{+}_{\vec{k}\lambda}\right)\left(\widehat{b}^{+}_1\widehat{b}_2 + \widehat{b}^{+}_2\widehat{b}_1\right) \ ,$$

where $\omega_k = kc$ does not depend on the parameter λ. For simplicity we denote:

$$\hbar g_{\vec{k}} = |e|\sqrt{\frac{\hbar\omega_k}{2\varepsilon_0 V}}\left(\vec{D}_{12}\cdot\vec{e}_{\vec{k}\lambda}\right) \ .$$

Then we have:

$$\widehat{H}_{\text{int}} = \mathrm{i}\hbar g_{\vec{k}}\left(\widehat{a}_{\vec{k}\lambda} - \widehat{a}^{+}_{\vec{k}\lambda}\right)\left(\widehat{b}^{+}_1\widehat{b}_2 + \widehat{b}^{+}_2\widehat{b}_1\right) \ . \tag{4.128}$$

Further analysis requires employing matrix elements of the operator $\widehat{H}_{\text{int}}$ over the states $|i, n_{\vec{k}\lambda}\rangle$. Among them we have only four nonzero elements:

$$\langle 2, n_{\vec{k}\lambda} - 1|\,\widehat{H}_{\text{int}}\,|1, n_{\vec{k}\lambda}\rangle \ , \quad \langle 1, n_{\vec{k}\lambda}|\,\widehat{H}_{\text{int}}\,|2, n_{\vec{k}\lambda} - 1\rangle \ ,$$
$$\langle 2, n_{\vec{k}\lambda}|\,\widehat{H}_{\text{int}}\,|1, n_{\vec{k}\lambda} + 1\rangle \ , \quad \langle 1, n_{\vec{k}\lambda} + 1|\,\widehat{H}_{\text{int}}\,|2, n_{\vec{k}\lambda}\rangle \ .$$

As an example, we calculate the first matrix element:

$$\begin{aligned}\langle 2, n_{\vec{k}\lambda} - 1|\,\widehat{H}_{\text{int}}\,|1, n_{\vec{k}\lambda}\rangle &= \mathrm{i}\hbar g_{\vec{k}}\,\langle 2, n_{\vec{k}\lambda} - 1|\left(\widehat{a}_{\vec{k}\lambda} - \widehat{a}^{+}_{\vec{k}\lambda}\right)|2, n_{\vec{k}\lambda}\rangle = \\ &= \mathrm{i}\hbar g_{\vec{k}}\,\langle 2, n_{\vec{k}\lambda} - 1|\,\widehat{a}_{\vec{k}\lambda}\,|2, n_{\vec{k}\lambda}\rangle = \\ &= \mathrm{i}\hbar g_{\vec{k}}\sqrt{n_{\vec{k}\lambda}}\,\langle 2, n_{\vec{k}\lambda} - 1|2, n_{\vec{k}\lambda} - 1\rangle = \mathrm{i}\hbar g_{\vec{k}}\sqrt{n_{\vec{k}\lambda}}\,. \end{aligned} \tag{4.129}$$

The values of the rest are easy to obtain in a similar way:

$$\begin{aligned}&\langle 2, n_{\vec{k}\lambda}|\,\widehat{H}_{\text{int}}\,|1, n_{\vec{k}\lambda} + 1\rangle = \mathrm{i}\hbar g_{\vec{k}}\sqrt{n_{\vec{k}\lambda} + 1} = -\langle 1, n_{\vec{k}\lambda} + 1|\,\widehat{H}_{\text{int}}\,|2, n_{\vec{k}\lambda}\rangle\ , \\ &\langle 1, n_{\vec{k}\lambda}|\,\widehat{H}_{\text{int}}\,|2, n_{\vec{k}\lambda} - 1\rangle = -\mathrm{i}\hbar g_{\vec{k}}\sqrt{n_{\vec{k}\lambda}}\,. \end{aligned} \tag{4.130}$$

Next, we consider four successively conceivable cases with met interactions between an atom in the ground or excited state and an electromagnetic wave or heat radiation.

Interaction Between an Atom in the Ground State and an Electromagnetic Wave

Substituting the state vector (4.125) into the Schrödinger equation (4.124), we obtain:

$$\begin{aligned}\mathrm{i}\hbar\frac{\partial}{\partial t}\left(c_1(t)\,|1, n_{k\vec{\lambda}}\rangle + c_2(t)\,|2, n_{\vec{k}\lambda} - 1\rangle\right) &= \left(\widehat{H}_0 + \widehat{H}_{\text{int}}\right)\left(c_1(t)\,|1, n_{\vec{k}\lambda}\rangle + \right. \\ &\left. + c_2(t)\,|2, n_{\vec{k}\lambda} - 1\rangle\right)\ . \end{aligned} \tag{4.131}$$

Next we scalarly multiply the equation (4.131) by the bra vectors $\langle 1, n_{\vec{k}\lambda}|$ and $\langle 2, \vec{n}_{\vec{k}\lambda} - 1|$. The account of formulas (4.129) and (4.130) yields a set of differential equations to define the functions $c_i(t)$:

$$\begin{aligned}\mathrm{i}\hbar\frac{\partial}{\partial t}c_1(t) &= (\varepsilon_1 + \hbar\omega_k n_{\vec{k}\lambda})c_1(t) - \mathrm{i}\hbar g_{\vec{k}}\sqrt{n_{\vec{k}\lambda}}\,c_2(t)\ , \\ \mathrm{i}\hbar\frac{\partial}{\partial t}c_2(t) &= \left[\varepsilon_2 + \hbar\omega_k(n_{\vec{k}\lambda} - 1)\right]c_2(t) + \mathrm{i}\hbar g_{\vec{k}}\sqrt{n_{\vec{k}\lambda}}\,c_1(t)\ . \end{aligned} \tag{4.132}$$

It should be solved with the initial condition $c_1(0) = 1$, $c_2(0) = 0$.

The system (4.132) can be simplified by substituting:

$$c_1(t) = \exp\left\{-\frac{\mathrm{i}}{\hbar}\left[\varepsilon_1 + \hbar\omega_k n_{\vec{k}\lambda}\right]t\right\}s_1(t), \quad c_2(t) = \exp\left\{-\frac{\mathrm{i}}{\hbar}[\varepsilon_2 + \hbar\omega_k(n_{\vec{k}\lambda} - 1)]t\right\}s_2(t).$$

After that, it becomes:

$$\begin{aligned}\frac{\partial}{\partial t}s_1(t) &= -g_{\vec{k}}\sqrt{n_{\vec{k}\lambda}}\exp\left\{-\mathrm{i}\left(\omega_{21} - \omega_k\right)t\right\}s_2(t)\ , \\ \frac{\partial}{\partial t}s_2(t) &= g_{\vec{k}}\sqrt{n_{\vec{k}\lambda}}\exp\left\{\mathrm{i}\left(\omega_{21} - \omega_k\right)t\right\}s_1(t)\ , \end{aligned} \tag{4.133}$$

where $\omega_{21} = (\varepsilon_2 - \varepsilon_1)/\hbar$. The initial condition for the new system coincides with the former one:

$$s_1(0) = 1\ , \quad s_2(0) = 0\ . \tag{4.134}$$

Using the first equation (4.133), we find the derivative $\partial s_2/\partial t$ and plug it into the second equation. This gives a closed equation for calculating $s_1(t)$:

$$\frac{\partial^2}{\partial t^2}s_1(t) + g_{\vec{k}}^2 n_{k\vec{\lambda}} s_1(t) + \mathrm{i}\gamma\frac{\partial}{\partial t}s_1(t) = 0\,, \tag{4.135}$$

where $\gamma = \omega_{21} - \omega_k$. The solution of the equation (4.135) appears as:

$$s_1(t) = a_1\exp(k_1 t) + a_2\exp(k_2 t)\,,$$

where $k_{1,2} = -(\mathrm{i}/2)[\gamma \mp \beta]$, $\beta = \sqrt{\gamma^2 + 4g_{\vec{k}}^2 n_{\vec{k}\lambda}}$. We express the function $s_2(t)$ via $s_1(t)$ from the first equation (4.133). The multiplicative constants a_1 and a_2 can be determined from the initial conditions (4.134):

$$a_1 = \frac{1}{2}\left(1 + \frac{\gamma}{\beta}\right)\,, \quad a_2 = \frac{1}{2}\left(1 - \frac{\gamma}{\beta}\right)\,.$$

As a result, we arrive at the following formula:

$$\begin{aligned} s_1(t) &= \exp\left(-\frac{\mathrm{i}\gamma}{2}t\right)\left[\cos\frac{\beta t}{2} + \frac{\mathrm{i}\gamma}{\beta}\sin\frac{\beta t}{2}\right]\,,\\ s_2(t) &= \frac{2g_{\vec{k}}\sqrt{n_{\vec{k}\lambda}}}{\beta}\exp\left(\frac{\mathrm{i}\gamma}{2}t\right)\sin\frac{\beta t}{2}\,. \end{aligned} \tag{4.136}$$

By the means of formulas (4.136), we derive the probability of transitioning the "atom + radiation" system from the state $|1, n_{\vec{k}\lambda}\rangle$ to the state $|2, n_{\vec{k}\lambda} - 1\rangle$:

$$|c_2(t)|^2 = |s_2(t)|^2 = \frac{4g_{\vec{k}}^2 n_{\vec{k}\lambda}}{\beta^2}\sin^2\frac{\beta t}{2} = \left(1 - \frac{\gamma^2}{\beta^2}\right)\sin^2\frac{\beta t}{2}\,,$$

as well as the reverse transition probability:

$$|c_1(t)|^2 = |s_1(t)|^2 = \cos^2\frac{\beta t}{2} + \frac{\gamma^2}{\beta^2}\sin^2\frac{\beta t}{2}\,.$$

The probability of finding the system in any of the states $|1, n_{\vec{k}\lambda}\rangle$ or $|2, n_{\vec{k}\lambda} - 1\rangle$ corresponds to a sure event and is equal to unity. This is in full accordance with the normalization condition (4.127) of the functions $c_i(t)$.

A change of filling the atom's levels as a result of its interaction with the electromagnetic wave is described by a periodic function of time:

$$|c_1(t)|^2 - |c_2(t)|^2 = \beta^{-2}\left[\gamma^2 + 4g_{\vec{k}}^2 n_{\vec{k}\lambda}\cos(\beta t)\right]\,, \tag{4.137}$$

i.e., there are so called *Rabi oscillations*. When Rabi oscillations take place, an atom under the influence of an electromagnetic wave makes periodic transitions between the ground $|1\rangle$ and excited $|2\rangle$ states. In other words, the atom periodically absorbs and emits photons with the energy $h\nu = \varepsilon_2 - \varepsilon_1$.

It should be especially emphasized that the "oscillatory" behavior of the atom (4.137) can be observed only under very special conditions. The fact is that the calculation that led to formula (4.137) is generally inconsistent. It is based on the assumption that the interaction between an individual atom and the radiation field is never interrupted, and is always described by the atom-single radiation mode interaction Hamiltonian (4.128). Indeed, the atom interacts with all of the electromagnetic field modes. Moreover, the radiation field is subject to the influence of the atom changing itself. Even in a vacuum, an atom interacts with zero fluctuations of an electromagnetic field. As a result, each of its energy level splits into a narrow band of width $\Delta\varepsilon_0$. The latter is related to the lifetime τ_0 of the excited atom by the uncertainty relation:

$$\Delta\varepsilon_0 \cdot \tau_0 \geq \hbar \,.$$

The broadening of the spectral lines of atoms due to vacuum fluctuations is called the *natural broadening*.

The quantum mechanical reason mentioned above is not the only reason why the Rabi oscillations are interrupted. There are other, more serious reasons for that: thermal motion of atoms and their mutual collisions. The Rabi oscillations can be observed only when the atom passes between the $|1\rangle$ and $|2\rangle$ states many times before interrupting its "oscillatory" behavior by the above processes (formulas (4.136) and (4.137) do not take them into account). In what follows, we call a time τ_{coll}, during which our calculation is valid, the time between collisions of the atom.

Thus, the Rabi oscillations occur only in the collisionless regime, when the oscillation frequency is much higher than the frequency of collisions:

$$\beta = \sqrt{\gamma^2 + 4g^2_{\vec{k}} n_{\vec{k}\lambda}} \gg \tau^{-1}_{\text{coll}} \,. \tag{4.138}$$

A more precise formulation of the conditions necessary for the Rabi oscillations to be observed requires estimating the parameter

$$4g^2_{\vec{k}} n_{\vec{k}\lambda} = \frac{2e^2\omega_k}{\varepsilon_0 V\hbar} \left(\vec{D}_{12} \cdot \vec{e}_{\vec{k}\lambda}\right)^2 n_{\vec{k}\lambda}$$

in formula (4.138). Given that the mean square of the electric field amplitude of the wave, as the number of photons is large, satisfies the relationship:

$$\langle n|\, \widehat{E}^2 \,|n\rangle \approx \frac{2\hbar\omega_k}{\varepsilon_0 V} n_{\vec{k}\lambda} \,,$$

and the matrix element $|\vec{D}_{12}|$ is of the order of the atomic Bohr radius a_{B}, we get:

$$4g^2_{\vec{k}} n_{\vec{k}\lambda} \sim e^2 a^2_{\text{B}} \hbar^{-2} \langle n|\, \widehat{E}^2 \,|n\rangle \,. \tag{4.139}$$

It follows that the inequality (4.138) is fulfilled at high intensities of the radiation field:

$$\frac{|e|}{\hbar} a_{\text{B}} \sqrt{\langle n|\, E^2 \,|n\rangle} \gg \tau^{-1}_{\text{coll}} \,. \tag{4.140}$$

With the advent of lasers emitting high intensity electromagnetic radiation with large quantity $\langle n|\widehat{E}^2|n\rangle$, it is currently possible to reveal the Rabi oscillations experimentally in a laboratory.

At the same time, the vast majority of modern optoelectronics devices, including laser resonators, use the collision regime. In this case, the difference in the atomic energy levels $(\varepsilon_2 - \varepsilon_1)$ is not precisely determined and so great that, in the formula for the probability of transitioning the "atom + radiation" system from the state $|1, n_{\vec{k}\lambda}\rangle$ to the state $|2, n_{\vec{k}\lambda} - 1\rangle$, we can put:

$$\beta = \sqrt{\gamma^2 + 4g_{\vec{k}}^2 n_{\vec{k}\lambda}} \approx \gamma = \omega_{21} - \omega_k \, .$$

As a result, it simplifies and becomes:

$$|c_2(t)|^2 = g_{\vec{k}}^2 n_{\vec{k}\lambda} \frac{\sin^2[(\omega_{21} - \omega_k)t/2]}{[(\omega_{21} - \omega_k)/2]^2} \, , \quad \omega_{21} = \frac{\varepsilon_2 - \varepsilon_1}{\hbar} \, . \tag{4.141}$$

It should be stressed that formula (4.141) can be applicable only for periods of time t less than the time between collisions: $t < \tau_{\text{coll}}$. Therefore, the Rabi oscillations are never observed in reality.

When uncertainty in the atomic energy difference $\varepsilon = (\varepsilon_2 - \varepsilon_1)$ is relatively large and is not exactly known, the *line shape function* appears to be useful. For this, the quantity $g(\varepsilon)\mathrm{d}\varepsilon$ should be entered; it yields the probability of finding the value of ε in the range between ε and $\varepsilon + \mathrm{d}\varepsilon$. Since the probability of finding ε in the entire energy axis corresponds to a trustworthy event, we have:

$$\int_{-\infty}^{+\infty} g(\varepsilon)\, \mathrm{d}\varepsilon = 1 \, .$$

The use of the line shape function is no more than a good trick to mathematically take into account the inaccuracy in the atomic energy difference. Rather than looking into the atomic transitions from the rigorously defined state $|1\rangle$ to the corresponding allowed state $|2\rangle$, we focus on atomic transitions into a group of states close in energy value to the difference $\varepsilon = \varepsilon_2 - \varepsilon_1$. Mathematically, this reduces the integration of the equality (4.141) over all the possible values of ε with the weight function $g(\varepsilon)$:

$$\left\langle |c_2(t)|^2 \right\rangle \equiv \int_{-\infty}^{\infty} |c_2(t)|^2 \, g(\varepsilon)\mathrm{d}\varepsilon = 2\pi t g_{\vec{k}}^2 n_{\vec{k}\lambda} \int_{-\infty}^{\infty} \mathrm{d}\varepsilon \delta\left(\frac{\varepsilon}{\hbar} - \omega_k, t\right) g(\varepsilon) \, , \tag{4.142}$$

where

$$\delta(\omega, t) = \frac{1}{2\pi} \frac{\sin^2(\omega t/2)}{\omega^2 t/4} \, . \tag{4.143}$$

Figure 4.13 represents the plots of the functions $\delta(\varepsilon/\hbar - \omega_k, t)$ and $g(\varepsilon)$.

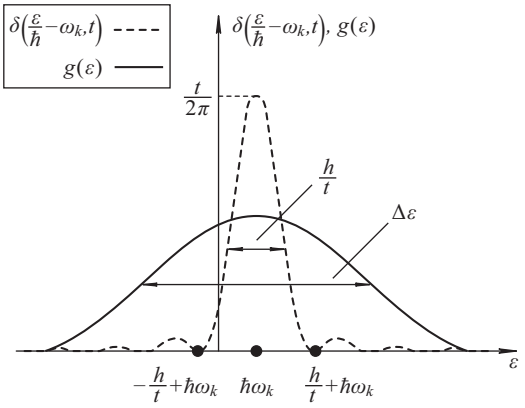


Fig. 4.13: Plots of the functions $\delta(\varepsilon/\hbar - \omega_k, t)$ (a dashed line) and $g(\varepsilon)$ (a solid line).

As can be seen from the figure above, both of them have a dome shape, but the "width" $\Delta\varepsilon$ of the function $g(\varepsilon)$ is much larger than the "width" h/t of the function $\delta(\varepsilon/\hbar-\omega_k, t)$ over periods of the real times $t \sim \tau_{\text{coll}}$:

$$h/t \ll \Delta\varepsilon \,. \tag{4.144}$$

This circumstance makes it possible to approximately evaluate the integral in (4.142).

Putting forward the argument that the function $g(\varepsilon)$ is a smooth function, we expand it into a Taylor series near the maximum of the "acute" function $\delta(\varepsilon/\hbar - \omega_k, t)$:

$$g(\varepsilon) = g(\hbar\omega_k) + (\varepsilon - \hbar\omega_k)\left.\frac{\mathrm{d}g(\varepsilon)}{\mathrm{d}\varepsilon}\right|_{\varepsilon=\hbar\omega_k} + \cdots \,.$$

Given that:

$$\int\limits_{-\infty}^{\infty} \delta\left(\frac{\varepsilon}{\hbar} - \omega_k, t\right)\mathrm{d}\varepsilon = \underbrace{\frac{\hbar}{\pi}\int\limits_{-\infty}^{\infty}\frac{\sin^2\xi}{\xi^2}\mathrm{d}\xi}_{\pi} = \hbar \,, \tag{4.145}$$

and confining ourselves to the first term of the expansion, we come to the formula:

$$\left\langle |c_2(t)|^2 \right\rangle = 2\pi t\hbar g^2_{\vec{k}} n_{\vec{k}\lambda} g(\hbar\omega_k) \,. \tag{4.146}$$

As a result, we have a linear time dependence that holds true only for times t satisfying the limitation:

$$\left\langle |c_2(t)|^2 \right\rangle \ll 1 \,, \tag{4.147}$$

because $|c_1(t)|^2 \approx 1$ and $|c_2(t)|^2 + |c_1(t)|^2 = 1$.

Finally, in perturbation theory, the applicability conditions of calculations in the collisional regime offer the top and bottom limitations for the time t:

$$2\pi\hbar g^2_{\vec{k}} n_{\vec{k}\lambda} g(\hbar\omega_k) \ll 1/t \ll \Delta\varepsilon/h \,. \tag{4.148}$$

Utilizing the estimate (4.139) for the parameter $g_{\vec{k}}$ and taking into account that $g(\hbar\omega_k) \sim 1/\Delta\varepsilon$, we transform the inequality (4.148):

$$\pi e^2 \langle n|\, \widehat{E}^2\, |n\rangle\, a_{\mathrm{B}}/(2\hbar\Delta\varepsilon) \ll 1/t \ll \Delta\varepsilon/h\ .$$

Here, we have replaced the function $g(\varepsilon)$ by a rectangular step of the width $\Delta\varepsilon$ and height of $1/\Delta\varepsilon$. From the foregoing, we can reach the conclusion that the above calculation is valid for relatively low intensities of the radiation field.

Interestingly, formula (4.146) for the transition probability is often derived by passing to the limit $t \to \infty$ in (4.142). In doing so, the function $\delta(\varepsilon/\hbar - \omega_k, t)$, having a narrow and very high maximum (see Figure 4.13), becomes the Dirac delta function:

$$\lim_{t\to\infty} \delta\left(\frac{\varepsilon}{\hbar} - \omega_k, t\right) = \delta\left(\frac{\varepsilon}{\hbar} - \omega_k\right)\ . \tag{4.149}$$

Due to the delta function, the integral in the right-hand side of (4.142) can be easily calculated over the variable ε. This results in the valid expression (4.146) for the transition probability.

Attention should be drawn to the fact that the limit $t \to \infty$ is impossible to go over to in this problem because all of the formulas hold true only for the periods of time $0 < t < \tau_{\mathrm{coll}}$. Replacing the function of classical mathematical analysis by the generalized function (4.149) is but a formal subterfuge for easy calculations. The more accurate calculations conducted above may serve as its justification. The formal approach for calculating integrals of the type (4.142) is based on two comments:

1. The "width" of the function $g(\varepsilon)$ is greater than the "width" of the function $\delta(\varepsilon/\hbar - \omega_k, t)$.
2. The area beneath the graph of the "acute" function $\delta(\varepsilon/\hbar - \omega_k, t)$ is time independent.

All the problems discussed below meet these conditions. Therefore, to simplify the calculations, we resort to the formal approach.

Next, the probability of passing the "atom + radiation" system from the state $|1, n_{\vec{k}\lambda}\rangle$ to the state $|2, n_{\vec{k}\lambda} - 1\rangle$ per unit of time is our concern. The probability itself of such a transition should be left aside:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\langle |c_2(t)|^2 \right\rangle = \frac{\left\langle |c_2(t)|^2 \right\rangle}{t} = 2\pi\hbar g_{\vec{k}}^2 n_{\vec{k}\lambda} g(\hbar\omega_k) = \frac{\pi e^2 \omega_k}{\varepsilon_0 V} n_{\vec{k}\lambda} g(\hbar\omega_k) \left(\vec{D}_{12} \cdot \vec{e}_{\vec{k}\lambda}\right)^2\ . \tag{4.150}$$

The transition rate (4.150), fortunately, does not depend on time, so its calculation by perturbation theory gives good results.

To complete the calculation, we average the transition rate found (4.150) over random orientations of the atoms. Suppose the dipole moment vector $\vec{D}_{12}$ of the atom lies within the solid angle $\mathrm{d}\Omega = \sin\theta \mathrm{d}\theta \mathrm{d}\varphi$, where θ is an angle between the vector $\vec{D}_{12}$ and the unit vector of polarization of the electromagnetic wave $\vec{e}_{\vec{k}\lambda}$. The angle φ defines the position of the projection of the vector $\vec{D}_{12}$ in a plane perpendicular to the vector $\vec{e}_{\vec{k}\lambda}$ (Figure 4.14.).

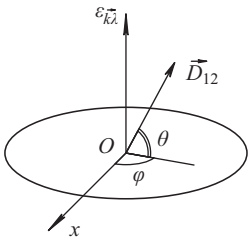


Fig. 4.14: The vector $\vec{D}_{12}$ in a spherical coordinate system.

Since the total solid angle is equal to 4π, the average value of $\cos^2\theta$ for all the equiprobable orientations of the vector $\vec{D}_{12}$ is given by the formula:

$$\left\langle \cos^2\theta \right\rangle = \frac{1}{4\pi}\int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \cos^2\theta \sin\theta \mathrm{d}\theta = \frac{1}{3}\,.$$

Therefore, averaged over the orientations of the vector $\vec{D}_{12}$, the transition rate of the atom undergoing single photon absorption from the beam containing $n_{\vec{k}\lambda}$ photons has the form:

$$W_{1\to 2} = \frac{\pi e^2 \omega_k}{\varepsilon_0 V}\left|\vec{D}_{12}\right|^2 \left\langle \cos^2\theta \right\rangle n_{\vec{k}\lambda} g(\hbar\omega_k) = \frac{\pi e^2 \omega_k}{3\varepsilon_0 V}\left|\vec{D}_{12}\right|^2 n_{\vec{k}\lambda} g(\hbar\omega_k)\,. \qquad (4.151)$$

We have come to the result (4.151) for the beam of photons, whose initial state is $|1, n_{\vec{k}\lambda}\rangle$. In this state, the number of photons are accurately defined. In fact, it can be shown that the outcome in the general quantum mechanical state is the same for the photon beam. However, it should be understood that, in this case, the quantity $n_{\vec{k}\lambda}$ means the average number of photons.

In particular, when the number of photons are large, the result holds also for coherent states, which correspond to an electromagnetic wave of classical physics. For the coherent states, the above formula may be rewritten in terms of Maxwell's classical theory. Moreover, it can be generalized to the case of propagating the electromagnetic wave through a medium containing a lot of atoms in the ground state.

The quantity $\hbar n_{\vec{k}\lambda}\omega_k/V$ in Maxwell's theory corresponds to the average energy of an electromagnetic wave per unit volume of the medium. Multiplied by the speed of light $\tilde{c}$ in a medium, the energy yields radiation intensity I, i.e., the average flow of energy through a unit area perpendicular to the direction of wave propagation:

$$I = \hbar\omega_k n_{\vec{k}\lambda}\tilde{c}/V\,.$$

In isotropic media for relatively weak electromagnetic fields, slowly varying in space and time, the simplest constitutive equations are valid:

$$\vec{D} = \varepsilon_0\varepsilon\vec{E}\,, \quad \vec{B} = \mu_0\mu\vec{H}\,,$$

where ε, μ = const are the relative dielectric and magnetic permitivities of a medium, respectively. Then, the speed of light in a medium is determined by the formula:

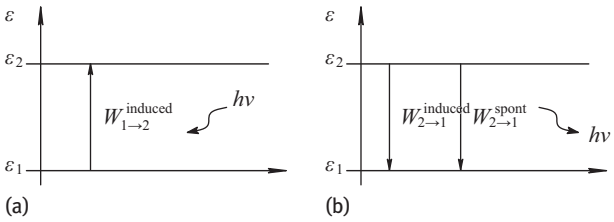


Fig. 4.15: Transition of an atom from a ground state to an excited state, induced by an electromagnetic wave field (a); induced and spontaneous transition of an atom from an excited state to a ground one (b).

$\tilde{c} = c/\sqrt{\varepsilon\mu}$, where $c = (\varepsilon_0\mu_0)^{-1/2}$ is the speed of light in a vacuum. The quantity $n = \sqrt{\varepsilon\mu}$ is called the refractive index of a medium. For nonmagnetic media discussed below: $\mu = 1$, $n = \sqrt{\varepsilon} > 1$.

Apart from the foregoing, to take the influence of a medium into account, we should replace the fundamental constant ε_0 in a vacuum by the dielectric constant $\varepsilon_0\varepsilon$ of a medium in formula (4.151) for the transition rate $W_{1\to 2}$.

Finally, the transition rate of one of the atoms of the medium from the ground state to an excited one, due to its interaction with the electromagnetic wave propagating through the medium, takes the form (see Figure 4.15 (a).):

$$W_{1\to 2} = \frac{\pi e^2 \left|\vec{D}_{12}\right|^2}{3\varepsilon_0\varepsilon\hbar\tilde{c}} Ig(\hbar\omega_k) \,. \tag{4.152}$$

We will further discuss in detail the classical theory of electromagnetic wave propagation in a medium, and show how to generalize the above formulas to the case of temporal dispersion of the medium, which plays an important role in lasers.

Interaction of an Atom in the Ground State with Thermal Radiation

Recall that Einstein postulated the following formulas for the probabilities of transitions per atom balanced by thermal radiation per second:

$$W_{2\to 1} = B\rho(\nu) + A \,, \quad W_{1\to 2} = B\rho(\nu) \,.$$

Let us look into a two level atom interacting with all modes of a quasicontinuous spectrum of the thermal radiation, and find an expression for the Einstein coefficient B. For this purpose, we should return to the original formula (4.141) for the probability of passing an atom from the state $|1\rangle$ to the state $|2\rangle$. For further analysis, this equality can be conveniently written in the form:

$$\frac{|c_2(t)|^2}{t} = \frac{\pi e^2 \omega_k n_{\vec{k}\lambda}}{\varepsilon_0 V\hbar} \left(\vec{D}_{12}\cdot\vec{e}_{\vec{k}\lambda}\right)^2 \delta\left(\frac{\varepsilon_2-\varepsilon_1}{\hbar} - \omega_k, t\right) \,. \tag{4.153}$$

To obtain the desired result, we do the following:

– As before, we average the quantity $(\vec{D}_{12} \cdot \vec{e}_{\vec{k}\lambda})^2$ over various positions of the vector $\vec{D}_{12}$, with respect to the polarization vector $\vec{e}_{\vec{k}\lambda}$ of a single radiation mode. This leads to the substitution:

$$\left(\vec{D}_{12} \cdot \vec{e}_{\vec{k}\lambda}\right)^2 \to \frac{\left|\vec{D}_{12}\right|^2}{3} .$$

– Recall that $\rho(\nu)\mathrm{d}\nu$ is the thermal radiation energy per unit volume for the frequency range of $(\nu, \nu + \mathrm{d}\nu)$. Suppose V is the volume of the cavity with thermal radiation, and that the cavity concentrates the radiation energy $V\rho(\nu)\mathrm{d}\nu$ within the above frequency interval. Because the atom at hand interacts with all modes of the thermal radiation, formula (4.153) needs to be modified by the replacement: $\omega_k \to 2\pi\nu$, $\hbar\omega_k n_{\vec{k}\vec{\lambda}} \to V\rho(\nu)\mathrm{d}\nu$. Next, we integrate the result over all possible frequencies from 0 to ∞. Finally, the formula for the transition rate of the atom from the state $|1\rangle$ to the state $|2\rangle$ acquires the appearance:

$$W_{1\to 2} = \frac{|c_2(t)|^2}{t} = \frac{\pi e^2}{3\varepsilon_0 \hbar^2}\left|\vec{D}_{12}\right|^2 \int_0^\infty \mathrm{d}\nu\rho(\nu)\delta\left[\frac{\varepsilon_2 - \varepsilon_1}{\hbar} - 2\pi\nu, t\right] . \tag{4.154}$$

– For the integral to be calculated, we formally approximate the quantity $\delta[(\varepsilon_2 - \varepsilon_1)/\hbar - 2\pi\nu, t]$ with the Dirac delta function that is time independent:

$$\delta\left[\frac{\varepsilon_2 - \varepsilon_1}{\hbar} - 2\pi\nu, t\right] = \frac{1}{2\pi}\delta\left(\nu - \frac{\varepsilon_2 - \varepsilon_1}{h}\right) . \tag{4.155}$$

The justification of such a procedure is still the same. However, it is not worth forgetting that the delta function is an even function and satisfies the identity $\delta(ax) = \delta(x)/|a|$. After replacing (4.155), the integral can be easily computed, and we ultimately arrive at:

$$W_{1\to 2} = \frac{e^2\left|\vec{D}_{12}\right|^2}{6\varepsilon_0\hbar^2}\rho(\nu) . \tag{4.156}$$

where $\nu = (\varepsilon_2 - \varepsilon_1)/h$.

The multiplier on the right-hand side of (4.156), before the spectral density of thermal radiation $\rho(\nu)$, is the Einstein coefficient required. The latter governs the induced atom's transition from the ground state to the excited state:

$$B = \frac{e^2\left|\vec{D}_{12}\right|^2}{6\hbar^2\varepsilon_0} . \tag{4.157}$$

Interaction of an Atom in an Excited State with an Electromagnetic Wave

Within this section we regard energy transitions of an atom from an excited state to a ground one as a result of its interaction with a single electromagnetic mode containing $n_{\vec{k}\lambda}$ photons.

Substitute the approximation (4.126) into the Schrödinger equation (4.124):

$$\begin{aligned} \mathrm{i}\hbar\frac{\partial}{\partial t}\left(c_1(t)\left|1, n_{\vec{k}\vec{\lambda}}+1\right\rangle + c_2(t)\left|2, n_{\vec{k}\vec{\lambda}}\right\rangle\right) = \\ = \left(\widehat{H}_0 + \widehat{H}_{\mathrm{int}}\right)\left(c_1(t)\left|1, n_{\vec{k}\lambda}+1\right\rangle + c_2(t)\left|2, n_{\vec{k}\lambda}\right\rangle\right) . \end{aligned} \tag{4.158}$$

Next, we can scalarly multiply (4.158) by $\langle 2, n_{\vec{k}\vec{\lambda}}|$ and $\langle 1, n_{\vec{k}\vec{\lambda}}+1|$. Using formulas (4.125), (4.126) and (4.127), we can obtain a set of differential equations to calculate the functions $c_{1,2}(t)$:

$$\begin{aligned} \mathrm{i}\hbar\frac{\partial}{\partial t}c_2(t) &= (\varepsilon_2 + \hbar\omega_k n_{\vec{k}\lambda})c_2(t) + \mathrm{i}\hbar g_{\vec{k}}\sqrt{n_{\vec{k}\lambda}+1}\,c_1(t) , \\ \mathrm{i}\hbar\frac{\partial}{\partial t}c_1(t) &= \left[\varepsilon_1 + \hbar\omega_k(n_{\vec{k}\lambda}+1)\right]c_1(t) - \mathrm{i}\hbar g_{\vec{k}}\sqrt{n_{\vec{k}\lambda}+1}\,c_2(t) , \end{aligned} \tag{4.159}$$

This set of differential equations needs to be solved with the initial condition: $c_2(0)=1$, $c_1(0)=0$.

It is easy to see that the set (4.159) and, as a consequence, its solution can be deduced from the system (4.132) by the formal change:

$$c_2(t) \to c_1(t) , \quad c_1(t) \to c_2(t) , \quad \varepsilon_1 + \hbar\omega_k n_{\vec{k}\lambda} \to \varepsilon_2 + \hbar\omega_k n_{\vec{k}\lambda} ,$$
$$\varepsilon_2 + \hbar\omega_k\left(n_{\vec{k}\lambda}-1\right) \to \varepsilon_1 + \hbar\omega_k\left(n_{\vec{k}\lambda}+1\right) , \quad g_{\vec{k}}\sqrt{n_{\vec{k}\lambda}} \to -g_{\vec{k}}\sqrt{n_{\vec{k}\lambda}+1} .$$

Therefore, the formula for the probability of transition of an atom from an excited state to a ground one should be:

$$|c_1(t)|^2 = 4g_{\vec{k}}^2(n_{\vec{k}\lambda}+1)\beta^{-2}\sin^2\frac{\beta t}{2} \approx 2\pi t g_{\vec{k}}^2(n_{\vec{k}\lambda}+1)\delta\left(\frac{\varepsilon_2-\varepsilon_1}{\hbar}-\omega_k, t\right) . \tag{4.160}$$

The new multiplier $(n_{\vec{k}\lambda}+1)$ in equation (4.160) appears instead of the former one $n_{\vec{k}\lambda}$ due to the noncommutative operators $\widehat{a}^+_{\vec{k}\lambda}$, $\widehat{a}_{\vec{k}\lambda}$. When the electromagnetic field is not quantized, there is no additional unity.

The modifier $(n_{\vec{k}\lambda}+1)$ allows two terms in the expression for the probability of the atom's transition from the excited state to the ground one. The term linear in $n_{\vec{k}\lambda}$ corresponds to the induced atomic transitions. The calculation for the induced transition rate $W^{\mathrm{induced}}_{2\to1}$ is the same as for the rate $W^{\mathrm{induced}}_{1\to2}$ (see the previous section). Thus, we have $W^{\mathrm{induced}}_{2\to1} = W^{\mathrm{induced}}_{1\to2}$.

Instead of the atom's interaction with a single electromagnetic mode, we are interested in what happens in the event of interaction between the atom and thermal radiation. The resulting interaction confirms Einstein's hypothesis that the rate of radiation induced thermal transitions of the atom from the ground state to the excited state can be identically written as the rate of the reverse induced transitions of the atom.

However, let us return to the atom-electromagnetic wave interaction. Interestingly, a photon born with the induced radiation has exactly the same values of frequency, the wave vector, and polarization as the photons before the radiation. Moreover, the new photon correlates in phase with the wave photons. This fact is important because it opens up the possibility, in principle, of generating powerful

electromagnetic pulses with identical polarization and wave vectors of the photons in the pulse. For this, it suffices to pass a weak linearly polarized wave with the frequency $\nu = (\varepsilon_2 - \varepsilon_1)/h$ through the medium whose atoms are in the excited state $|2\rangle$. When affected by the wave, all of the medium atoms move into the state $|1\rangle$ like an avalanche. The photons of the same energy, polarization, and direction of propagation enrich the initial pulse.

Being independent of the number of photons $n_{\vec{k}\lambda}$, the summand in formula (4.160) corresponds to spontaneous emission. Such transitions take place even in the absence of any external radiation. They are initiated by the interaction between the atom and zero point electromagnetic vacuum fluctuations. Only the quantum theory can describe the spontaneous emission. A photon spontaneously emitted can have any orientation of its wave vector, arbitrary polarization, and a random phase. In this respect, it differs significantly from stimulated emitted photons.

Calculation of the Einstein Coefficient A

To start with, we assume that there are no photons in the cavity. Next, we calculate the Einstein coefficient A that determines the probability of a spontaneous transition of an atom per second from an excited state to its ground state. Since we have an interest in the transitions set forth above, we put $n_{\vec{k}\lambda} = 0$ in (4.160) and rewrite the formula in a form suitable for further analysis:

$$\frac{|c_1(t)|^2}{t} = \frac{\pi e^2 \omega_k}{\varepsilon_0 V \hbar} \left(\vec{D}_{12} \cdot \vec{e}_{\vec{k}\lambda}\right)^2 \delta\left(\frac{\varepsilon_2 - \varepsilon_1}{\hbar} - \omega_k, t\right) . \tag{4.161}$$

In the cavity, electromagnetic modes, as a result of the spontaneous emission of the atom, can occupy only permitted energy states. Their number for the modes with frequencies in the interval $(\nu, \nu + \mathrm{d}\nu)$ is evaluated by the formula:

$$VD(\nu)\mathrm{d}\nu = \frac{8\pi\nu^2}{c^3} V \mathrm{d}\nu , \tag{4.162}$$

where V is the volume of the cavity. To compute the spontaneous transition rate $W_{2\to1}^{\text{spont}}$, formula (4.161) should be updated as follows:

1. As mentioned earlier, it first needs to average the quantity $(\vec{D}_{12} \cdot \vec{e}_{\vec{k}\lambda})^2$ over various orientations of the vector $\vec{D}_{12}$ relative to the polarization vector $\vec{e}_{\vec{k}\lambda}$ of a single radiation mode. This yields the substitution:

$$\left(\vec{D}_{12} \cdot \vec{e}_{\vec{k}\lambda}\right)^2 \to \frac{\left|\vec{D}_{12}^2\right|}{3} .$$

2. Formula (4.161) should be multiplied by the number of allowed photon states (4.162) contained in the frequency range $(\nu, \nu + \mathrm{d}\nu)$, and the result should be integrated over the frequencies between 0 and ∞. Then, we arrive at the spontaneous transition rate:

$$W_{2\to1}^{\text{spont}} = \frac{\left\langle |c_1(t)|^2 \right\rangle}{t} = \frac{16\pi^3 e^2}{3\varepsilon_0 \hbar c^3} \left|\vec{D}_{12}\right|^2 \int_0^\infty \mathrm{d}\nu\, \nu^3 \delta\left[\frac{\varepsilon_2 - \varepsilon_1}{\hbar} - 2\pi\nu, t\right] .$$

3. To calculate the above integral, we formally approximate the quantity $\delta[(\varepsilon_2 - \varepsilon_1)/\hbar - 2\pi\nu, t]$ by the time independent delta function:

$$\delta\left[\frac{\varepsilon_2 - \varepsilon_1}{\hbar} - 2\pi\nu, t\right] = \frac{1}{2\pi}\delta\left(\nu - \frac{\varepsilon_2 - \varepsilon_1}{h}\right).$$

Ultimately, we come up with the formula for the spontaneous transition rate (Figure 4.15b):

$$W_{2\to1}^{\text{spont}} = \frac{8(\pi e)^2 \nu^3}{3\varepsilon_0 \hbar c^3}\left|\vec{D}_{12}\right|^2, \tag{4.163}$$

where $\nu = (\varepsilon_2 - \varepsilon_1)/h$. This is just the Einstein coefficient:

$$A = 1/t_{\text{spont}} = W_{2\to1}^{\text{spont}}.$$

For the thermal radiation in the cavity, we find the ratio of the theoretically calculated coefficients A and B:

$$\frac{A}{B} = \frac{8\pi h\nu^3}{c^3}.$$

The finding secured coincides exactly with the ratio of the phenomenological coefficients A and B, postulated by Einstein for Planck's formula.

4.10 Absorption and Amplification of Directed Plane-Parallel Flux by Matter

As previously mentioned, a single two level atom of a medium, under the influence of an external electromagnetic field, is forced to make transitions between its states. The rates of the induced transitions are equal:

$$W_{1\to2} = W_{2\to1} = \frac{\pi e^2\left|\vec{D}_{12}\right|^2}{3\tilde{c}\varepsilon\varepsilon_0\hbar} Ig(\hbar\omega_k) \equiv W^{\text{induced}}. \tag{4.164}$$

The relation (4.164) is very difficult to experimentally verify. This is because it is first necessary to know the wave function of a many electron atom to calculate the square of the modulus of the matrix element $|\vec{D}_{12}|^2$. Together with that, the rate of the spontaneous transition is related to the lifetime of the excited atom as follows: $W_{2\to1}^{\text{spont}} = A = 1/t_{\text{spont}}$. The lifetime can be easily measured in practice.

We have theoretically calculated the Einstein coefficient A (4.163), as well as the lifetime of the atom in a vacuum t_{spont}. The lifetime of an atom in an isotropic medium without dispersion comes from the lifetime of the atom in a vacuum by the formal replacement: $c \to \tilde{c}$, $\varepsilon_0 \to \varepsilon\varepsilon_0$. As a result, we arrive at an expression that determines the lifetime of an excited atom t_{spont} in a medium:

$$t_{\text{spont}}^{-1} = \frac{8\pi^2 e^2 \nu^3\left|\vec{D}_{12}\right|^2}{3\tilde{c}^3\varepsilon\varepsilon_0\hbar}. \tag{4.165}$$

Combining the relations (4.164) and (4.165), we get a formula convenient for experimental verification of the rates of the atomic transitions induced by an external and/or medium field:

$$W^{\text{induced}} = \frac{\widetilde{\lambda}^3 g(\hbar\omega_k)}{8\pi\widetilde{c}t_{\text{spont}}} I\,, \tag{4.166}$$

where $\widetilde{\lambda} = \widetilde{c}/\nu$ is the wavelength of radiation in a medium. Next we show how to generalize the relation (4.166) to the case of isotropic nonmagnetic media with time dispersion.

In formula (4.166), the line shape function $g(\varepsilon)$ models processes that make the difference between the energies ε_2 and ε_1 of the atom indefinite. Particular calculations require choosing either a Lorentzian or Gaussian curve as a line shape function.

The most common mechanisms responsible for the shape of the line are collisions of atoms of a medium with each other, or spontaneous transitions of the atoms in other states left aside by our calculations (there are more possible quantum states of an atom than two). The Lorentz line describes well such processes (Figure 4.16):

$$g(\varepsilon) = \frac{1}{\pi}\left(\frac{\Delta\varepsilon}{2}\right)\frac{1}{(\varepsilon-\varepsilon_0)^2 + (\Delta\varepsilon/2)^2}\,. \tag{4.167}$$

The quantity $\Delta\varepsilon$ called the conditional width of the curve depends on the concentration of colliding atoms.

In addition to the foregoing processes, there is a Doppler broadening of the lines due to thermal motion of the atoms [13]. This broadening is best modeled by a Gaussian function, which does not depend on the concentration of colliding atoms (Figure 4.17):

$$g(\varepsilon) = \frac{1}{\delta\sqrt{\pi}}\exp\left(-\frac{(\varepsilon-\varepsilon_0)^2}{\delta^2}\right)\,. \tag{4.168}$$

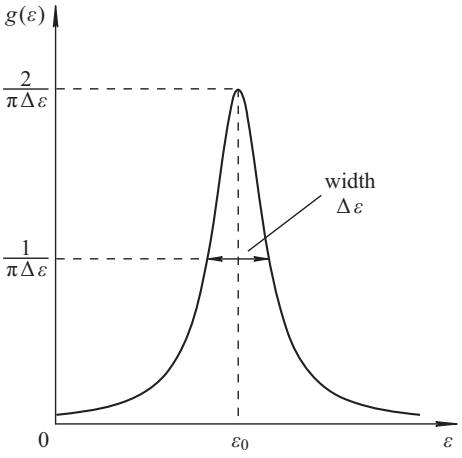


Fig. 4.16: The Lorentzian line shape function.

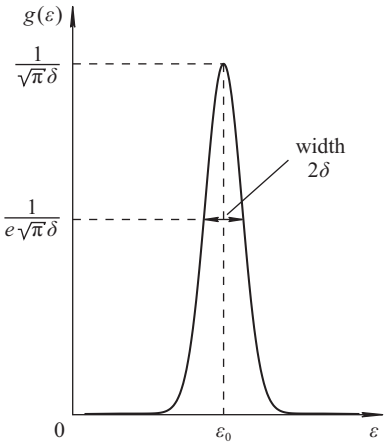


Fig. 4.17: The Gaussian line shape function.

The areas underneath both curves (4.167) and (4.168) are equal to unity. So the more the height of their peaks is greater, the less the width of the line. The Gaussian line is more acute compared to the Lorentzian one and quickly decays as $|\varepsilon| \to \infty$. Whatever the mechanisms are, the line width is proportional to $\sqrt{T}$.

Let two statistically independent mechanisms for line broadening act simultaneously. Separately, they are described by the functions $\varphi_1(\varepsilon)$ and $\varphi_2(\varepsilon)$. The function of the resulting line shape is taken as the convolution:

$$\varphi(\varepsilon) = \int_{-\infty}^{+\infty} \mathrm{d}x \varphi_1(x)\varphi_2(\varepsilon + \varepsilon_0 - x) \,. \tag{4.169}$$

Here, $\omega_0 = \varepsilon_0/\hbar$ is a common center frequency for the two distributions.

Now we will discuss what happens when a plane electromagnetic wave with intensity I and frequency ν passes through a medium whose atoms are distributed over the energy levels ε_1 and ε_2 ($\varepsilon_2 > \varepsilon_1, \nu \approx \varepsilon_2 - \varepsilon_1$). Suppose the unit volume contain N_1 atoms with the energy ε_1 and N_2 atoms with the energy ε_2. The medium is not necessarily equilibrium (Figure 4.18).

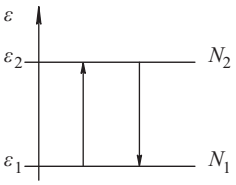


Fig. 4.18: A two level system of energy transitions.

The number of atomic transitions of the type $|2\rangle \to |1\rangle$ per unit time, under the influence of the electromagnetic wave, is equal to:

$$N_{2\to1} = N_2 W^{\text{induced}} \,.$$

The number of the opposite transitions ($|1\rangle \to |2\rangle$) is given by the expression:

$$N_{1\to2} = N_1 W^{\text{induced}} \,.$$

The difference $(N_1 - N_2)W^{\text{induced}}$ is an excess between the numbers of the upward and downward transitions (Figure 4.18). With each up transition, the electromagnetic wave loses a photon with the energy $h\nu = \varepsilon_2 - \varepsilon_1$. With each down transition, the electromagnetic wave gains a single photon. The total energy acquired by the electromagnetic wave per unit time is equal to

$$-h\nu\,(N_1 - N_2)\,W^{\text{induced}} \,. \tag{4.170}$$

We compute the same values in another way. Consider a small volume of matter in the form of a cube, as shown in Figure 4.19.

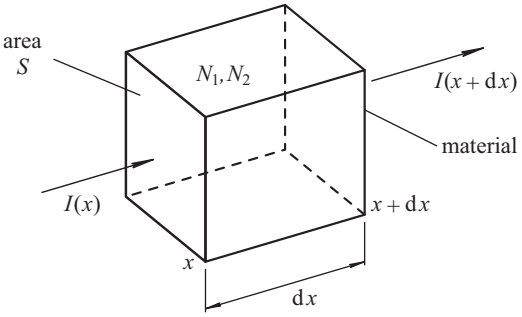


Fig. 4.19: Emission power balance for an infinitesimal volume of matter. The power $h\nu(N_1 - N_2)W^{\text{induced}}S\mathrm{d}x$ transmitted by an electromagnetic radiation field to the atoms of matter must be equal to the reduction in the power of the passing wave.

Let an electromagnetic wave with the intensity $I(x)$ be normally incident on the face of the cube (with the coordinate x) and come out through the opposite face (with the coordinate $x + \mathrm{d}x$), its output intensity being $I(x + \mathrm{d}x)$. Under steady state conditions, a power increase in the electromagnetic wave in the volume of the cube amounts to:

$$S\,[I(x + \mathrm{d}x) - I(x)] = \frac{\mathrm{d}I}{\mathrm{d}x}S\mathrm{d}x\,,$$

where S is the area of the side face of the cube. Together with the above, this gain, according to formula (4.170) is equal to $-h\nu(N_1 - N_2)W^{\text{induced}}S\mathrm{d}x$. Thus, we arrive at the equation:

$$\frac{\mathrm{d}I}{\mathrm{d}x} = -h\nu\,(N_1 - N_2)\,W^{\text{induced}}\,.$$

Substituting the explicit expression for the rate of the induced transitions (4.166) into the right side of this equation, we obtain:

$$\frac{\mathrm{d}I}{\mathrm{d}x} = -\,(N_1 - N_2)\,\frac{(\tilde{\lambda})^2 h}{8\pi t_{\text{spont}}}g(h\nu)I\,. \tag{4.171}$$

Let us discuss two cases.

Case 1: A medium is in thermal equilibrium. Before passing an electromagnetic wave through the medium, the ratio of the atomic levels is determined by the Boltzmann factor:

$$\frac{N_2}{N_1} = \exp\left(-\frac{\varepsilon_2 - \varepsilon_1}{k_{\text{B}}T}\right)\,.$$

Hence it follows that $N_2 < N_1$. When the wave passes through the medium, N_2 and N_1 change negligibly so it can be thought as N_2, $N_1 \approx$ const. Then, equation (4.171) is easy to integrate:

$$I(x) = I(0)\exp(-\alpha x)\,,$$

where

$$\alpha = (N_1 - N_2)\,\frac{(\tilde{\lambda})^2 h}{8\pi t_{\text{spont}}}g(h\nu) > 0\,. \tag{4.172}$$

Thus, all thermodynamically equilibrium media absorb energy (Figure 4.20b). The quantum theory has helped us to calculate the absorption coefficient α (4.172).

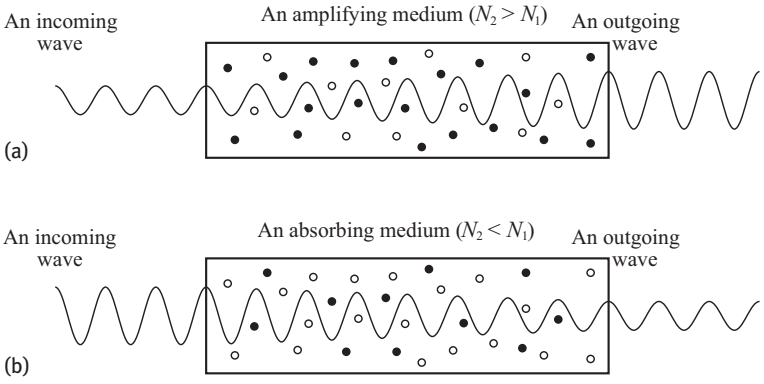


Fig. 4.20: An electromagnetic wave amplifies in a medium with an inverted population (a) and is absorbed by the medium with an ordinary level population (b). The atoms in the excited state are indicated by black circles, the atoms in the ground state are indicated by open circles.

Case 2: A medium is in a nonequilibrium state with a population inversion: $N_2 > N_1$. Then, after passing through the medium with the inverted population, the wave strengthens rather than weakens (Figure 4.20a). Therefore, instead of the absorption coefficient α, the amplification coefficient γ should be entered:

$$\gamma = (N_2 - N_1)\,\frac{(\tilde{\lambda})^2 h}{8\pi t_{\text{spont}}} g\,(h\nu) > 0\,. \tag{4.173}$$

The numbers N_2 and N_1 suffer significant changes when passing the wave through the medium, so equation (4.171) cannot be integrated. The complete theory must contain additional equations which take the population changes of N_2 and N_1 into account.

4.11 The Concept of Time and Spatial Dispersions of a Medium

Next, we consider nonmagnetic media and assume that there are no external charges and currents in the media, which means that we deal with dielectrics. An electromagnetic field in such media is described by the following set of Maxwell's equations:

$$\begin{aligned} &\operatorname{div}\vec{B} = 0\,, && \operatorname{div}\vec{D} = 0\,,\\ &\vec{B} = \mu_0\vec{H}\,, && \vec{D} = \varepsilon_0\vec{E} + \vec{P}\,,\\ &\operatorname{rot}\vec{E} = -\frac{\partial\vec{B}}{\partial t}\,, && \operatorname{rot}\vec{H} = \frac{\partial\vec{D}}{\partial t}\,. \end{aligned} \tag{4.174}$$

In contrast to the problem of propagation of electromagnetic waves in a vacuum, the system (4.174) is not closed and must be supplemented by constitutive equations, which relate the dipole moment vector $\vec{P}$ per unit volume of the medium to electric field strength $\vec{E}$. Seeking such a relationship is the problem with microscopic theory.

However, some conclusions can be made without resorting to microscopic theory and can be linked up with the latter afterwards.

If an external electric field $\vec{E}$ is weak, as compared to internal circumatomic fields, we may restrict ourselves to a linear relationship between the $\vec{E}$ and $\vec{P}$ fields. In general, it can be written as:

$$P_i = \int_{-\infty}^{t} \mathrm{d}t' \int \mathrm{d}^3\vec{r}\,' \alpha_{ij}(\vec{r}, \vec{r}\,', t, t')E_j(\vec{r}\,', t') \,. \tag{4.175}$$

Formula (4.175) takes into account the fact that the relationship between $\vec{E}$ and $\vec{P}$ is nonlocal. The dipole moment at the point with radius vector $\vec{r}$ at a time t is defined by the values of the coordinates of a spatial region around this point (*spatial dispersion*), and depends on the evolution in time (*time dispersion*).

In formula (4.175), the limits of integration over the time t' are arranged by the *principle of causality*: the field $\vec{P}(\vec{r}, t)$ at the time t is prepared by the events that occurred either before or at the same time.

The temporal dispersion is associated with charge inertia and field relaxation processes in a medium, and spatial dispersion represents the transfer of the field action from one point of the medium to another during the transport and/or vibrations of charged particles.

Imagine the properties of the medium be unchangeable in time. Then any process would be independent on the time origin. Given the uniformity of time, we get:

$$\alpha_{ij}(\vec{r}, \vec{r}\,', t, t') = \alpha_{ij}(\vec{r}, \vec{r}\,', t - t') \,.$$

If the medium is also spatially uniform, we get:

$$\alpha_{ij}(\vec{r}, \vec{r}\,', t, t') = \alpha_{ij}(\vec{r} - \vec{r}\,', t - t') \,.$$

Next, we look at only the local relationship over the spatial coordinates, when $\alpha_{ij}(\vec{r} - \vec{r}\,', t - t') = \alpha_{ij}(t - t')\delta^3(\vec{r} - \vec{r}\,')$. This is true for the fields that vary slowly in space. Then the dipole moment is related to the electric field strength as follows:

$$P_i(\vec{r}, t) = \int_{-\infty}^{t} \mathrm{d}t' \alpha_{ij}\left(t - t'\right) E_j\left(\vec{r}, t'\right) \,. \tag{4.176}$$

Formula (4.176) saves the nonlocal (integral) relationship over time. The reality of such a relationship is a delay in the responses of the medium on the external field or on its after action. It is time dispersion of a medium that plays a crucial role for theoretically describing lasers. Time dispersion of a laser medium is ultimately due to the fact that the difference in energy levels of atoms of the medium and frequency of an electromagnetic wave propagating inside the medium are related through the formula $\varepsilon_2 - \varepsilon_1 = h\nu$. This leads to a resonant response of the medium on the electromagnetic wave, with all the after field effects rising drastically.

Moreover, to simplify the further analysis, we assume that the medium is isotropic. Then we have that $\alpha_{ij}(t-t') = \delta_{ij}\alpha(t-t')$, and the constitutive equation coupling the $\vec{E}$ and $\vec{P}$ fields takes the more simple form:

$$P_i(\vec{r},t) = \int_{-\infty}^{t} \mathrm{d}t'\,\alpha(t-t')E_i(\vec{r},t')\,. \tag{4.177}$$

The problem of propagation of electromagnetic fields in a medium boils down to the problem of propagation of plane monochromatic waves; this is due to the completeness of a set of the latter. For monochromatic waves, the time dependence of the field has the following appearance:

$$\vec{E}(\vec{r},t') = \vec{E}(\vec{r},\omega)\exp\left(-\mathrm{i}\omega t'\right) + \text{c.c.}\,. \tag{4.178}$$

Formula (4.178) indicates no explicit dependence of the electromagnetic wave on spatial coordinates. Plugging (4.178) into (4.177), we can rewrite the latter in the vector form:

$$\begin{aligned}\vec{P}(\vec{r},t) &= \int_{-\infty}^{t} \mathrm{d}t'\,\alpha\left(t-t'\right)\left(\vec{E}(\vec{r},\omega)\exp\left(-\mathrm{i}\omega t'\right) + \text{c.c.}\right) = \\ &= \exp\left(-\mathrm{i}\omega t\right)\int_{-\infty}^{t} \mathrm{d}t'\,\alpha\left(t-t'\right)\exp\left[-\mathrm{i}\omega(t'-t)\right]\vec{E}(\vec{r},\omega) + \text{c.c.} \equiv \\ &\equiv \exp\left(-\mathrm{i}\omega t\right)\vec{P}(\vec{r},\omega) + \text{c.c}\,.\end{aligned} \tag{4.179}$$

Now we separately calculate the integral over time:

$$\begin{aligned}\int_{-\infty}^{t} \mathrm{d}t'\,\alpha\left(t-t'\right)\exp\left\{-\mathrm{i}\omega\left(t-t'\right)\right\} &= \updownarrow \tau = t-t' \updownarrow = -\int_{+\infty}^{0} \mathrm{d}\tau\exp\left(\mathrm{i}\omega\tau\right)\alpha\tau = \\ &= \int_{0}^{\infty} \mathrm{d}\tau\exp\left(\mathrm{i}\omega\tau\right)\alpha\left(\tau\right)\,.\end{aligned}$$

We denote the Fourier component of the response function $\alpha(t)$ through $\alpha(\omega)$:

$$\alpha\left(\omega\right) = \int_{0}^{\infty} \mathrm{d}\tau\quad\exp\left(\mathrm{i}\omega\tau\right)\alpha\left(\tau\right)\,.$$

Note that, by virtue of the reality of the function $\alpha(\tau)$, the equality

$$\alpha^{*}\left(\omega\right) = \alpha\left(-\omega\right) \tag{4.180}$$

is true. Since $\vec{D} = \varepsilon_0\vec{E} + \vec{P}$, the field $\vec{D}(\vec{r},t)$ should be expanded as follows:

$$\vec{D}(\vec{r},t) = \vec{D}\left(\vec{r},\omega\right)\exp\left(-\mathrm{i}\omega t\right) + \text{c.c.}\,,$$

where

$$\vec{D}(\vec{r},\omega) = \varepsilon_0 \vec{E}(\vec{r},\omega) + \vec{P}(\vec{r},\omega) = \varepsilon_0 \vec{E}(\vec{r},\omega) + \alpha(\omega)\vec{E}(\vec{r},\omega) \ ,$$

or it is equivalent to:

$$\vec{D}(\vec{r},\omega) = \varepsilon_0 \varepsilon(\omega)\, \vec{E}(\vec{r},\omega) \ , \tag{4.181}$$

where $\varepsilon_0\varepsilon(\omega) = \varepsilon_0 + \alpha(\omega)$.

Thus, the relationship (4.181) between the Fourier components $\vec{D}(\vec{r},\omega)$ and $\vec{E}(\vec{r},\omega)$ is still local, but $\varepsilon(\omega)$ is a complex function. Let us discuss what consequences we may encounter.

The solution of Maxwell's equations:

$$\begin{aligned} &\operatorname{div}\vec{B} = 0 \ , && \operatorname{div}\vec{D} = 0 \ , \\ &\vec{B} = \mu_0 \vec{H} \ , && \vec{D} = \varepsilon_0 \vec{E} + \int\limits_{-\infty}^{t} \mathrm{d}t' \alpha\left(t - t'\right) \vec{E}(\vec{r}, t') \ , \\ &\operatorname{rot}\vec{E} = -\frac{\partial \vec{B}}{\partial t} \ , && \operatorname{rot}\vec{H} = \frac{\partial \vec{D}}{\partial t} \ , \end{aligned} \tag{4.182}$$

needs to be sought in the form of an expansion over plane monochromatic waves, with the spatial coordinate dependence already being clearly highlighted:

$$\vec{E}(\vec{r}, t) = \vec{E}_0(\omega) \exp\left[-\mathrm{i}\omega t + \mathrm{i}\vec{k}\cdot\vec{r}\right] + \text{c.c.} \ , \tag{4.183}$$

$$\vec{H}(\vec{r}, t) = \vec{H}_0(\omega) \exp\left[-\mathrm{i}\omega t + \mathrm{i}\vec{k}\cdot\vec{r}\right] + \text{c.c.} \ . \tag{4.184}$$

Inserting expressions (4.183) and (4.184) into the system of integral differential equations (4.182), and setting equal the coefficients of linearly independent harmonics (exponents), we obtain the system of algebraic equations:

$$(\vec{k}\cdot\vec{E}_0) = 0 \ , \qquad \updownarrow \text{ from } \operatorname{div}\vec{D} = 0 \updownarrow \ , \tag{4.185a}$$

$$(\vec{k}\cdot\vec{H}_0) = 0 \ , \qquad \updownarrow \text{ from } \operatorname{div}\vec{B} = 0 \updownarrow \ , \tag{4.185b}$$

$$[\vec{k}\times\vec{H}_0] = -\omega\varepsilon_0\varepsilon(\omega)\,\vec{E}_0 \ , \quad \updownarrow \text{from rot}\,\vec{H} = \frac{\partial \vec{D}}{\partial t} \updownarrow \ , \tag{4.185c}$$

$$[\vec{k}\times\vec{E}_0] = \mu_0\omega\vec{H}_0 \ , \qquad \updownarrow \text{from rot}\,\vec{E} = -\frac{\partial \vec{B}}{\partial t} \updownarrow \ . \tag{4.185d}$$

In writing formulas (4.185a)–(4.185d), we have taken the relation (4.181) into account.

Vectorially multiplying the equation (4.185c) by the vector $\vec{k}$ yields:

$$\begin{aligned} [\vec{k}\times[\vec{k}\times\vec{H}_0]] &= \vec{k}\underbrace{(\vec{k}\cdot\vec{H}_0)}_{=0} - \vec{H}_0\cdot\vec{k}^2 = -\omega\varepsilon_0\varepsilon(\omega)\,[\vec{k}\times\vec{E}_0] = \\ &= \updownarrow\, [\vec{k}\times\vec{E}_0] = \mu_0\omega\vec{H}_0 \,\updownarrow = -\omega^2\varepsilon_0\mu_0\varepsilon(\omega)\,\vec{H}_0 \ . \end{aligned}$$

Given the relationship $c^2 = 1/\varepsilon_0\mu_0$, and canceling the common factor, we arrive at:

$$k^2 = \frac{\omega^2\varepsilon(\omega)}{c^2} \ . \tag{4.186}$$

Processes in lasers occur at a given real frequency ω of an electromagnetic wave. In this case, the wave vector obeys the laws of electromagnetism and the properties of the medium. According to the relation (4.186), the wave vector, as well as the function $\varepsilon(\omega)$, turn out to be complex.

We write $\sqrt{\varepsilon(\omega)}$ as:

$$\sqrt{\varepsilon(\omega)} = \eta(\omega) + \mathrm{i}\kappa(\omega) \ . \tag{4.187}$$

Then we have $k = (\omega/c)(\eta(\omega) + \mathrm{i}k(\omega))$.

Now, it is convenient to introduce a real unit vector $\vec{n}$ that indicates the direction of propagation of the electromagnetic wave. The expression for the wave vector becomes:

$$\vec{k} = \frac{\omega}{c}(\eta + \mathrm{i}\kappa)\vec{n} \ . \tag{4.188}$$

Suppose that a real unit vector $\vec{e}$ characterizes the polarization of the field $\vec{E}$:

$$\vec{E}_0 = E_0(\omega)\,\vec{e} \ . \tag{4.189}$$

By the virtue of (4.185a), the vectors $\vec{e}$ and $\vec{n}$ are orthogonal.

The vector $\vec{H}_0$ can be expressed through $\vec{E}_0$ from the equation (4.185d):

$$\vec{H}_0 = \sqrt{\frac{\varepsilon_0}{\mu_0}}(\eta + \mathrm{i}\kappa)[\vec{n} \times \vec{e}]E_0 \ ,$$

or:

$$\vec{H}_0 = \sqrt{\frac{\varepsilon_0}{\mu_0}}(\eta + \mathrm{i}\kappa)[\vec{n} \times \vec{E}_0] \ . \tag{4.190}$$

The vectors $\vec{e}$, $[\vec{n} \times \vec{e}]$ and $\vec{n}$ form a right-handed triple (Figure 4.21).

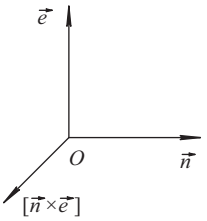


Fig. 4.21: The relative orientation of the vectors $\vec{e}$, $[\vec{n} \times \vec{e}]$ and $\vec{n}$. The vectors $\vec{e}$, $[\vec{n} \times \vec{e}]$ describe the polarization of the fields $\vec{E}$ and $\vec{H}$. The vector $\vec{n}$ characterizes the direction of propagation of electromagnetic wave.

To clarify this result, we consider an electromagnetic wave propagating along the axis Ox ($\vec{n} = (1, 0, 0)$):

$$\vec{E}(\vec{r}, t) = \vec{E}_0(\omega)\exp[-\mathrm{i}\omega t + \mathrm{i}kx] + \text{c.c.} = \vec{E}_0(\omega)\exp\left[-\mathrm{i}\omega t + \frac{\mathrm{i}\omega\eta}{c}x - \frac{\omega\kappa}{c}x\right] + \text{c.c.} =$$

$$= \vec{E}_0(\omega)\exp\left[-\frac{\alpha x}{2}\right]\exp[-\mathrm{i}\omega t + \mathrm{i}k_0 x] + \text{c.c.} \ .$$

In the above formula, the quantity $k_0 = \omega\eta/c = \omega/\tilde{c}$ refers to the real wave number for an electromagnetic wave in a disperse medium. The parameter $\tilde{c} = c/\eta$ describes

the wave velocity, with $\eta(\omega)$ being the refraction index of the disperse medium. The parameter $\alpha = 2\omega \mathrm{k}(\omega)/c$ characterizes attenuation (growth) of amplitude of the wave as it propagates along the axis Ox.

The macroscopic electrodynamics defines an energy flux of an electromagnetic wave as the energy passing per second through the unit area perpendicular to the direction of the wave propagation. The flux of energy is a vector physical quantity (the *Umov–Poynting vector*):

$$\vec{S} = [\vec{E} \times \vec{H}] \,. \tag{4.191}$$

The scalar product $\vec{S} \cdot \vec{n}$ is referred to as the instantaneous intensity of radiation. Here, $\vec{n}$ is the unit vector in the direction of propagation of the wave. The instantaneous energy density of an electromagnetic wave is given by the formula:

$$w = \frac{1}{2}\left[\vec{E} \cdot \vec{D} + \vec{B} \cdot \vec{H}\right] \,. \tag{4.192}$$

By their nature, devices and the human eye are not able to register fast, time varying, instantaneous intensity of radiation and instantaneous energy density of electromagnetic waves. The above tools can detect only average in time values of the given quantities:

$$I \equiv \left\langle \vec{S} \cdot \vec{n} \right\rangle = \frac{1}{T} \int_0^T \vec{S} \cdot \vec{n} \mathrm{d}t \,, \tag{4.193}$$

$$W \equiv \langle w \rangle = \frac{1}{T} \int_0^T w \mathrm{d}t \,, \tag{4.194}$$

where $T = 2\pi/\omega$ is the period of oscillations. The time average instantaneous intensity of radiation is called *the radiation intensity* and denoted by the letter I.

Given the field polarizations (4.183) and (4.184) and the relation $[\vec{e} \times [\vec{n} \times \vec{e}]] = \vec{n}$, formula (4.191) for an electromagnetic wave energy flux can be written as:

$$\vec{S} = \vec{n} E(\vec{r}, t) H(\vec{r}, t) \,, \tag{4.195}$$

where

$$\begin{gathered} E(\vec{r}, t) = E\left(\vec{r}, \omega\right) \exp\left(-\mathrm{i}\omega t\right) + \text{c.c.} \,, \quad E\left(\vec{r}, \omega\right) = E_0\left(\omega\right) \exp\left(\mathrm{i}\vec{k} \cdot \vec{r}\right) \,. \\ H(\vec{r}, t) = \sqrt{\frac{\varepsilon_0}{\mu_0}} (\eta + \mathrm{i}\kappa) E\left(\vec{r}, \omega\right) \exp\left(-\mathrm{i}\omega t\right) + \text{c.c.} \,. \end{gathered} \tag{4.196}$$

It is easy to show that the time average value of the product of two functions of the form

$$\begin{aligned} A(t) &= A\left(\omega\right) \exp\left(\mathrm{i}\omega t\right) + \text{c.c.} \,, \\ B(t) &= B\left(\omega\right) \exp\left(-\mathrm{i}\omega t\right) + \text{c.c.} \,. \end{aligned}$$

is calculated by the formula:

$$\langle A(t)B(t)\rangle = 2\mathrm{Re}\left[A(\omega)B^*(\omega)\right] .$$

Using this result and the relations (4.195) and (4.196), we find the intensity of the radiation:

$$\begin{aligned} I \equiv \left\langle \vec{S}\cdot\vec{n}\right\rangle &= 2\mathrm{Re}\left[E(\vec{r},\omega)H^*(\vec{r},\omega)\right] = 2\left|E(\vec{r},\omega)\right|^2\sqrt{\frac{\varepsilon_0}{\mu_0}}\eta = \\ &= 2\left|E_0(\omega)\right|^2\sqrt{\frac{\varepsilon_0}{\mu_0}}\eta\exp\left[-\frac{2\omega}{c}\kappa(\vec{n}\cdot\vec{r})\right] . \end{aligned} \tag{4.197}$$

In calculating this, we have taken into account that $\mathrm{i}(\vec{k}\cdot\vec{r}) = (\omega/c)(\mathrm{i}\eta - \mathrm{k})(\vec{n}\cdot\vec{r})$.

In particular, when the wave propagates along the axis Ox ($\vec{n} = (1,0,0)$), its intensity changes, depending on the coordinate x as follows:

$$I(x) = I(0)\exp\left[-\frac{2\omega\kappa}{c}x\right] \equiv I(0)\exp(-\alpha x) , \tag{4.198}$$

where $I(0)$ is the value of the intensity at the coordinate $x = 0$;

$$\alpha = \frac{2\omega\kappa(\omega)}{c} . \tag{4.199}$$

Comparing the formula of the macroscopic electrodynamics (4.199) with that previously obtained from the microscopic calculations, we find:

$$\alpha = \frac{2\omega\kappa(\omega)}{c} = (N_1 - N_2)\frac{\widetilde{\lambda}^2 hg(h\nu)}{8\pi t_{\mathrm{spont}}} , \tag{4.200}$$

where $\widetilde{\lambda}$ is the length of the electromagnetic wave in the medium and $\widetilde{\lambda} = \widetilde{c}/\nu$, $\widetilde{c} = c/\eta$ is the speed of light in the medium.

Thus, we have already calculated the parameter κ using quantum theory. The refraction index η can also be found. However, we will not dwell on this here, but will briefly discuss the scheme of the quantum mechanical calculation of η later.

Let us compute the average energy density of the electromagnetic wave:

$$\begin{aligned} W &= \frac{1}{2}\left[\left\langle\vec{E}\cdot\vec{D}\right\rangle + \left\langle\vec{B}\cdot\vec{H}\right\rangle\right] = \\ &= \mathrm{Re}\left[E^*(\vec{r},\omega)\varepsilon_0\varepsilon(\omega)E(\vec{r},\omega)\right] + \mathrm{Re}\left[\varepsilon_0(\eta^2+\kappa^2)E^*(\vec{r},\omega)E(\vec{r},\omega)\right] = \\ &= 2\left|E(\vec{r},\omega)\right|^2\varepsilon_0\eta^2 . \end{aligned}$$

It is worth pointing out that $\varepsilon(\omega) = \eta^2 - \kappa^2 + 2\mathrm{i}\eta\kappa$. The term with the $(\eta^2+\kappa^2)$ is derived from the product H^*H.

If one accepts that $\widetilde{c} = c\eta^{-1} = (\varepsilon_0\mu_0)^{-1/2}\eta^{-1}$ and compares the resulting expression with formula (4.198), we come up with:

$$W = \frac{I}{\widetilde{c}} ,$$

or it is equivalent to:

$$I = \widetilde{c}W \,. \tag{4.201}$$

Expression (4.201) between the intensity of electromagnetic waves and the average energy density has been used previously.

4.12 Intrinsic Oscillations of Optical Laser

The word "laser" is made up from the initial letter of each word from the phrase, "light amplification by stimulated emission of radiation."

The first laser was built on a ruby crystal in 1960. In 1964, N.G. Basov, N.M. Prokhorov, and Charles H. Townes were awarded the Nobel Prize in Physics for the realization of the idea of the laser. Before the advent of lasers, there were no monochromatic radiation sources in optics. Due to exceptional monochromatic radiation, lasers are irreplaceable for performing especially scrupulous measurements of atomic and molecular spectra, as well as for creating a carrier wave in fiber optic communication lines.

The advantages of laser over light sources preceding it are:

- High temporal coherence: measured at different times, correlations between the phases of a laser wave are preserved for periods of time, up to 10^{-1} sec (a wave of conventional sources has the time of correlations of 10^{-10} sec).
- High spatial coherence: a laser wave has an almost perfectly flat front. The front of laser radiation preserves correlations between the phases and polarizations at distances up to $5 \cdot 10^{-2}$ cm.

Since a laser beam diverges very little as it travels away from the laser, the laser radiation can be concentrated in a small volume of the third degree of the radiation wave λ^3, with its intensity maintenance being very high. The intensity of the standard pulsed laser amounts to 10^6–10^{10} W/cm^2. At the same time, the intensity value of a modern high powered laser is 10^{20} W/cm^2. In this case, the electric field strength reaches gigantic values in the order of 10^{11} V/cm^2, which is an order of magnitude higher than that of intra–atomic fields. Such conditions radically change the nature of interactions of optical radiation with matter. This circumstance is widely used in science and engineering. In particular, the laser light makes it possible to generate plasma so that hot thermonuclear fusion can begin. The field of laser radiation reveals unusual nonlinear properties of a medium, such as the generation of multiple harmonics, the self-focusing of the radiation, and the formation of particle like energy bunches (solitons).

Lasers are widespread in medicine, with them being used as light scalpels, for selective destruction of cancer cells, and irradiation of poorly healing wounds or human blood, etc.

Finally, laser light retains its polarization, which is important for carrying out all possible kinds of interference experiments in optical media.

The Operating Principle of a Simple Laser

Consider the operation of the apparatus shown schematically in Figure 4.22. A medium with an inverted population is placed in an optical cavity (a Fabry–Perot etalon) equipped by two parallel semitransparent mirrors facing each other.

Suppose a linearly polarized plane electromagnetic wave fell left into the resonator from a vacuum (point one), its electric field strength being:

$$\vec{E}_i = \vec{e}\,\{E_i \exp[-\mathrm{i}\omega t + \mathrm{i}kx] + \text{c.c.}\}\ , \tag{4.202}$$

where $\vec{e}$ is the unit polarization vector. In the medium, the real wave number $k = \omega/c$ changes and becomes complex:

$$k \to k' = \frac{\omega}{c}(\eta + \mathrm{i}\kappa)\ . \tag{4.203}$$

As the wave travels through the cavity, other phase factor characterizes its space time change:

$$\exp\left[-\mathrm{i}\omega t + \mathrm{i}k'x\right]\ .$$

We write down the imaginary part of the wave number k' in the following form:

$$\mathrm{Im}k' \equiv \frac{\omega\kappa\,(\omega)}{c} = \frac{\alpha_0 - \gamma}{2}\ .$$

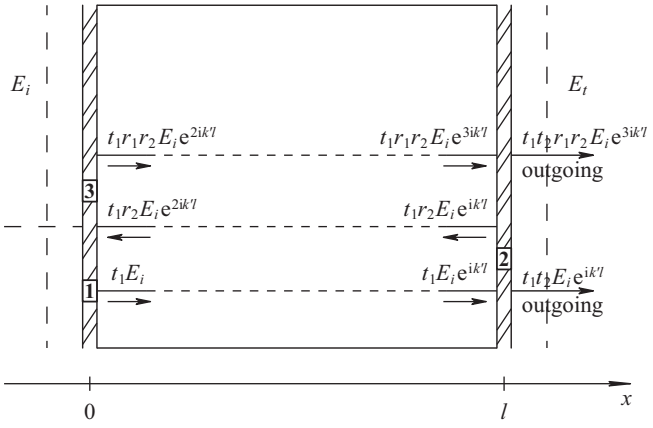


Fig. 4.22: Scheme of a laser generator. The center is an optical resonator made of two semitransparent mirrors, with an inverted population medium placing between them.

The term $-\gamma/2$ governs the $2 \leftrightarrow 1$ energy transitions of atoms of the medium:

$$\gamma = (N_2 - N_1)\frac{\widetilde{\lambda}^2 hg\,(h\nu)}{8\pi t_{\text{spont}}}\,, \tag{4.204}$$

where $\widetilde{\lambda} = \widetilde{c}/\nu$ is the wavelength in the medium and $\widetilde{c} = c/\eta$ is the speed of light in the medium. The phenomenological parameter $\alpha_0 > 0$ is responsible for the wave absorption by the medium, due to all other energy loss mechanisms except for the transitions $2 \leftrightarrow 1$.

We designate the ratio of the amplitude of the electric field transmitted through the left mirror to the amplitude of the incident electric field as t_1, and as t_2 for the right mirror. Suppose $r_1(r_2)$ is the ratio of the amplitude of the electric field reflected within the medium to the amplitude of the incident electric field for the left (right) mirror.

Now we will discuss how the wave (4.202) passes through the cavity. At point one, the wave with the amplitude $t_1 E_i$ falls left into the medium through the first mirror. After going through that mirror and reaching the second mirror, the wave acquires an additional phase factor $\exp(\mathrm{i}lk')$, where l is the resonator length. Furthermore, the wave is partially reflected back (point two) and partially transmitted through the second mirror. The multipliers r_2 and t_2, respectively take these wave amplitude portions into account. Once crossed again the resonator, the reflected wave, gains the multiplier $\exp(2\mathrm{i}lk')$. After that, it is partially reflected by (point three) and partially penetrates through the mirror. However, we are interested only in the resulting wave that comes out from the right side of the cavity. For brevity, we will not discuss the resulting wave at the left output of the cavity; that is why it is not shown in the figure. After multiplying the wave coming into point three by the coefficient r_1, we obtain the amplitude of the reflected wave at the same point. Then the wave reaches the right mirror with the phase factor $\exp(3\mathrm{i}lk')$, and its part t_2 leaves the cavity. It can be seen that we have arrived at a geometric progression of amplitudes at the output of the resonator (right): $t_1 t_2 E_i \exp(\mathrm{i}lk')$, $t_1 t_2 r_1 r_2 E_i \exp(3\mathrm{i}lk')\ldots$

We sum up the geometric progression to find the outgoing wave:

$$\begin{aligned}\vec{E}_t(r,t) = \vec{e}\Big\{t_1 t_2 E_i \exp\left(-\mathrm{i}\omega t + \mathrm{i}k'l\right)\left[1 + r_1 r_2 \exp\left(2\mathrm{i}k'l\right) + \right.\\ \left. + r_1^2 r_2^2 \exp\left(4\mathrm{i}k'l\right) + \cdots\right] + \text{c.c.}\Big\} =\\ = \vec{e}\left\{\frac{t_1 t_2 E_i \exp\left(-\mathrm{i}\omega t + \mathrm{i}k'l\right)}{1 - r_1 r_2 \exp\left(2\mathrm{i}k'l\right)} + \text{c.c.}\right\}\,.\end{aligned}$$

Generally speaking, the calculation should make allowances for changes of the electromagnetic wave phase by $\pm\pi$, due to this reflecting from the denser medium. However, the given problem may ignore this circumstance because the wave experiences an even number of reflections before leaving the resonator.

When the denominator in the formula for the outgoing wave

$$\vec{E}_t(r,t) = \vec{e}\left\{\frac{t_1 t_2 E_i \exp\left(-\mathrm{i}\omega t + \mathrm{i}k'l\right)}{1 - r_1 r_2 \exp\left(2\mathrm{i}k'l\right)} + \text{c.c.}\right\} \tag{4.205}$$

vanishes, even a negligible wave entering the resonator provides a giant wave amplitude at its output. We have come up with a condition that initiates the intrinsic oscillations inside the resonator:

$$r_1 r_2 \exp\left(2ik'l\right) = 1\ , \tag{4.206}$$

where, according to formula (4.203): $k' = k_0 + i(\alpha_0 - \gamma)/2$, $k_0 = \omega\eta/c = \omega/\tilde{c}$. We split the complex relation (4.206) on two conditions: a phase and amplitude one.

The phase condition has the appearance $2k_0 l = 2m\pi$, where m is an integer. This is a normal condition for interference amplification of the multiply reflected waves between the mirrors:

$$k_0 l = m\pi\ . \tag{4.207}$$

Given that $\omega = 2\pi\nu$, we can rewrite equation (4.207) in the form:

$$\frac{2\nu l}{\tilde{c}\,(\nu)} = m\ . \tag{4.208}$$

The frequency ν in (4.208) cannot be arbitrary because it is related to the energies of atoms of the medium: $\nu = (\varepsilon_2 - \varepsilon_1)/h$. However, condition (4.208) can be easily satisfied by changing the resonator length l.

The amplitude condition $r_1 r_2 \exp[l(\gamma(\omega) - \alpha_0)] = 1$ can be rewritten as a formula for the threshold gain:

$$\gamma_{\text{threshold}}\,(\omega) = \alpha_0 - \frac{1}{l}\ln\,(r_1 r_2)\ . \tag{4.209}$$

The explicit form of the factor $\gamma(\omega)$ (4.204) yields the critical density of the inversed population of the atomic levels. Once it has exceeded its value, we can excite intrinsic oscillations in the laser resonator:

$$N_{\text{threshold}} = (N_2 - N_1)_{\text{threshold}} = \frac{8t_{\text{spont}}\pi}{hg\,(h\nu)\,\tilde{\lambda}^2}\left[\alpha_0 - \frac{1}{l}\ln(r_1 r_2)\right]\ . \tag{4.210}$$

Upon excitation of intrinsic oscillations inside the resonator, the laser takes energy from the inversed population medium and emits it in the optical range. Without producing "pumping," or in other words, without transferring the new atoms of the medium from the ground state to excited states, the inverted population decreases and the intrinsic laser oscillations decay away. This happens at the time when the inversed population of the atomic levels falls below the calculated critical value $N_{\text{threshold}}$. If the atoms of the medium in the resonator are always "pumped" to level two, the intrinsic oscillations can be achieved uninterrupted. Hence, *lasers* can be classified as operating in either *continuous* or *pulsed mode*.

4.13 A Pulsed Ruby Laser

The first quantum generator of light was a ruby laser built in 1960, which operates, for example, in pulse mode.

A working substance of a ruby laser is a ruby crystal that is aluminum oxide (Al_2O_3) known as corundum, doped by chromium when grown. The red color of ruby is caused by the presence of the positive ions of Cr^{3+}. An Al^{3+} ion in the Al_2O_3 crystal lattice substitutes for a Cr^{3+}-ion. As a consequence, the crystal has two absorption bands: one is in the green part of the spectrum and another is shifted in the blue part (Figure 4.24). The density of the red color of ruby depends on the concentration of Cr^{3+}ions: the higher the concentration is, the richer the red color. The ion concentration in dark red ruby reaches 1%.

A ruby laser uses energy transitions of Cr^{3+} ions. Pumping is done by irradiating the crystal ruby with an intense flash of light from a powerful xenon filled tube (Figure 4.23). The latter is similar to that applied when photographing. The tube produces a burst of light as a current pulse passes through it. The current pulse instantly heats the xenon gas to several thousand Kelvin. The pumping is impossible to maintain uninterrupted, because the tube cannot withstand continuous operation over a long period at such a high temperature.

The tube light is not monochromatic, so it is important that the Cr^{3+} ion-spectrum has quasicontinuous bands 3 and 4 (Figure 4.24). It is towards these energy states that the Cr^{3+} ions initially move, absorbing photons of light from the xenon filled tube.

As can be seen from Figure 4.24, the vertical lines are responsible for the transitions pumping. The wavy line indicates a laser transition. The inclined lines show extremely fast radiationless transitions (within a time order of 10^{-7} sec). In this case,

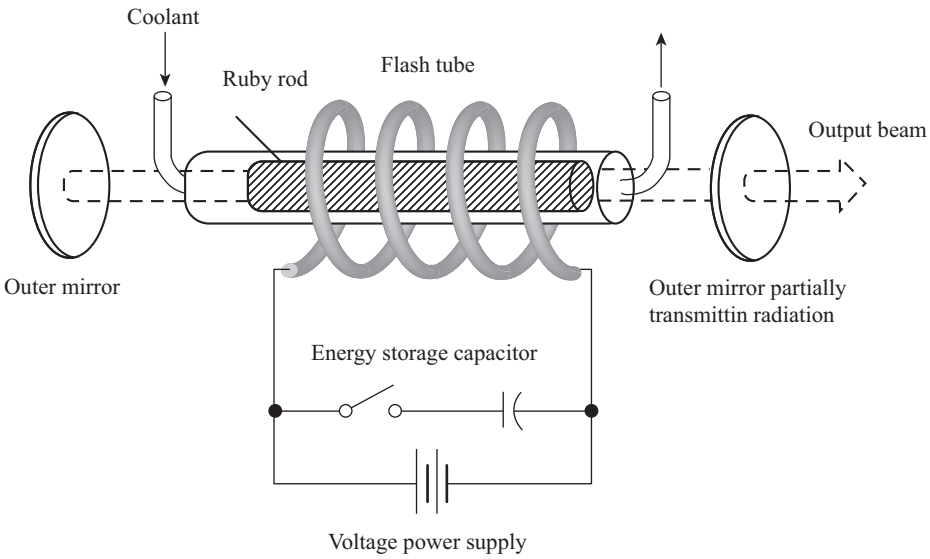


Fig. 4.23: A typical pulse ruby laser with two outside mirrors, pumped by a flash lamp.

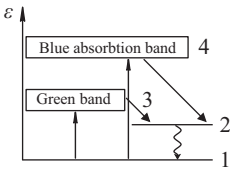


Fig. 4.24: Energy levels used in a ruby laser.

the excess energy is converted into the energy of lattice vibrations, which leads to the heating of ruby. Therefore, the ruby rod should be continuously cooled.

Life time of the Cr^{3+} ions at the level two is relatively great (of the order of $3 \cdot 10^{-3}$ sec). Therefore, the pumping transfers most of the ions from the ground (first) level to the upper metastable level (2nd). Intrinsic oscillations in a ruby laser are excited when the population difference is $N_2 - N_1 > N_{\text{threshold}}$.

The polarization radiation characteristics of a ruby laser depends, first of all, on the orientation of the ruby crystal. Ruby is a uniaxial crystal. Therefore, when the laser geometrical and optical axes are perpendicular to each other, a linearly polarized light can be obtained. In contrast, if they are parallel, the laser radiation is unpolarized.

If a lasing medium is isotropic (for example, a cubic crystal), a polarized light can be produced by reflecting an output beam. For this purpose, instead of a semitransparent mirror, the resonator output (right side of the operating medium) is built of a biprism with two mirrors placed on its bottom and right edges (Figure 4.25). The cross section of the biprism is an isosceles triangle; its base corresponds to the transparent left edge of the biprism. The angle at the base of the triangle and the turn of the biprism are chosen so that the light rays inside the biprism should be incident and reflected from the mirrors at right angles. Such a system divides the light beam coming out from the operating medium into two parts. Refracted in the biprism and reflected from the mirrors, one part of the beam, as expected, returns to the resonator. Another part, as a result of reflection (refraction) of the output radiation at the Brewster angle on the left side of the biprism, becomes linearly polarized. This part of the beam leaves the resonator.

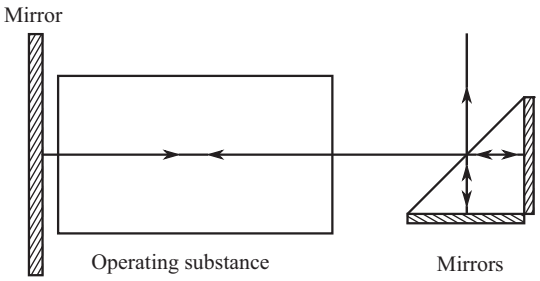


Fig. 4.25: Scheme for obtaining polarized radiation with a biprism.

The general principle of producing polarized radiation is as follows. For this, the direction of preferential orientation should be created in the operating lasing medium by, for example, adding dyes whose fluorescent light is polarized. This causes a sharp rise in the degree of emitted radiation polarization over time, even when the initial polarization is negligible.

During laser radiation processes, the population difference decreases and the laser beam is interrupted when the degree of inversion of the medium becomes smaller than the threshold value.

4.14 Heterolasers

On 10 October 2000, the Nobel Committee announced that the annual prize in physics was going to be awarded to Zhores I. Alferov (The Ioffe Physico – Technical Institute, the Russian Academy of Sciences, St. Petersburg, Russia), Herbert Kroemer (University of California, Santa-Barbara, USA), and Jack S. Kilby (Texas Instruments Company, Dallas, USA). The scientists were awarded for the invention and development of fast optoelectronic and microelectronic components, based on so called heterojunctions and heterostructures.

A heterojunction is defined as the contact of two semiconductors different in chemical composition. As a rule, at the interface of the semiconductors, a bandgap, mobility of charge carriers, their effective mass, and other important physical characteristics of the materials are changed. However, a "dramatic" heterojunction changes the semiconductor properties at a distance comparable to, or less, than the width of the space charge region. Combinations of different heterojunctions such as *p*-*n*-, *n*-*n*-, *p*-*p*-types can form such heterostructures.

As far back as distant 1957, Herbert Kroemer, an American of German origin, invented the first heterostructured transistor unique in its properties. Six years later he and Zh. I. Alferov (independently of one another) offered the tenets underlying the design of innovative, heterostuctured lasers. In the same year, Zh. Alferov patented an injection quantum generator that became famous afterwards. The third laureate physicist made an enormous contribution to the development of integrated circuits on heterostructures. The fundamental works of these researchers made it possible to create subminiature devices with fantastic properties for modern optoelectronics and microelectronics.

In 1962, Zh. Alferov, jointly with a small group of enthusiasts, began investigations of semiconductor heterostructures. At this time, scientists had to have amazing intuition and scientific courage to leave a well trodden track of physics of semiconductor monojunctions and take the research path of heterostructures, where no one has found anything before. There were very few who believed in the possibility of fabricating an ideal heterojunction. As a matter of fact, a heterojunction with useful properties is possible to be made only in the event of coinciding types, orientations, and periods

of crystal lattices of matched materials. Moreover, an ideal heterojunction interface needs to be free of structural defects and mechanical strains. It can even be said that the interface of semiconductors with different conductivity types is the technical device itself.

Manufacturing of ideal heterojunctions and heterostructures has become possible due to the development of methods of epitaxial growth of semiconductor crystals. Zh. Alferov put forward and implemented a brilliant idea: to grow other crystals with the same lattice parameters, but with a completely different energy spectrum of electrons and holes, on the face of a crystal with certain lattice parameters.

All materials to produce heterostructures refer to the central part of the periodic system of elements (Groups III-V). The middle of the periodic table (Group IV) includes germanium (Ge) and silicon (Si). However, in practice, quite a lot of effort was required to develop heterostructures based on germanium -silicon alloys because of the 4% difference of silicon and germanium lattice constants.

Silicon and germanium play the same important role in electronic device technologies as steel in technology of structural materials. However, similar to modern metallurgy that uses other materials besides steel, electronics employs, along with silicon and germanium, semiconductor solid solutions. Each element of Group III can bind any element of group V, with compounds of type $A^{III}B^{V}$ forming.

Gallium and aluminum belong to Group III of the periodic table, and arsenic belongs to the fifth one. GaAs (gallium arsenide) and an $Al_xGa_{1-x}As$ solid solution (aluminum gallium arsenide) were found to be the most suitable for creating semiconductor heterostructures, wherein a portion of gallium atoms are replaced by aluminum atoms. The value of x usually varies between 0.15 and 0.35. The band gap of gallium arsenide amounts to 1.5 eV and grows with increasing x in $Al_xGa_{1-x}As$ solid solutions. So, in AlAs compounds ($x = 1$), the band gap is 2.2 eV.

For the components of the systems listed above, the periods and thermal expansion coefficients of the crystal lattices coincide with great accuracy. This ensures no defects and stresses in them and, consequently, high quality surfaces of interfaces in heterostructures.

Mobility of electrons or holes is a characteristic that is responsible for how easily the charge carriers move through a matter. At low temperatures, when the scattering by phonons can be neglected, the mobility of the charge carriers in today's heterostructures of $GaAs/Al_xGa_{1-x}As$ is 1000 times higher than that of silicon, and is equal in order of magnitude to $10^7 cm^2/(V{\cdot}s)$. In this case, the mean free path of an electron before it scatters is approximately 0.2 mm, meaning that the electron passes by a million atoms without scattering.

The discovery of ideal heterojunctions by Alferov, as well as the development of technologies for their production, entailed successful research of unique electrical and optical properties of heterostructures. In particular, many new effects were predicted. Among these are:

(a) The superinjection effect represents abruptly increasing the density of carriers injected into the narrow band gap layer of a semiconductor of heterostructures as compared to the carrier density of the wide band gap emitter.
(b) The electron confinement effect that is due to changing the width of a forbidden energy band and provides localization of charge carriers in the thin narrow band gap layer of a semiconductor.

Both effects are extremely important, for example, to create inverse population in the active layer of a semiconductor laser.

(c) The optical confinement effect that is associated with a difference in refractive indices of the layers and allows one to create spatial localization of electromagnetic waves. This fact is essential, for example, in generating a powerful electromagnetic output pulse of the active layer of a laser without energy losses.

These effects enriched opportunities to control the motion of charge carriers and light fluxes in heterostructures. This enabled one to dramatically improve the parameters of most known semiconductor devices, to create fundamentally new devices, especially promising for applications in optical and quantum electronics. Over a very short time (in the late 60s/early 70s), Zh. Alferov and his colleagues have designed a first low threshold heterolaser that operates in a continuous mode at room temperature, highly effective light emitting diodes, photodiodes, phototransistors, and solar cells. Manufacturers gained the technology of producing new types of power diodes, transistors, and thyristors based on heterostructures.

The simplest injection heterolaser includes two heterojunctions. One of them, of the *p*-*n*-type, injects electrons. In fact, it is a diode switched in the forward direction. The second heterojunction, of the *p*-*p*-type, limits the spreading of charge carriers from the active middle layer. It is important that the middle layer should consist of a material having a smaller band gap and a larger dielectric constant than the outer layers. Two plane parallel faces of the sandwich of the semiconductor layers, perpendicular to the planes of the *p*-*p*- and *p*-*n*-junctions, serve as mirrors of the optical resonator (Figure 4.26).

Active particles in a laser are free charge carriers; electrons and holes. Population inversion is achieved by passing a high forward current through a *p*-*n*-junction. In this case, excess charge carriers inject into the active layer.
Due to potential barriers at the boundaries of the active layer (Figure 4.27), the electron hole plasma turns out to be trapped in a potential well in the middle layer. Therefore, there are no recombination energy losses in the outer layers.

The coherent lasing is brought about by means of quantum transitions in the middle layer between permitted bands rather than discrete energy levels as in a ruby laser. Due to the significant difference in the dielectric constants, the middle layer acts as a high quality waveguide. So, radiation losses in the outer layers are negligible (the phenomenon of total internal reflection of electromagnetic waves takes place).

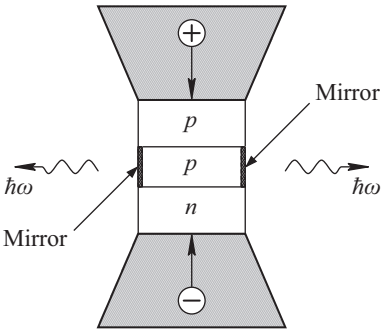


Fig. 4.26: Overall scheme of an injection heterolaser.

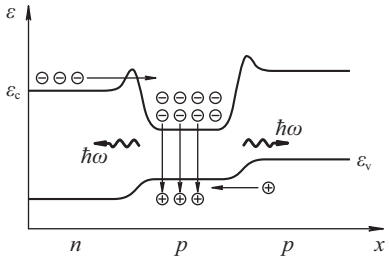


Fig. 4.27: Energy diagram of laser heterostructures: ε_c and ε_v are the edges of the conduction and valence bands.

Quantum well lasers can be tuned in wavelength: the frequency generated by the laser increases with decreasing the well depth.

A current that corresponds to the beginning of the generation is called a threshold one. In the heterolasers, the threshold current is record small – twenty times lower than in conventional semiconductor lasers. An important achievement is to provide continuous generation of radiation without cooling heterolasers, which is impossible for single junction semiconductor lasers operating on a similar scheme.

Other attractive features of the heterolasers are: small dimensions of the active medium, low inertia and high efficiency of conversion of electrical energy into light energy (it amounts to 60% as compared to that of 2–3% for single junction semiconductor lasers).

In the early 70s, Zh. Alferov, with his staff, formulated the tenets of ideal heterostructures using multicomponent solid solutions; in particular, InGaAsP-heterostructures (quaternary or more). The lattice period of such solutions is found to be slightly correctable as their composition changes, while the bandgap width ranges widely. This circumstance made it possible to expand the radiation range of heterolasers into both the infrared and visible spectral regions. It is the same lasers that are now being used as a radiation source in fiber optic communication lines at extended distances.

In the 80s/90s, Zh. Alferov and his team continued studies of solar cells based on heterostructures, which led to the creation of photovoltaic cells operating under concentrated solar radiation. The efficiency coefficient of the new generation of devices approached the theoretical one and was about 30–40%. For comparison, the efficiency of the first semiconductor solar cells accounted to only 1%. Similar converters established a good basis for works on cosmic and solar energetics. Solar cells of this type turned out to be extremely efficient and reliable. For example, they had been

working aboard the Russian Space Station Mir for many years (about ten years) without noticeable power losses.

Although it sounds fantastic, heterostructures give the possibility of designing new types of materials with desired properties at interatomic distances. In the future, these materials will significantly improve characteristics of many devices and systems, making the latter subminiaturized. The first fruits of the new electronics have already become part of our lives. Laser diodes based on heterostructures can be found in CD-ROM drivers. Fiber optic communication lines, including the Internet, became available only due to new properties of materials based on heterostructures. High performance transistors, based on heterostructures, are being used today in satellite communications, mobile phones, etc.

4.15 Formalism of the Density Matrix and Semiclassical Theory of the Propagation of Electromagnetic Waves in a Two Level Atom Medium

The formalism of the density matrix is widely used for describing quantum systems consisting of a large number of identical particles. It is convenient to carry out averaging over the ensemble of particles.

Let us illustrate the method on a concrete example. Consider an interaction of N two level atoms with an electromagnetic wave. The electromagnetic wave is assumed to contain many photons, so it can be described in terms of Maxwell's theory. We restrict ourselves to the electric dipole approximation, which takes into account the main interaction between the atoms and the electromagnetic wave field.

For a nonmagnetic medium ($\vec{M} = \vec{0}$, $\vec{B} = \mu_0 \vec{H}$), Maxwell's equations acquire the form:

$$\begin{aligned} \operatorname{div} \vec{E} &= \frac{1}{\varepsilon_0} \operatorname{div} \vec{P} \,, \\ \operatorname{div} \vec{H} &= 0 \,, \\ \operatorname{rot} \vec{H} &= \vec{j}_{\text{ext}} + \frac{\partial}{\partial t} \left(\varepsilon_0 \vec{E} + \vec{P} \right) \,, \\ \operatorname{rot} \vec{E} &= -\mu_0 \frac{\partial \vec{H}}{\partial t} \,, \end{aligned} \tag{4.211}$$

where $\vec{P}$ is the dipole moment per unit volume of the medium. The density of the external electric current

$$\vec{j}_{\text{ext}} = \sigma \vec{E} \tag{4.212}$$

is introduced to allow certain problems to be solved, such as power losses due to any absorbing background medium, and also due to wave diffraction and mirror transmission in lasers.

From Maxwell's equations it is easy to derive an formula that defines the distribution of the electric field $\vec{E}$ in the medium:

$$\text{rot rot}\,\vec{E} = -\Delta\vec{E} + \vec{\nabla}(\text{div}\,\vec{E}) = -\mu_0\sigma\frac{\partial\vec{E}}{\partial t} - \frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2} - \mu_0\frac{\partial^2\vec{P}}{\partial t^2}\,.$$

It should be noted that $\text{div}\,\vec{E}$, in general, is different from zero. In the resulting equation, we move all the terms containing $\vec{E}$ to the left-hand side to get:

$$-\Delta\vec{E} + \vec{\nabla}(\text{div}\,\vec{E}) + \frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2} + \mu_0\sigma\frac{\partial\vec{E}}{\partial t} = -\mu_0\frac{\partial^2\vec{P}}{\partial t^2}\,. \tag{4.213}$$

For definiteness, consider transverse electromagnetic waves that propagate along the axis Ox. Suppose they are linearly polarized along the Oz-axis (Figure 4.28). Then $\vec{E} = (0, 0, E(x, t))$, $\vec{P} = (0, 0, P(x, t))$, and we can simplify equation (4.213) up to:

$$\frac{\partial^2}{\partial x^2}E - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} - \mu_0\sigma\frac{\partial E}{\partial t} = \mu_0\frac{\partial^2}{\partial t^2}P\,. \tag{4.214}$$

Next, we represent a two level atom as a quantum system in terms of energetic spin. A Hamiltonian that describes the state of the atom and its interaction with the electric field $\vec{E}$ has the form:

$$\widehat{H} = \widehat{H}_{\text{atom}} + \widehat{H}_{\text{int}}\,, \tag{4.215}$$

$$\widehat{H}_{\text{atom}} = \frac{\varepsilon_1 + \varepsilon_2}{2}\widehat{I} + \frac{\varepsilon_1 - \varepsilon_2}{2}\widehat{R}_3\,, \tag{4.216}$$

$$\widehat{H}_{\text{int}} = -\widehat{\vec{p}}\cdot\vec{E}\,, \tag{4.217}$$

where $\widehat{\vec{p}} = -|e|\vec{D}_{12}\widehat{R}_1$ is the operator of the dipole moment of the atom. Recall that:

$$\begin{aligned}
&\widehat{I} = |1\rangle\langle 1| + |2\rangle\langle 2|\,, && \widehat{R}_1 = |2\rangle\langle 1| + |1\rangle\langle 2|\,,\\
&\widehat{R}_2 = \mathrm{i}\,(|2\rangle\langle 1| - |1\rangle\langle 2|)\,, && \widehat{R}_3 = |1\rangle\langle 1| - |2\rangle\langle 2|\,,\\
&\widehat{R}_s\widehat{R}_p = \mathrm{i}\varepsilon_{spq}\widehat{R}_q + \delta_{sp}\widehat{I}\,,
\end{aligned} \tag{4.218}$$

where ε_{spq} in the unit is a completely antisymmetric pseudotensor ($\varepsilon_{123} = 1$); the repeated index implies summation over the index range.

Suppose $|\Psi\rangle$ is a state of an isolated atom. The vector of $|\Psi\rangle$ satisfies the Schrödinger equation:

$$\mathrm{i}\hbar\frac{\partial}{\partial t}|\Psi\rangle = \widehat{H}\,|\Psi\rangle \tag{4.219}$$

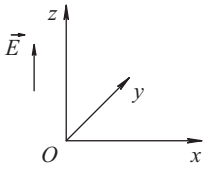


Fig. 4.28: The polarization direction of the electric field propagating along the axis Ox.

or

$$-\mathrm{i}\hbar\frac{\partial}{\partial t}\langle\Psi| = \widehat{H}\langle\Psi| \tag{4.220}$$

in virtue of the fact that the operator $\widehat{H}$ is Hermitian, i.e., $\widehat{H}^{+} = \widehat{H}$.

The ket-vector of $|\Psi\rangle$ can be expanded in an arbitrary complete set of vectors. We choose as such vectors the orthonormal eigenvectors of the operator $\widehat{H}_{\text{atom}}$:

$$\widehat{H}_{\text{atom}}|i\rangle = \varepsilon_i|i\rangle\ , \quad i = 1, 2\ , \tag{4.221}$$

$$\langle i|j\rangle = \delta_{ij}\ . \tag{4.222}$$

Then $|\Psi\rangle = c_1(t)|1\rangle + c_2(t)|2\rangle$, where $c_i = \langle i|\Psi\rangle = \langle\Psi|i\rangle^{*}$. Next, we consider the state $|\Psi\rangle$ as normalized:

$$\langle\Psi\Psi\rangle = |c_1|^2 + |c_2|^2 = 1\ . \tag{4.223}$$

From quantum mechanics, it is known that, in the state $|\Psi\rangle$, the average value of any observable quantity corresponding to the Hermitian operator $\widehat{A}$ is calculated by the formula:

$$\langle\Psi|\widehat{A}|\Psi\rangle = \sum_{i,j=1,2}\langle\Psi|i\rangle\langle i|\widehat{A}|j\rangle\langle j|\Psi\rangle = \sum_{i,j=1,2} c_i^{*}c_j A_{ij}\ . \tag{4.224}$$

In deriving the equality, we have used the condition of completeness of the states $\{|i\rangle\}$:

$$\sum_i |i\rangle\langle i| = \widehat{I}\ . \tag{4.225}$$

The average value calculated by (4.224) determines the observed quantity A for the isolated atom. In practice, most problems deal with systems that consist of a large number of identical atoms. In fact, we measure, not the quantity $\langle\Psi|\widehat{A}|\Psi\rangle$, but its value additionally averaged over all the N atoms of the medium. Denote the appropriate average as $\langle\langle\widehat{A}\rangle\rangle$:

$$\begin{aligned}\left\langle\left\langle\widehat{A}\right\rangle\right\rangle &= \frac{1}{N}\sum_k N_k \sum_{i,j=1,2} c_i^{(k^{*})}c_j^{(k)}A_{ij} = \sum_k w_k \sum_{i,j=1,2}\left\langle\Psi^{(k)}\middle|i\right\rangle\langle i|\widehat{A}|j\rangle\left\langle j\middle|\Psi^{(k)}\right\rangle = \\ &= \sum_{j=1,2}\langle j|\left(\sum_k w_k\left|\Psi^{(k)}\right\rangle\left\langle\Psi^{(k)}\right|\sum_{i=1,2}|i\rangle\langle i|\widehat{A}\right)|j\rangle = \sum_{j=1,2}\langle j|\widehat{\rho}\widehat{A}|j\rangle = \mathrm{Sp}\left(\widehat{\rho}\widehat{A}\right)\ .\end{aligned} \tag{4.226}$$

Here, N_k is the number of atoms of the medium in the state $|\Psi^{(k)}\rangle$ and $\sum_k N_k = N$, $w_k = N_k/N$ is the probability of realizing the quantum mechanical state $|\Psi^{(k)}\rangle$ in the ensemble of N atoms. The density operator is given by:

$$\widehat{\rho} = \sum_k w_k\left|\Psi^{(k)}\right\rangle\left\langle\Psi^{(k)}\right|\ . \tag{4.227}$$

It is easy to see that $\widehat{\rho}^{+} = \widehat{\rho}$ (the operator $\widehat{\rho}$ is Hermitian) and:

$$\mathrm{Sp}\,\widehat{\rho} = \sum_k w_k = 1\ . \tag{4.228}$$

Let us find equations to be solved to get the density operator $\hat{\rho}$. For this purpose, we calculate the total time derivative of expression (4.227):

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho} = \sum_k w_k \left(\left(\frac{\partial}{\partial t} \left| \Psi^{(k)} \right\rangle \right) \left\langle \Psi^{(k)} \right| + \left| \Psi^{(k)} \right\rangle \left(\frac{\partial}{\partial t} \left\langle \Psi^{(k)} \right| \right) \right) .$$

Using formulas (4.219) and (4.220), we replace the quantities $\partial|\Psi^{(k)}\rangle/\partial t$ and $\partial\langle\Psi^{(k)}|/\partial t$ by their values, and finally obtain:

$$\frac{\mathrm{d}\hat{\rho}}{\mathrm{d}t} = \frac{1}{\mathrm{i}\hbar} \left[\widehat{H}, \hat{\rho} \right] . \tag{4.229}$$

It is worth noting that the operator $\hat{\rho}$ is a specific operator, instead of a wave function. Therefore, unlike other operators of the Schrödinger picture, it varies over time.

We introduce the representation of the values observed in the basis $\{|i\rangle\}$ where the operators $\widehat{H}$ and $\hat{\rho}$ correspond to 2×2 matrices. The operator $\hat{\rho}$ corresponds to the density matrix:

$$\hat{\rho} = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} , \tag{4.230}$$

Where $\rho_{ij} = \langle i|\hat{\rho}|j\rangle$, and by virtue of hermiticity of $\hat{\rho}$,

$$\rho_{11} = \rho_{11}^* , \quad \rho_{22} = \rho_{22}^* , \quad \rho_{12} = \rho_{21}^* . \tag{4.231}$$

The Hamiltonian (4.216) acquires the form:

$$\widehat{H} = \frac{\varepsilon_1 + \varepsilon_2}{2} \hat{I} + \frac{\varepsilon_1 - \varepsilon_2}{2} \hat{\sigma}_3 - \hat{\vec{p}} \cdot \vec{E} , \tag{4.232}$$

where $\hat{\vec{p}} = -|e|\vec{D}_{12}\hat{\sigma}_1$ and $\hat{\sigma}_i$ is Pauli matrices. Recall the appearance and algebra of the matrices:

$$\hat{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \quad \hat{\sigma}_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} , \quad \hat{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \quad \hat{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} , \tag{4.233}$$

$$\hat{\sigma}_s \hat{\sigma}_p = \mathrm{i}\varepsilon_{spk}\hat{\sigma}_k + \delta_{sp}\hat{I} .$$

Here, as previously, ε_{spk} is the Levi–Civita symbol.

We write down a matrix corresponding to the operator Hamilton (4.232) in detail, using the fact that the given problem addresses $\vec{E} = (0, 0, E)$, $\vec{D}_{12} = (0, 0, D_{12})$:

$$\widehat{H} = \begin{pmatrix} \varepsilon_1 & |e| D_{12} E \\ |e| D_{12} E & \varepsilon_2 \end{pmatrix} . \tag{4.234}$$

Let us elucidate the physical meaning of the elements of the density matrix $\hat{\rho}$. The quantity

$$\rho_{11} = \sum_k w_k \left| \left\langle 1 | \Psi^{(k)} \right\rangle \right|^2 . \tag{4.235}$$

describes the probability of detecting a group of atoms with energies ε_1 in the system of N atoms, i.e., the atoms in the ground state. This group contains $N_1 = N\rho_{11}$ atoms.

Similarly, the quantity ρ_{22} makes it possible to calculate the number of atoms N_2 with the energies $\varepsilon_2 > \varepsilon_1$, being in the excited state:

$$N_2 = N\rho_{22} \, .$$

To reveal the physical meaning of the off diagonal elements of the matrix $\hat{\rho}$, we calculate the average value of the dipole moment of a single atom in the ensemble of N atoms:

$$\begin{aligned} \left\langle \left\langle \hat{\vec{p}} \right\rangle \right\rangle &= \mathrm{Sp}\left(\hat{\rho}\hat{\vec{p}}\right) = -|e|\,\vec{D}_{12}\,\mathrm{Sp}\left(\rho\sigma_1\right) = -|e|\,\vec{D}_{12}\left(\rho_{12}+\rho_{21}\right) = \\ &= -|e|\,\vec{D}_{12}\left(\rho_{21}+\rho_{21}^{*}\right) \, . \end{aligned} \tag{4.236}$$

Thus, the off diagonal elements of the density matrix determines the mean dipole moment of the individual atoms of the medium. Therefore, the dipole moment per unit volume of the medium is:

$$\vec{P} = n\left\langle \left\langle \hat{\vec{p}} \right\rangle \right\rangle = -n\,|e|\,\vec{D}_{12}\left(\rho_{21}+\rho_{21}^{*}\right) \, , \tag{4.237}$$

where $n = N/V$ is the number of atoms per unit volume of the medium.

As a result, the Maxwell equation (4.214) that defines the electric field distribution in the medium takes the form:

$$\frac{\partial^2}{\partial x^2}E - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}E - \mu_0\sigma\frac{\partial E}{\partial t} = \mu_0 n\,|e|\,D_{12}\frac{\partial^2}{\partial t^2}\left(\rho_{21}+\rho_{21}^{*}\right) \, . \tag{4.238}$$

Here, we have taken into account that $\vec{E} = (0, 0, E)$, $\vec{D}_{12} = (0, 0, D_{12})$ in the problem considered.

In the basis $\{|i\rangle\}$, the equations for the density operator $\hat{\rho}$ boil down to equations for the elements of the density matrix $\hat{\rho}$. These equations are called the Bloch equations. Using the formula:

$$\left[\hat{H}, \hat{\rho}\right] = \begin{pmatrix} \gamma(\rho_{21}-\rho_{12}) & \gamma(\rho_{22}-\rho_{11})+\rho_{12}(\varepsilon_1-\varepsilon_2) \\ \gamma(\rho_{11}-\rho_{22})+\rho_{21}(\varepsilon_2-\varepsilon_1) & \gamma(\rho_{12}-\rho_{21}) \end{pmatrix} \, , \tag{4.239}$$

where $\gamma = |e|D_{12}E$, $\rho_{21} = \rho_{12}^{*}$, it is easy to get a set of equations for the density matrix elements (the Bloch equations):

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\rho_{21} &= \frac{1}{\mathrm{i}\hbar}\left(\left(\varepsilon_2-\varepsilon_1\right)\rho_{21} + |e|\,D_{12}E\left(\rho_{11}-\rho_{22}\right)\right) \, , \\ \frac{\mathrm{d}}{\mathrm{d}t}\rho_{11} &= \frac{1}{\mathrm{i}\hbar}\,|e|\,D_{12}E\left(\rho_{21}-\rho_{21}^{*}\right) \, , \\ \frac{\mathrm{d}}{\mathrm{d}t}\rho_{22} &= -\frac{1}{\mathrm{i}\hbar}\,|e|\,D_{12}E\left(\rho_{21}-\rho_{21}^{*}\right) \, . \end{aligned} \tag{4.240}$$

From the definition of matrix element

$$\rho_{21} = \frac{1}{N}\sum_k N_k c_2^{(k)*} c_1^{(k)} \, ,$$

we see that ρ_{21} is the sum of complex numbers whose phases differ from atom to atom, and dramatically change after each collision of the atoms. From this, we can conclude that in the absence of an external field, the quantity ρ_{21} must vanish after a few collisions of atoms with each other.

In addition, it is intuitively clear that, in the absence of an external field, the diagonal elements of the matrix $\widehat{\rho}$ must relax over time to their equilibrium values $\rho_{11}^{(0)}$ and $\rho_{22}^{(0)}$.

The qualitative speculations can be put into a mathematical form if one formally introduces it into the Bloch equation (4.240); the so called relaxation terms:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\rho_{21} &= \frac{1}{\mathrm{i}\hbar}\left((\varepsilon_2-\varepsilon_1)\rho_{21}+|e|\,D_{12}E(\rho_{11}-\rho_{22})\right)-\frac{\rho_{21}}{T_2}\,,\\
\frac{\mathrm{d}}{\mathrm{d}t}\rho_{11} &= \frac{1}{\mathrm{i}\hbar}|e|\,D_{12}E\left(\rho_{21}-\rho_{21}^{*}\right)-\frac{\left(\rho_{11}-\rho_{11}^{(0)}\right)}{T_1}\,,\\
\frac{\mathrm{d}}{\mathrm{d}t}\rho_{22} &= -\frac{1}{\mathrm{i}\hbar}|e|\,D_{12}E\left(\rho_{21}-\rho_{21}^{*}\right)-\frac{\left(\rho_{22}-\rho_{22}^{(0)}\right)}{T_1}\,.
\end{aligned} \tag{4.241}$$

The time T_2 bears the name of the transverse relaxation time. During this time, the individual atoms of the medium "forget" their phase coherence. In the absence of an external field, the diagonal elements of the density matrix recover their equilibrium values during the period T_1.

Maxwell's equations (4.238) and the *Bloch equations* (4.241) with relaxation terms constitute a closed system suitable for the analysis of many phenomena, including nonlinear ones such as the propagation of electromagnetic waves through a medium with active atoms. In this case, the atoms are described quantum mechanically, and the electromagnetic wave is represented in a classic way (the quasiclassical approximation). The approach set forth above is insufficient only if the quantum properties of electromagnetic waves in a medium are of our concern.

Note that, with this method, it is easy to calculate the dielectric susceptibility of a medium to define a relationship between harmonics $\vec{P}(\omega)$ and $\vec{E}(\omega)$:

$$\vec{P}\,(\omega)=\alpha\,(\omega)\,\vec{E}\,(\omega)\ .$$

The formula means that the complex refractive index of the active medium

$$\sqrt{\varepsilon\,(\omega)}=\eta+\mathrm{i}\kappa=\sqrt{\alpha\,(\omega)+\varepsilon_0}$$

can be completely found. The connection between the imaginary part κ of this index and the microscopic theory was already discussed earlier.

The semiclassical approach not only reproduces all the formulas received earlier, but also automatically calculates the shape of the line. The line shape appears to be Lorentzian and assigned by the T_1 and T_2 parameters. Thus, the relaxation terms postulated in the Bloch equations explain quite a certain physical mechanism responsible for the line shape. Namely, they account for the processes involving collisions between atoms of the medium.

4.16 Self-Induced Transparency and the Concept of Strongly Nonlinear Particle-like Excitations (Solitons)

Between 1965–1970, an interesting effect was predicted theoretically, and then experimentally studied. It was found that under certain conditions, ultrashort pulses of laser light pass through the optical medium, not being absorbed as if through a fully transparent system. This effect was given the name of *self-induced transparency*; it is associated with the formation of the so called *optical solitons*. We will now illustrate this statement.

Suppose that atoms of a medium are in a ground state with the energy ε_1, and short laser pulses with the frequency $\nu = (\varepsilon_2 - \varepsilon_1)/h$ pass through the medium. For simplicity, we assume that each atom has only two energy levels: ε_2 and ε_1 ($\varepsilon_2 > \varepsilon_1$). The front edge of the pulse causes the electrons of the atoms of the medium to move from the level ε_1 to the higher level ε_2. This is the reason for weakening the laser pulse, due to spending some portion of its energy.

When passing through the medium with excited atoms, the back edge of the pulse returns the electrons to the previous levels. In this case, induced electromagnetic radiation with the frequency $\nu = (\varepsilon_2 - \varepsilon_1)/h$ occurs and replenishes the laser pulse energy lost earlier (Figure 4.29).

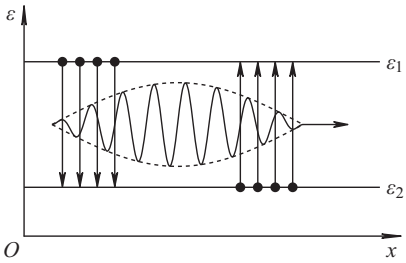


Fig. 4.29: Energy diagram. The process of energy exchange between atoms of a medium and a laser pulse.

The shape and speed of the laser pulse adjust themselves so that the absorption and emission energy processes should arise simultaneously. As a result, the laser pulse takes the form of a particle like envelope soliton. Such solitons transfer the energy bunches of a certain shape through the medium without losses. The phenomenon of self-induced transparency can be described in terms of the semiclassical theory expounded. Let us dwell on this in more detail.

The possibilities of modern technologies allow one to easily design a device that generates laser pulses with a duration of a nanosecond (10^{-9} sec) or even a picosecond (10^{-10} sec). When propagated through a medium as a dilute gas, the nanosecond pulse duration proves to be much shorter than the relaxation times of the atoms of the medium:

$$\tau \ll T_1, T_2 \,, \tag{4.242}$$

In this case, we can neglect the dissipative terms in the Maxwell and Bloch equations. The relaxation processes described by such terms are too slow, and so, have no noticeable influence on the propagation of the laser pulse through the medium or its self-localization.

For further analysis, it is more convenient to use the following notation

$$\rho_3 = \rho_{11} - \rho_{22} \tag{4.243}$$

for the quantity that characterizes the difference in populations between the energy levels of the atoms of the medium. When $\rho_3 > 0$, the number of atoms in the energy levels ε_1 is greater than that in the energy levels ε_2. In particular, when $\rho_3 = 1$, we have a thermodynamically equilibrium medium in a ground state. Inversely filled levels are typical of the medium with $\rho_3 < 0$. When $\rho_3 = -1$, all the medium atoms are in an excited state with the energy $\varepsilon_2 > \varepsilon_1$.

In addition, for convenience, we separate the real and imaginary parts of the complex function:

$$\rho_{21} = \frac{1}{2}\left[\rho_1 + \mathrm{i}\rho_2\right] . \tag{4.244}$$

Given that the dipole moment of the atom $\vec{p} = -|e|\vec{D}_{12}$ lies codirectionally along the vector of the electric field $\vec{E}$, we get the following notation of the *Maxwell–Bloch equations:*

$$\frac{\partial}{\partial t}\rho_3 = -\frac{2\left|\vec{p}\right|E}{\hbar}\rho_2 , \tag{4.245a}$$

$$\frac{\partial}{\partial t}\rho_1 = \omega_{21}\rho_2 , \tag{4.245b}$$

$$\frac{\partial}{\partial t}\rho_2 = -\omega_{21}\rho_1 + \frac{2\left|\vec{p}\right|E}{\hbar}\rho_3 , \tag{4.245c}$$

$$\left[\frac{\partial^2}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]E = \mu_0 n\left|\vec{p}\right|\frac{\partial^2}{\partial t^2}\rho_1 , \tag{4.245d}$$

where $\omega_{12} = (\varepsilon_2 - \varepsilon_1)/\hbar$.

Using the relation (4.245c), we exclude the parameter ρ_2 from the equation (4.245b),

$$\frac{\partial^2}{\partial t^2}\rho_1 = -\omega_{21}^2\left[\rho_1 - \frac{2\left|\vec{p}\right|E}{\hbar\omega_{21}}\rho_3\right] . \tag{4.246}$$

Applying the equation (4.245b), we leave aside the parameter ρ_2 in (4.245a):

$$\frac{\partial}{\partial t}\rho_3 = -\frac{2\left|\vec{p}\right|E}{\hbar\omega_{21}}\frac{\partial}{\partial t}\rho_1 . \tag{4.247}$$

Also, we transform (4.245d) through the calculated derivative (4.246):

$$\left[\frac{\partial^2}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]E = -\mu_0 n\left|\vec{p}\right|\omega_{21}^2\left[\rho_1 - \frac{2\left|\vec{p}\right|E}{\hbar\omega_{21}}\rho_3\right] . \tag{4.248}$$

Next, we proceed to the dimensionless field $\vec{E}$ in the obtained equations (4.246)–(4.248), highlighting its characteristic amplitude A:

$$E = A\widetilde{E}\,, \tag{4.249}$$

and introduce the dimensionless small parameter:

$$\varepsilon = \frac{A\,|\vec{p}|}{\hbar\omega_{21}} \ll 1\,. \tag{4.250}$$

Suppose that the density of a gas through which the laser pulse travels is so small that the estimate

$$\frac{c^2\mu_0 n\,|\vec{p}|}{2A} = \varepsilon\gamma \tag{4.251}$$

holds. Here, $\gamma = O(1)$. In the new notations, the Maxwell–Bloch equations take the form convenient for them to be approximately solved:

$$\begin{aligned}
&\left[\frac{\partial^2}{\partial t^2} + \omega_{21}^2\right]\rho_1 = 2\omega_{21}^2\varepsilon\widetilde{E}\rho_3\,,\\
&\frac{\partial}{\partial t}\rho_3 = -2\varepsilon\widetilde{E}\frac{\partial}{\partial t}\rho_1\,,\\
&\left[\frac{\partial^2}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\widetilde{E} = -\frac{2\varepsilon\gamma}{c^2}\omega_{21}^2\left[\rho_1 - 2\varepsilon\widetilde{E}\rho_3\right]\,.
\end{aligned} \tag{4.252}$$

When $\varepsilon \ll 1$, the approximate solution of the system (4.252) can be arrived at by employing the so called method of multiscale expansions.

Now, our concern is to find the Maxwell equation solutions that resemble a traveling wave with a fast space time dependence of the type:

$$\widetilde{E} \cong E_0\cos\left(\omega_{21}\left(t - \frac{x}{c}\right)\right)\,. \tag{4.253}$$

In writing the above equation, we have taken two facts into account:

1. Only waves with the frequency $\omega_{21} = (\varepsilon_2 - \varepsilon_1)/\hbar$ can intensively interact with two level atoms of the medium.
2. In the absence of medium, expression (4.253) is an exact solution of Maxwell's equation.

Apparently, the presence of the medium only brings relatively slow time space wave amplitude modulations and small corrections to (4.253). For a mathematical description, the slow spatial and temporal variables:

$$X = \frac{\varepsilon\omega_{21}}{c}x\,, \quad T = \varepsilon\omega_{21}\left(t - \frac{x}{c}\right)\,, \tag{4.254}$$

are needed for seeking the solution to be expanded for the field $\widetilde{E}$:

$$\widetilde{E} = E_0(X, T)\cos\left(\omega_{21}\left(t - \frac{x}{c}\right)\right) + \varepsilon E_1(x, t, X, T) + \cdots\,. \tag{4.255}$$

If one carefully analyzes the form of the equations (4.252), we can make sure that such a structure for the expansion of $\vec{E}$ involves only the following expansions for ρ_1 and ρ_2:

$$\rho_1 = P_0(X, T)\sin\left(\omega_{21}\left(t - \frac{x}{c}\right)\right) + \varepsilon\rho_{11}(x, t, X, T) + \cdots , \tag{4.256}$$

$$\rho_3 = \rho_{30}(X, T) + \varepsilon\rho_{31}(x, t, X, T) + \cdots . \tag{4.257}$$

Substituting the expansions (4.255)–(4.257) into the Maxwell–Bloch equations (4.252), and setting equal the coefficients of the different powers of ε to zero, we obtain a chain of equations. The choice of the initial terms of the expansion is justified because of automatically satisfying the first ε^0-order equations of the chain. For the terms $O(\varepsilon)$, we have:

$$\left[\frac{\partial^2}{\partial t^2} + \omega_{21}^2\right]\rho_{11} = 2\omega_{21}^2\cos\theta\left[E_0\rho_{30} - \frac{\partial}{\partial T}P_0\right] , \tag{4.258a}$$

$$\frac{\partial}{\partial t}\rho_{31} = -\omega_{21}\left[\frac{\partial}{\partial T}\rho_{30} + 2E_0P_0\underbrace{\cos^2\theta}_{(1+\cos 2\theta)/2}\right] , \tag{4.258b}$$

$$\left[\frac{\partial^2}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]E_1 = -\frac{2\omega_{21}^2}{c^2}\left[\frac{\partial}{\partial X}E_0 + \gamma P_0\right]\sin\theta . \tag{4.258c}$$

Here, $\theta = \omega_{21}(t - x/c)$.

It is not hard to see that the integration of the equation (4.258b) over the fast variable t (the slow variables are considered constant parameters) yields no growing terms in t provided that the following condition is met:

$$\frac{\partial}{\partial T}\rho_{30} = -E_0P_0 . \tag{4.259}$$

Under the condition (4.259), the equation (4.258b) gives a small correction ρ_{31} not required for further analysis.

It is also easy to check that:

$$\begin{aligned}\left(\frac{\partial^2}{\partial t^2} + \omega_{21}^2\right)\begin{pmatrix}\sin\theta\\ \cos\theta\end{pmatrix} = 0 ,\\ \left(\frac{\partial^2}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\begin{pmatrix}\sin\theta\\ \cos\theta\end{pmatrix} = 0 .\end{aligned} \tag{4.260}$$

Ultimately, this leads to the fact that if one leaves the summands proportional to $\cos\theta$ and $\sin\theta$ in the right-hand sides of equations (4.258a) and (4.258b), their solutions will contain growing terms of the type $x\cos\theta$, $x\sin\theta$, $t\cos\theta$, $t\sin\theta$ as $x \to \infty$ and $t \to \infty$. For the calculations to be self-consistent, a requirement needs to met; that the terms proportional to $\cos\theta$ and $\sin\theta$ should be omitted in the right-hand sides of equations (4.258a) and (4.258b). Then, we obtain the following equations to calculate the

functions of the slow variables:

$$\frac{\partial}{\partial T}P_0 = E_0\rho_{30}\,, \quad \frac{\partial}{\partial X}E_0 = -\gamma P_0\,. \tag{4.261}$$

Under the restrictions of (4.261), the fast variable dependent corrections ρ_{11} and E_1 can be easily calculated. They are not worth writing out because the equations obtained for the functions ρ_{30}, P_0, and E_0 already form a closed system, with it being a simplified model that describes the main effect of the mutual influence of the electromagnetic field and the atoms of the medium:

$$\begin{aligned}
&\frac{\partial}{\partial T}P_0 = E_0\rho_{30}\,,\\
&\frac{\partial}{\partial T}\rho_{30} = -E_0P_0\,,\\
&\frac{\partial}{\partial X}E_0 = -\gamma P_0\,.
\end{aligned} \tag{4.262}$$

It should be emphasized that the perturbation theory cannot always be reduced to the simplified model (4.262). This happens only when making the right choice of the slow variables that reflect the characteristic spatial and temporal processes in a system.

From the first two equations of (4.262) we find:

$$\frac{\partial}{\partial T}\left(P_0^2 + \rho_{30}^2\right) = 0\,.$$

Without loss of generality, we may put:

$$P_0^2 + \rho_{30}^2 = 1\,.$$

This allows one to parameterize P_0 and ρ_{30}:

$$P_0 = \pm\sin\Phi\,, \quad \rho_{30} = \pm\cos\Phi\,. \tag{4.263}$$

Such alternation of signs is compatible with the first two equations (4.262), which also imply that:

$$E_0 = \frac{\partial}{\partial T}\Phi\,. \tag{4.264}$$

If one plugs the results of (4.263) and (4.264) into the last equation of (4.262), we derive a differential equation for the calculation of Φ:

$$\frac{\partial}{\partial X}\frac{\partial}{\partial T}\Phi = \mp\gamma\sin\Phi\,. \tag{4.265}$$

Analysis of the solutions of equation (4.265) with different signs shows that the upper sign describes the propagation of electromagnetic waves through an equilibrium medium. The lower sign corresponds to the propagation of electromagnetic waves

through a medium with inverted population. A medium with inverted population gives unstable solutions of Φ.

Next, we limit ourselves to the discussion of solutions of the equation:

$$\frac{\partial}{\partial X}\frac{\partial}{\partial T}\Phi = -\gamma \sin \Phi \,, \tag{4.266}$$

that describes the effect of self-induced transparency of an initially equilibrium medium. The solutions of (4.266) define the electric field $E_0 = \partial\Phi/\partial T$ in a medium and the quantity $\rho_3 = \cos\Phi$ that characterizes the degree of inversion of the medium when passing an electromagnetic pulse through it. We are interested in a strongly disturbed, initially equilibrium medium with $\rho_3 = 1$ before passing the wave, and hence $\Phi = 0$ or $\Phi = 2\pi n$, where n is an integer.

Earlier, we looked at cases of weakly excited states of the medium when the governing equations could be linearized near its stable equilibrium position. In the given problem, it is no use to linearize (4.265) since this leads to the equation:

$$\frac{\partial}{\partial X}\frac{\partial}{\partial T}\Phi = -\gamma\Phi \,. \tag{4.267}$$

As the traditional perturbation theory claims, when expanding $\sin\Phi = \Phi - \Phi^3/3! + \cdots$, no finite order in powers of Φ for equation (4.266) gives particle like energy bunches (solitons). The power expansion is not suitable because it violates the periodicity property of the function $\sin\Phi$. We will discuss this issue in more detail.

This book deals with different condensed media. In specific examples, we have ascertained that the real media can always be described by nonlinear differential equations in partial derivatives. Generally speaking, many condensed media have a highly nontrivial, strongly nonlinear ground state (for example, a domain structure). However, we have limited ourselves to the discussion of relatively simple media with a spatially homogeneous stable ground state. Under low external impacts on such a medium, its behavior near the homogeneous ground state can always be described by linearizing original nonlinear equations. After the linearization, the medium's physical properties are theoretically described by linear differential equations in partial derivatives. Their complete set of solutions, as a rule, is represented in the form of low amplitude waves. In media without energy losses, the small amplitude waves are juxtaposed with quasiparticles. For example, thermal vibrations of a crystal are governed by phonons.

It is important that linear differential equations satisfy the principle of superposition. The superposition principle facilitates the description of linear wave processes extremely well. An arbitrarily complicated wave field can be represented as the sum of the simplest normal modes using the Fourier transform as a powerful mathematical tool. An arbitrary wave packet, when moving, changes its shape. By finding a change in characteristics of individual harmonics that form the packet, any change in its shape can be easily calculated.

Waves to be arisen in a medium subjected to large external influences are described by nonlinear differential equations for which the superposition principle does not hold. A distinctive feature of nonlinear waves is processes of their interaction. The nonlinear waves can transform beyond recognition when interacting.

For example, when two nonlinear waves meet, they may disappear, giving rise to the third wave whose frequency is equal to the sum or difference of the initial wave frequencies. After the meeting of two waves with close frequencies, one of them may be forced to change its frequency (the phenomenon of self-synchronization of waves).

Here, it would be appropriate to quote I.R. Shen, a well known expert in nonlinear optics: "Physics would be dull and life most unfulfilling if all physical phenomena around us were linear. Fortunately, we are living in a nonlinear world. While linearization beautifies physics, nonlinearity provides excitement in physics." Against linear waves, the parameters of even the simplest periodic nonlinear waves (their frequency and velocity) are wave amplitude dependent. This leads, in particular, to compression of nonlinear waves. It is well known that nonlinear periodic waves on water become steeper as their amplitude grows; their crests are folded back and overturned (Figure 4.30).

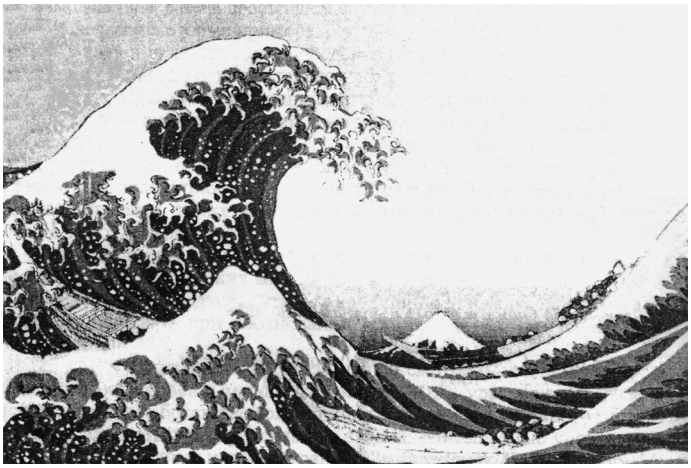


Fig. 4.30: A nonlinear world.

It has been found that, under strong external impacts in condensed matter where there is dispersion but small friction losses, a complete set of nonlinear normal modes includes (apart from waves arising under weak influences) particle like waves also (*solitons*). Moreover, it is solitons that govern the main physical properties of condensed matter under strong external influences.

At first glance, a soliton looks like a wave packet of linear theory, but it is not the case. To clarify the mechanism of soliton formation, we recall that waves with dif-

ferent wave numbers propagate in dispersive media with different phase velocities. Therefore, as linear theory states, a wave packet involving different harmonics always splits into its wave components over a certain period of time. It should be emphasized that the wave packet energy, of course, is retained despite blurring of the packet itself. The amplitude of the wave packet gradually diminishes and becomes hardly noticeable. It is no wonder that the translation of the word *dispersion* from Latin (dispertio) means scatter or breaking apart. However, even weak *nonlinearity of a medium* causes interactions between harmonic waves the wave packet is made up from. They tend to compress the wave packet or to increase the slope of the resulting wave front. The multiscale expansion method uses slow variables to treat these interactions as a nonlinear transfer of energy from one harmonic to the others, which linear theories never produce. The balancing of two competing effects occurs. The former is compression of the wave packet due to the nonlinear harmonic interactions, and the latter is blurring of the packet due to the dispersion. As a result, special particle like waves, solitons, are formed.

As particles, the solitons are localized in space and retain their shape and velocity, not only when moving, but even after colliding with solitary waves similar to themselves. The solitons may be regarded as stable formations. Upon colliding, the solitons scatter away like billiard balls, restoring their initial (before the collision) shape and velocity.

It is interesting and important to note that an arbitrary wave pulse of sufficiently large amplitude splits into a number of stable solitons over time. In this respect, the origin of the word *soliton* is the same as of a solo performer: solitary states. That is why they define all the basic physical properties of condensed matter under extreme external influences. At the same time, initially small amplitude wave pulses, which ignore nonlinearity of the medium, form no solitons, but are blurred due to dispersion.

General Conclusions

1. In the absence of nonlinearity of a condensed matter, dispersion prevents solitons from emerging due to the blurring of waves.
2. If there is no dispersion but nonlinearity, the possibility of forming solitons is also excluded because of the continuous transfer of energy from one harmonic to the others. In most cases, this feature appears as a steepening of the resulting wave front or compression of wave pulses.
3. Soliton formation is given by the basic properties of all condensed matters, where energy losses due to friction are little and the balance between dispersion and nonlinearity is possible.
4. For this reason, many nonlinear systems of different physical nature have proved to be possible to theoretically describe within universal theoretical models. To integrate these nonlinear models, scholars have succeeded in gaining powerful

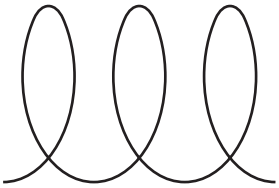


Fig. 4.31: An example of topologically stable solitons in a wire.

analytical methods of generalizing the Fourier transforms for linear systems. Of course, these methods are not applicable to all nonlinear partial differential equations. Simultaneously, the Fourier method also does not allow all linear differential equations to be integrated. New mathematics rests on support of different versions of the so called inverse scattering method. With the latter, it has first become possible to theoretically detail many highly nonlinear states of a medium. The properties of such states are unusual and cannot be explained within linear theory. For systems close to ones that are integrable by the inverse scattering problem, special versions of the soliton perturbation theory have been developed. The results secured also cannot be reproduced in any finite order of the traditional perturbation theory.

5. Analytical mechanics claim that, according to the Liouville–Arnold theorem, the equations of motion of a dynamic system with N degrees of freedom can be integrated in quadratures only if the system has N so called first integrals of motion. From this perspective, nonlinear partial differential equations, such as equation (4.266), have an infinite number of degrees of freedom. It has turned out that the integrability of some of them by the inverse scattering problem is due to an infinite number of conservation laws. Amazing dynamic stability of each soliton is guaranteed by an infinite series of conservation laws.
6. It is remarkable that, among the variety of solitons, there are ones whose dynamic stability is reinforced by topological reasons. let us explain this statement with an example of wire or fishing line loops (Figure 4.31).

If an endless wire lies in a plane, the loop soliton cannot be created or destroyed. This gives us the right to call such a soliton topological. Recall that topology studies properties of figures that remain unchanged when the figures are continuously deformed. When lying in a plane, the endless wire loop cannot be eliminated by continuous deformations. The wire loops are an example of topologically stable solitons; the most nonlinear and the most stable ones, which have already been discussed. In the general, the topological solitons can be described by fields with geometric peculiarities, which cannot be eliminated without destroying the ordered state in a large volume of matter. It is not hard to verify that wire loops not only can move but also rest. The example mentioned above illustrates the possibility of existing static solitons or soliton like topological defects.

Today, scientists are studying solitons in crystals, superconductors, in the atmosphere of the Earth, and other planets [17, 18]. Soliton representations are used to delineate patterns of dents on the surface of loaded plates and shells [19]. Apparently, solitons have played, and continue to play an important role in the evolution of the Universe. Modern theories of elementary particles try to interpret the elementary particles as solitons. Here, it would be appropriate to quote Louis de Broglie, one of the founders of quantum mechanics: "In my present views, I came to the thought that the wave-particle duality requires developing wave mechanics based on nonlinear equations, in relation to which linear equations were but only approximate forms holding true under certain circumstances."

It is interesting that magnetic materials are rich in the variety of their internal structure and in properties of magnetic solitons. At first, no one has even noticed that the motion of the well-known domain walls in ferromagnetic materials is similar to particle motion. Today we know that the domain wall is a topologically stable soliton. Investigations of nonlinear properties of magnetic materials have experienced significant progress owing to the soliton theory involved [20–22]. Curiously, even the peculiarities of the internal structure of the domain walls and domains can currently be more adequately outlined in terms of soliton like topological defects, magnetic vortices, and spirals. The vortex and spiral structures are the closest relatives of solitons [21].

A complete interpretation of experimental data under extreme external impact on a medium is impossible to run without carrying out an analysis of soliton states.

Let us return to the equation:

$$\frac{\partial}{\partial X}\frac{\partial}{\partial T}\Phi = -\gamma \sin \Phi \,,$$

which describes the self-induced transparency effect. This equation has been investigated by the inverse scattering method in detail. Its soliton solutions are found explicitly. The simplest soliton solutions, of course, can be obtained by traditional methods of integration. However, explicit analytical formulas, which delineate the motion of a system of N solitons, can be derived only by methods related to the inverse scattering method. To verify this, we present the analytical formula that characterizes pair collisions of solitons, including their motion between collisions and recovery of their shape after colliding:

$$\Phi = -4 \arg\left[\det(I + \upsilon)\right] \,, \tag{4.268}$$

$$\upsilon_{nm}(X, T) = \frac{c_m}{\lambda_n + \lambda_m} \exp \sqrt{\gamma}\left[2\mathrm{i}\lambda_m X + \frac{\mathrm{i}}{2\lambda_m} T\right] \,. \tag{4.269}$$

There are two types of solitons:

1. kinks and antikinks (bends)
2. breathers (pulsating solitons)

Each soliton of the first type (a bend) is described by a real number c_m, and a purely imaginary number $\lambda_m = i\kappa_m$; $\kappa_m > 0$.

Each soliton of the second type (a breather) is parameterized by a pair of numbers λ_s and $-\lambda_s^*$, symmetrically arranged with respect to the imaginary axis and a pair of complex conjugate numbers c_s and c_s^*, which corresponds to the former. Here, $\mathrm{Im}\lambda_s > 0$.

Thus, a system that comprises K kinks and antikinks and L breathers ($N = K + L$) is characterized by a matrix v with $(K + 2L) \times (K + 2L)$ dimension.

Let us look at the simplest solution. Suppose there is only one bend. In this case v is the number equal to:

$$v = \frac{c}{2i\kappa} \exp \varphi \, ; \tag{4.270}$$

$$\Phi = -4 \arg(1 + v) = 4\sigma \operatorname{arctg} \exp(\varphi + \delta) \, , \tag{4.271}$$

where $\varphi = \sqrt{\gamma}(-2\kappa X + T/2\kappa)$, $\sigma = \operatorname{sign} c$, and $\delta = \ln|c/2\kappa|$. The parameter δ may be eliminated by choosing the origin of the coordinate system, so further, we put that $\delta = 0$. The result (4.271) holds true, and this is easily checked.

The shape of the curves of the functions $\Phi(T)$ for $X = \mathrm{const}$ (Figure 4.32) substantiates the name of the solutions (the bends): there is a kink when $\sigma = 1$ and an antikink when $\sigma = -1$.

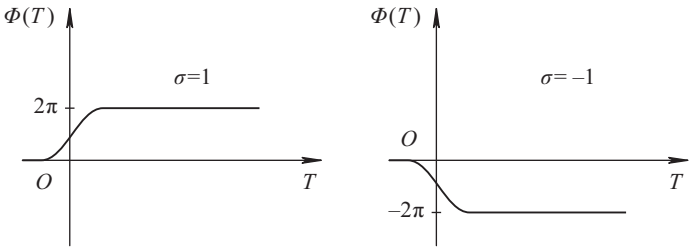


Fig. 4.32: Dependence of $\Phi(T)$ for $X = \mathrm{const}$.

The bends of the function $\Phi(T)$ correspond to the electric field bunches localized in the space:

$$\tilde{E} = \frac{\partial}{\partial T}\Phi = \sigma \frac{\sqrt{\gamma}}{\kappa} \frac{1}{\operatorname{ch} \varphi} \, . \tag{4.272}$$

Going over from the slow variables X and T to the initial variables x and t, we find:

$$\Phi = 4\sigma \operatorname{arctg} \exp\left(\frac{\varepsilon\omega_{21}\alpha}{2}\left(t - \frac{x}{V}\right)\right) \, , \tag{4.273}$$

$$\tilde{E} = \frac{\sigma\alpha}{\operatorname{ch}\left[\frac{\varepsilon\omega_{21}\alpha}{2}\left(t - \frac{x}{V}\right)\right]} \, . \tag{4.274}$$

Here $\alpha = \sqrt{\gamma}/\kappa$ is a parameter that characterizes the amplitude of the soliton. The quantity $2/(\varepsilon\omega_{21}\alpha)$ has the meaning of the duration of the soliton pulse. The velocity of a moving soliton as a whole is less than the speed of light:

$$V = \frac{c}{1 + 4\kappa^2} < c \tag{4.275}$$

Measured in different experiments, the velocities of picosecond and nanosecond pulses turned out to be two to four orders of magnitude smaller than the speed of light in a vacuum.

It can be seen from the formulas given that the soliton increases its amplitude and velocity as the duration shortens. This rule is common for many solitons. To keep it firmly in mind, the following comedic phrase is helpful for remembering the rule: "A tall and thin man moves faster than a short and fat one."

The solution of Φ as $x \to \pm\infty$ tends to 0 and $2\pi\sigma$, which corresponds to the ground state of the medium in the regions where there are no electromagnetic pulses: $\rho_3 = \cos\Phi \to 1$ as $|x| \to \infty$.

Another elementary formation is a breather (a pulsating soliton). The general formula defines it by two parameters λ associated with reduction:

$$\lambda_1 = r + \mathrm{i}\kappa\,, \quad \lambda_2 = -r + \mathrm{i}\kappa\,, \tag{4.276}$$

where $\kappa > 0$, and by two parameters c:

$$c_1 = a\,, \quad c_2 = a^*\,, \tag{4.277}$$

where a is an arbitrary complex number. The breather has the form:

$$\Phi = -4\mathrm{arctg}\left(\frac{\kappa}{r}\cdot\frac{\sin\left[r\sqrt{\gamma}\left(2X + T/\left(r^2+\kappa^2\right)\right) + \varphi_0\right]}{\mathrm{ch}\left[\kappa\sqrt{\gamma}\left\{2(X - X_0) - T/\left(r^2+\kappa^2\right)\right\}\right]}\right)\,, \tag{4.278}$$

where

$$\varphi_0 = \arg\left(\frac{a}{\kappa}\right)\,, \quad X_0 = \frac{1}{2\kappa\sqrt{\gamma}}\ln\left|\frac{ra}{2\kappa(r + \mathrm{i}\kappa)}\right|\,.$$

The solution (4.278) describes a localized bunch of the electric field $\widetilde{E} = \partial\Phi/\partial T$; the bunch moves as a whole and pulsates. Hence, its name comes from a breather (a "breathing" soliton).

Naturally, such solitons may be used for transmitting information over optical fibers. The advantages are obvious:

- high rate of transmission of vast amount of information because of the short pulse duration
- low power consumption
- high reliability

In 1991, impressive results were achieved in this area. Nakazawa, jointly with his employees, brought about the transmission of information over optical fiber at a speed of 10 Gbit per second; a distance of 10^6 km.

In general, by maintaining solitons, their form and velocity and stability and spatial localization are very valuable properties for any digital data transmission. It is one of the most promising directions of applied science. The future generation of computers using superconductors is expected to employ logic and memory devices, in which a role of a soliton plays a fluxon; a magnetic flux quantum. It can be shown that the fluxon motion in narrow superconducting (Josephson) contacts can also be theoretically described by equation (4.266).

5 Dislocations and Martensitic Transitions

5.1 Ordered Macroscopic States of a Crystal and a Nonlinear Theory of Elasticity

A theoretical description of condensed matter often requires resorting to both microscopic degrees of freedom and macroscopic objects. The former are discussed from the standpoint of quantum theory, and the latter are introduced phenomenologically, in the language of classical physics. Among them, for example, are vortex filaments in superconductors and superfluid liquids, *dislocations*, *grain boundaries* in crystals, and domain walls in magnetic materials. The consistent quantum theory holds that their formation is the result of spontaneous symmetry breaking in a system, and of condensation of a huge number of Bose quasiparticles corresponding to the collective excitations of the system. The appearance of such excitations emanates from the Goldstone theorem. According to this theorem, an ordered macroscopic state that spontaneously breaks symmetry of a microscopic Hamiltonian of a system is maintained by long range correlations created self-consistently by the system itself. Quanta corresponding to the correlations obey the Bose–Einstein statistics. Their energy tends to zero as the momentum tends to zero. In this case, a large number of Goldstone bosons are possible to condensate into extended macroscopic objects with different symmetry. In the system, supracondensate bosons tend to partially restore the symmetry spontaneously broken by the ordered macrostates formed.

This chapter deals with the physical properties of macroscopic objects with classical behavior, dislocations, and *interphase boundaries* in crystals. A spontaneous violation of the total translational invariance of the microscopic Hamiltonian of the systems causes the formation of the above objects, followed by the condensation of a large number of acoustic phonons.

The fields of dislocations and interphase boundaries have topological features. Each dislocation meets a line, and each interphase boundary corresponds to a two-dimensional surface of topological singularities. It can be argued that all macroscopic defects of condensed matter, with topological singularities, are always formed as a result of the condensation of gapless quasiparticles (Goldstone bosons) [23]. Macroscopic objects with classical behavior, such as vortices in superconductors, dislocations, and interphase boundaries in crystals, are formed in quantum systems. Therefore, they interact with other quasiparticles of the matter, changing their quantum numbers. Their interaction can be described in terms of the effective potential acting on the quasiparticles. For example, electrons in the core region of the vortex filament of a superconductor are affected by the potential of the filament. As a consequence, the electron energy decreases, which explains the nature of the so called solid core of the vortex. Near the dislocation or the interphase boundary, the potential of these

https://doi.org/10.1515/9783110586183-005

macro-objects acts on phonons of the crystal. As a result, special phonon states localized along the dislocation line or near the interphase boundary appear.

Dislocations and interphase boundaries occupy regions which are extended enough; their dimensions are much greater than the interatomic distance. Therefore, the continuum theory of elasticity ensures their fruitful semiclassical discussion. Such an approach delineates well the peculiarities of a crystal subjected to deformation to form moving dislocations, as well as nonlinear effects near martensitic transitions.

There are two main ways of theoretically describing changes in the positions of material particles of a body during its deformation. As a "mark" of the material point of the body, the Lagrangian approach uses Cartesian coordinates of the position of this point in the undeformed state of the body. Then we watch the evolution of the marked material points of the body. In contrast, the Eulerian approach traces goings on in a fixed area of space as time goes by. Such a description is ideally suitable for studying liquids, where a rapidly deforming mass often comes from nowhere and leaves for nowhere. So it would be preferable to consider what is happening here and now, before our eyes. However, being convenient kinematically, the Eulerian description becomes clumsy when it comes to the laws of dynamics and leads to difficulties. The point is that the dynamic equations include quantities related to material particles of a body, rather than to spatial regions occupied temporarily by these particles. Some relations, being obvious under the Lagrangian approach, require complicated reasoning when using Eulerian coordinates [24]. We will adhere to the Lagrangian description when discussing dislocations and martensitic transitions in crystals.

Let us use an undeformed body as a frame of reference: x_k is the Cartesian coordinates of a material point of the body before deformation, $X_k = x_k + u_k(\vec{x}, t)$ are the coordinates of the same point after deformation, and $\vec{u}(\vec{x}, t)$ is the displacement vector of medium. In the undeformed state of the body, the vector that connects two close particles of the medium can be written in the form:

$$\mathrm{d}\vec{x} = (\mathrm{d}x_1, \mathrm{d}x_2, \mathrm{d}x_3) \ .$$

Therefore, the square of the distance between the particles is given by:

$$(\mathrm{d}\vec{x})^2 = \mathrm{d}x_k \mathrm{d}x_k \ .$$

Furthermore, unless otherwise stated, the double indices imply summation.

After deforming the body, the vector connecting these close particles becomes different:

$$\mathrm{d}\vec{X} = (\mathrm{d}X_1, \mathrm{d}X_2, \mathrm{d}X_3) \ .$$

The distance between the particles changes as follows:

$$\left(\mathrm{d}\vec{X}\right)^2 = \frac{\partial X_k}{\partial x_s}\frac{\partial X_k}{\partial x_p}\mathrm{d}x_s \mathrm{d}x_p \ .$$

The quantity

$$g_{sp} = \frac{\partial X_k}{\partial x_s}\frac{\partial X_k}{\partial x_p}$$

characterizes the metric properties of the deformed medium; it is invariant with respect to rotation of the body as a whole: $X_k \to D_{ks}X_s$. Here, D is a matrix of the rotation group $SO(3)$ that does not depend on the coordinates $\vec{x}$:

$$D_{is}D_{ik} = \delta_{sk}\,, \quad \det D = 1\,,$$

where δ_{sk} is the Kronecker symbol.

Since, upon deformation of the body, the change in distance between its close particles can be written in the form:

$$\left(\mathrm{d}\vec{X}\right)^2 - \left(\mathrm{d}\vec{x}\right)^2 = \left[g_{is} - \delta_{is}\right]\mathrm{d}x_i\mathrm{d}x_s \equiv 2\eta_{is}\mathrm{d}x_i\mathrm{d}x_s\,, \tag{5.1}$$

the Lagrangian strain tensor

$$\eta_{ij} = \frac{1}{2}\left[g_{ij} - \delta_{ij}\right]$$

is usually chosen as a measure of the nonlinearly elastic deformation of medium. In terms of the displacement field $\vec{u}(\vec{x}, t)$, it appears as:

$$\eta_{ij} = \frac{1}{2}\left[\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} + \frac{\partial u_k}{\partial x_i}\frac{\partial u_k}{\partial x_j}\right]\,. \tag{5.2}$$

Note that the tensor η_{ij} is symmetric with respect to the indices i, j. Let us explain the relationship between the tensors and the quantities observed.

If the deformations are sufficiently small, we may make the replacement:

$$\left(\mathrm{d}\vec{X}\right)^2 - \left(\mathrm{d}\vec{x}\right)^2 \approx 2\left|\mathrm{d}\vec{x}\right|\left(\left|\mathrm{d}\vec{X}\right| - \left|\mathrm{d}\vec{x}\right|\right)\,.$$

Then, from (5.1), we derive an approximate formula for the relative elongation of a sample in the vicinity of a point with a radius vector $\vec{x}$:

$$\frac{\left|\mathrm{d}\vec{X}\right| - \left|\mathrm{d}\vec{x}\right|}{\left|\mathrm{d}\vec{x}\right|} \approx \frac{\eta_{is}\mathrm{d}x_i\mathrm{d}x_s}{\left(\mathrm{d}\vec{x}\right)^2}\,.$$

Next, we trace the change in the volume element occupied by the same material particles of the medium. Suppose the particles are in a small volume $\mathrm{d}V_{\mathbf{x}} = \mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3$ before deformation of the body. When deformed, the body changes its initial volume:

$$\mathrm{d}V_{\mathbf{X}} = \mathrm{d}X_1\mathrm{d}X_2\mathrm{d}X_3 = \det J\mathrm{d}V_{\mathbf{x}}\,.$$

However, it always contains an unchanged mass of material particles. Here, J is the matrix with the elements:

$$J_{ij} = \frac{\partial X_i}{\partial x_j} = \delta_{ij} + \frac{\partial u_i}{\partial x_j}\,.$$

We denote the initial and final densities of the medium by $\rho_{\mathbf{x}}$ and $\rho_{\mathbf{X}}$. Then, by the law of conservation of mass, we have:

$$\rho_{\mathbf{X}}\mathrm{d}V_{\mathbf{X}} = \rho_{\mathbf{x}}\mathrm{d}V_{\mathbf{x}}\,.$$

Hence, a useful relation determines the change in the density of the medium during deformation of the body:

$$\frac{\rho_{\mathbf{x}}}{\rho_{\mathbf{X}}} = \frac{\mathrm{d}V_{\mathbf{X}}}{\mathrm{d}V_{\mathbf{x}}} = \det J \,. \tag{5.3}$$

When deformations are small, i.e., ($|\partial u_i/\partial x_j| \ll 1$), from (5.3), we obtain simple expressions in terms of the displacement field for relative changes in density and small volume near the point with the radius vector $\vec{X}$:

$$\frac{\mathrm{d}\rho_{\mathbf{X}} - \mathrm{d}\rho_{\mathbf{x}}}{\mathrm{d}\rho_{\mathbf{x}}} \approx -\frac{\partial u_i}{\partial x_i}\,; \quad \frac{\mathrm{d}V_{\mathbf{X}} - \mathrm{d}V_{\mathbf{x}}}{\mathrm{d}V_{\mathbf{x}}} \approx \frac{\partial u_i}{\partial x_i}\,. \tag{5.4}$$

In the nonlinear theory of finite deformations [25], the density of the elastic energy of a sample can be expanded over the strain tensor components η_{ij}. The expansion coefficients are determined by the requirements of energy invariance under transformations of the symmetry group of an undeformed matter. The latter concerns only the Lagrangian coordinates x_i of the material particles of the body. For an isotropic medium, there are only three independent invariants of the tensor η_{ij}, through which the remaining invariants can be expressed [25, 26]:

$$I_1 = \eta_{mm}\,, \quad I_2 = \eta_{mp}\eta_{pm}\,, \quad I_3 = \eta_{mp}\eta_{ps}\eta_{sm}\,.$$

With these remarks in mind, the expression for the elastic energy of an isotropic nonlinear body can be represented in the form:

$$W = \int\limits_{V_0} \varphi \mathrm{d}^3\vec{x}\,; \quad \varphi = \frac{\lambda}{2}I_1^2 + \mu I_2 + \frac{A}{3}I_3 + BI_1I_2 + \frac{C}{3}I_1^3 + \cdots\,. \tag{5.5}$$

Here, φ is the energy per unit volume of the body before deformation, λ, μ, A, B, C ... are the elastic moduli of the material, and the integration is performed over the volume V_0 of the undeformed body. The coefficients λ and μ of the terms, quadratic in η_{ij} in (5.5), are called linear moduli of elasticity. The coefficients of higher powers of η_{ij} are referred to as nonlinear moduli. This nonlinearity is usually named "physical" because it is associated with the nonlinearity of the forces of intermolecular interaction in a medium and is different for different materials. The source of the second type of nonlinearity proves to be the very definition of the strain tensor (5.2) that contains products of derivatives of the displacement field. This relationship, which is independent of the physical properties of the deformed body, is called "geometric nonlinearity."

Nonlinear elastic modules are unknown for most materials, and their known values are not reliable. Hereinafter, for definiteness, the elastic moduli are thought to be comparable in order of magnitude. Such an approximation is valid for many materials [27]. For further analysis, it is important that the terms in (5.5) that contain higher order invariants contribute less to the deformation of the medium since, usually, $\max|\eta_{ik}| < 1$.

Note that the physical nonlinearity of media with a microstructure can be easily taken into account in the framework of this approach. The influence of the microstructure can be simulated by additional gradient terms in the expression for the elastic energy density of the body. We will return to this issue later in the discussion of martensitic transitions.

When the deformation processes are adiabatic, the dynamic equations for a nonlinearly elastic body can be obtained from the Hamiltonian principle:

$$\delta S_0 + \int_{t_0}^{t_1} \delta A \mathrm{d}t = 0, \quad S_0 = \int_{t_0}^{t_1} [K - U]\, \mathrm{d}t\,. \tag{5.6}$$

The kinetic energy K of the system has the form:

$$K = \int_{V_0} \frac{\rho_0}{2} \left(\frac{\partial u_i}{\partial t} \right)^2 \mathrm{d}^3\vec{x}\,,$$

where ρ_0 is the density of a material in the undeformed state (henceforth, we will assume that $\rho_0 = \mathrm{const}$), and the integration is performed over the volume V_0 of the undeformed body.

The potential energy U consists of the elastic energy W of the body, and the energy W_1 of its interaction with external mass forces:

$$W_1 = -\int_{V_0} \rho_0 P_i u_i \mathrm{d}^3\vec{x}\,.$$

Here, P_i is an external mass force, and $\rho_0 P_i$ is a force acting per unit volume of the body before deformation.

Let us represent the force acting on the oriented surface element $(\mathrm{d}S_1, \mathrm{d}S_2, \mathrm{d}S_3)$ of the deformed body in the form $(T^{\mathrm{ext}}_{1j}, T^{\mathrm{ext}}_{2j}, T^{\mathrm{ext}}_{3j})\mathrm{d}S_j$. The elements T^{ext}_{ij} form an external stress tensor. In this notation, the work of external surface forces is written as:

$$\delta A = \int_S \delta u_i T^{\mathrm{ext}}_{ij} \mathrm{d}S_j\,. \tag{5.7}$$

It is worth emphasizing that the external forces are applied to the surface S of the deformed body. To make use of the Hamilton principle (5.6), we should replace the integration over S in the relation (5.7) by integration over the surface σ of the undeformed body. The transformation needed:

$$\int_S \delta u_i T^{\mathrm{ext}}_{ij} \mathrm{d}S_j = \int_\sigma \delta u_i T^{\mathrm{ext}}_{ij} \frac{\partial \det J}{\partial J_{js}} \mathrm{d}\sigma_s\,,$$

is achieved by means of the identity:

$$\mathrm{d}S_j = \frac{1}{2}\varepsilon_{jkp}\mathrm{d}X_k \wedge \mathrm{d}X_p = \frac{1}{2}\overbrace{\left[\varepsilon_{j'kp}\frac{\partial X_{j'}}{\partial x_s}\frac{\partial X_k}{\partial x_m}\frac{\partial X_p}{\partial x_n}\right]}^{\varepsilon_{smn}\det J}\frac{\partial x_s}{\partial X_j}\mathrm{d}x_m \wedge \mathrm{d}x_n =$$
$$= \det J\frac{\partial x_s}{\partial X_j}\frac{1}{2}\varepsilon_{smn}\mathrm{d}x_m \wedge \mathrm{d}x_n = \det J\frac{\partial x_s}{\partial X_j}\mathrm{d}\sigma_s = \frac{\partial \det J}{\partial J_{js}}\mathrm{d}\sigma_s \,,$$

where $\mathrm{d}\sigma_s = \mathrm{d}x_m \wedge \mathrm{d}x_n$, $\wedge$ is the symbol of the external product, and ε_{pmn} is an absolutely antisymmetric unit tensor ($\varepsilon_{123} = 1$). Here, we have used the relations:

$$\frac{\partial X_{j'}}{\partial x_s}\frac{\partial x_s}{\partial X_j} = \delta_{j'j}\,; \quad \varepsilon_{smn}\det J = \varepsilon_{j'kp}\frac{\partial X_{j'}}{\partial x_s}\frac{\partial X_k}{\partial x_m}\frac{\partial X_p}{\partial x_n}\,,$$

and the representation of the inverse matrix:

$$\left(J^{-1}\right)_{sj} = \frac{\partial x_s}{\partial X_j} = \frac{\partial \ln\det J}{\partial J_{js}}\,.$$

The Hamilton principle (5.6) produces the necessary equations of dynamics and boundary conditions. The equations of dynamics have the form:

$$-\rho_0\frac{\partial^2 u_i}{\partial t^2} + \frac{\partial P_{is}}{\partial x_s} + \rho_0 P_i = 0\,; \quad P_{is} = \frac{\partial W}{\partial\left(\partial u_i/\partial x_s\right)}\,. \tag{5.8}$$

In general, the *Piola–Kirchhoff tensor* P_{is} is not symmetric with respect to the indices i, s.

When varying the action S_0 over the fields u_i, integrals over the surface σ arise to lead to two types of boundary conditions. A part σ' of the body surface, where external forces are given, reveals the following nonlinear boundary conditions:

$$P_{is}n_s|_{\sigma'} = T_{ij}^{\mathrm{ext}}\det J\frac{\partial x_s}{\partial X_j}n_s\Big|_{\sigma'} = T_{ij}^{\mathrm{ext}}\frac{\partial \det J}{\partial J_{js}}n_s\Big|_{\sigma'}\,.$$

Here, $\vec{n}$ is the vector of the unit normal to the surface σ'. In another part σ'' of the body surface where the displacement vector is given, when varying the action, it must be supposed that $\delta u_s = 0$. In addition, $\delta u_s = 0$ at the ends of the time interval $[t_0, t_1]$.

In the deformed body, the i-th component of the momentum flux density goes through a unit area orthogonal to the j-th one of the Cartesian axis. Both components define the element T_{ij} of the stress tensor. The expression $T_{ij}\mathrm{d}S_j$ involves the i-th component of the force acting on the oriented area $(\mathrm{d}S_1, \mathrm{d}S_2, \mathrm{d}S_3)$ of the deformed body from the directions of the surrounding parts of the medium. The law of conservation of the angular momentum guarantees the symmetry property of the tensor with respect to the indices i, j. The stress tensor is related to the Piola–Kirchhoff tensor and the elastic energy density of the body by a chain of equalities [25]:

$$T_{ij} = \frac{1}{\det J}P_{ik}\frac{\partial X_j}{\partial x_k} = \frac{1}{\det J}P_{jk}\frac{\partial X_i}{\partial x_k} = \frac{1}{\det J}\frac{\partial X_j}{\partial x_m}\frac{\partial \varphi}{\partial \eta_{mn}}\frac{\partial X_i}{\partial x_n}\,. \tag{5.9}$$

According to (5.9), the terms I_1 in the energy density φ would lead to the appearance of constant stresses even in an undeformed body. That is why there are no such terms in the expansion (5.5).

When the body is slightly deformed, i.e., ($|\partial u_i/\partial x_j| \ll 1$), the geometric nonlinearity of the strain tensor can be neglected:

$$\eta_{ij} \approx \varepsilon_{ij} = \frac{1}{2}\left[\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right] . \tag{5.10}$$

In this case, we leave only the first two terms in the expansion of the elastic energy:

$$\varphi = \mu\,(\varepsilon_{ik})^2 + \frac{\lambda}{2}\,(\varepsilon_{ii})^2 = \mu\left(\varepsilon_{ik} - \frac{1}{3}\varepsilon_{ll}\delta_{ik}\right)^2 + \frac{1}{2}\left(\lambda + \frac{2}{3}\mu\right)(\varepsilon_{ll})^2 . \tag{5.11}$$

If one chooses the tensor ε_{ik}, in such a way that $\varepsilon_{ll} = 0$, the right-hand side of (5.11) saves only the first term. If one chooses the tensor of the form $\varepsilon_{ik} = \delta_{ik}\text{const}$, only the second term remains. Hence, it follows that the elastic energy of a nondeformed body is minimum if $\lambda + 2\mu/3 > 0$, $\mu > 0$.

The approximation of (5.10) and (5.11) allows us to make no distinctions between the volumes and surfaces of the body, before and after deformation: $\mathrm{d}V_{\mathbf{X}} \approx \mathrm{d}V_{\mathbf{x}}$, $\mathrm{d}S_i \approx \mathrm{d}\sigma_i$. The Piola–Kirchhoff and stress tensors coincide and are proportional to the strain tensor:

$$T_{ij} \approx P_{ij} \approx \sigma_{ij} = \frac{\partial \varphi}{\partial \varepsilon_{ij}} = \lambda\varepsilon_{pp}\delta_{ij} + \mu\varepsilon_{ij} . \tag{5.12}$$

The stress tensor σ_{ij} in (5.12) is symmetric in the indices i, j. The boundary conditions on the sample surface subjected to external stresses are simplified:

$$\sigma_{is}n_s|_{\sigma'} = \sigma_{is}^{\text{ext}}n_s\Big|_{\sigma'} .$$

The dynamic equations in (5.8) become linear and, in the absence of mass forces, take the form:

$$-\rho_0\frac{\partial^2 u_i}{\partial t^2} + c_{ijkl}\frac{\partial^2 u_k}{\partial x_j \partial x_l} = 0 . \tag{5.13}$$

The above equations are suitable for describing even anisotropic crystals. For an isotropic medium, the tensor of elastic constants is equal to:

$$c_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) , \tag{5.14}$$

and equation (5.13) appear as:

$$-\rho_0\ddot{\vec{u}} + (\lambda + \mu)\,\vec{\nabla}\,\mathrm{div}\,\vec{u} + \mu\Delta\vec{u} = 0 . \tag{5.15}$$

Particular solutions of model (5.15) are longitudinal and transverse sound waves (Goldstone modes) moving in the sample independently of each other with velocities of:

$$c_{\mathrm{l}} = \sqrt{(\lambda + 2\mu)/\rho_0}\,; \quad c_{\mathrm{t}} = \sqrt{\mu/\rho_0}\,.$$

The validity of the assertion is easiest to verify by applying the divergence and rotor operations to the equality (5.15). Then, taking into account the identities:

$$\operatorname{div}\vec{\nabla}f = \Delta f\,; \quad \operatorname{rot}\vec{\nabla}f = 0\,,$$

we immediately obtain closed wave equations:

$$\left(\frac{\partial^2}{\partial t^2} - \frac{1}{c_{\mathrm{l}}^2}\Delta\right)\operatorname{div}\vec{u} = 0\,; \quad \left(\frac{\partial^2}{\partial t^2} - \frac{1}{c_{\mathrm{t}}^2}\Delta\right)\operatorname{rot}\vec{u} = 0\,.$$

They control the propagation of longitudinal ($\operatorname{div}\vec{u}$) and transverse ($\operatorname{rot}\vec{u}$) deformations in the medium.

We demonstrate that, in the longwave approximation, the equations of motion for atoms in a crystal (2.17) are reduced to the equations of the linear theory of elasticity (5.13). Suppose that a unit cell of a crystal contained one atom, and only three acoustic modes propagate in the crystal. In the harmonic approximation, the equations of dynamics for the crystal lattice take the form:

$$M\ddot{U}_\alpha(l) = -\sum_{l',\alpha'}\Phi_{\alpha\alpha'}\left(l-l'\right)U_{\alpha'}\left(l'\right)\,.$$

The time dependent vector $\vec{U}(l)$ describes small oscillations of an atom with a mass M, near the equilibrium position $\vec{R}^l = l_1\vec{a}_1 + l_2\vec{a}_2 + l_3\vec{a}_3$; $\Phi_{\alpha\alpha'}(l)$ is a matrix of force constants. A set of three numbers $l = \{l_1, l_2, l_3\}$ enumerates the unit cells, and the indices $\alpha, \alpha' = 1, 2, 3$ specify the vector components. When discussing the dynamics of a crystal, we will explicitly indicate all the summations.

Suppose that the characteristic spatial scale of the displacement field change was much larger than the lattice constant. In this case we can assume that the coordinates of the atoms run through a series of continuous values and the displacements are continuous functions of the coordinates: $\vec{U}(l) = \vec{U}(\vec{x})$. We confine ourselves to the first terms in the expansion of the function $U_{\alpha'}(l') = U_{\alpha'}(\vec{x}')$ in a series, with respect to a point with the radius vector $\vec{x}$:

$$U_{\alpha'}\left(\vec{x}'\right) \approx U_{\alpha'}(\vec{x}) + \sum_{\gamma=1}^{3}\frac{\partial U_{\alpha'}(\vec{x})}{\partial x_\gamma}R_\gamma^{l'-l} + \frac{1}{2}\sum_{\gamma,\delta=1}^{3}\frac{\partial^2 U_{\alpha'}(\vec{x})}{\partial x_\gamma \partial x_\delta}R_\gamma^{l'-l}R_\delta^{l'-l}\,.$$

Here, $\vec{x}' - \vec{x} \approx \vec{R}^{l'} - \vec{R}^l = \vec{R}^{l'-l}$. Such an approximation is justified if the atoms interact weakly at large distances. As a result, we find:

$$-\sum_{l',\alpha'}\Phi_{\alpha\alpha'}\left(l-l'\right)U_{\alpha'}\left(l'\right) \approx \sum_{\alpha'}C_{\alpha\alpha'}U_{\alpha'}(\vec{x}) + \sum_{\alpha',\gamma}C_{\alpha\alpha'\gamma}\frac{\partial U_{\alpha'}(\vec{x})}{\partial x_\gamma} + \sum_{\alpha',\gamma,\delta}C_{\alpha\alpha'\gamma\delta}\frac{\partial^2 U_{\alpha'}(\vec{x})}{\partial x_\gamma \partial x_\delta}\,.$$

The coefficient

$$C_{\alpha\alpha'} = -\sum_{l'}\Phi_{\alpha\alpha'}\left(l-l'\right) = 0$$

is equal to zero by the property (2.25) of the force matrix (a shift of the crystal as a whole should produce no forces acting on the atoms of the crystal). For crystals with an inversion center (almost all crystals have an inversion center), the second coefficient is also zero:

$$C_{\alpha\alpha'\gamma} = -\sum_{l'} \Phi_{\alpha\alpha'}\left(l - l'\right) R_{\gamma}^{l'-l} = 0\,.$$

As a result, the equation of dynamics for a crystal is rewritten in the form:

$$-\rho_0 \frac{\partial^2}{\partial t^2} U_{\alpha}(\vec{x}) + \sum_{\gamma,\alpha',\delta'} c_{\alpha\gamma\alpha'\delta} \frac{\partial^2 U_{\alpha'}(\vec{x})}{\partial x_{\gamma} \partial x_{\delta}} = 0\,, \tag{5.16}$$

where $\rho_0 = M/V_a$ is the density of a medium and V_a is the volume of a unit cell of a crystal:

$$c_{\alpha\gamma\alpha'\delta} = -\frac{1}{2V_a} \sum_{l'} \Phi_{\alpha\alpha'}\left(l - l'\right) R_{\gamma}^{l'-l} R_{\delta}^{l'-l} = \frac{1}{V_a} C_{\alpha\alpha'\gamma\delta}\,. \tag{5.17}$$

Equations (5.13) and (5.16) have the same algebraic structure, although they use different notations for vector fields, tensor indices, and summations. For cubic crystals, the constant tensor (5.17) is close to the tensor of elastic moduli of an isotropic medium (5.14). In general, the transformations of the point symmetry group of the crystal identify the form of the coefficients in the expansion in powers of η_{ij} of the elastic energy of the crystal [28], and lead to the coincident of equations (5.13) and (5.16).

5.2 Dislocations in a Crystal

The *mechanical properties* of solids are called those, that determine the ability of solids to change their shape under external mechanical impacts (pressure, force, etc.) and resist destruction by these forces.

Deformation of a solid body is the result of changes in the mutual arrangement of the particles of which the body consists. *Elastic deformation* of a crystal refers to the fact that the atoms are slightly displaced from their equilibrium positions under the action of external forces until the equilibrium is restored. In this case, the equilibrium is established between the forces of attraction and repulsion of atoms, on the one hand, and external forces, on the other hand. Elastic deformation is characterized by reversibility; it disappears after relieving the stresses deforming the body. It is for this reason that Hooke's law holds for elastic deformation (5.12). *Hooke's law* corresponds to a linear relationship between deformation and stress.

As *plastic deformation* of a crystal occurs, the crystal layers slide relative to each other. Plastic deformation is irreversible. After removing an external stress, the body does not restore its original shape. To shift crystalline planes relatively to each other, forces tangent to these planes are needed. No matter which way the external force is directed, there are always atomic crystalline planes, along which some component

of this force acts. So, there exists a possibility of a shear even under a tensile external stress. According to modern concepts, the carriers of elementary acts of plastic shear along the crystalline planes are special macroscopic defects of the crystal, or dislocations. Although the mechanism of formation and motion of dislocations can be explained only on the basis of atomistic concepts, the classical theory of elasticity outlines many of the dislocation associated strength and plastic properties of materials. For this, it is sufficient to introduce sources into the theory to simulate changes in the microstructure and topological properties of a crystal as dislocations are formed. Now we expound the microscopic interpretations underlying this approach.

Dislocations are defects of crystals. They represent lines along and near which the proper arrangement of the crystal atomic plane is violated. There are two simplest types of dislocations: edge and screw ones.

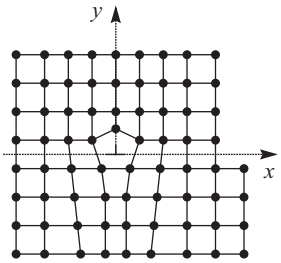


Fig. 5.1: Scheme of the arrangement of atoms near an edge dislocation.

Figure 5.1 shows the commonly accepted scheme of an *edge dislocation* in a single cubic crystal. The edge dislocation coincides with the *Oz*-axis of the Cartesian coordinate system, and represents a straight line on which the edge of the extra atomic half-plane inserted inside a perfect crystal terminates. Dislocations of such a type can be caused by uneven motion of atoms during sliding. Their appearance is similar to the formation of wrinkles of a carpet due to uneven motion of its individual parts. The dislocations can also be formed during crystal growth from the melt. The atoms are strongly displaced from their equilibrium positions in a perfect crystal and localized along a cylindrical region, whose characteristic radius is equal to only a few interatomic distances. This region is called the *core of a dislocation*. It is important to note that, outside the dislocation core, the crystal remains almost perfect and is subjected only to elastic deformations.

The deformations away from the dislocation can be detected by traversing along a closed contour on the lattice around the dislocation in the *Oxy*-plane (Figure 5.1). If one introduces the displacement vector $\vec{u}$ of each atom, directed from its position in a perfect crystal lattice, the total increment of the vector is different from zero, and equal to the lattice period along the *Ox*-axis. It is this topological singularity of the displacement field that is regarded as the initial one for detecting dislocations in a crystal

macroscopically. Most of the existing physical properties of dislocations bear no relationship to their microscopic models. Therefore, they can be described phenomenologically within the theory of elasticity, on the basis of a similar determination.

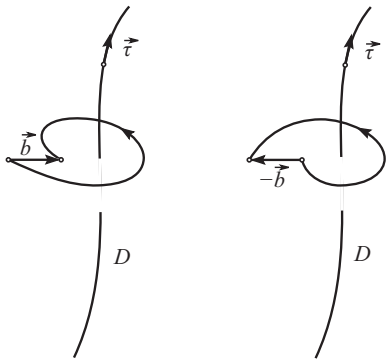


Fig. 5.2: Increment of the displacement field when traversing a line D along a closed contour for dislocations with different directions of the Burgers vector.

Thus, a dislocation in a crystal is called a special line D, characterized by the following property. In the coordinate system $Oxyz$, when traversing around a line D along a closed contour γ, the displacement field $\vec{u}(\vec{x}, t)$ of the deformed body acquires a given increment $\vec{b}$, which is equal in magnitude and direction to one of the main vectors of the crystal lattice:

$$\oint_{\gamma} \mathrm{d}u_k = b_k \,. \tag{5.18}$$

Furthermore, for definiteness, we assume that the direction taken while traversing the contour obeys the rule of the right screw with respect to the direction of the tangent vector $\vec{\tau}$ to the line D (see Figure 5.2). The vector $\vec{b}$ is called the *Burgers vector* of a dislocation. Its possible values are defined by a crystallographic structure of the body. For the edge dislocation, the vectors $\vec{b}$ and $\vec{\tau}$ are orthogonal. An edge dislocation is usually denoted by the symbol $\perp$. The vertical line indicates from which side an extra half-plane of atoms is inserted.

A screw dislocation is a straight line parallel to the vector $\vec{b}$. The presence of a screw dislocation in a crystal turns the crystal lattice planes into a helicoidal surface similar to a spiral staircase. Figure 5.3 displays the scheme of the arrangement of atomic planes in the presence of a screw dislocation that coincides with the line OO', parallel to the Oz-axis. Any crystal plane containing the OO'-axis can be a plane of comparatively slight mechanical motion of a screw dislocation. The exit point of the screw dislocation on the outer surface of the sample appears as a step (Figure 5.3). During the process of crystallization, the atoms of a medium, depositing out of vapor or a solution easily tack onto such a step, which triggers spiral growth of the crystals.

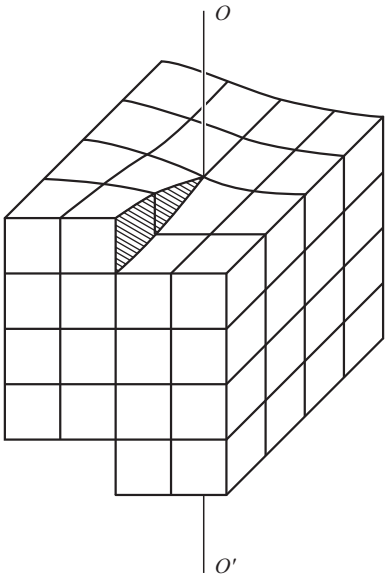


Fig. 5.3: Part of a crystal with a screw dislocation.

When using literature data for calculating the Burgers vector value, it is necessary to follow the sign in formula (5.18). We have chosen the sign "+" on the right-hand side of (5.18). In this case, the dislocations, which are codirectional to the Oz-axis, have the Burgers vectors: $\vec{b} = (b, 0, 0)$ and $\vec{b} = (0, 0, b)$, where $b > 0$ (see Figures 5.1 and 5.3).

The dislocations can be merged in the event of there being two dislocations with opposite (identical) directions of the Burgers vector and identical (opposite) directions of the dislocation lines. Such dislocations annihilate when their lines coincide. As a result, the two defects disappear and the perfect crystal lattice is restored (see Figure 5.4). The excess of energy is released in the form of thermal lattice vibrations and acoustic emission [29].

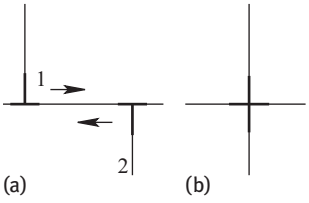


Fig. 5.4: Annihilation of edge dislocations: (a) two "extra" atomic half-planes; (b) forming a regular atomic plane as dislocations merge.

When subjected to external forces, the edge dislocation can move relatively easily only in its *slip plane*. The latter is a crystal plane passing through the Burgers vector and the dislocation line. As a result of the dislocation moving, it breaks and reconnects bonds between, not all atoms in the slip plane, but only between those near the dislocation line at a given moment (Figure 5.5). Therefore, the local shear strain can be observed at relatively low external stresses.

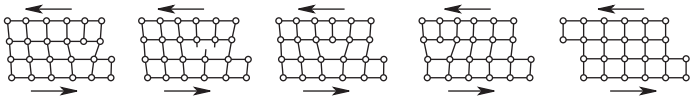


Fig. 5.5: The scheme of motion of an edge dislocation in the slip plane.

Imagine that the dislocation shown in Figure 5.1 moves in its slip plane towards the left edge of the crystal. It is easy to see (Figure 5.5) that when it reaches this edge, an irreversible plastic deformation of the body occurs. The entire upper half of the body shifts relative to the lower half by a distance equal to the lattice constant. The body in the final state has a regular crystal lattice, which means it turns out to be unstressed. It is important to note that this dislocation displacement requires a stress two or three orders of magnitude less than for disrupting all interatomic bonds in the slip plane, and shifting the entire upper half of the crystal with respect to the lower half on the lattice constant. The problem of the plastic shear of the upper part of a crystal with respect to the lower part is in some sense analogous to the problem of moving a carpet along the floor surface. It is not easy to move a carpet as a whole even for a short distance. But it is not difficult to obtain the same result by forming small wrinkles in the carpet and then to shift them from one edge of the carpet to the other. As mentioned earlier, the wrinkles on the carpet are analogous to the dislocations. It can be said that *dislocations* are elementary *carriers of plasticity*.

At high temperatures, the edge dislocation makes it possible to move perpendicular to the slip plane. This displacement is slow and limited, so it is called a *dislocation climb*. The climb of the edge dislocation is associated with the diffusion mass transfer and is accomplished, either by detaching atoms from the edge of the extra half-plane, or by attaching them to it (Figure 5.6).

Due to the fact that the Burgers vector along the dislocation line is constant (see definition (5.18)), the curvilinear dislocation can consist of segments of edge, screw, or mixed types. The topological restriction (5.18) also guarantees nonbreaking off the dislocation line inside the crystal. It either ends at the surface of the crystal or forms a closed loop.

The dislocation line can branch out into several separate dislocations. Suppose the dislocation line l_1 with the Burgers vector $\vec{b}_1$ branches at a point A into two dis-

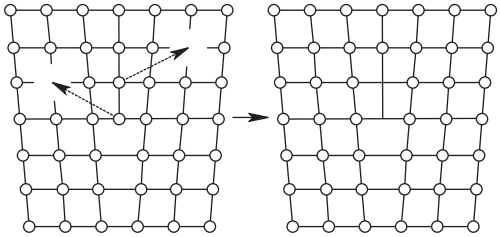


Fig. 5.6: Climb of an edge dislocation: atoms of the extra half-plane pass into vacant lattice sites.

location lines, l_2 and l_3, with Burgers vectors, $\vec{b}_2$ and $\vec{b}_3$, respectively (Figure 5.7). In this case, the dislocations l_1, l_2, and l_3 are said to form a *node*. Now we use the definition (5.18) and choose a single contour γ near the point M_1, and another contour near the point M_2. Since the complete Burgers vector is preserved along the dislocation, we obtain $\vec{b}_1 = \vec{b}_2 + \vec{b}_3$. More than three dislocations can converge in one node.

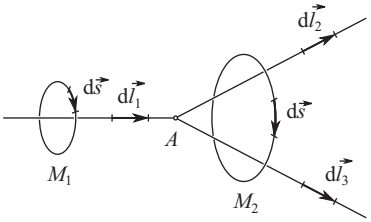


Fig. 5.7: A node formed by three dislocations.

Elastic dislocation fields are long range, which causes regular dislocation structures to form. The point is that, in equilibrium configurations of the type of regular walls or networks the dislocations consist of, the elastic fields from different dislocation lines are compensated and do not propagate over large distances. As a result, the energy of the entire dislocation system drastically diminishes. The dislocation network usually contains a large number of nodes and can be two- and three-dimensional. The Burgers vectors of dislocations forming a triple node always lie in a single plane. Therefore, when it is possible for only triple nodes to exist, the dislocation network is topologically two-dimensional. However, at great distances, it can be distorted and can intersect itself.

From microscopic examination, it directly follows that the dislocation loop can move relatively easily without breaking the crystal continuity only along a certain surface, which is called the *slip surface*. The latter is a cylindrical surface Σ_g whose generatrices are parallel to the vector $\vec{b}$, and a closed dislocation line D serves as a guiding line (Figure 5.8). The shift of the dislocation loop by the vector $\vec{b}$ along the surface Σ_g changes no interatomic spacing near Σ_g. In this case, plastic deformation

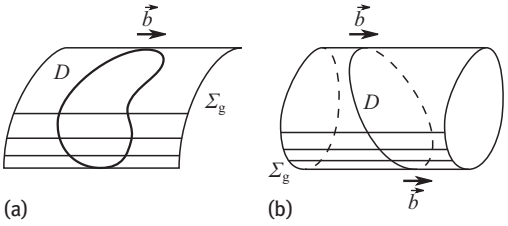


Fig. 5.8: The slip surfaces of dislocation loops: (a) the loop D can be contracted to a point by sliding; (b) the dislocation line D does not contract to a point.

is not accompanied by local changes in the volume of the medium. After the local shift, the crystal is again in equilibrium, albeit in an internally stressed state. Such a displacement of dislocations under the action of external forces is called *sliding*.

Depending on their position on the slip surface, the dislocation loops are divided into two types. A dislocation loop of the first type can be contracted to a point by sliding. Slip of a dislocation loop of the second type around the surface Σ_g never makes the loop infinitely small. The second type dislocation loop has the smallest dimensions when it lies in a plane perpendicular to the vector $\vec{b}$. At the same time, there can be scenarios for the dislocations transition to another slip surface by sliding (or to another slip plane, when it comes to a flat dislocation loop). Let us explain the assertion.

Imagine that a part of the dislocation line is lying in its slip plane, and that it has a screw type segment on which $\vec{\tau} \| \vec{b}$ (Figure 5.9 (a)). Also, recall that a screw dislocation can generate many slip planes. Any crystal plane parallel to the vector $\vec{b}$ can be taken as the slip plane of the screw segment of the dislocation loop. This plane can cross the slip plane of the remaining part of the dislocation at an angle. The screw segment δl_0 of finite length slides in a new sliding plane, taking the form of an arc

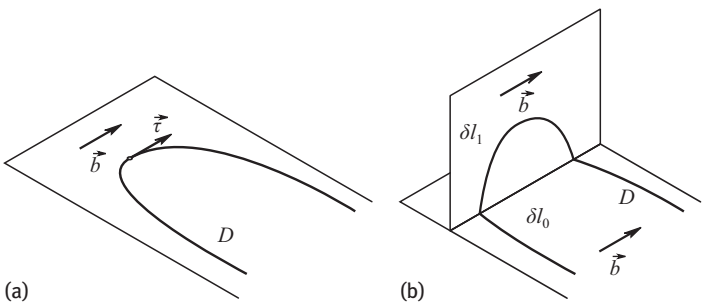
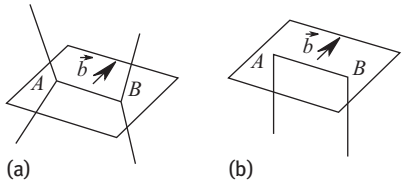


Fig. 5.9: Transverse slip: (a) the initial slip plane of dislocations; (b) the outcrop of the screw segment δl_0 of dislocations onto the transverse slip plane.

δl_1 (Figure 5.9 (b)). A similar dislocation emergence from the old slip plane is called *transverse sliding*.

Sources of Dislocations

To ensure significant plastic deformation in a crystal, mechanisms for dislocations to nucleate and multiply must exist. The main ones are the Frank–Read sources and double transverse sliding.



(a) (b)

Fig. 5.10: A dislocation segment fixed in two points.

The *Frank–Read source* can be a segment of the dislocation secured at the ends at points A and B. This may be a segment between two nodes of a dislocation network (see Figure 5.10 (a)) or between two breakpoints on a single dislocation (Figure 5.10 (b)). Under the applied stress, the initially straight line segment one (Figure 5.11) is bent into loop two, then into loop three. Two loop elements in line four closely approach each other. After merging, they annihilate. The next configuration consists of the closed loop five and short AB segment of the dislocation. It is worth pointing out that the regular crystal lattice is restored in the merge region. Now, loop five freely extends as a separate dislocation line. The short segment again begins to

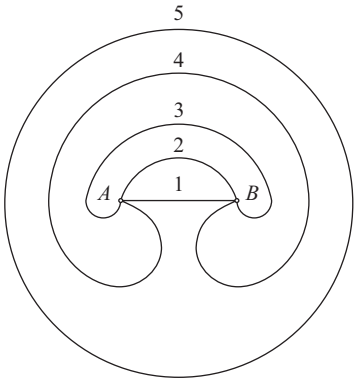


Fig. 5.11: The Frank–Read source.

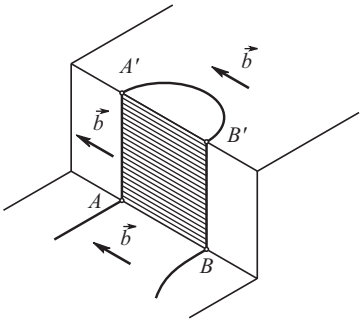


Fig. 5.12: Scheme of double transverse sliding.

pass successive configurations 1, 2, 3, 4,... The dislocation *AB* segment that gives rise to a sequence of closed dislocation loops is called the Frank–Read source.

The Frank–Read source is often formed as a result of so called *double transverse sliding*. Suppose that the screw *AB* segment suffers transverse sliding (Figure 5.12) to form the loop $AA'B'B$ in a new slip plane. Consider such a dislocation. Being purely screw, the segment $A'B'$ can pass into the new slip plane, parallel to the original one by double transverse sliding. If the points A' and B' are unmovable, the dislocation segment $A'B'$ can act as the Frank–Read source.

Having ascertained the reason for the weakening of the strength of crystals with dislocations, two alternative ways of obtaining materials with increased strength can be indicated. One of them is the use of crystals that are almost, or completely, free from dislocations. However, this method is not widely used in producing composite reinforced materials with high strength and heat resistance.

The strength and plasticity of crystals depends on how easily dislocations move in them. Therefore, an opposite method has become in demand. It is the creation of materials with a large number of defects in the crystalline structure. Defects, which can also be other dislocations, impede the dislocation motion. Surrounding the dislocations, the elastically distorted region of the crystal scatters phonons, conduction electrons, and other quasiparticles. The scattering processes take away some of the energy from the moving dislocations and make them decelerate.

Hardening and Plastic Deformation

As Frank and Read have shown, the plastic deformation process causes a huge number of new dislocations. When moving, the dislocations penetrate one through the other. Since each dislocation in the crystal is a place of the violated order, attraction and repulsion forces arise between the dislocations. As a result, in some metals, the resistance to the dislocation motion increases by more than one hundred times. That is what accounts for the hardening of crystals during plastic deformation.

Dislocations and Impurities

The addition of impurities to a substance usually enhances the strength and hardness by impeding the motion of dislocations. Let us clarify this statement using an example of an edge dislocation (Figure 5.1). The lattice above the dislocation line is somewhat compressed, and below it, it is stretched. Because of this, the impurity atoms form a cluster mainly in the region of the dislocations. Namely, atoms whose dimensions are smaller than the sizes of the main lattice atoms are located in the compressed part, i.e., above the dislocation line. The larger atoms are arranged under it in the stretched part. The accumulation of impurity atoms near dislocations requires a greater stress to initiate the process of the dislocation traveling. This means hardening of the body. As calculations show, a rather small number of impurity atoms accumulated at the dislocations (of the order of 0.01 percent or less) are enough for hardening. This explains the strong influence of impurities on the mechanical properties of solids.

Simultaneously, if an alloy is a mixture of two kinds of almost identically sized atoms (for example, gold and silver), the hardening is very weakly expressed. In such cases, the impurities do not accumulate at dislocations; in any other place of the lattice, their energy is about the same.

Dislocations and Heat Treatment

Long term high temperature treatment followed by slow cooling of bodies increases the mobility of dislocations. Is happens due to the growth of the kinetic energy of atoms at a high temperature. Impurity atoms clustered near dislocations diffuse, and partially leave their sites at the dislocations to equalize their concentration throughout the body. The heat treated dislocations move more freely, and plastic deformation of the body is possible to observe at lower stresses.

Special heat treatment processes, including heating to a certain high temperature, followed by immediate cooling of bodies are more complicated. After such heat treatments, crystals of one, or a few, strengthening phases precipitate in the alloy. These phases strongly violate the regularity of the crystal structure. The resulting fine crystalline multiphase structure exerts great resistance to the motion of dislocations. At high temperatures, diffusion processes may lead to enlargement of the fine crystalline structure of the alloy, to a dissolution of precipitates of the strengthening phases, or to a disappearance of distortions of the crystal lattice. In high temperature alloys, the diffusion processes proceed slowly enough, which ensures the long term preservation of the strengthened state at high temperatures.

Dislocations and Destruction

Under certain conditions, local elastic stresses caused by large dislocation clusters can become significant. As a result of the stress concentration, a disruption of the bonds linking the atoms of the neighboring crystal planes and the initiation of microcracks can occur. Further growth and merging of several microcracks lead to the destruction of the crystal. One such scenario was proposed and analyzed by Stroh [30].

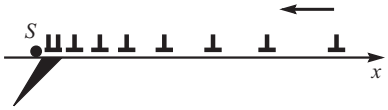


Fig. 5.13: The formation of a wedging crack as head dislocations unite into the cluster.

Let us look at a set of a large number of identical rectilinear edge dislocations located parallel to each other in one and the same slip plane. Suppose an external shear stress pushes the dislocations towards a stopper S located there too (Figure 5.13). The stopper can be, for example, a grain boundary or the inclusion of a strengthening phase. With sufficient magnitude of the external stress, the head dislocations of the cluster approach each other at a distance in the order of the interatomic distance. Then, the internal stresses near the stopper below the slip plane stretch the lattice extensively. They increase so much that they may disrupt the bonds between adjacent atomic planes and form a wedging crack.

5.3 Basic Equations of the Theory of Dislocations

Having gained the atomic understandings underlying the theory of dislocations, we go over to the discussion of its quasiclassical equations. Condition (5.18) stands for a topological singularity of the displacement field of a crystal as dislocations occur. The topological singularity is associated with the ambiguity of a field $\vec{u}(\vec{x}, t)$. The function $\vec{u}(\vec{x}, t)$ receives a given increment $\vec{b}$ when traversing around the dislocation line. Certainly, there is no physical ambiguity because the increment $\vec{b}$ corresponds to an additional displacement of the crystal atoms by one of the lattice periods. Such a shift does not change the crystal states.

A general property of the functions with topological singularities is that some of their derivatives do not commute with each other in the domain of existence of the singularities. It is not hard to persuade ourselves that this statement holds true if one rewrites the definition of dislocation (5.18) in the differential form. We use the Stokes' theorem and transform the integral along the contour γ into an integral over the surface S spanned by this contour:

$$\oint_{\gamma} \mathrm{d}u_k = \oint_{\gamma} \mathrm{d}x_p \partial_p u_k = \int_{S} \mathrm{d}S_p \varepsilon_{pmn} \partial_m \partial_n u_k = b_k \,. \tag{5.19}$$

In what follows, the contracted notations $\partial_p = \partial/\partial x_p$ adopted in the field theory are used to simplify the formulas. We enumerate the spatial vector components by the letters of the Latin alphabet: $p, m, \ldots = 1, 2, 3$. The letters of the Greek alphabet will appear later when listing all the space time coordinates: x_μ ($\mu = 0, 1, 2, 3$). Here time is written as $t = x_0$.

The right-hand side of (5.19) differs from zero, since $b_k \neq 0$. The tensor ε_{pmn} is antisymmetric: $\varepsilon_{pmn} = -\varepsilon_{pnm}$, therefore the left-hand side of (5.19) does not vanish, only when the second derivatives of the field $\vec{u}(\vec{x}, t)$ are not commutative:

$$\partial_m \partial_n u_k - \partial_n \partial_m u_k \equiv [\partial_m, \partial_n] u_k \neq 0 .$$

Here and below, all the derivatives are interpreted in accordance with the theory of generalized functions.

The space time domain, where

$$[\partial_\mu, \partial_\nu] u_k \neq 0 ; \quad \mu, \nu = 0, 1, 2, 3 ; \quad k = 1, 2, 3 , \tag{5.20}$$

is called a domain of the topological singularity of the function $\vec{u}(\vec{x}, t)$.

In the problem at hand, the derivatives $\partial_m u_k$ and $\partial_0 u_k$ correspond to the observable quantities and, consequently, must be single valued functions with higher commuting derivatives over the coordinates x_μ:

$$\partial_\mu \partial_\nu \partial_\alpha u_k - \partial_\nu \partial_\mu \partial_\alpha u_k \equiv [\partial_\mu, \partial_\nu] \partial_\alpha u_k = 0 . \tag{5.21}$$

The relations (5.20) and (5.21) express general mathematical conditions for the existence of topological singularities. We show that the definition of a dislocation (5.19) allows us to specify them.

The dislocation line can be defined through a spatial parameter σ and time t: $\vec{l} = \vec{l}(\sigma, t)$, $\sigma_1 \leq \sigma \leq \sigma_2$. We determine the *tensor of dislocation density*:

$$D_{ik}(\vec{x}, t) = b_k \int_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \frac{\partial l_i(\sigma, t)}{\partial \sigma} \delta^{(3)}\left(\vec{x} - \vec{l}(\sigma, t)\right) , \tag{5.22}$$

where $\delta^{(3)}(\vec{x})$ is the spatial delta function. Then the right-hand side of (5.19) can be represented as an integral of D_{ik} over the surface S, spanned by the contour γ:

$$\int_S \mathrm{d}S_p D_{pk} = b_k \int_S \mathrm{d}S_p \int_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \frac{\partial l_p(\sigma, t)}{\partial \sigma} \delta^{(3)}\left(\vec{x} - \vec{l}(\sigma, t)\right) = b_k . \tag{5.23}$$

Let us elucidate the finding. The dislocation line intersects the surface S only at a single point. The argument of the delta function at this point vanishes. The integral of the delta function over the small volume $\Delta S_p \Delta l_p$ containing the indicated point is equal to unity. The remaining elements of the surface S and the line $\vec{l}(\sigma, t)$ do not contribute

to the integration. Hence, formula (5.23) follows. By the arbitrariness of the contour, the equality of two integrals:

$$\int\limits_S \mathrm{d}S_p \varepsilon_{pmn} \partial_m \partial_n u_k = \int\limits_S \mathrm{d}S_p D_{pk} \,,$$

implies the equality of the integrands:

$$D_{pk} = \varepsilon_{pmn} \partial_m \partial_n u_k \,. \tag{5.24}$$

From (5.24), using the identity:

$$\varepsilon_{rsp} \varepsilon_{pmn} = \delta_{rm} \delta_{sn} - \delta_{rn} \delta_{sm} \,, \tag{5.25}$$

we find an explicit relationship between the noncommutative derivatives of the field $\vec{u}(\vec{x}, t)$ over the spatial coordinates:

$$\partial_r \partial_s u_k - \partial_s \partial_r u_k = \varepsilon_{rsp} D_{pk} \,, \tag{5.26}$$

where the tensor D_{pk} is determined by expression (5.22).

Combining formulas (5.21) and (5.24), we get the following important equations:

$$\partial_p D_{pk} = 0 \,; \quad \partial_0 D_{pk} + \varepsilon_{pmn} \partial_m I_{nk} = 0 \,, \tag{5.27}$$

where $I_{nk} = \partial_n \partial_0 u_k - \partial_0 \partial_n u_k$.

The first of the equations (5.27) expresses the constancy of the Burgers vector along the dislocation line. We verify that it determines more precisely the geometric properties of the dislocation line in representation (5.22). Using (5.22), we calculate the quantity $\partial_p D_{pk}$:

$$\begin{aligned}
\partial_p D_{pk}(\vec{x}, t) &= b_k \int\limits_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \frac{\partial l_p(\sigma, t)}{\partial \sigma} \frac{\partial}{\partial x_p} \delta^{(3)}\left(\vec{x} - \vec{l}(\sigma, t)\right) = \\
&= -b_k \int\limits_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \frac{\partial l_p}{\partial \sigma} \frac{\partial}{\partial l_p} \delta^{(3)}\left(\vec{x} - \vec{l}(\sigma, t)\right) = \\
&= -b_k \int\limits_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \frac{\partial}{\partial \sigma} \delta^{(3)}\left(\vec{x} - \vec{l}(\sigma, t)\right) = \\
&= b_k \left[\delta^{(3)}\left(\vec{x} - \vec{l}(\sigma_1, t)\right) - \delta^{(3)}\left(\vec{x} - \vec{l}(\sigma_2, t)\right)\right] \,.
\end{aligned}$$

Here, $\vec{l}(\sigma_1, t)$ and $\vec{l}(\sigma_2, t)$ are the edge points of the dislocation line at a time t. It is seen from the relationship obtained that the first of the continuity equations (5.26) is satisfied inside the body only when the line of dislocation is closed, or emerges on the surface of the body.

Comment

Having emerged on the external surface of the body, the dislocation changes the boundary conditions. The elastic fields inside large crystals are usually said to depend weakly on the edge points of dislocations on the surface of the body. Such an assumption may be violated in samples of small crystal sizes [29].

The second of the continuity equations expresses the conservation law of the Burgers vector for the motion of a dislocation. Since the form of the tensor D_{pk} is known, an explicit expression for the density tensor of the dislocation flow I_{nk} can be found from the second equation (5.27). We recall that the quantity I_{nk} characterizes the noncommutativity of the derivatives of the field $\vec{u}(\vec{x}, t)$ over the spatial coordinates and time. Using the representation (5.22), we compute $\partial_0 D_{pk}$:

$$\partial_0 D_{pk}(\vec{x}, t) = b_k \int_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \left[\frac{\partial^2 l_p(\sigma, t)}{\partial\sigma\partial t} \delta^{(3)}\left(\vec{x} - \vec{l}(\sigma, t)\right) + \frac{\partial l_p(\sigma, t)}{\partial\sigma} \frac{\partial}{\partial t} \delta^{(3)}\left(\vec{x} - \vec{l}(\sigma, t)\right) \right] =$$

$$= b_k \int_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \left[\frac{\partial^2 l_p}{\partial\sigma\partial t} \delta^{(3)}(\vec{x} - \vec{l}) - \frac{\partial l_p}{\partial\sigma} \frac{\partial l_m}{\partial t} \frac{\partial}{\partial x_m} \delta^{(3)}(\vec{x} - \vec{l}) \right] .$$

We transform the first term on the right-hand side of the equation by integration by parts:

$$\int_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \frac{\partial^2 l_p}{\partial\sigma\partial t} \delta^{(3)}(\vec{x} - \vec{l}) = \frac{\partial l_p(\sigma, t)}{\partial t} \delta^{(3)}\left(\vec{x} - \vec{l}(\sigma, t)\right)\Big|_{\sigma=\sigma_1}^{\sigma=\sigma_2} - \int_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \frac{\partial l_p}{\partial t} \frac{\partial}{\partial\sigma} \delta^{(3)}(\vec{x} - \vec{l}) =$$

$$= -\int_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \frac{\partial l_p}{\partial t} \frac{\partial l_m}{\partial\sigma} \frac{\partial}{\partial l_m} \delta^{(3)}(\vec{x} - \vec{l}) =$$

$$= \int_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \frac{\partial l_p}{\partial t} \frac{\partial l_m}{\partial\sigma} \frac{\partial}{\partial x_m} \delta^{(3)}(\vec{x} - \vec{l}) .$$

In the case of closed dislocation loops, the extra integral terms vanish. For dislocations emerging on the body surface, these terms also vanish if one takes the values of $\vec{x}$ inside the body. Applying identity (5.25), we transform the final result:

$$\partial_0 D_{pk}(\vec{x}, t) = b_k \frac{\partial}{\partial x_m} \int_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \left[\frac{\partial l_p}{\partial t} \frac{\partial l_m}{\partial\sigma} - \frac{\partial l_p}{\partial\sigma} \frac{\partial l_m}{\partial t} \right] \delta^{(3)}(\vec{x} - \vec{l}) =$$

$$= \varepsilon_{pmn} \frac{\partial}{\partial x_m} \left[b_k \varepsilon_{nrs} \int_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \frac{\partial l_r}{\partial t} \frac{\partial l_s}{\partial\sigma} \delta^{(3)}(\vec{x} - \vec{l}) \right] . \tag{5.28}$$

Comparing (5.28) with the second equality (5.27), we derive a formula of the relationship between mixed derivatives of the field $\vec{u}(\vec{x}, t)$ over the spatial coordinates and

time:

$$I_{nk} = \partial_n \partial_0 u_k - \partial_0 \partial_n u_k = b_k \varepsilon_{nsr} \int_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \frac{\partial l_s}{\partial \sigma} \frac{\partial l_r}{\partial t} \delta^{(3)}(\vec{x} - \vec{l}) \,. \tag{5.29}$$

When a dislocation moves along the slip surface (such a motion is called *conservative*), the condition $I_{nn} = 0$ is fulfilled, since in this case, the vectors b_k, $\partial l_s/\partial\sigma$, $\partial l_r/\partial t$ lie in one plane.

The basic dynamic equation of the elasticity theory (5.13) in the new notations appears as:

$$-\rho_0 \partial_0^2 u_i + c_{ijkl} \partial_j \partial_l u_k = 0 \,. \tag{5.30}$$

Next, we discuss the case of an isotropic medium with the tensor of elastic constants (5.14):

$$c_{ijkl} = c_{jikl} = c_{klij} = c_{ijlk} = \lambda \delta_{ij} \delta_{kl} + \mu \left(\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk} \right) \,. \tag{5.31}$$

To calculate the properties of crystals with dislocations, instead of the ambiguous functions $\vec{u}_i(\vec{x}, t)$, it is more convenient to use the distortion tensor $\beta_{ni} = \partial_n u_i$, and the displacement velocity $v_i = \partial_0 u_i$ of the medium particles. This is because they are single valued functions of spatial coordinates x_k and time x_0. The equations for calculating the fields $\beta_{ni}(\vec{x}, t)$ and $v_i(\vec{x}, t)$ are obtained by differentiating the equation (5.30) over the variables x_k and x_0, respectively. After differentiation, the derivatives are rearranged in accordance with the rules (5.21), (5.26) and (5.29). This leads to the appearance of the sources:

$$D_{ik}(\vec{x}, t) = b_k \int_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \frac{\partial l_i(\sigma, t)}{\partial \sigma} \delta^{(3)} \left(\vec{x} - \vec{l}(\sigma, t) \right) \,;$$

$$I_{nk}(\vec{x}, t) = b_k \varepsilon_{nsr} \int_{\sigma_1}^{\sigma_2} \mathrm{d}\sigma \frac{\partial l_s(\sigma, t)}{\partial \sigma} \frac{\partial l_r(\sigma, t)}{\partial t} \delta^{(3)} \left(\vec{x} - \vec{l}(\sigma, t) \right)$$

in the resulting equations:

$$\begin{aligned} -\rho_0 \partial_0^2 \beta_{ni} + c_{ijkl} \partial_j \partial_l \beta_{nk} &= \rho_0 \partial_0 I_{ni} - c_{ijkl} \varepsilon_{nlp} \partial_j D_{pk} \,; \\ -\rho_0 \partial_0^2 v_i + c_{ijkl} \partial_j \partial_l v_k &= c_{ijkl} \partial_j I_{lk} \,. \end{aligned} \tag{5.32}$$

These sources simulate changes in the topological properties and microstructure of the medium due to the presence of dislocations. The attractiveness of this approach lies in excluding the ambiguous function $\vec{u}_i(\vec{x}, t)$ from consideration. The single valued and continuous fields $\beta_{ni}(\vec{x}, t)$ and $v_i(\vec{x}, t)$ are primary; almost all observed quantities may be expressed through them.

A Screw Dislocation and Its Equation of Motion

Let us examine the characteristic features of the theory of dislocations using an example of a screw dislocation in a crystal. The elastic field of the screw dislocation is

particularly simple and admits an illustrative interpretation in the language of electrodynamics. This allows us to explain a number of key results of the theory without resorting to tedious computations.

Consider a rectilinear screw dislocation perpendicular to the plane Oxy of an unbounded body:

$$\vec{l}\,(\sigma, t) = (l_1(t), l_2(t), \sigma)\ , \quad -\infty < \sigma < \infty\,; \quad \vec{b} = (0, 0, b)\ .$$

Because of the symmetry of the problem, the displacement field of a dislocation has the form $\vec{u} = (0, 0, u)$, where u depends only on the coordinates x_1, x_2 and the time x_0. We compare the nonzero derivatives $\beta_{13} = \partial_1 u$, $\beta_{23} = \partial_2 u$, $\partial_0 u$ and the vectors $\vec{E} = (\beta_{23}, -\beta_{13}, 0)$ and $\vec{H} = (0, 0, \partial_0 u/c)$, where $c \equiv c_t = \sqrt{\mu/\rho_0}$ is the speed of transverse sound in a medium. In this case equations (5.31) and (5.32) coincide exactly with the equations of classical electrodynamics for the electric and magnetic field strengths that accompany the motion of a rectilinear charged filament:

$$-\frac{1}{c^2}\frac{\partial^2 \vec{E}}{\partial t^2} + \Delta\vec{E} = 4\pi\vec{\nabla}\rho + \frac{4\pi}{c^2}\frac{\partial \vec{j}}{\partial t}\,; \quad -\frac{1}{c^2}\frac{\partial^2 \vec{H}}{\partial t^2} + \Delta\vec{H} = -\frac{4\pi}{c}\,\mathrm{rot}\,\vec{j}\,, \tag{5.33}$$

where

$$\Delta = \partial_1^2 + \partial_2^2\ , \quad \vec{\nabla}\rho = \left(\partial_1\rho, \partial_2\rho, 0\right)\ , \quad \mathrm{rot}\,\vec{j} = (0, 0, \partial_1 j_2 - \partial_2 j_1)\ .$$

The charge ($\rho(\vec{r}, t)$) and current ($\vec{j}(\vec{r}, t)$) densities of the filament are expressed in terms of the dislocation parameters:

$$\rho = \frac{b}{4\pi}\delta^{(2)}\left(\vec{r} - \vec{l}_\perp(t)\right)\ , \quad \vec{j} = \frac{b}{4\pi}\,(V_1(t), V_2(t), 0)\,\delta^{(2)}\left(\vec{r} - \vec{l}_\perp(t)\right)\ .$$

Here, $\delta^{(2)}(\vec{r})$ is the two-dimensional delta function of the coordinates $\vec{r} = (x_1, x_2)$ and the vectors $\vec{l}_\perp = (l_1(t), l_2(t))$ and $\partial\vec{l}_\perp/\partial t = (V_1, V_2)$ specify the filament position in the plane Oxy and the speed of its movement. The coincidence of the two problems becomes clearer if one rewrites relations (5.26) and (5.29) through $\vec{E}, \vec{H}, \vec{j}, \rho$ and uses them to replace equations (5.33) by the equivalent system of Maxwell equations:

$$\mathrm{div}\,\vec{E} = 4\pi\rho\ , \qquad \mathrm{rot}\,\vec{E} = -\frac{1}{c}\frac{\partial \vec{H}}{\partial t}\ ,$$

$$\mathrm{rot}\,\vec{H} = \frac{4\pi}{c}\vec{j} + \frac{1}{c}\frac{\partial \vec{E}}{\partial t}\ , \quad \mathrm{div}\,\vec{H} = 0\ .$$

Due to the symmetry of the problem, the last equation is satisfied identically. It is given for the complete coincidence of the theory of a screw dislocation with two-dimensional electrodynamics [31].

Well known results of electrostatics allow us to immediately come to the nonzero components of the distortion tensor of the screw dislocation:

$$\vec{E}_\perp = \left(\beta_{23}, -\beta_{13}\right) = \left(\partial_2 u, -\partial_1 u\right) = \frac{b}{2\pi}\frac{\vec{r}}{r^2}\ . \tag{5.34}$$

The integration of equation (5.34) gives the displacement field $\vec{u}(\vec{r})$, that is written up to integration constants, in the form:

$$u = \frac{b}{2\pi} \operatorname{arctg} \frac{x_2}{x_1}$$

outside the singular line $\vec{r} = 0$. The above formula is not quite appropriate because the function $\operatorname{arctg}(x_2/x_1)$, by definition, has the narrow range of variation: $-\pi/2 \leq \operatorname{arctg}(x_2/x_1) \leq +\pi/2$ and, therefore, performs a double jump by $\pm\pi$ when moving the vector $\vec{r}$ along the plane Oxy. In fact, the designation $\operatorname{arctg}(x_2/x_1)$ implies a branch of the much valued logarithmic function:

$$\varphi(\vec{r}) \equiv \frac{1}{2\mathrm{i}} \ln \frac{x_1 + \mathrm{i}x_2}{x_1 - \mathrm{i}x_2} \,. \tag{5.35}$$

This function is single valued and continuous in the Oxy-plane with a cut. Usually, the half-plane $\{x_2 = 0, x_1 \geq 0\}$ is chosen as the cut, and the branch is specified by the condition of making expression (5.35) vanish on the upper edge of the cut, where $x_2 = +0$, $x_1 \geq 0$. Then formula (5.35) determines the polar angle ranged within $0 \leq \varphi \leq 2\pi$, when the vector $\vec{r}$ runs through points of the Oxy-plane with a cut. From the mathematical point of view, the choice of the cut is completely arbitrary and defines only the direction the polar angle is measured from. However, difficulties of a physical nature can arise. On the cut, the displacement field undergoes a jump by a lattice constant. Such a jump does not violate the correct structure and continuity of the crystal only when the cut coincides with the slip plane or the slip surface of the dislocation (see Figure 5.3). Therefore, to make the displacement field unambiguous, we should always choose the cuts coincided with a part of the dislocation slip surface.

In the linear theory of elasticity, the tensor of internal elastic stresses is related to the strain tensor $\varepsilon_{ij} = (\beta_{ij} + \beta_{ji})/2$ of a medium by Hooke's law:

$$\sigma_{ij} = c_{ijkl}\varepsilon_{kl} = \lambda\varepsilon_{pp}\delta_{ij} + 2\mu\varepsilon_{ij} \,. \tag{5.36}$$

In the presence of a screw dislocation, only the shear stress components are nonzero:

$$\sigma_{13} = \sigma_{31} = \mu\beta_{13} = -\frac{b\mu x_2}{2\pi r^2} \,; \quad \sigma_{23} = \sigma_{32} = \mu\beta_{23} = \frac{b\mu x_1}{2\pi r^2} \,. \tag{5.37}$$

According to (5.34) and (5.37), the distortions and stresses generated by the screw dislocation decrease slowly with increased distance (by the law $\sim r^{-1}$). The mentioned feature is typical of all rectilinear dislocations.

It is useful to calculate the energy of an elastic field created by a rectilinear dislocation in a crystal. The elastic energy per unit dislocation length is given by the integral:

$$W_{\mathrm{D}} = \frac{1}{2} \int \varepsilon_{ik}\sigma_{ik} \mathrm{d}x_1 \mathrm{d}x_2 = \int \left[\frac{\lambda}{2} \varepsilon_{pp}^2 + \mu\varepsilon_{ik}^2 \right] \mathrm{d}x_1 \mathrm{d}x_2 \,. \tag{5.38}$$

For a screw dislocation, the integral (5.38) is equal to:

$$W_{\mathrm{D}} = \frac{\mu}{2} \int \left(\beta_{13}^2 + \beta_{23}^2 \right) \mathrm{d}x_1 \mathrm{d}x_2 = \frac{\mu b^2}{4\pi} \int \frac{\mathrm{d}r}{r} \,. \tag{5.39}$$

The integration over r must be, in general, made within the limits $(0, \infty)$. However, the integral diverges logarithmically on both limits. The divergence for $r = 0$ is due to the inapplicability of the continuum theory of elasticity at interatomic distances. As the lower limit of integration, we should take the value of r_0 of the order of the lattice constant a. In this case, we omit the dislocation core energy W_0, whose maximum value is not difficult to estimate. Indeed, in the dislocation core with the cross sectional area $\sim b^2$, the relative displacement of atoms is of the order of unity. Hence, we have:

$$W_0 \sim \mu b^2 \,. \tag{5.40}$$

The upper cutoff parameter of the divergent integral (5.39) is determined by the magnitude of R in the order of the crystal size. This leaves us with:

$$W_{\mathrm{D}} = \frac{\mu b^2}{4\pi} \ln \frac{R}{r_0} \,. \tag{5.41}$$

Comparing (5.40) and (5.41), we conclude that, under the condition $\ln(R/r_0) > 1$, formula (5.41) yields the main part of the dislocation energy. At the same time, the "big parameter" $\ln(R/r_0)$ is only theoretically too large. Taking the ratio $R/r_0 \sim 10^5 - 10^6$ (although, it is often smaller), we come to a conclusion that $\ln(R/r_0)$ is slightly different from 4π. Therefore, for a rough estimate, the energy per unit length of any dislocation is accepted as being equal to $W_{\mathrm{D}} \approx \mu b^2$. The weak logarithmic dependence of the energy, and other observable quantities on the cutoff parameters, does not come across any serious difficulties in theoretically describing dislocations.

It is interesting to note that a dislocation in crystals of finite dimensions can experience attraction from their surface. Consider a screw dislocation near the boundary of a crystal occupied by the region $x \geq 0$. Suppose that the dislocation is parallel to the axis Oz and intersects the plane Oxy at a point $(x_0, 0)$. In other words, it is at a distance x_0 from the surface of the sample (Figure 5.14). External forces being absent, the boundary conditions must be satisfied on the surface of the body:

$$\sigma_{is} n_s|_{\sigma} = 0 \,, \tag{5.42}$$

where $\vec{n}$ is the vector of the unit normal to the surface σ. Within the problem under consideration, the boundary conditions (5.42) have a simple form:

$$\sigma_{31}|_{x=0} = \mu \beta_{13}\big|_{x=0} = 0 \,.$$

The same boundary conditions appear in electrostatics when calculating the electric field strength $\vec{E}$ of a charged filament near the boundary of a conductor. The tangential components of the field $\vec{E}$ must vanish on the surface of the conductor.

As in electrostatics, the distortion field of a screw dislocation in an elastic halfspace ($x \geq 0$) can be computed by using the image method. In the region $x \geq 0$, the distortion field coincides with the field of a system of two dislocations in an unbounded medium. The position and Burgers vector of the first of these are the same as

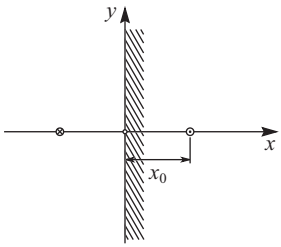


Fig. 5.14: A screw dislocation near the surface of a semi–infinite crystal and its mirror reflection outside the crystal.

for a real dislocation inside the body. The second fictitious dislocation is placed outside of the body. Its position is a mirror reflection of the position of the real dislocation relative to the body surface. The Burgers vector direction of the fictitious dislocation is opposite. The superposition of the fields of two dislocations inside the body (for $x \geq 0$)

$$\beta_{13} = -\frac{b}{2\pi}\left[\frac{y}{(x-x_0)^2+y^2} - \frac{y}{(x+x_0)^2+y^2}\right];$$
$$\beta_{23} = \frac{b}{2\pi}\left[\frac{x-x_0}{(x-x_0)^2+y^2} - \frac{x+x_0}{(x+x_0)^2+y^2}\right].$$

satisfies equations (5.31) and (5.32) with the correct boundary conditions.

The charged filament is attracted to the surface of the conductor in the same way, as if its mirror image with an opposite charge was behind the surface. Thus, we conclude that the dislocation near the body surface might experience the force of attraction from the surface. Usually, there are various barriers to keep the dislocation from moving to the boundary of the crystal. The dislocation emergence on the crystal surface is equivalent to annihilation of the real and fictitious dislocations, and should be accompanied by a burst of acoustic emission. The authors of the work [32] detected sound radiation caused by the emergence of dislocations on the free crystal surface.

If a screw dislocation with an axis Oz is at the center of an unbounded plate with a thickness of d, the influence of the plate surfaces ($y = \pm d/2$) on the dislocation is equivalent to interaction with a sequence of fictitious dislocations with alternating directions of the Burgers vector. The positions of the fictitious dislocations outside the plate are determined by a series of reflections from both plate surfaces (Figure 5.15).

It is not difficult to complete the summation of the fields by means of formulas:

$$\frac{y}{x^2+y^2} = \frac{1}{2}\left(\frac{1}{y+\mathrm{i}x} + \frac{1}{y-\mathrm{i}x}\right); \quad \frac{x}{x^2+y^2} = \frac{1}{2\mathrm{i}}\left(\frac{1}{y-\mathrm{i}x} - \frac{1}{y+\mathrm{i}x}\right);$$
$$\operatorname{ctg}\xi = \sum_{n=-\infty}^{\infty}\frac{1}{\xi+n\pi}. \tag{5.43}$$

We achieve the final result:

$$\beta_{13} = -\frac{b\,\mathrm{ch}\,(\pi x/d)\sin(\pi y/d)}{d\,[\mathrm{ch}\,(2\pi x/d) - \cos(2\pi y/d)]}, \quad \beta_{23} = \frac{b\,\mathrm{sh}\,(\pi x/d)\cos(\pi y/d)}{d\,[\mathrm{ch}\,(2\pi x/d) - \cos(2\pi y/d)]}. \tag{5.44}$$

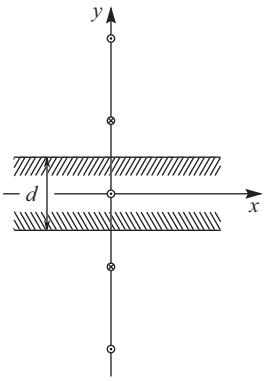


Fig. 5.15: A screw dislocation in the center of a plate and a series of its mirror reflections with alternating directions of the Burgers vectors.

At small distances ($r < d/\pi$) from the dislocation line, the distortion field (5.44) is the same as for a dislocation in an unbounded medium (5.34). However, as the distance grows, beginning from a distance of the order of the plate thickness (in the region $|x| > d/\pi$), the elastic dislocation field in the plate exponentially tends to zero. In the slip plane (when $y = 0$), the dislocation produces a short range but nonlocal shear stress:

$$\sigma_{23} = \frac{\mu b}{2d \operatorname{sh}(\pi x/d)} .$$

This example shows that the real or fictitious dislocations, combined into periodic structures, strongly screen the elastic fields.

In the external electromagnetic field $\vec{E}^{\text{ext}} = (E_1^{\text{ext}}, E_2^{\text{ext}}, 0)$, $\vec{H}^{\text{ext}} = (0, 0, H^{\text{ext}})$, the Lorentz force acts per unit length of the charged filament:

$$\vec{F} = \kappa \vec{E}^{\text{ext}} + \frac{\kappa}{c} \left[\vec{V} \times \vec{H}^{\text{ext}} \right] ,$$

where κ and $\vec{V}$ are the linear density of the electrical charge and velocity of the filament respectively. A similar force must act on the screw dislocation in the deformed crystal, in the given external fields $\vec{E}^{\text{ext}} = (\beta_{23}^{\text{ext}}, -\beta_{13}^{\text{ext}}, 0)$, $\vec{H}^{\text{ext}} = (0, 0, \partial_0 u^{\text{ext}}/c)$. In this case, the analog of the charge density is the quantity $\kappa = \mu b$. It is worth noting that the deformation rate of the body and the velocity of dislocation motion are usually small in comparison with the speed of sound c. Therefore, the second term in the expression for the force acting on the dislocation is always neglected:

$$\vec{F} \approx \kappa \vec{E}^{\text{ext}} = b \left(\sigma_{23}^{\text{ext}}, -\sigma_{13}^{\text{ext}}, 0 \right) . \tag{5.45}$$

Inside the body, the components of the external stress tensor (σ_{13}^{ext}, σ_{23}^{ext}) and distortion fields (β_{13}^{ext}, β_{23}^{ext}) are related by Hooke's law (5.36). Next, we derive a general expression for the force that acts per unit of length of not only the screw dislocation, but any under the external stress.

The equations of the screw dislocation like the equations of two-dimensional electrodynamics are invariant under Lorentz transformations. The limiting speed of mo-

tion of the screw dislocation is the speed of transverse sound c. As already mentioned earlier, when analyzed, the dislocation motion, due to its typical velocity and smallness, allows the relativistic effects to be ignored. Nevertheless, it is reasonable to assume that the intrinsic elastic energy per unit length of the screw dislocation (5.41) is equivalent to its mass, in accordance with the analog of the Einstein formula:

$$m = \frac{W_{\mathrm{D}}}{c^2} = \frac{\rho_0 b^2}{4\pi} \ln \frac{R}{r_0} .$$

Then, the equation of the screw dislocation dynamic under a given shear stress field coincides with the equation of motion of a nonrelativistic charged particle in two-dimensional electrodynamics:

$$m\left(\dot{V}_1, \dot{V}_2, 0\right) = \kappa \vec{E}^{\mathrm{ext}} = b\left(\sigma_{23}^{\mathrm{ext}}, -\sigma_{13}^{\mathrm{ext}}, 0\right) . \tag{5.46}$$

Attention should be paid to the fact that equation (5.46) contains only nondissipative forces of elastic origin. A real dislocation may be also subjected to additional forces, which are inelastic in nature.

The works [33] and [34] give a theory of the intended approach. In these works equations of motion of a dislocation loop, under both a field of given external stresses and intrinsic dislocation stresses, are obtained. The inertial properties of the dislocation loop are characterized by a nonlocal mass tensor. After the derivation of the equation of motion for a single dislocation, it is also not difficult to write down a closed system of equations in order to completely determine the evolution of dislocation loops and an elastic field in a solid. Unfortunately, this system is too complicated, and therefore, rarely used to solve practical problems at present.

The Green's Function Method in the Theory of Dislocations

In the case of an unbounded medium, the basic equations of the theory of dislocations may be solved in quadratures [33–38]. In this section effective formulas are obtained for elastic fields of all possible dislocation systems, which are convenient for comparing the predictions of the theory with experimental results.

To calculate the elastic fields of a moving dislocation, we introduce the dynamic Green function; a decaying at infinity solution of the singular equation:

$$\rho_0 \partial_0^2 G_{im}\left(\vec{x} - \vec{x}', t - t'\right) - c_{ijkl}\partial_j \partial_l G_{km}\left(\vec{x} - \vec{x}', t - t'\right) = \delta_{im}\delta^{(3)}(\vec{x} - \vec{x}')\delta^{(1)}\left(t - t'\right) .$$

This solution must satisfy the causality condition: $G_{im}(\vec{x} - \vec{x}', t - t') = 0$ for $t < t'$. Because of the properties of the homogeneity of time and the medium, the Green function depends on the differences of the arguments. Its explicit form is much easier to seek

by the Fourier transform method. The further analysis uses the representation of [38]:

$$4\pi\rho_0 G_{ij}(\vec{x},t) = \frac{1}{c_t^2 R}\delta_{ij}\delta^{(1)}\left(t-\frac{R}{c_t}\right) + \\ + \partial_i\partial_j\left\{\frac{1}{R}\left[\left(t-\frac{R}{c_l}\right)\Theta\left(t-\frac{R}{c_l}\right)-\left(t-\frac{R}{c_t}\right)\Theta\left(t-\frac{R}{c_t}\right)\right]\right\}, \quad (5.47)$$

$R = |\vec{x}|$, $\Theta(t)$ is the Heaviside step function:

$$\Theta(t) = \begin{cases} 1, & t > 0 \\ 0, & t < 0. \end{cases}$$

Using the Green function, after simple integrations, we find the solution of the system (5.32):

$$\beta_{nm} = \int\limits_{-\infty}^{\infty} dt' \int d^3\vec{x}'\varepsilon_{njh}c_{ijkl}\partial_l G_{km}\left(\vec{x}-\vec{x}',t-t'\right)D_{hi}\left(\vec{x}',t'\right) - \\ - \rho_0\int\limits_{-\infty}^{\infty} dt' \int d^3\vec{x}'\partial_0 G_{im}\left(\vec{x}-\vec{x}',t-t'\right)I_{ni}\left(\vec{x}',t'\right); \quad (5.48)$$

$$v_m = -\int\limits_{-\infty}^{\infty} dt' \int d^3\vec{x}' c_{ijkl}\partial_l G_{km}\left(\vec{x}-\vec{x}',t-t'\right)I_{ij}\left(\vec{x}',t'\right).$$

Here and below, the sequence order of the indices sometimes changes to account for the symmetry properties of the tensor (5.31).

When discussing acoustic radiation from a moving dislocation, the elastic modulus tensor (5.31) is conveniently expressed in terms of the velocities of the longitudinal $c_l = \sqrt{(\lambda+2\mu)/\rho_0}$ and transverse $c_t = \sqrt{\mu/\rho_0}$ sound waves:

$$c_{ijkl} = \rho_0\left(c_l^2 - 2c_t^2\right)\delta_{ij}\delta_{kl} + \rho_0 c_t^2\left(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}\right).$$

The delta function involving the expressions for the densities D_{ik} and I_{nk} (5.22), (5.29), the elastic fields (5.48) may be written in the form of contour integrals along the dislocation line $x_h'(\sigma,t')$:

$$\beta_{nm} = c_{ijkl}b_i\varepsilon_{njh}\int\limits_{-\infty}^{\infty} dt' \int\limits_{\sigma_1}^{\sigma_2} d\sigma\frac{\partial x_h'(\sigma,t')}{\partial\sigma}\partial_l G_{km}\left(\vec{x}-\vec{x}',t-t'\right) - \\ - \rho_0 b_i\varepsilon_{nlh}\int\limits_{-\infty}^{\infty} dt' \int\limits_{\sigma_1}^{\sigma_2} d\sigma\frac{\partial x_l'}{\partial\sigma}\frac{\partial x_h'}{\partial t'}\partial_0 G_{im}\left(\vec{x}-\vec{x}',t-t'\right); \quad (5.49)$$

$$v_m = -c_{ijkl}b_i\varepsilon_{jnh}\int\limits_{-\infty}^{\infty} dt' \int\limits_{\sigma_1}^{\sigma_2} d\sigma\frac{\partial x_n'}{\partial\sigma}\frac{\partial x_h'}{\partial t'}\partial_l G_{km}\left(\vec{x}-\vec{x}',t-t'\right).$$

Formulas (5.49) appear as integrals along the dislocation line and clearly illustrate the fact that the dislocation is a linear topological singularity of the elastic field of the crystal.

For a motionless dislocation, in formulas (5.48) and (5.49), we should put that $D_{hi} = D_{hi}(\vec{x}')$, $I_{ni} = 0$, $x'_h = x'_h(\sigma)$ and take into account that the integral

$$\int_{-\infty}^{\infty} \mathrm{d}t' G_{km}\left(\vec{x} - \vec{x}', t - t'\right) = G_{km}(\vec{x} - \vec{x}')$$

gives a static Green's function:

$$G_{ij}(\vec{x}, t) = \frac{1}{8\pi\mu}\left[\frac{2\delta_{ij}}{R} - \frac{\lambda + \mu}{\lambda + 2\mu}\partial_i\partial_j R\right] . \tag{5.50}$$

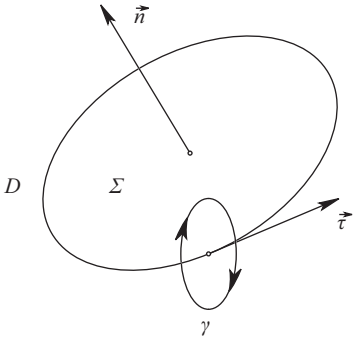


Fig. 5.16: Mutual orientation of the vectors $\vec{\tau}$, $\vec{n}$ and the traversing direction along the contour γ.

The theory of dislocations often applies an alternative definition of a dislocation, which is sometimes more convenient than the original one. A microscopic examination implies that the displacement field $\vec{u}(\vec{x}, t)$ of the dislocation loop undergoes a jump by a lattice constant along a part Σ of the slip surface bounded by the dislocation line. This circumstance makes it possible to define the dislocation by the relation:

$$\vec{u}_-(\vec{x}, t) - \vec{u}_+(\vec{x}, t) = \vec{b}\,; \quad \vec{x} \in \Sigma\,, \tag{5.51}$$

where $\vec{u}_+$ and $\vec{u}_-$ are values of $\vec{u}(\vec{x}, t)$ respectively on the upper and lower sides of the surface Σ. The "upward" direction (a positive direction) is given by the direction of the normal $\vec{n}$ toward the surface Σ. The directions of the vectors $\vec{n}$ and $\vec{\tau}$ are connected by the right screw rule (Figure 5.16). Rather than seeking an ambiguous displacement field, we will seek the single valued and continuous function $\vec{u}^G(\vec{x}, t)$ outside the cut of Σ, with a given jump (5.51) at the cut itself.

With the jump (5.51) happening, the distortion tensor of the field $\vec{u}^G(\vec{x}, t)$ must have a singularity on the surface. The singular part is written using the delta function:

$$\beta^P_{ik}(\vec{x}, t) = -\int_{\Sigma} \mathrm{d}S_i b_k \delta^{(3)}(\vec{x} - \vec{l})\,. \tag{5.52}$$

Here, the integration is over the surface Σ bounded by the dislocation line $\vec{l}(\sigma, t)$. The singular distortion β_{ik}^P simulates a plastic shear of the crystal by a lattice constant, and is not caused by elastic stresses. This approach yields the total distortion tensor of the medium as the sum:

$$\beta_{ik}^G \equiv \partial_i u_k^G = \beta_{ik} + \beta_{ik}^P \,. \tag{5.53}$$

Here, the elastic part β_{ik} of the distortion tensor is a single valued and continuous function of coordinates. It determines internal stresses in a crystal with a dislocation: $\sigma_{ij} = c_{ijkl}\beta_{lk}$. Therefore, the new notations bring the basic equation (5.29) of the elasticity theory to the form:

$$-\rho_0 \partial_0^2 u_i^G + c_{ijkl}\partial_j \beta_{lk} = 0 \,.$$

Hence, taking (5.53) into account, we obtain a closed equation for calculating the single valued but discontinuous function $\vec{u}^G(\vec{x}, t)$:

$$-\rho_0 \partial_0^2 u_i^G + c_{ijkl}\partial_j \partial_l u_k^G = c_{ijkl}\partial_j \beta_{lk}^P \,. \tag{5.54}$$

It is interesting to note that the plastic distortion (5.52) is related to the dislocation density tensor (5.22) by:

$$D_{pk} = -\varepsilon_{pmn}\partial_m \beta_{nk}^P \,. \tag{5.55}$$

Formula (5.55) expresses the well known Stokes theorem.

Relation (5.55) can also be regarded as the result of integration of the second continuity equation (5.27) over time:

$$\partial_0 D_{pk} + \varepsilon_{pmn}\partial_m I_{nk} = 0 \,.$$

From here, we receive another useful relation at once:

$$\beta_{nk}^P(\vec{x}, t) = \int_{-\infty}^{t} \mathrm{d}t' I_{nk}\left(\vec{x}, t'\right) \,. \tag{5.56}$$

After traveling through the dislocation between two material points initially displaced by the vector $(\mathrm{d}x_1, \mathrm{d}x_2, \mathrm{d}x_3)$, these are shifted along the coordinate axes by the segments $\mathrm{d}s_k^P = \beta_{ik}^P \mathrm{d}x_i$. Therefore, the quantity β_{ik}^P is said to describe the kinematics of plastic deformation. According to (5.4), the local volume changes as the body deforms are governed by the quantity $\partial_i u_i^G = \beta_{ii} + \beta_{ii}^P \neq 0$. Relation (5.56) shows that there are no volume changes due to the plastic deformation of the crystal when the dislocation moves conservatively ($I_{ii} = 0$): $\beta_{ii}^P = 0$.

The solution of equation (5.54) for the displacement field can easily be found using the Green functions (5.47) and (5.50). After simple integrations, we obtain:

$$u_m^G = -\int_{-\infty}^{\infty} \mathrm{d}t' \int \mathrm{d}^3\vec{x}' c_{ijkl}\partial_l G_{km}\left(\vec{x} - \vec{x}', t - t'\right)\beta_{ji}^P\left(\vec{x}', t'\right) \,.$$

Or, subject to (5.52), we get the more detailed:

$$u_m^G = -\int_{-\infty}^{\infty} \mathrm{d}t' \int_{\Sigma(t')} \mathrm{d}S'_j c_{ijkl} \partial_l G_{km} \left(\vec{x} - \vec{x}', t - t'\right) b_i \,. \tag{5.57}$$

For a stationary dislocation, the surface Σ is independent of t.

An Edge Dislocation

As an example, let us consider the problem of calculating the elastic fields of a rectilinear edge dislocation. We choose the axis Oz along a dislocation line, and the axis Ox along the Burgers vector: $\vec{b} = (b, 0, 0)$. It follows from the symmetry of the problem that the vector $\vec{u}^G$ lies in the plane Oxy and is independent of x_3. After simple, but tedious, computations we find from (5.57) that:

$$u_1^G = \frac{b}{2\pi}\left[\operatorname{arctg}\frac{y}{x} + \frac{1}{2(1-\nu)}\frac{xy}{r^2}\right]; \quad u_2^G = \frac{b}{4\pi(1-\nu)}\left[(1-2\nu)\ln\frac{r}{R} + \frac{x^2}{r^2}\right], \tag{5.58}$$

where $x_1 = x$, $x_2 = y$, $r^2 = x^2 + y^2$; $\nu = \lambda[2(\lambda+\mu)]^{-1}$ is the Poisson coefficient. For all materials, it ranges within $0 \le \nu \le 1/2$. While regularizing the logarithmically divergent integral in the second formula of (5.58), we have entered the cutoff parameter R (the size of the crystal).

The components of the stress tensor have the form:

$$\begin{aligned} \sigma_{11} &= -D\frac{y(3x^2+y^2)}{r^4}, \quad & \sigma_{22} &= D\frac{y(x^2-y^2)}{r^4}, \\ \sigma_{12} &= D\frac{x(x^2-y^2)}{r^4}, & \sigma_{33} &= \nu(\sigma_{11}+\sigma_{22}), \end{aligned} \tag{5.59}$$

where $D = \mu b[2\pi(1-\nu)]^{-1}$. In contrast to the stresses of a screw dislocation, the stresses (5.59) are anisotropic in the Oxy-plane.

The average hydrostatic pressure created in the medium by the edge dislocation

$$p = -\frac{1}{3}\sigma_{kk} = \frac{2}{3}(1+\nu)D\frac{y}{r^2}$$

has a simple physical meaning. Above the Ozx-plane ($y > 0$) where there is an "extra" inserted half-plane in the crystal (Figure 5.1), the medium is compressed and, therefore, $p > 0$. Beneath the Ozx-plane ($y < 0$), we have $p < 0$ (the medium is stretched in all directions).

Dislocation Systems

We will now write down the formulas for the tensors (5.22) and (5.29) of a single dislocation in a form convenient for generalizations. Suppose the coordinates in these

formulas are in a small neighbourhood of the point $\vec{l}(\sigma_0, t)$ of the dislocation line. We introduce a local Cartesian coordinate system in this region, pointing one of its axis along the dislocation line. Then

$$\delta^{(3)}\left(\vec{x} - \vec{l}\,(\sigma, t)\right) = \delta^{(1)}\left[(\sigma - \sigma_0)\left|\partial\vec{l}/\partial\sigma_0\right|\right]\delta^{(2)}\left(\vec{\xi}\right) ,$$

where $\vec{\xi}$ is a two-dimensional radius vector measured from the dislocation axis in a plane perpendicular to the tangent vector to the dislocation line:

$$\vec{\tau}\,(\sigma_0, t) = \frac{\partial\vec{l}}{\partial\sigma_0}\left|\frac{\partial\vec{l}}{\partial\sigma_0}\right|^{-1} .$$

After integration, the tensor (5.22) acquires the simple form of:

$$D_{ik} = \tau_i b_k \delta^{(2)}\left(\vec{\xi}\right) . \tag{5.60}$$

The relation (5.29) admits an analogous expression:

$$I_{nk} = \varepsilon_{nsr}\tau_s b_k V_r \delta^{(2)}\left(\xi\right) , \tag{5.61}$$

where $V_r = \partial l_r(\sigma_0, t)/\partial t$ is the velocity of the dislocation length element.

In the case of continuous distribution of the dislocations, instead of (5.60) and (5.61), we can enter the tensors:

$$D_{ik}(\vec{x}, t) = \tau_i b_k n , \quad I_{nk}(\vec{x}, t) = \varepsilon_{nsr}\tau_s b_k V_r n , \tag{5.62}$$

where n is the number of dislocations in the direction of the vector $\vec{\tau}$ (each dislocation has the Burgers vector $\vec{b}$), intersecting a unit area perpendicular to $\vec{\tau}$ at a given point. If there are sets of dislocations with other n, $\vec{b}$, $\vec{\tau}$, the total dislocation density and the total dislocation flux density are obtained by summing over all the sets.

The tensors (5.62) are connected by the continuity condition:

$$\partial_0 D_{ik} + \varepsilon_{ilm}\partial_l I_{mk} = 0 , \tag{5.63}$$

which expresses the law of conservation of the Burgers vector in a medium. To show it, let us integrate the relation (5.63) over a surface spanned by a closed line L. Introducing the complete Burgers vector $\vec{B}$ of dislocations covered by the line L, and using the Stokes theorem, we come to:

$$\frac{\partial B_k}{\partial t} = -\oint_L \mathrm{d}x_i I_{ik} .$$

It is clear to see from this equality that the integral on the right-hand side determines the magnitude of the Burgers vector carried away by dislocations that cross the line L. The element I_{ik} governs the flux of the k-th component of the Burgers vector through a unit area parallel to the i-th Cartesian axis.

The basic equations of the dislocation theory are linear. Therefore, the elastic fields of the dislocation system represents a superposition of fields from individual dislocations of the system. After redefinition of the tensors D_{ik}, I_{ik}, the formulas for calculating the elastic fields of the dislocation system retain the form (5.48).

5.4 Interaction of Dislocations with a Stress Field

Consider a dislocation loop D in a field of external (with respect to the dislocation) elastic stresses σ_{ik}^{ext} and find the force acting on the dislocation [29]. To do this, the work δA spent by the external forces upon a small displacement of the loop should be calculated. One can represent the work in the form:

$$\delta A_D = \oint_D F_i \delta X_i \mathrm{d}l \,, \tag{5.64}$$

where δX_i is the displacement of an element of a dislocation line and F_i are the components of the force the elastic stresses per unit dislocation length act with.

Suppose that the dislocation displacement generates a change δu_i in the displacement field. The work performed by the external stresses in a volume V_0 of the body, due to the dislocation displacement, being:

$$\delta A = \int_S \sigma_{ik}^{\text{ext}} \delta u_k \mathrm{d}S_i \,, \tag{5.65}$$

where S is the surface bounding the volume V_0 of the body.

In the given case, it is convenient to use transformations of the integral (5.65) to apply the Gauss theorem. Therefore, the displacement field $\vec{u}$ around the dislocation should be regarded as a single valued function of coordinates. With this approach, the field $\vec{u}$ has a discontinuity along the surface Σ spanned by the dislocation loop. To exclude discontinuity points, we surround the loop with a closed surface $\widetilde{S}$, outside which (in a volume V') the function $\vec{u}(\vec{x}, t)$ is continuous. Using the Gauss theorem, we find:

$$\begin{aligned}\delta A &= \int_{V'} \partial_l \left(\sigma_{ik}^{\text{ext}} \delta u_k\right) \mathrm{d}^3\vec{x} - \oint_{\widetilde{S}} \sigma_{ik}^{\text{ext}} \delta u_k \mathrm{d}S_i = \\ &= \int_{V'} \left(\partial_l \sigma_{ik}^{\text{ext}}\right) \delta u_k \mathrm{d}^3\vec{x} + \int_{V'} \sigma_{ik}^{\text{ext}} \delta \varepsilon_{ik} \mathrm{d}^3\vec{x} - \oint_{\widetilde{S}} \sigma_{ik}^{\text{ext}} \delta u_k \mathrm{d}S_i \,. \end{aligned} \tag{5.66}$$

In the second term, given the symmetry property $\sigma_{ik}^{\text{ext}} = \sigma_{ki}^{\text{ext}}$, we replace the distortion field $\partial_i u_k$ by the deformation field: $\varepsilon_{ik} = (\partial_i u_k + \partial_k u_i)/2$. As a normal to the surface S_0, the external normal with respect to the volume V' is taken.

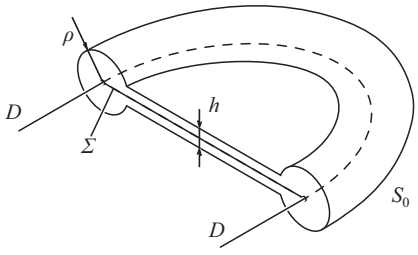


Fig. 5.17: Scheme of a surface S_0 and a cut Σ.

Inside the crystal, the external stress field satisfies the equilibrium conditions: $\partial_i \sigma_{ik}^{\text{ext}} = 0$. Therefore, the first integral on the right-hand side of (5.66) disappears. To calculate the last two integrals, we choose a surface that passes along the upper and lower banks of the cut Σ (with a gap h), and which contains a thin tube S_0 of radius ρ around the dislocation line D (scheme of the surface S_0 is shown in Figure 5.17).

When the cut edges come closer to each other ($h \to 0$) and the tube radius decreases ($\rho \to 0$), the integral over the volume V' (because of the continuity of the fields σ_{ik}^{ext} and ε_{ik}) turns into an integral over the volume V_0:

$$\int\limits_{V'} \sigma_{ik}^{\text{ext}} \delta\varepsilon_{ik} \mathrm{d}^3\vec{x} \to \int\limits_{V_0} \sigma_{ik}^{\text{ext}} \delta\varepsilon_{ik} \mathrm{d}^3\vec{x} \,.$$

The surface integral in (5.66) is also simplified. First of all, we should note that the integral over the tube surface vanishes as $\rho \to 0$, since the dislocation displacement field possesses the following property: $\lim_{\rho\to 0} \rho u_i = 0$. On the banks of the cut Σ still remaining in the integral, the values of the continuous functions σ_{ik}^{ext} in the limit are the same, and the limiting values of $\vec{u}$ are different in a constant value $\vec{b}$. Therefore, instead of (5.66), we obtain:

$$\delta A = \int\limits_{V_0} \widetilde{\sigma}_{ik}^{\text{ext}} \delta\widetilde{\varepsilon}_{ik} \mathrm{d}^3\vec{x} + \frac{1}{3} \int\limits_{V_0} \sigma_{ll}^{\text{ext}} \delta\varepsilon_{kk} \mathrm{d}^3\vec{x} + b_k \delta \int\limits_{\Sigma} \sigma_{ik}^{\text{ext}} \mathrm{d}\Sigma_i \,, \tag{5.67}$$

where $\widetilde{\sigma}_{ik}^{\text{ext}}$ and $\widetilde{\varepsilon}_{ik}$ are the deviators of stresses and deformations. For example, $\widetilde{\varepsilon}_{ik} = \varepsilon_{ik} - \delta_{ik}\varepsilon_{ll}/3$. Based on the obvious identity

$$\sigma_{ik}^{\text{ext}} \delta\varepsilon_{ik} = \left(\sigma_{ik}^{\text{ext}} - \frac{1}{3}\delta_{ik}\sigma_{ll}^{\text{ext}}\right) \delta\varepsilon_{ik} + \frac{1}{3}\sigma_{ll}^{\text{ext}} \delta\varepsilon_{kk} = \widetilde{\sigma}_{ik}^{\text{ext}} \delta\widetilde{\varepsilon}_{ik} + \frac{1}{3}\sigma_{ll}^{\text{ext}} \delta\varepsilon_{kk} \,,$$

the first two terms in (5.67) replace the integral over the volume in (5.66).

In the last term on the right-hand side of (5.67), the symbol of the variation δ is taken outside the integral sign since the stress distribution σ_{ik}^{ext} is assumed to be given.

If the element of the dislocation line $\mathrm{d}l$ is displaced by $\delta\vec{X}$, the area of the surface Σ changes by the magnitude:

$$\delta\Sigma_k = \varepsilon_{ksp} \delta X_s \tau_p \mathrm{d}l \,. \tag{5.68}$$

Relation (5.68) should be used when transforming the last integral into (5.67). In a case when the displacement $\delta\vec{X}$ lies in the dislocation slip plane, expression (5.68) completely characterizes the changes occurring in the crystal. However, if the displacement is perpendicular to the slip plane, additional conditions that arise from the medium continuity become important. The displacement of the dislocation in a direction perpendicular to the plane of its slip causes a local change ($\delta\varepsilon_{kk}\mathrm{d}^3\vec{x}$) in the volume (5.4), the value of which is evaluated purely geometrically:

$$\delta\varepsilon_{kk} \mathrm{d}^3\vec{x} = \varepsilon_{imn} \tau_m b_n \delta X_i \mathrm{d}l \,. \tag{5.69}$$

Taking into account (5.68) and (5.69), we separate out δA in the work, from the part that is responsible for inelastic deformation of the medium as the dislocation displacement migrates:

$$\delta A = \int\limits_{V_0} \widetilde{\sigma}_{ik}^{\text{ext}} \delta \widetilde{\varepsilon}_{ik} \mathrm{d}^3 \vec{x} + \oint\limits_D \varepsilon_{imn} \tau_n \left(\sigma_{mk}^{\text{ext}} - \frac{1}{3} \sigma_{ll}^{\text{ext}} \delta_{mk} \right) b_k \delta X_i \mathrm{d}l \,. \tag{5.70}$$

The first integral in (5.70) is equal to an increase in the elastic energy of the medium in the volume of the body. The linear integral over the dislocation loop

$$\delta A_D = \oint\limits_D \varepsilon_{imn} \tau_n \left(\sigma_{mk}^{\text{ext}} - \frac{1}{3} \sigma_{ll}^{\text{ext}} \delta_{mk} \right) b_k \delta X_i \mathrm{d}l \tag{5.71}$$

determines the work done for the dislocation displacement. Comparing (5.71) with (5.64), we obtain an expression for the force acting per unit of dislocation length:

$$F_i = \varepsilon_{imn} \tau_n \widetilde{\sigma}_{mk}^{\text{ext}} b_k \,. \tag{5.72}$$

A formula like (5.72), where the stress deviator $\widetilde{\sigma}_{mk}^{\text{ext}}$ is replaced by the stress tensor σ_{mk}^{ext}, was first obtained by Peach and Koehler [39]. The necessity of taking the inelastic change in volume into account when moving the dislocation, i.e., the necessity to substitute $\sigma_{mk}^{\text{ext}} - \sigma_{ll}^{\text{ext}} \delta_{mk}/3$ for σ_{mk}^{ext}, has been indicated by Weertman [40]. Such a replacement, in particular, allows us to make a conclusion that the external hydrostatic pressure $\sigma_{mk}^{\text{ext}} = p\delta_{mk}$ has no influence on the dislocation.

For the parallel Oz-axis of a screw dislocation with the Burgers vector $\vec{b}(0, 0, b)$, formula (5.72) reproduces expression (5.45) found earlier:

$$F_1 = b\sigma_{23}^{\text{ext}} \,, \quad F_2 = -b\sigma_{13}^{\text{ext}} \,. \tag{5.73}$$

For the parallel Oz-axis of an edge dislocation with the Burgers vector $\vec{b} = (0, 0, b)$, the nonzero components of the force are equal to:

$$F_1 = b\sigma_{12}^{\text{ext}} \,, \quad F_2 = -b\widetilde{\sigma}_{11}^{\text{ext}} \,. \tag{5.74}$$

Polygonization: A Dislocation Model of a Grain Boundary

An analysis of the interaction of two rectilinear dislocations leads to a number of interesting conclusions. Consider first the case of two screw dislocations. Suppose that a force acting per unit of length of a screw dislocation was caused by another dislocation. This force is determined by formulas (5.73) and (5.37). It has a radial direction:

$$\vec{F} = \frac{\mu b_1 b_2 \vec{r}}{2\pi r^2} \,,$$

where $\vec{r} = (x, y)$ is the radius vector joining the dislocation centers in the Oxy-plane. As charged filaments, dislocations of the same sign ($b_1 b_2 > 0$) repel, and dislocations of different signs ($b_1 b_2 < 0$) attract.

The force of the interaction of two edge dislocations is not radial, but is characterized by a significant angular dependence. Imagine that slip planes of the dislocations are parallel. We will choose the Oxz-plane parallel to the slip plane and direct the Oz-axis parallel to the dislocation lines. Then the force can be calculated directly using formulas (5.74) and (5.59). If a dislocation coincides with the Oz-axis (Figure 5.18), it acts on another dislocation passing through the point $x = r\cos\theta$, $y = r\sin\theta$ in the Oxy-plane with a force whose projection onto the slip plane has the form:

$$F_1 = \frac{\mu b_1 b_2 \cos\theta \cos 2\theta}{2\pi(1-\nu)r}. \tag{5.75}$$

This projection of the force is of great interest since the dislocation can mechanically displace only in the slip plane. The disappearance of the projection F_1 corresponds to a configuration of two dislocations, equilibrium with respect to the slip of the configuration. From formula (5.75), it follows that $F_1 = 0$ for $\theta = \pi/4$ and $\theta = \pi/2$. It is easy to verify that the first of the variants (Figure 5.19 (a)) corresponds to the condition of stable equilibrium of two dislocations of opposite signs ($b_1 b_2 < 0$). The result explains the formation of *dipoles from* edge *dislocations*. The second variant (Figure 5.19 (b)) corresponds to the condition of stable equilibrium of two dislocations of the same sign ($b_1 b_2 > 0$). Straight line edge dislocations with identical Burgers vectors lying in parallel slip planes have a tendency to collect in one plane perpendicular to their slip planes.

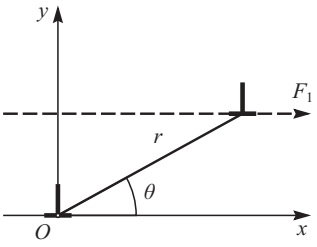


Fig. 5.18: A component of the interaction force of edge dislocations, parallel to a slip plane.

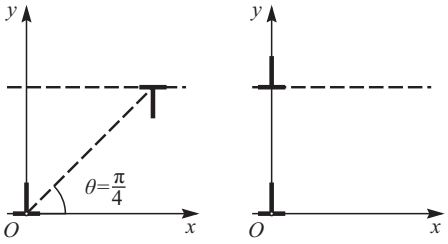


Fig. 5.19: A stable configuration of two edge dislocations of opposite signs (a) and the same signs (b).

The set of a large number of parallel edge dislocations located in a plane perpendicular to their Burgers vectors is called a *dislocation wall*. The ability of dislocations to align in the walls is referred to as *polygonization*.

Now, we will calculate shear stresses created by an infinitely extended dislocation wall. The element σ_{12} of the stress tensor produced by all wall dislocations is given by the sum:

$$\sigma_{12} = Dx \sum_{n=-\infty}^{\infty} \frac{x^2 - (y - nw)^2}{\left[x^2 + (y - nw)^2\right]^2} ,$$

where $w > b$ is the distance between dislocations in the wall. For $x > w$, this expression reduces to the form [41]:

$$\sigma_{12} \approx 4\pi^2 D \frac{x}{w^2} \exp\left(-\frac{2\pi x}{w}\right) \cos\frac{2\pi y}{w} .$$

Thus, when moving away from the wall, the stresses decrease exponentially. Against the field of a single dislocation, the elastic field of the wall is concentrated in a layer of thickness in the order of the distance between dislocations.

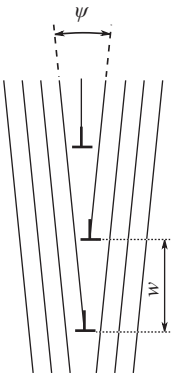


Fig. 5.20: Dislocation model of the boundary between crystallites.

A dislocation wall, if any, leads to a misorientation between two parts of the crystal by the angle $\psi \approx b/w$ (Figure 5.20). Hence, it follows that the dislocation wall is a model of the boundary of two blocks, or subgrains, of a crystal with a small misorientation. The formation of a regular boundary between crystallites includes both the slip of dislocations near the boundary (often over long distances) and their creep along the normal to the slip plane over small distances, which is required by the condition for a uniform dislocation distribution along the boundary.

5.5 An Expansion in Multipole Moments of Fields Created by Dislocation Systems

There are many experimental studies that evidence the existence of intense sound emission when plastically deforming crystals; it is caused by migrating dislocations and various dislocation reactions. The research results allow the acoustic emissions to be seen as a promising, nondestructive testing and control method.

Elastic fields produced by moving dislocations are represented in quadratures (5.48). Therefore, to compare the predictions of the theory with the experimental data, it is necessary to evaluate appropriate integrals. In most cases, this cannot be done because of the complexity of the integrands. Therefore, attempts have been made to calculate the asymptotics of the elastic fields for small and large distances r from the dislocation system, under limitations of the size d of the region containing the dislocations and the radiation wavelength λ. In the dipole approximation, the sound radiation of a number of dislocation systems has been studied by [31, 35, 42–45]. However, the dipole approximation does not provide a complete understanding of the acoustic emissions of dislocations. For example, in the case of a closed moving dislocation loop of unchanged dimensions and shape, there is no dipole radiation [45]. Getting higher approximations by expanding exact expressions for the elastic fields over the small parameters d/λ, $\lambda/r < 1$ faces serious computational difficulties. In addition, solving a wide range of problems in the theory of strength and plasticity requires more detail, not limited to the radiation zone information on fields created by arbitrarily moving dislocations. Obtaining such information is an extremely difficult task, however.

A systematic analysis of the elastic fields from dislocations may be carried out by expanding Green's function of a problem at hand, in a system of orthogonal functions that take into account the symmetry of the field source distribution [46]. Such a technique has been shown [47] to make it possible to write down any term in the expansion of the elastic fields without resorting to cumbersome calculations. The expansion itself is valid, both for small, large, or intermediate distances from the dislocation system, for any size of the region that contains the dislocations, and for an arbitrary variation frequency of the fields. Let us present the results of this approach.

For further analysis, it is useful to write the elastic fields $\beta_{nm}(\vec{x}, t)$, $v_m(\vec{x}, t)$ and their sources $D_{hi}(\vec{x}, t)$, $I(\vec{x}, t)$ in the form of Fourier integrals in time. For example:

$$v_m(\vec{x}, t) = \int_{-\infty}^{\infty} \mathrm{d}\omega e^{-\mathrm{i}\omega t} v_m(\vec{x}, \omega) \ .$$

The integral representations of the elastic fields (5.48) have the form of convolutions. Therefore, after the Fourier transform, the nonlocal time coupling between the elastic fields and their sources is replaced by a local connection between the corresponding Fourier images. Formulas for the Fourier components of the elastic field are expressed in the form of derivatives of a small number of fundamental integrals over spatial co-

ordinates:

$$v_m(\vec{x},\omega) = -\sum_{\alpha=a,b} \widehat{N}^{(\alpha)}_{mfj}(\omega,\nabla)\int \mathrm{d}^3\vec{x}\,' I_{(fj)}(\vec{x}\,',\omega)\,\frac{e^{\mathrm{i}k_\alpha|\vec{x}-\vec{x}\,'|}}{|\vec{x}-\vec{x}\,'|}\;;$$

$$\beta_{nm}(\vec{x},\omega) = \sum_{\alpha=a,b}\left\{\widehat{N}^{(\alpha)}_{mfj}(\omega,\nabla)\int \mathrm{d}^3\vec{x}\,' P_{n(fj)}(\vec{x}\,',\omega)\,\frac{e^{\mathrm{i}k_\alpha|\vec{x}-\vec{x}\,'|}}{|\vec{x}-\vec{x}\,'|} + \right. \tag{5.76}$$

$$\left. +\mathrm{i}\omega\widehat{M}^{(\alpha)}_{mj}(\omega,\nabla)\int \mathrm{d}^3\vec{x}\,' I_{nj}(\vec{x}\,',\omega)\,\frac{e^{\mathrm{i}k_\alpha|\vec{x}-\vec{x}\,'|}}{|\vec{x}-\vec{x}\,'|}\right\}\;,$$

where $k_a = \omega/c_\mathrm{l}$, $k_b = \omega/c_\mathrm{t}$, and the operators $\widehat{N}^{(\alpha)}_{mfj}$ and $\widehat{M}^{(\alpha)}_{mj}$ have the form:

$$\widehat{N}^{(a)}_{mij}(\omega,\nabla) = \frac{1}{2\pi}\left[\left(\frac{1}{2}-\frac{c_\mathrm{t}^2}{c_\mathrm{l}^2}\right)\delta_{ij}\partial_m - \frac{c_\mathrm{t}^2}{\omega^2}\partial_m\partial_i\partial_j\right];\quad \widehat{N}^{(b)}_{mij}(\omega,\nabla) = 2c_\mathrm{t}^2\partial_j\widehat{M}^{(b)}_{mi}(\omega,\nabla)\;;$$

$$M^{(a)}_{mj}(\omega,\nabla) = -\frac{1}{4\pi\omega^2}\partial_m\partial_j\;;\quad M^{(b)}_{mj}(\omega,\nabla) = \frac{1}{4\pi}\left[\frac{\delta_{mj}}{c_\mathrm{t}^2}+\frac{1}{\omega^2}\partial_m\partial_i\right].$$

The Fourier images of the sources $I_{fj}(\vec{x},\omega)$ and $P_{nij}(\vec{x},\omega) = \varepsilon_{njh}D_{hj}(\vec{x},\omega)$ are combined in

$$I_{(fj)} = \frac{1}{2}\left(I_{fj}+I_{jf}\right)\;;\quad P_{n(ij)} = \frac{1}{2}\left(P_{nij}+P_{nji}\right)\;.$$

In obtaining formulas (5.76), we have taken into account the relation:

$$\left(\Delta+k^2\right)\frac{e^{\mathrm{i}k|\vec{x}-\vec{x}\,'|}}{|\vec{x}-\vec{x}\,'|} = -4\pi\delta^{(3)}(\vec{x}-\vec{x}\,')\;,$$

where Δ is the three-dimensional Laplace operator.

Because of the connection (5.27), the tensors $D_{fj}(\vec{x},\omega)$ and $I_{fj}(\vec{x},\omega)$ are not independent:

$$D_{fj}(\vec{x},\omega) = \frac{1}{\mathrm{i}\omega}\varepsilon_{fkl}\partial_k I_{lj}(\vec{x},\omega)\;,\quad \omega\neq 0\;. \tag{5.77}$$

A consequence of the equality turns out to be the connection of the Fourier images $\beta_{nk}(\vec{x},\omega)$ and $v_k(\vec{x},\omega)$ for $\omega\neq 0$:

$$\beta_{nk}(\vec{x},\omega) = -\frac{1}{\mathrm{i}\omega}\left[\partial_n v_k(\vec{x},\omega) - I_{nk}(\vec{x},\omega)\right]\;. \tag{5.78}$$

Formulas (5.76) to (5.78) are basic for further calculations.

A System of Rectilinear Dislocations

We make use of the linearity of the basic equations and, in particular, consider systems of screw and edge dislocations. The dislocations are assumed to be moving in their slip planes. Hence, the condition $I_{kk} = 0$ is fulfilled. Suppose that a coordinate system is

chosen so that the dislocation lines is orthogonal to the plane $x_3 = 0$. Because of the symmetry of the problem, the elastic fields and their sources depend only on three variables: x_1, x_2, t. Therefore, in (5.76), we can calculate the integral over x_3:

$$\int\limits_{-\infty}^{\infty} \mathrm{d}x_3' \frac{e^{\mathrm{i}k_\alpha|\vec{x}-\vec{x}'|}}{|\vec{x}-\vec{x}'|} = \pi \mathrm{i} H_0^{(1)}\left(k_\alpha \left|\vec{r}-\vec{r}'\right|\right) ,$$

where $\vec{r} = (x_1, x_2)$ is a two-dimensional vector in the Oxy-plane, and $0 \le \arg k_\alpha \le \pi$, $H_0^{(1)}(x)$ is the Bessel function of a third kind [48].

For a system of screw dislocations, the densities of the Burgers vector and its flux in the Oxy-plane have the form:

$$\rho(\vec{r}, t) = n(\vec{r}, t) b_3 , \quad \vec{i}(\vec{r}, t) = n(\vec{r}, t) b_3 \vec{V} .$$

Here, $n(\vec{r}, t)$ is the average number of dislocations with the Burgers vector $\vec{b} = (0, 0, b_3)$ and velocity $\vec{V} = (V_1(t), V_2(t))$ in the Oxy-plane, which intersects with the unit area of the plane near a point with the radius vector $\vec{r}$. According to (5.77), their Fourier components are related by

$$-\mathrm{i}\omega\rho(\vec{r}, \omega) + \operatorname{div}\vec{i}(\vec{r}, \omega) = 0 . \tag{5.79}$$

Equation (5.79) is used for transforming the coefficients of asymptotic expansions.

The nonzero components of the elastic fields (5.76) have the form:

$$(\beta_{23}(\vec{r}, \omega), -\beta_{13}(\vec{r}, \omega)) = -\vec{\nabla}A_0 + \mathrm{i}k_b\vec{A} ; \quad v_3(\vec{r}, \omega) = c_\mathrm{t}(\partial_1 A_2 - \partial_2 A_1) , \tag{5.80}$$

where $k_b = \omega/c_\mathrm{t}$; $A_0 = A_0(\vec{r}, \omega)$ and $\vec{A}(\vec{r}, \omega) = (A_1(\vec{r}, \omega), A(\vec{r}, \omega))$ are the scalar and vector potentials in the problem:

$$\left(A_0, \vec{A}\right) = \frac{\mathrm{i}}{4}\int \mathrm{d}^2\vec{r}' H_0^{(1)}\left(k_b\left|\vec{r}-\vec{r}'\right|\right)\left(\rho\left(\vec{r}', \omega\right), \frac{1}{c_\mathrm{t}}\vec{i}\left(\vec{r}', \omega\right)\right) . \tag{5.81}$$

Taking equations (5.79) and (5.81) into account, it is easy to verify that the components of the trivector $(A_0, \vec{A})$ satisfy the equations

$$-\left(\Delta + k_b^2\right)\left(A_0, \vec{A}\right) = \left(\rho, \vec{i}/c_\mathrm{t}\right) ; \quad -\mathrm{i}k_b A_0 + \operatorname{div}\vec{A} = 0 .$$

The latter coincide in form with the equations for the Fourier components of the potentials of a system of uniformly charged parallel filaments in electrodynamics [46].

Here, Δ is the two-dimensional Laplace operator. For $\omega \neq 0$, these equations allow us to express the distortion field only through the vector potential $\vec{A}$ and write down the result as (5.78):

$$(\beta_{23}, -\beta_{13}) = -\frac{1}{\mathrm{i}k_b}\vec{\nabla}\operatorname{div}\vec{A} + \mathrm{i}k_b\vec{A} = -\frac{1}{\mathrm{i}\omega}\left[\partial_2 v_3 - i_1, -\partial_1 v_3 - i_2\right] . \tag{5.82}$$

Suppose that d is the characteristic size of the region that contains the dislocations, and the origin of the coordinate system was selected somewhere inside the dislocation distribution. We can study the elastic fields outside the region. We represent the quantity $|\vec{r} - \vec{r}'|$ through the polar coordinates in the plane $x_3 = 0$ ($r' < d$, $r > d$):

$$\left|\vec{r} - \vec{r}'\right| = \sqrt{r^2 + r'^2 - 2rr' \cos(\psi - \psi')} .$$

For the expansion of the fields $\beta_{nm}(\vec{x}, \omega)$, $v_m(\vec{x}, \omega)$ into a series in multipole moments, we use the addition theorem for cylindrical functions [48]:

$$H_0^{(1)}(M) = \sum_{m=-\infty}^{+\infty} J_m(p) H_m^{(1)}(s) e^{im(\psi-\psi')} , \tag{5.83}$$

where $M = \sqrt{p^2 + s^2 - 2ps\cos(\psi - \psi')}$, and $J_m(p)$ and $H_m^{(1)}(s)$ are the Bessel functions of the first and third kind, respectively. Using (5.83) from (5.80) to (5.82), we find an asymptotic series for the elastic fields (for $\omega \neq 0$) in a region without dislocations:

$$v_3 = \frac{ic_t}{4} \sum_{n=-\infty}^{+\infty} \left(b_1^{(n)}\partial_2 - b_2^{(n)}\partial_1\right) H_n^{(1)}(k_b r)\, e^{in\psi} ; \quad \beta_{k3} = -\frac{\partial_k v_3}{i\omega}, \; k = 1, 2 . \tag{5.84}$$

It should be emphasized that the formulas for the coefficients of the expansion

$$\vec{b}^{(n)}(\omega) = \left(b_1^{(n)}, b_2^{(n)}\right) = \int d^2\vec{r} e^{-in\psi} J_n(k_b r)\, \vec{i}(\vec{r}, \omega)$$

impose no restrictions on the frequency ω and size of the region d, which contains dislocations.

To compare the approach set forth above with the results of other works, we take the traditional limitation of these works. Suppose that the wavelength of the acoustic oscillations was much larger than the dimensions of the dislocation system: $d \ll k_b^{-1}$; the expressions for the multipole coefficients would become simpler:

$$\begin{aligned} \vec{b}^{(m)} &\approx \vec{\beta}^{(m)} \frac{1}{m!}\left(\frac{k_b}{2}\right)^m , \qquad \vec{\beta}^{(m)} = \int d^2\vec{r}\, \vec{i}(\vec{r}, \omega)\, r^m e^{-im\psi} ; \\ \vec{b}^{(-m)} &\approx \vec{\beta}^{(m)+} \frac{(-1)^m}{m!}\left(\frac{k_b}{2}\right)^m , \quad \vec{\beta}^{(m)+} = \int d^2\vec{r}\, \vec{i}(\vec{r}, \omega)\, r^m e^{im\psi} ; \quad m \geq 0 . \end{aligned} \tag{5.85}$$

In the particular case of one oscillating screw dislocation, the first term in the expansion (5.84) in the approximation (5.85) reproduces the result of [31].

Formulas (5.84) and (5.85) give a complete picture of the multipole radiation of screw dislocations in the wave zone distinguished by the condition: $d \ll k_b^{-1} \ll r$. After simple transformations and integrations by parts, the asymptotic series (given

that (5.79) holds) reduces to the form:

$$(\beta_{23}, -\beta_{13}) = \frac{ik_b}{2}\sqrt{\frac{k_b}{2\pi r}}\exp\left(ik_b r - \frac{i\pi}{4}\right)\left[\vec{p} - \vec{n}\,(\vec{n}\cdot\vec{p}) + m\,(-n_2, n_1) - \right.$$
$$\left. -\frac{ik_b}{4}\left(\vec{Q} - \vec{n}\left(\vec{n}\cdot\vec{Q}\right)\right) + \cdots\right];$$
$$v_3 = \frac{ik_b c_t}{2}\sqrt{\frac{k_b}{2\pi r}}\exp\left(ik_b r - \frac{i\pi}{4}\right)\left[n_1 p_2 - n_2 p_1 + m - \frac{ik_b}{4}\left(n_1 Q_2 - n_2 Q_1\right) + \cdots\right], \tag{5.86}$$

where $\vec{n} = \vec{r}/r$, $Q_j = K_{ji}n_i$; the lower spatial indices i, j run through the values 1, 2; the summation is performed over twice-occurring indices. The dipole ($\vec{p}$) and quadrupole (K_{ij}) moments of the dislocation system, as well as the analog of the magnetic dipole moment m of electrodynamics [46], have the form:

$$\vec{p} = \int d^2\vec{r}\,\rho\,(\vec{r}, \omega)\,\vec{r}\,; \quad K_{ij} = \int d^2\vec{r}\,\rho\,(\vec{r}, \omega)\left(2r_i r_j - \delta_{ij}r^2\right);$$
$$m = \frac{1}{2c_t}\int d^2\vec{r}\left[i_2\,(\vec{r}, \omega)\,r_1 - i_1\,(\vec{r}, \omega)\,r_2\right].$$

In general, the sound energy flux density is given by the expression [49]:

$$\Pi_i(\vec{x}, t) = -\sigma_{ik}(\vec{x}, t)v_k(\vec{x}, t).$$

As in the case of the spectral expansion of fluctuations [50], the spectral expansion of acoustic emission over frequencies is found using the formula:

$$d\varepsilon d\omega = -2\pi n_i\left[\sigma_{ik}\,(\omega)\,v_k\,(-\omega) + \sigma_{ik}\,(-\omega)\,v_k\,(\omega)\right]d\omega dS, \tag{5.87}$$

which gives an average energy flux over the period of oscillations through the site dS with the normal $\vec{n}$ within the frequency interval from ω to $\omega + d\omega$. The multiplier 2π is introduced to maintain the correct relations between the averaged quantities written in the t and ω variables.

Using (5.86) and (5.87), we compute the angular distribution of the differential intensity of the dipole radiation per unit dislocation length:

$$\frac{d\varepsilon_p\,(\omega, \psi)}{d\psi} = \frac{\rho}{2}c_t^3 k_b^3\left|n_1 p_2 - n_2 p_1\right|^2.$$

Providing that all the components of the vector $\vec{p}(\omega)$ have the same phase, we get:

$$\frac{d\varepsilon_p\,(\omega, \xi)}{d\xi} = \frac{\rho}{2}c_t^3 k_b^3 \sin 2\xi\left|\vec{p}\,(\omega)\right|^2, \tag{5.88}$$

where ξ is an angle between $\vec{r}$ and $\vec{p}$. Integrating over ξ, we find the total spectral density of the dipole radiation:

$$\varepsilon_p\,(\omega) = \frac{\pi}{2}\rho c_t^3 k_b^3\left|\vec{p}\,(\omega)\right|^2. \tag{5.89}$$

For a single oscillating screw dislocation, as well as in the case of annihilation of two screw dislocations, formulas (5.88) and (5.89) lead to the findings of [51] and [44].

For the dipole (m) and quadrupole (K_{ij}) moments, a similar calculation yields:

$$\frac{\mathrm{d}\varepsilon_m(\omega,\psi)}{\mathrm{d}\psi} = \frac{\rho}{2}c_{\mathrm{t}}^3 k_b^3 |m|^2 ; \frac{\mathrm{d}\varepsilon_K(\omega,\psi)}{\mathrm{d}\psi} = \frac{\rho c_{\mathrm{t}}^3 k_b^5}{32}|n_1 Q_2 - n_2 Q_1|^2 .$$

In contrast to the radiation of the dipole $\vec{p}$, the radiation of the dipole m has a cylindrical symmetry.

The features of calculating the elastic fields in the near zone ($d \ll r \ll k_b^{-1}$) may be explained by the example of a screw dislocation system. A complete analysis of other dislocation systems is given in [47]. In the near zone, to take the terms proportional to $\delta^{(1)}(\omega)$ into account, we should employ the relations (5.80). In the region ($d \ll r \ll k_b^{-1}$), the distortion field mainly contributes to the scalar potential. Therefore, the basic calculation formulas are:

$$(\beta_{23}, -\beta_{13}) \approx -(\partial_1 A_0, \partial_2 A_0) ; \quad \upsilon_3 = c_{\mathrm{t}}(\partial_1 A_2 - \partial_2 A_1) ;$$
$$\begin{pmatrix} A_0 \\ c_{\mathrm{t}}\vec{A} \end{pmatrix} \cong \frac{1}{2\pi}\left[\begin{pmatrix} \alpha^{(0)} \\ \vec{\beta}^{(0)} \end{pmatrix} \ln\frac{L}{r} + \sum_{m=1}^{\infty}\frac{1}{2mr^m}\left\{\begin{pmatrix} \alpha^{(m)} \\ \vec{\beta}^{(m)} \end{pmatrix} e^{\mathrm{i}m\psi} + \begin{pmatrix} \alpha^{(m)+} \\ \vec{\beta}^{(m)+} \end{pmatrix} e^{-\mathrm{i}m\psi}\right\}\right] ;$$
$$\alpha^{(m)} = \int \mathrm{d}^2\vec{r}\,\rho(\vec{r},\omega)\, r^m e^{-\mathrm{i}m\psi} ; \quad \alpha^{(m)+} = \int \mathrm{d}^2\vec{r}\,\rho(\vec{r},\omega)\, r^m e^{\mathrm{i}m\psi} .$$

We have entered the constant L (the sample size) that is insignificant for determining the observable quantities under the sign of the logarithm, in order to make its argument dimensionless.

After simple transformations, the asymptotic series are written in the form:

$$(\beta_{23}, -\beta_{13}) = \frac{B\vec{n}}{2\pi r}\delta^{(1)}(\omega) + \frac{2\vec{n}(\vec{n}\cdot\vec{p}) - \vec{p}}{2\pi r^2} - \frac{1}{4\pi}\vec{\nabla}\frac{K_{ij}n_i n_j}{r^2} + \cdots ;$$
$$\upsilon_3 = \frac{\mathrm{i}k_b c_{\mathrm{t}}}{2\pi r}(n_1 p_2 - n_2 p_1) + \frac{\mathrm{i}k_b c_{\mathrm{t}}}{4\pi r^2}(n_1 Q_2 - n_2 Q_1) + \cdots .$$

Here, $B\delta^{(1)}(\omega) = \int \mathrm{d}^2\vec{r}\rho(\vec{r},\omega)$, $(0,0,B)$ is the total Burgers vector of the dislocation system. The first term in the distortion expansion coincides with the field of an immovable screw dislocation. In the near zone, the formulas for elastic fields do not contain the multipole factor m. The fundamental allowance for the dipole approximation is related to the quadrupole moment K_{ij} of the screw dislocations system.

For a system of parallel edge dislocations in the region $r > d$, the multipole expansions of the elastic fields are:

$$\upsilon_m(\vec{x},\omega) = -\pi\mathrm{i}\sum_{\alpha=a,b}\widetilde{N}_{mfj}^{(\alpha)}(\omega,\nabla)\sum_{n=-\infty}^{+\infty}H_n^{(1)}(k_\alpha r)\, e^{\mathrm{i}n\psi} I_{(fj)n}(\omega,\alpha) ; \quad \beta_{k3} = -\frac{\partial_k \upsilon_3}{\mathrm{i}\omega} . \tag{5.90}$$

Here, the lower indices run through 1, 2 and the frequency is $\omega \neq 0$. The differentiation operators and the coefficients are determined by the formulas:

$$\widetilde{N}^{(a)}_{mfj} = -\frac{c_t^2}{2\pi\omega^2}\partial_m\partial_f\partial_j,\ \widetilde{N}^{(b)}_{mfj} = N^{(b)}_{mfj}\ ; \quad I_{(fj)n}(\omega,\alpha) = \int \mathrm{d}^2\vec{r}\, I_{(fj)}(\vec{r},\omega)\, J_n(k_\alpha r)\, e^{-in\psi}\ .$$

If the wavelength of the oscillations of the elastic field is much larger than the dimensions of the dislocation system (for $d \ll k_\alpha^{-1}$), the multipole coefficients are simplified:

$$I_{(fj)m}(\omega,\alpha) \approx \frac{1}{m!}\left(\frac{k_\alpha}{2}\right)^m \tau_{(fj)m}(\omega)\ , \quad I_{(fj)-m}(\omega,\alpha) \approx \frac{(-1)^m}{m!}\left(\frac{k_\alpha}{2}\right)^m \tau^+_{(fj)m}(\omega)\ ;$$
$$\tau_{(fj)m}(\omega) = \int \mathrm{d}^2\vec{r}\, I_{(fj)m}(\vec{r},\alpha)\, r^m e^{-\mathrm{i}m\psi}\ , \quad \tau^+_{(fj)m}(\omega) = \int \mathrm{d}^2\vec{r}\, I_{(fj)m}(\vec{r},\alpha)\, r^m e^{\mathrm{i}m\psi}\ , \tag{5.91}$$

where $m = 0, 1, 2, \ldots$. For a single oscillating edge dislocation, the first term in the expansion (5.90) in the approximation (5.91) comes to the result of [42].

In the wave zone ($d \ll k_\alpha^{-1} \ll r$) for the elastic fields, we obtain the series:

$$\begin{pmatrix} v_p \\ \beta_{sp} \end{pmatrix} = -c_t^2\sqrt{\frac{\omega}{2\pi r}} \sum_{\alpha=a,b} \begin{pmatrix} -1 \\ n_s c_\alpha^{-1} \end{pmatrix} \frac{\Phi^{(\alpha)}_{pfj}}{c_\alpha^{5/2}} \exp\left(\mathrm{i}k_\alpha r - \frac{\mathrm{i}\pi}{4}\right) \cdot$$
$$\cdot \left[\tau_{(fj)0} + \sum_{m=1}^{+\infty} \frac{1}{m!}\left(\frac{-\mathrm{i}k_\alpha}{2}\right)^m \left(\tau_{(fj)m} e^{\mathrm{i}m\psi} + \tau^+_{(fj)m} e^{-\mathrm{i}m\psi}\right)\right]\ ;$$
$$\Phi^{(a)}_{ikm} = n_i n_k n_m\ , \quad \Phi^{(b)}_{ikm} = (\delta_{ik} - n_i n_k)\, n_m\ , \quad \vec{n} = \vec{r}/r\ .$$

The spectral density of the dipole radiation has the form:

$$\frac{\mathrm{d}\varepsilon(\omega,\psi)}{\mathrm{d}\psi} = 2\rho|\omega|\left[n_l \tau_{(lm)0}(\omega)\, \tau^*_{(ms)0}(\omega)\, n_s + \left|n_l \tau_{(lm)0}(\omega)\, n_m\right|^2 \left(\frac{c_t^4}{c_l^4} - 1\right)\right]\ ,$$

where ψ is an angle between the vector $\vec{r}$ and the Ox-axis. As the dislocations move in a slip plane (along the Ox-axis), only the component (I_{12}) of the dislocation flux density tensor differs from zero. In this case, we have:

$$\frac{\mathrm{d}\varepsilon(\omega,\psi)}{\mathrm{d}\psi} = 2\rho|\omega|\left[\cos^2 2\psi + \sin^2 2\psi \frac{c_t^4}{c_l^4}\right]\left|\tau_{(12)0}(\omega)\right|^2\ ;$$
$$\varepsilon(\omega) = 2\pi\rho|\omega|\left|\tau_{(12)0}(\omega)\right|^2\left(1 + \frac{c_t^4}{c_l^4}\right)\ . \tag{5.92}$$

Comparison of expressions (5.88), (5.89) and (5.92) indicates a significant difference in the frequency and angular distributions of radiation for the screw and edge dislocations. This means that, according to the acoustic emission data, it is possible to determine which dislocations are responsible for the process of plastic deformation of the crystal.

The calculation of the fields of rectilinear dislocations admits a useful generalization. Consider a system of dislocations whose lines are not rectilinear, but are

stretched out along the Oz-axis. We expand the elastic fields $v_m(\vec{x}, \omega)$, $\beta_{nm}(\vec{x}, \omega)$ into Fourier integrals (series) in the variable x_3:

$$v_m(\vec{r}, \kappa, \omega) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \mathrm{d}x_3 e^{-\mathrm{i}\kappa x_3} v_m(\vec{x}, \omega); \quad \beta_{nm}(\vec{r}, \kappa, \omega) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \mathrm{d}x_3 e^{-\mathrm{i}\kappa x_3} \beta_{nm}(\vec{x}, \omega),$$

where $\vec{r} = (x_1, x_2)$ is a two-dimensional vector in the Oxy-plane. Taking formulas (5.76) and (5.78) into account, and the relation [52]:

$$\int\limits_{-\infty}^{\infty} \mathrm{d}x_3 \frac{e^{\mathrm{i}k_\alpha|\vec{x}|-\mathrm{i}\kappa x_3}}{2\pi|\vec{x}|} = -\frac{\mathrm{i}}{2} H_0^{(1)}\left(r\sqrt{k_\alpha^2 + (|\kappa| e^{\mathrm{i}\pi/2})^2}\right),$$

we may make a conclusion that, when conservatively moving the dislocations, the multipole expansions of the Fourier components have the form of expression (5.90). The indices run through the values 1, 2, 3 and so we should offer the following replacement:

$$\begin{aligned}
v_m(\vec{r}, \omega) &\Rightarrow v_m(\vec{r}, \kappa, \omega), \quad \beta_{nm}(\vec{r}, \omega) \Rightarrow \beta_{nm}(\vec{r}, \kappa, \omega); \\
I_{(fj)m}(\omega, \alpha) &\Rightarrow I_{(fj)m}(\omega, \kappa, \alpha) = \int \mathrm{d}^2\vec{r} I_{(fj)}(\vec{r}, \kappa, \omega) J_m\left(r\sqrt{k_\alpha^2 + (|\kappa| e^{\mathrm{i}\pi/2})^2}\right) e^{-\mathrm{i}m\psi}; \\
H_m^{(1)}(k_\alpha r) &\Rightarrow \frac{1}{2\pi} H_m^{(1)}\left(r\sqrt{k_\alpha^2 + (|\kappa| e^{\mathrm{i}\pi/2})^2}\right); \\
J_m(rk_\alpha) &\Rightarrow J_m\left(r\sqrt{k_\alpha^2 + (|\kappa| e^{\mathrm{i}\pi/2})^2}\right); \quad \partial_3 \Rightarrow \mathrm{i}\kappa.
\end{aligned} \tag{5.93}$$

In the particular case of a single oscillating dislocation fixed by equidistant points, under the condition $|d\sqrt{k_\alpha^2 - \kappa^2}| \ll 1$, the first term of the relations corresponding to (5.90) and (5.93) coincides with the formula found in [43] by another method.

A System of Dislocation Loops

Let us examine a system of dislocation loops that perform conservative motion ($I_{kk} = 0$) in a region with a characteristic size d. Suppose that the origin of the coordinate system lies inside the dislocation distribution, and that the fields v_m, β_{nm} are considered in a region not containing dislocations. For $\omega \neq 0$, the multipole expansions of the fields (5.76) and (5.78):

$$\begin{gathered}
v_m(\vec{x}, \omega) = -4\pi\mathrm{i} \sum_{\alpha=a,b} \sum_{l=0}^{+\infty} \sum_{p=-l}^{l} k_\alpha \widetilde{N}_{mfj}^{(\alpha)}(\omega, \nabla) h_l^{(1)}(k_\alpha r) Y_{lp}(\theta, \varphi) I_{(fj)lp}(\omega, \alpha); \\
\beta_{nm} = -\frac{\partial_n v_m}{\mathrm{i}\omega}
\end{gathered} \tag{5.94}$$

are obtained using the relation [46]:

$$\frac{e^{ik|\vec{x}-\vec{x}'|}}{|\vec{x}-\vec{x}'|} = 4\pi ik \sum_{l=0}^{+\infty} \sum_{p=-l}^{l} j_l\left(k\left|\vec{x}'\right|\right) h_l^{(1)}\left(k\left|\vec{x}\right|\right) Y_{lp}^*\left(\theta', \varphi'\right) Y_{lp}\left(\theta, \varphi\right) .$$

Here, $j_l(z)$, $h_l^{(1)}(z)$ are the spherical Bessel functions; $Y_{lp}(\theta, \varphi)$ are spherical harmonics; $|\vec{x}|$, θ, φ are spherical coordinates; the operators $\widetilde{N}_{mfj}^{(\alpha)}(\omega, \nabla)$ have been defined earlier; the lower spatial indices run through the values 1, 2, 3; the summation is made over twice occurring indices:

$$I_{(fj)lp}\left(\omega, \alpha\right) = \int \mathrm{d}^3\vec{x}\, I_{(fj)}\left(\vec{x}, \omega\right) j_l\left(k_\alpha\left|\vec{x}\right|\right) Y_{lp}^*\left(\theta, \varphi\right) .$$

If the wavelength of the oscillations of the elastic field is much greater than the dimensions of the dislocation system ($d \ll k_\alpha^{-1}$), the exact values of the coefficients may be replaced by approximate ones:

$$I_{(fj)lp}\left(\omega, \alpha\right) \approx \frac{k_\alpha^l}{(2l+1)!!}\tau_{(fj)lp}\left(\omega\right) , \quad \tau_{(fj)lp}\left(\omega\right) = \int \mathrm{d}^3\vec{x}\, I_{(fj)}\left(\vec{x}, \omega\right)\left|\vec{x}\right|^l Y_{lp}^*\left(\theta, \varphi\right) .$$

With this restriction, the first term in expression (5.94) ($l = 0$) reproduces the result of [35].

In the wave zone (for $d \ll k_\alpha^{-1} \ll |\vec{x}|$) we find:

$$\begin{pmatrix} v_m \\ \beta_{sm} \end{pmatrix} = -2c_{\mathrm{t}}^2 \sum_{\alpha=a,b} \begin{pmatrix} -1 \\ n_s\alpha^{-1} \end{pmatrix} \frac{\Phi_{mfj}^{(\alpha)}}{|\vec{x}|\,\alpha^2} \exp\left(ik_\alpha\left|\vec{x}\right|\right) \sum_{l=0}^{+\infty} \sum_{p=-l}^{l} \frac{(-ik_\alpha)^{l+1}}{(2l+1)!!}\tau_{(fj)lp}\left(\omega\right) Y_{lp}\left(\theta, \varphi\right) ;$$
$$\Phi_{ikm}^{(a)} = n_i n_k n_m , \quad \Phi_{ikm}^{(b)} = \left(\delta_{ik} - n_i n_k\right) n_m , \quad n_i = x_i/\left|\vec{x}\right| . \tag{5.95}$$

The terms corresponding to the values $l = 0, 1$ agree with the formulas obtained in [35, 45] after tedious calculations.

The above results allow us to go beyond the traditional limitations: $d \ll k_\alpha^{-1} \ll |\vec{x}|$. For example, using (5.94), we derive an expression for the spectral density of the dipole radiation of the system of dislocation loops:

$$\begin{aligned} \frac{\mathrm{d}\varepsilon\left(\omega, \Omega\right)}{\mathrm{d}\Omega} &= \frac{4\rho\omega^2}{c_{\mathrm{t}}}\left[\left(\frac{c_{\mathrm{t}}}{c_{\mathrm{l}}}\right)^5 \left|n_l I_{(lm)00}\left(\omega, a\right) n_m\right|^2 + n_l I_{(lm)00}\left(\omega, b\right) I_{(ms)00}^*\left(\omega, b\right) n_s - \right. \\ &\quad \left. - \left|n_l I_{(lm)00}\left(\omega, b\right) n_m\right|^2\right] ; \\ \varepsilon\left(\omega\right) &= \frac{16\pi\rho\omega^2}{5c_{\mathrm{t}}}\left[\frac{2}{3}\left(\frac{c_{\mathrm{t}}}{c_{\mathrm{l}}}\right)^5 \left|I_{(lm)00}\left(\omega, a\right)\right|^2 + \left|I_{(lm)00}\left(\omega, b\right)\right|^2\right] . \end{aligned} \tag{5.96}$$

It is valid in a wide area, where the weaker inequalities $k_\alpha^{-1} \ll |\vec{x}|$, $d < |\vec{x}|$ are fulfilled. Here, $\mathrm{d}\Omega$ is a solid angle element. In the wave zone (5.96), they agree with the results of [35, 45]. Formulas (5.96), in particular, deal with the radiation that arises when annihilating double kinks in the dislocation line (Figure 5.21) without restrictions of $g \ll k_\alpha^{-1}$ on the kink height g as in [45].

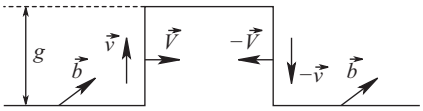


Fig. 5.21: Annihilation of kinks in the line of dislocation.

For kinks of rectangular form, in expressions (5.96), we should put that:

$$I_{(fk)00}(\omega, \alpha) \cong \frac{1}{8\pi\sqrt{2}} \left(\varepsilon_{fhn} b_k + \varepsilon_{khn} b_f\right) \frac{u_n \nu_h}{\mathrm{i}\omega k_\alpha} \left[\sqrt{\frac{\pi}{2}} - k_\alpha g J_{1/2}(k_\alpha g)\, S_{-3/2,-1/2}(k_\alpha g) - \right.$$
$$\left. - k_\alpha g J_{-1/2}(k_\alpha g)\, S_{-1/2,1/2}(k_\alpha g) \right],$$

Here, ν_h is the unit tangent vector to the kink line, u_n is the relative rate of kinks at the moment of the collision, $J_\mu(z)$ is the Bessel function, and $S_{\mu,\upsilon}(z)$ is the Lommel function [48, 52].

5.6 The Peierls Model of a Dislocation Core

Crystal deformations over large distances from a dislocation are well described by the linear theory of elasticity. However, the region of a dislocation core requires special investigation. In the core, significant displacements of atoms from equilibrium positions occurs, along with switching of bonds between atoms. The structure of the core is responsible for the dislocation mobility in crystals. The distribution of atoms in the dislocation core depends on the periodic potential generated by atomic planes adjacent to the dislocation slip surface, as well as on deformations of the crystal in the outer region relative to the core.

The simplest model of the dislocation core was proposed by Peierls [53]. Its sequential generalization of a dynamic case is given in [54]. According to [53, 54], the dislocation core is "smeared" in the slip plane. In the core region, the energy of interaction between neighboring atomic planes (the Peierls energy E_P) is taken into account for the nearest neighbor approximation:

$$E_\mathrm{P} = \frac{1}{\Delta S} \int_A \Phi(\vec{w}) n_j \mathrm{d}S_j \,. \tag{5.97}$$

Here $\vec{w}(x, t) = \vec{u}_+(x, t) - \vec{u}_-(x, t)$ is the relative displacement of atoms along two sides of a slip plane A with a normal n_j. The integration is performed over the slip plane and ΔS is the area per atom. The form of the potential energy $\Phi(\vec{w})$ per atom in plane A is determined by the symmetry of the crystal. Figure 5.22 illustrates the approach of Peierls on the example of an edge dislocation in a cubic crystal.

The crystal halves above and below plane A are described in the approximation of an elastic continuum. In plane A, the stresses σ_{ij}, based on the relations of the linear theory of elasticity, must be balanced by inelastic forces from the lattice. This

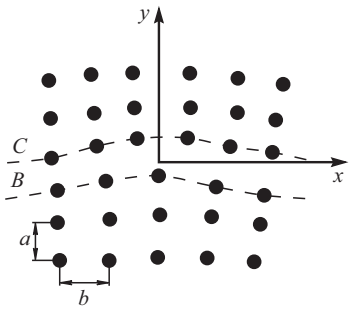


Fig. 5.22: The Peierls model of an edge dislocation in a cubic crystal. The material above ($y > 0$) and below ($y < 0$) the slip plane is described in the approximation of an elastic continuum. The interaction between the atomic planes B and C is taken into account in the nearest neighbor approximation.

condition defines the dynamics of the dislocation core and has the form [54]:

$$\sigma_{ij}(\vec{x}, t)n_j - \frac{1}{\Delta S}\frac{\partial \Phi}{\partial w_i(\vec{x}, t)} = 0, \vec{x} \in A \, . \tag{5.98}$$

The elastic stresses can be represented as:

$$\sigma_{ij}(\vec{x}, t) = \sigma_{ij}^{\text{ext}}(\vec{x}, t) + \sum_{\upsilon=0}^{\infty} \int \Phi_{ijkl}^{(\upsilon)}(\vec{x} - \vec{x}')\partial_t^{\upsilon}\beta_{kl}^{P}\left(\vec{x}', t\right) \mathrm{d}^3\vec{x}' \, , \tag{5.99}$$

where $\sigma_{ij}^{\text{ext}}(\vec{x}, t)$ is the stress due to forces applied to the outer surface of the body. The second term represents the expansion of the elastic stresses from the dislocation. For an unbounded medium, the kernels $(\Phi_{ijkl}^{(\upsilon)}(\mathrm{r} - \mathrm{r}'))$ of integral operators are calculated in [37]. They come from formulas (5.48) and (5.56). The plastic distortion $\beta_{kl}^{P}(\vec{x}, t)$ differs from zero only on the slip dislocation surface:

$$\beta_{kl}^{P}(\vec{x}, t) = w_l(\vec{x}, t)n_k \, , \quad \vec{x} \in A \, .$$

The nonlinear equations, (5.98) and (5.99), completely trace the evolution of a displacement $\vec{w}(\vec{x}, t)$ under the action of external forces.

Suppose that λ and r_0 are the characteristic spatial scale of a weakly changeable elastic field and the core size of the Peierls dislocation, respectively, and that they satisfy the condition: $\lambda \gg r_0$. Then, if the dislocation migrates quasistationarily, i.e., its velocity is low, equations (5.98) and (5.99) may be replaced by averaged ones, as is done when describing a flat cluster of moving dislocations [29]. This section covers dynamic equations obtained by averaging for the straight Peierls dislocations in an unbounded medium and in a plate. We consider the motion of a dislocation in a glide plane, which passes through the center of the plate parallel to its surfaces. Let us examine, for definiteness, a dislocation in a cubic crystal. Suppose that the dislocation is parallel to the Oz-axis, and $y = 0$ is its slip plane. This would lead to equations (5.98) and (5.99) acquiring the form:

$$m\partial_t \mathrm{v} + \beta \mathrm{v} = b\left(-M\widehat{\mathrm{L}}[\rho] + \sigma^{\text{ext}} - \frac{Nb}{2\pi a}\sin\frac{2\pi w}{b}\right) \, , \quad \partial_t \rho + \partial_x(\rho \mathrm{v}) = 0 \, . \tag{5.100}$$

Here, $\rho = -\partial_x w$, $M = \mu/2(1-\nu)$, $N = 2\mu(3-2\nu)^{-1}$ and $M = N = \mu/2$ for an edge and screw dislocation, respectively. $w(x, t)$ is a nonzero component of the relative displacement vector of atoms, b is a lattice constant in the slip plane, a is a lattice constant in the orthogonal direction, $m > 0$ is the effective dislocation mass [29], σ^{ext} is an external shear stress field, and v is the displacement velocity in the slip plane. The quantity $-\beta$v simulates the dissipative force of the medium resistance to the dislocation motion ($\beta > 0$).

At a great distance from the dislocation, the crystal lattice should be perfect. This is possible under the condition:

$$w(x = +\infty, t) - w(x = -\infty, t) = \int_{-\infty}^{-\infty} \mathrm{d}x \partial_x w(x, t) = -b\,. \tag{5.101}$$

The validity of (5.101) for an edge dislocation is obvious, as can be seen in Figure 5.1. According to (5.101), the inhomogeneous relative displacement $w(x, t)$ can be considered equivalent to the presence of small fictitious dislocations distributed along the Ox-axis. Fictitious dislocations located in the interval from x to $x + \mathrm{d}x$ govern the plastic shear $\rho(x, t)\mathrm{d}x = -\partial_x w \mathrm{d}x$ in the plane $y = 0$. The second equation (5.100) expresses the law of conservation of the plastic shear as the fictitious dislocations migrate. The entire set of the fictitious dislocations simulates the core of a real dislocation and carries a complete plastic shear b. The quantity $-M\hat{\mathrm{L}}[\rho]$ should be compared to the shear stress caused by the fictitious dislocations in the slip plane (for $y = 0$). Formulas (5.37), (5.44), and (5.59) for elastic stresses from a dislocation make it possible to immediately write the kernel of the integral operator $\hat{\mathrm{L}}$. Elastic stresses of an edge dislocation in a plate are found in [55].

For an unbounded medium, the operator $\hat{\mathrm{L}}$ coincides with the Hilbert transform:

$$\hat{\mathrm{L}}[\rho] \equiv \hat{\mathrm{H}}[\rho] = \frac{\text{V.p.}}{\pi} \int_{-\infty}^{+\infty} \frac{\mathrm{d}x'}{x' - x}\,. \tag{5.102}$$

The letters "V. p." stand for an understanding of the integral above as the Cauchy principal value. In the case of a screw dislocation, the dynamic equations (5.100) generalize (5.46).

In a plate with a thickness of d ($d \gg a$), the operator is $\hat{\mathrm{L}} \equiv \hat{\mathrm{T}}$:

$$\hat{\mathrm{T}}[\rho] = \frac{\text{V.p.}}{d} \int_{-\infty}^{\infty} \frac{\rho(x')\mathrm{d}x'}{\mathrm{sh}\left[\pi(x' - x)/d\right]}\,.$$

For the energy (5.97), Peierls took the simplest approximation:

$$E_{\text{P}} = \frac{N}{2a}\left(\frac{b}{\pi}\right)^2 \int_{-\infty}^{\infty} \sin^2\left(\frac{\pi w}{b}\right)\mathrm{d}x\,. \tag{5.103}$$

The coefficient N in (5.103) is chosen so that, at large distances from the dislocation core, the static solution of model (5.100) should reproduce the solution of the linear theory of elasticity. For example, for an edge dislocation in an unbounded medium, the approximate solution of model (5.100) with $|x| \gg |b|$, $\sigma^{\text{ext}} = 0$

$$\frac{w}{a} \approx \frac{Mb}{\pi N x} = \frac{(3-2\nu)\, b}{4\pi(1-\nu)x} = \partial_2 u_1^G\Big|_{y=0}$$

is consistent with formula (5.58).

It is necessary to note that the dynamic equations (5.100) are applicable, only for describing a single dislocation or synchronous motion of a periodic sequence of identical dislocations. This is because the relative motions of different dislocations do not need to be taken into account.

How well model (5.100) meets the experimental data can be judged only after constructing its solutions. The nonlocal term $\widehat{\mathrm{L}}[\partial_x w]$ and the strong nonlinearity $\sin(2\pi w/b)$ in equation (5.100) make it difficult to integrate. Within the Peierls model, an analytical description of the static configurations [55] is known: a) for one dislocation and two dislocations with opposite Burgers vectors in a constant field, and b) for several periodic chains of dislocations. The integration method was based on the Fourier transform and was not very constructive. The authors of [56] proposed an effective technique for solving the Peierls model explicitly, by methods used in soliton theory. Below, we offer the results of this approach, where dynamic generalizations of the static configurations are obtained. Furthermore, we show how the equations of the model and its solutions change when studying the motion of different dislocations.

The Peierls Dislocation in an Unbounded Medium

We show that in the case of an unbounded medium, the integration of model (5.100) can be carried out using methods of complex analysis. We use the substitution:

$$w = \frac{\mathrm{i}b}{2\pi} \ln \frac{f(x-v(t))}{f^*(x-v(t))}\,. \tag{5.104}$$

It can be demonstrated that, from the mathematical point of view, a meromorphic function $f(z)$ meets the solutions of the dislocation type. We recall that a function is said to be meromorphic in a given domain if it is an analytic function everywhere except for in a certain number of poles. Each complex zero (pole) $x = \nu_1 + \mathrm{i}\nu_2$ of the order n of the function $f(z)$ corresponds to a single dislocation with the Burgers vector $B = bn$ ($B = -bn$). Indeed, when

$$f(x) = \left[e^{\mathrm{i}\pi/2}\,(\nu_2 + \mathrm{i}(x-\nu_1))\right]^n\,,$$

where $\nu_k^* = \nu_k$ ($k = 1, 2$), we have:

$$w = -bn\left(\frac{1}{2} + \frac{1}{\pi}\,\mathrm{arctg}\,\frac{x-\nu_1}{\nu_2}\right)$$

and, consequently, $w(x=-\infty)-w(x=+\infty)=bn$. The parameters ν_1 and ν_2 have a clear physical meaning: they determine the coordinates of the dislocation center and the size of its core, respectively.

Applying Cauchy's theorem on residues, it is not difficult to verify that the identity

$$\widehat{\mathrm{H}}\left[\frac{1}{x-\nu}\right]=-\frac{\mathrm{i}\,\mathrm{sign}\,\mathrm{Im}\nu}{x-\nu} \tag{5.105}$$

holds. This property allows us to pass from the integrodifferential equations (5.100) to ordinary differential equations. For definiteness, suppose that the zeros and poles of $f(z)$ lie in the upper half-plane and $\partial_z \ln f(z)\to\infty$ is $|z|\to\infty$. Then, using (5.105), we obtain a useful relation:

$$\mathrm{i}\widehat{\mathrm{H}}\left[\partial_x \ln\frac{f(x)}{f^*(x)}\right]=\partial_x \ln[f(x)f^*(x)]\ . \tag{5.106}$$

Having substituted (5.104), (5.106) into (5.100), we obtain:

$$m\partial_t^2\nu_1+\beta\partial_t\nu_1=b\left(\frac{Mb}{2\pi}\left[\frac{\partial_x f}{f}+\frac{\partial_x f^*}{f^*}\right]+\sigma^a-\frac{Nb}{4\pi a\mathrm{i}}\left[\frac{f^*}{f}-\frac{f}{f^*}\right]\right)\ . \tag{5.107}$$

Setting equal (separately) the two linearly independent groups of terms in (5.107), i.e., a part containing poles and a part not containing them, we find:

$$\begin{gathered} m\partial_t^2\nu_1=b\sigma^a-\beta\partial_t\nu_1\ ,\\ f^*\left[-\frac{Mb}{2\pi}\partial_x f+\frac{Nb}{4\pi\mathrm{i}a}f^*\right]+f\left[-\frac{Mb}{2\pi}\partial_x f^*-\frac{Nb}{4\pi\mathrm{i}a}f\right]=0\ . \end{gathered} \tag{5.108}$$

For a single dislocation or a periodic sequence of identical dislocations, the function $f(z)$ is an entire function with simple zeros. Its explicit form can be found if $f(x)$ satisfies the simpler linear equation:

$$-\frac{Mb}{2\pi}\partial_x f+\frac{Nb}{4\pi\mathrm{i}a}f^*+\frac{\mathrm{i}\lambda Nb}{2\pi a}f=0\ . \tag{5.109}$$

Here, λ is a real parameter. The solution of equation (5.109), with the necessary analytic properties, has the form:

$$f(x)=\exp\left(\frac{\mathrm{i}\pi}{2}\right)\mathrm{sh}\left[\frac{\rho}{2}+\frac{\mathrm{i}N\,\mathrm{sh}\,\rho}{2Ma}x\right]\ ,\quad \lambda=\frac{\mathrm{ch}\,\rho}{2}\ ,$$

and describes a periodic chain of identical dislocations. In the chain, the size of each dislocation core $r_0=\rho L/(4\pi)$ depends on the distance ($L=4\pi Ma/(N\,\mathrm{sh}\,\rho)$) between the closest dislocations.

In the limit $\rho\to 0$, the chain period increases indefinitely, and we arrive at an expression corresponding to a single dislocation:

$$\begin{gathered} f(x)=\exp\left(\frac{\mathrm{i}\pi}{2}\right)\left(1+\frac{\mathrm{i}Nx}{Ma}\right)\ ;\\ w(x,t)=-b\left(\frac{1}{2}+\frac{1}{\pi}\,\mathrm{arctg}\,\frac{x-\nu_1(t)}{r_0}\right)\ ,\quad r_0=\frac{Ma}{N}\ . \end{gathered} \tag{5.110}$$

At an infinitely large distance from the dislocation, the crystal lattice should be perfect:

$$w(x = +\infty) = -b\,, \quad w(x = -\infty) = 0\,, \quad \lim_{|x|\to\infty} \sin^2 \pi w/b = 0\,.$$

We have chosen the integration constant in (5.110) so that this condition should be satisfied. The half width of the edge dislocation r_0 depends on the Poisson ratio and can take values in the interval $3a/4 \le r_0 \le a$.

The description of the motion of the dislocation center (5.108) is equivalent to the problem of the dynamics of a material point. If $v_1(0) = 0$, $\partial_t v_1 = v_0$, we have:

$$v_1(t) = \int_0^t \left[\frac{b}{m} \int_0^{t'} \sigma^a(t'')\mathrm{d}t'' + v_0 \right] \exp\left\{ -\frac{\beta}{m}(t - t') \right\} \mathrm{d}t'\,, \quad \beta m > 0\,. \tag{5.111}$$

Formulas (5.110) and (5.111) give an idea of the motion of the Peierls dislocation in the arbitrarily time dependent shear stress field. For a motionless dislocation, the obtained formulas become known earlier [55].

The discreteness of the crystal lattice must lead to a periodic resistance force against the dislocation motion. However, such a force is absent in model (5.100). This is due to averaging the description. The discreteness of the crystal lattice is smoothed out in formula (5.103), by replacing the summation of atomic bonds in the slip plane by integration:

$$\sum_{n=-\infty}^{+\infty} \Phi\,(x = nb) \quad \Rightarrow \quad \frac{1}{b} \int_{-\infty}^{+\infty} \Phi(x)\mathrm{d}x = \frac{N}{2a} \left(\frac{b}{\pi} \right)^2 \int_{-\infty}^{\infty} \sin^2 \left(\frac{\pi w}{b} \right) \mathrm{d}x = E_\mathrm{P}\,.$$

The simplest way to account for the discreteness of the system is as follows [57]. We use the solution (5.110) and calculate the energy of discrepancy of the crystal with a dislocation using the formula:

$$E = \sum_{n=-\infty}^{+\infty} \Phi\,(x = nb) = \frac{Nb}{2a} \left(\frac{b}{\pi} \right)^2 \sum_{n=-\infty}^{+\infty} \sin^2 \left[\frac{\pi}{b} w\,(x = nb) \right] =$$
$$= \frac{Nb}{2a} \left(\frac{b}{\pi} \right)^2 \sum_{n=-\infty}^{+\infty} \frac{r_0^2}{(nb - v_1)^2 + r_0^2}\,.$$

By means of (5.43), we perform the summation:

$$\sum_{n=-\infty}^{+\infty} \frac{r_0^2}{(nb - v_1)^2 + r_0^2} \approx \frac{\pi r_0}{b} \left(1 + 2e^{-2\pi r_0/b} \cos \frac{2\pi v_1}{b} \right)\,.$$

Hence, we come to an expression for the periodic *Peierls–Nabarro force* acting on a dislocation in a crystal:

$$F = -\frac{\partial E}{\partial v_1} = 2Mbe^{-2\pi r_0/b} \sin \frac{2\pi v_1}{b}\,.$$

The maximum value of this force determines those shear stresses that should be applied to the crystal so that the dislocation begins to move in its slip plane:

$$\sigma_s = \frac{1}{b}\max F = 2Me^{-2\pi r_0/b} \sim \left(10^{-2} - 10^{-4}\right)\mu\,.$$

Dislocations in a Plate

In the previous section, we have sought solutions of (5.100) in the class of functions of the form $f(z) = c\prod_k(z-\nu_k)/\prod_n(z-\mu_n)$ (the products can also be infinite). Choosing $f(z)$ in such a way, we not only take into account the presence of dislocations, but we also apply a remarkable mathematical property. If one expands the function $\widehat{\mathrm{H}}[\partial_x w]$ into simple fractions, we obtain the expression:

$$\sum_k\left(\frac{a_k}{x-\nu_k} + \frac{b_k}{x-\nu_k^*}\right) + \sum_n\left(\frac{c_n}{x-\mu_n} + \frac{d_n}{x-\mu_n^*}\right)\,.$$

It is essential that the same analytic structure belongs to the expansion of the function $\sin(2\pi w/b)$. Consequently, after substituting (5.104) into (5.100), we should set the coefficients of the linearly independent functions $(x-\nu_k)^{-1}$, $(x-\nu_k^*)^{-1}$, $(x-\mu_n)^{-1}$, $(x-\mu_n^*)^{-1}$ equal to zero. The condition (5.109) tacitly performs this operation. It is interesting to note that the requirement of the residues at the poles of equation (5.100) vanishing does not trace the dynamics of the dislocation centers because time in these conditions enters as a parameter.

Guided by the above qualitative considerations, we proceed to constructing solutions in a plate. We have the natural generalization (5.105):

$$\widehat{\mathrm{T}}\frac{1}{\mathrm{sh}\left[\pi\left(x-\nu\right)/d\right]} = -\frac{\mathrm{i}}{\mathrm{sh}\left[\pi\left(x-\nu\right)/d\right]}\,\mathrm{sign}\left[\sin\left(\frac{2\pi}{d}\mathrm{Im}\nu\right)\right]\,.$$

The solutions of the model should be sought in the form (5.104) in the class of meromorphic functions:

$$f(x) = \frac{\prod_k\left(\mathrm{sh}\,A_k - \mathrm{sh}\left(\mathrm{i}B_k\right)\right)}{\prod_n\left(\mathrm{sh}\,C_n - \mathrm{sh}\left(\mathrm{i}D_n\right)\right)}\,;$$

$$A_k = \frac{\pi}{d}\left(x-\nu_k^{(1)}\right)\,,\quad B_k = \frac{\pi}{d}\nu_k^{(2)}\,,\quad C_n = \frac{\pi}{d}\left(x-\mu_n^{(1)}\right)\,,\quad D_n = \frac{\pi}{d}\mu_n^{(2)}\,.$$

The real parameters $\nu_k^{(1)}$, $\mu_n^{(1)}$ and $\nu_k^{(2)}$, $\mu_n^{(2)}$ determine the coordinates of dislocation centers and the dimensions of their cores. In particular, we seek the solution for one dislocation in the following way:

$$f(x) = \mathrm{sh}\,A_1 - \mathrm{sh}\left(\mathrm{i}B_1\right)\,,\quad 0 < \nu_1^{(2)} < d/2\,,$$

Then:

$$\begin{aligned}
\partial_x w &= \frac{\mathrm{i}b}{2d}\left[\frac{1}{\operatorname{sh}(A_1-\mathrm{i}B_1)}-\frac{1}{\operatorname{sh}(A_1+\mathrm{i}B_1)}\right],\\
\hat{\mathrm{T}}\partial_x w &= \frac{b}{2d}\left[\frac{1}{\operatorname{sh}(A_1-\mathrm{i}B_1)}+\frac{1}{\operatorname{sh}(A_1+\mathrm{i}B_1)}\right],\\
\sin\frac{2\pi w}{b} &= -\mathrm{ith}\,(\mathrm{i}B_1)\left[\frac{1}{\operatorname{sh}(A_1-\mathrm{i}B_1)}+\frac{1}{\operatorname{sh}(A_1+\mathrm{i}B_1)}\right].
\end{aligned} \tag{5.112}$$

Plugging relations (5.112) into equations (5.100), and equating the coefficients for the linearly independent functions $1/\operatorname{sh}(A_1 \pm \mathrm{i}B_1)$, we find:

$$w = -b\left(\frac{1}{2}+\frac{1}{\pi}\operatorname{arctg}\frac{\operatorname{sh}\left[\pi\left(x-v_1(t)\right)/d\right]}{\sin\pi v_1^{(2)}/d}\right), \quad \operatorname{tg}\frac{\pi}{d}v_1^{(2)} = \frac{M\pi a}{Nd}.$$

The dependence of $v_1(t)$ is the same as in (5.111). The image forces from the plate surface cause the Peierls dislocation core to compress.

For periodic chains of identical dislocations, the function $f(x)$ involves an infinite product of factors of the type $\operatorname{sh} A_k - \operatorname{sh}(\mathrm{i}B_k)$. Moreover, the functions $\partial_x w$ and $\sin(2\pi w/b)$ are expanded into simple fractions of type (5.112). In [56], the authors have solved model (5.100) exactly. Its solutions correspond to synchronous oscillations of N identical periodicity dislocation chains inserted into each other.

Interaction of Dislocations

The previous sections have covered in full the dynamics of only a single dislocation. A description of dislocation chains is valid provided that all the dislocations have the same initial velocities and oscillate synchronously. To consistently describe the interaction between dislocations of a system, it is necessary to change equations (5.100). This can be done within the approach proposed in [54] if one regards plastic displacements w_i from individual dislocations as independent variables when varying the system's action.

Let us first look into a case of moving two dislocations with opposite Burgers vectors in the one slip plane in an unbounded medium. The energy dissipation is neglected. We write down the total plastic displacement in a form consistent with the known static solution of [55]:

$$\begin{aligned}
&w = \frac{b\mathrm{i}}{2\pi}\ln\left[e^{-\mathrm{i}\gamma(t)}\frac{(x-\nu)\left(x-\eta^*\right)}{\left(x-\nu^*\right)(x-\eta)}\right], \quad \nu = \nu_1(t)+\mathrm{i}\nu_2(t),\\
&\eta = \eta_1(t)+\mathrm{i}\nu_2(t), \quad \nu_2 > 0, \quad \nu_s^* = \nu_s \quad (s=1,2), \quad \eta_1^* = \eta_1, \quad \gamma^* = \gamma.
\end{aligned} \tag{5.113}$$

The variational principle of least action was proposed to describe the dynamics of dislocations [54]. We use (5.113) as a trial function and vary the functional of action [54] by the coordinates of dislocations. Suppose the motion is quasistationary. Then the

functions $v_2(t)$ and $\gamma(t)$ depend weakly on time and the effective mass of dislocations is constant. From this it follows that the equations of dynamics take the form:

$$\begin{aligned}\mu_1\partial_t^2 v_1 &= M\widehat{\mathrm{H}}\left[\partial_x w\right] + \sigma^{\mathrm{ext}}(t) - \frac{Nb}{2\pi a}\sin\frac{2\pi w}{b}\,,\\ -\mu_1\partial_t^2\eta_1 &= M\widehat{\mathrm{H}}\left[\partial_x w\right] + \sigma^{\mathrm{ext}}(t) - \frac{Nb}{2\pi a}\sin\frac{2\pi w}{b}\,.\end{aligned} \tag{5.114}$$

Here, $\mu_1 = m/b$, m is the effective mass of each dislocation. Simple calculations yield:

$$\partial_t^2\left(v_1+\eta_1\right) = 0\,,\quad \mu_1\partial_t^2\left(v_1-\eta_1\right) = \frac{Mb}{\pi}\frac{1}{v_1-\eta_1} + \sigma^{\mathrm{ext}}(t)\,,$$
$$v_2 = \frac{Ma}{N\cos\gamma}\,,\quad \gamma = \arcsin\frac{2Ma}{N\left(v_1-\eta_1\right)}\,.$$

Hence, it follows that the center of mass of the dislocations move with a constant velocity. The relative motion is determined by an external field and by the strength of the interaction between the dislocations ($\sim 1/(v_1-\eta_1)$). The sizes of the dislocation cores v_2 depend on the time varying distance $v_1-\eta_1$ between dislocations through the parameter γ. The time dependence v_2 is stronger, the smaller the distance between dislocations. Equations (5.114) can also be used to describe relative synchronous oscillations of two dislocation chains with opposite Burgers vectors. An appropriate exact solution has been found in [56].

It is worth noting that, when $|v_1-\eta_1| \gg 2Ma/N \sim 2v_2$, expression (5.113) for w is a superposition of single dislocations (5.110). The center of each dislocation moves in an effective field generated by other dislocations. The technique of approximately describing the dislocation system can be elaborated for the case of being the distance l between the dislocation much larger than the characteristic size r_0 of the dislocation core [56]. This statement can be exemplified through the interaction of two dislocations. Although the integration in (5.114) is carried out in infinite limits, the integral is mainly contributed to by the dislocation core region of the order size r_0 ($\partial_x w$,is a generalization of the sum of δ-functions of the linear theory). For $l \gg r_0$, nonsingular cores of integral equations can be expanded in powers of the small parameter r_0/l. Thus, for two dislocations located at the center of the plate, we have:

$$\begin{aligned}\mu_1\partial_t^2 v_1 &= M\widehat{\mathrm{T}}\left[\partial_x w_1\right] + \sigma^{(1)}_{eff}(t) - \frac{Nb}{2\pi a}\sin\frac{2\pi w_1}{b}\,,\quad \left(x\sim v_1(t)\right)\,,\\ \mu_2\partial_t^2\eta_1 &= M\widehat{\mathrm{T}}\left[\partial_x w_2\right] + \sigma^{(2)}_{eff}(t) - \frac{Nb}{2\pi a}\sin\frac{2\pi w_2}{b}\,,\quad \left(x\sim \eta_1(t)\right)\,,\\ \sigma^{(1,2)}_{eff}(t) &= \sigma^{\mathrm{ext}}(t) \mp \frac{Mb\varepsilon_{2,1}}{d\,\mathrm{sh}\left[\pi\left(v_1-\eta_1\right)/d\right]}\,,\end{aligned} \tag{5.115}$$

in the main approximation. Here, $b\varepsilon_1 = -\int_{-\infty}^{\infty}\mathrm{d}x\partial_x w_i$ ($\varepsilon_i = \pm 1, i = 1,2$), and μ_i are constant parameters related to the effective mass of the i-th dislocation as ($\mu_i \sim m_i/b\varepsilon_i$). We have also taken into account that, near the core of a dislocation with the number i ($x \approx v_1$ or $x \approx \eta_1$), the approximate equality $w(x) \approx w_i(x)+bk$ is satisfied; k is

an integer, $w_i(x)$ is the displacement caused by the i-th dislocation. Thus, the analysis of the interaction between dislocations boils down to the previously considered problem of the motion of one dislocation in an alternating external field (compare (5.100) and (5.115)). In the limit $|\nu_1 - \eta_1| \to \infty$, the system (5.115) splits into two equations (5.100) that delineate the motion of independent dislocations in an external field.

5.7 Weakly Nonlinear Soliton-like Excitations in a Two-Dimensional Martensitic Transition Model

The term *martensitic* pertains to diffusionless phase *transitions* of the displacement type. They can be described by shear deformations. Martensitic transformations take place due to ordered cooperative atomic motion in a crystal lattice. The relative displacement of neighboring atoms is small. They amount to a small part of the interatomic distance. However, these displacements result in rearranging the unit cell of the crystal lattice. So, the martensitic transformation can be considered to be reduced to deformation of the unit cell of the initial phase. For a wide class of crystals, a decrease in temperature causes a transition from a phase with a cubic lattice to a phase with an orthorhombic or tetragonal symmetry lattice. A high temperature phase is called *austenite*, and a low temperature phase is called *martensite*. Martensitic transitions, and phenomena and processes associated with them, are a challenge for applications because they are inherent in many structural materials. They significantly affect their physical properties and, to a certain extent, allow these properties to be controlled. Martensitic transformations are difficult to analytically depict.

For a 4 mm/2 mm (square/rectangle) martensitic phase transition, the authors of [58, 59] have proposed a model based on the nonlinear elasticity theory. It is analogous to the Ginsburg–Landau model. Similar models are successfully used for describing phase transitions in magnetic materials. Such a model serves as a good approximation for the martensitic transitions. Therefore, the latter are called ferroelastic transitions. Of course, the ferroelastic transitions do not exhaust the whole variety of martensitic transitions. The model appears to be a good approximation for the cubic tetragonal martensitic transition in a number of real systems (Nb_3Sn, V_3Si, $In_{0.76}Tl_{0.24}$, $Fe_{0.72}Pd_{0.28}$) [59]. A common feature of these compounds differing in structure is that the atoms in them are displaced along the distinguished planes in the square/rectangle type as the martensitic transition occurs. The phenomenological description of first-kind phase transitions in 2D systems requires keeping invariants in the expression for elastic energy up to higher powers of the order parameter (at least up to the sixth power, inclusively). This causes significant nonlinearity of the problem. In addition, the heterogeneity of the structure of interphase boundaries should be described by the inclusion of higher gradients of the displacement field into the elastic energy. Essential nonlinearity and dispersion (to be introduced to the theory)

would allow avoiding the use of fictitious dislocations of transformation [60] when analytically describing the crystal lattice rearrangement.

Let us describe the basic relationships for a 2D model of the 4 mm/2 mm martensitic transition. Suppose that X_i is the position of a material point in an elastically deformed medium, x_i is its position in an undeformed state, and $u_i = X_i(\vec{x}, t) - x_i$ is the displacement vector. The classical nonlinear theory of elasticity represents the elastic deformation energy in terms of the Lagrangian strain tensor:

$$\eta_{ij} = \frac{1}{2}\left[u_{i,j} + u_{j,i} + u_{r,i}u_{r,j}\right] , \quad u_{i,j} = \frac{\partial u_i}{\partial x_j} . \tag{5.116}$$

According to [59], the free energy for the 4 mm/2 mm phase transition has the form:

$$\begin{aligned} w = {} & \frac{1}{2}A_1 e_1^2 + \frac{1}{2}A_2 e_2^2 + \frac{1}{2}A_3 e_3^2 + \frac{1}{4}B_2 e_2^4 + \frac{1}{6}C_2 e_2^6 + \frac{1}{2}d_1\left(e_{1,1}^2 + e_{1,2}^2\right) + \\ & + \frac{1}{2}d_2\left(e_{2,1}^2 + e_{2,2}^2\right) + \frac{1}{2}d_3\left(e_{3,1}^2 + e_{3,2}^2\right) + d_4\left(e_{1,1}e_{2,1} - e_{1,2}e_{2,2}\right) + \\ & + d_5\left(e_{1,1}e_{3,2} + e_{1,2}e_{3,1}\right) + d_6\left(e_{2,1}e_{3,2} - e_{2,2}e_{3,1}\right) . \end{aligned} \tag{5.117}$$

Here, $A_i, B_2, C_2, d_j (i = 1, 2, 3; j = 1, 2, \ldots, 6)$ are phenomenological constants. The fields $e_3 = \eta_{12}$ and $e_1 = (\eta_{11} + \eta_{22})/\sqrt{2}$ are responsible for the shear deformation and the dilatation of the medium, respectively. The field $e_2 = (\eta_{11} - \eta_{22})/\sqrt{2}$ controls the extension of the square into the rectangle and is an order parameter in the Landau theory for a 4 mm/2 mm phase transition. In the context of Landau's theory, A_2 is assumed to be temperature dependent in the vicinity of the transition, with this dependence being such that $A_2 = 0$ at the transition point. As usual, we neglect the temperature dependence of other elastic moduli. To realize the first order phase transition, it is necessary to fulfill the conditions $B_2 < 0$, $C_2 > 0$. Expression (5.117) for the energy qualitatively reproduces the observed hysteresis phenomena during the transition [59]. Equations of dynamics can be derived by varying the Lagrangian in the fields u_i:

$$L = \int \mathrm{d}x_1 \mathrm{d}x_2 \left[\frac{\rho}{2}\partial_t u_i \partial_t u_i - w\right] , \tag{5.118}$$

where ρ is the density of the medium in the undeformed state. Furthermore, we put that $\rho = \text{const}$.

The formations of the type of flat domain walls and plane parallel domain structures were first investigated within the 2D martensitic transition model by [58]. However, as noted in [59], the expression for the energy, used in [58], differs from (5.117), as it contains no terms compatible with the crystal symmetry. Therefore, it is insufficient to determine the phenomenological constants from experimental phonon dispersion curves. The works by [59, 61–64] have examined one-dimensional solutions of the martensitic transition models, with the paper [59] taking the complete expression for the energy (5.117) into account. However, in these papers, solutions of the type of domain walls and plane parallel domain structures have been obtained by neglecting

nonlinear terms in the strain tensor (5.116). In this section, it is shown that the geometric nonlinearity must not be ignored, because it substantially changes the nature of the interaction of even small amplitude modes. Equations describing the interaction of phonon modes near the phase transition point are constructed. It is remarkable that these equations admit two-dimensional soliton solutions. Regions of physical parameter values, where a certain type of soliton is realized and can be stable, are indicated. The above findings have been secured in [65–67].

The Ground State of Crystals and the Spectrum of Linear Modes

Without an external load, the equilibrium state of crystals is characterized by homogeneous static deformations, $u_{i,j}^{(0)} = \text{const}$. The minimum energy conditions $\partial w/\partial u_{i,j}^{(0)} = 0$ give a 4 mm phase with a square lattice ($u_{i,j}^{(0)} = 0$) and two versions of a 2 mm phase with a rectangular lattice ($u_{i,j}^{(0)} \neq 0$). In this section, we consider statistical and dynamic formations in the 4 mm phase background. An analysis carried out in [59, 61] shows that the square lattice phase is metastable for $0 < \tau \equiv 16C_2A_2/(3B_2^2) < 1$ and thermodynamically stable for $1 < \tau < 4/3$.

Let us examine the propagation of elastic waves along a preferential direction in the (x_1, x_2)-plane. A variation of the wave field in the direction orthogonal to the preferential one is assumed to be weaker. To describe such waves, it is convenient to proceed from x_1, x_2 to the new variables:

$$\xi = x_1 \cos\varphi + x_2 \sin\varphi\,, \quad \eta = -x_1 \sin\varphi + x_2 \cos\varphi\,.$$

Suppose that ξ is the coordinate along the wave propagation direction and that the dependence of the field variables on η is weaker. It is not difficult to show that the phonon spectrum has two Goldstone branches ($u_i \sim \exp \mathrm{i}[\mathrm{k}_1\xi + k_2\eta - \omega_i(k_1, k_2)t]\,, \quad k_1 \gg k_2$):

$$\begin{aligned}
&\rho\omega_i^2(k_1,k_2) = \alpha_i^{(20)}k_1^2 - \alpha_i^{(11)}k_1k_2 + \alpha_i^{(02)}k_2^2 + \alpha_i^{(40)}k_1^4\,, \quad i = 1,2\,;\\
&\alpha_i^{(20)} = \frac{1}{2}(\alpha - \varepsilon_i\delta)\,; \quad \alpha_i^{(11)} = \frac{1}{2}\varepsilon_i\delta^{-1}\left(\mu^2 - \nu^2\right)\sin 4\varphi\,;\\
&\alpha_i^{(02)} = \frac{\alpha}{2} + \frac{1}{4}\varepsilon_i\delta^{-1}\left\{\left[\delta^{-1}\left(\mu^2-\nu^2\right)\sin 4\varphi\right]^2 - \mu^2 - \nu^2 - 3\left(\mu^2-\nu^2\right)\cos 4\varphi\right\}\,;\\
&\alpha_i^{(40)} = \frac{1}{2}\left\{\frac{1}{2}\left(d_1 + d_2 + 2d_4 + \frac{1}{2}d_3\right) + \left(\frac{1}{\sqrt{2}}d_5 - d_4\right)\sin^2 2\varphi - \frac{1}{2}\varepsilon_i\delta^{-1}\left[\nu\left(d_1+\right.\right.\right.\\
&\qquad \left.\left.\left.+d_2 + 2d_4 - \frac{1}{2}d_3\right)\cos^2 2\varphi + \mu\left(d_1 - d_2 + \frac{1}{2}d_3 + \sqrt{2}d_5\right)\sin^2 2\varphi\right]\right\}\,;\\
&\alpha = \frac{1}{2}\left(A_1 + A_2 + \frac{1}{2}A_3\right)\,, \quad \mu = \left(A_1 - A_2 + \frac{1}{2}A_3\right)\,, \quad \nu = \frac{1}{2}\left(A_1 + A_2 - \frac{1}{2}A_3\right)\,,\\
&\delta^2 = (\nu\cos 2\varphi)^2 + (\mu\sin 2\varphi)^2\,, \quad \varepsilon_i = \pm 1\,.
\end{aligned} \tag{5.119}$$

For the directions $\varphi = 0, \pm\pi/4 \pmod{\pi}$, the dispersion law (5.119) is diagonalized ($\alpha_i^{(11)}(\varphi) = 0$) and agrees with that given in [59].

Effective 2D + 1 Equations of Nonlinear Dynamics of Phonon Modes

To derive simplified equations for low amplitude nonlinear waves against the background of the ground state (4 mm phase), we will write down the dynamic equations for the deviation $\overline{u}_i$ of the displacement field from the equilibrium state, with accuracy up to quadratic terms in the deviation amplitudes:

$$\rho\partial_t \mathbf{V} + \mathbf{D}(\nabla) : \mathbf{V} + \mathbf{E}(\varphi)\partial_\xi : \partial_\xi \mathbf{V} \circ \partial_\xi \mathbf{V} = 0 \,. \tag{5.120}$$

Here, $\mathbf{V} = (\overline{u}_1, \overline{u}_2)$, the differential operator $\mathbf{D}(\nabla)$ and the coefficients $\mathbf{E}(\varphi)$ are cumbersome, but can easily be calculated from (5.118). With the dependence of the field variables on the coordinate η being weak, we retained the derivatives over the variable η in (5.120) only in the linear terms of the equations. In doing so, we have neglected the fourth-order derivatives over η even in the linear terms. To build effective equations for the dynamics of low amplitude waves that propagate along the directions met in the condition $\alpha_i^{(11)}(\varphi) = 0$, we can use the nonlinear perturbation theory. In this case, the interaction of the Goldstone modes corresponding to a branch of the spectrum with the index i is conveniently described in a reference frame moving with the velocity $s_i = \partial_k \omega_i|_{k=0}$. The solution of the equations should be sought in the form:

$$\mathbf{V} = \sum_{n=1}^{\infty} \varepsilon^n \mathbf{V}^{(0)}(x, y, \tau) \,. \tag{5.121}$$

Here, $x = \varepsilon(\xi + s_i t)$, $y = \varepsilon^2 \eta$, $\tau = \varepsilon^3 t$ are slow variables. The parameter ε is necessary for visual grouping of terms of the same magnitude order. At the end of the calculations, we set $\varepsilon = 1$. For the above directions of wave propagation, the solvability conditions for the equations of lower orders in the parameter ε are fulfilled automatically. The solvability requirement for a system arising in the order ε^5 of perturbation theory yields a closed effective equation that coincides with the Kadomtsev–Petviashvili (KP) equation (5.124). The KP model describes the propagation of waves in one direction at a speed close to the speed of sound s_i. For the purpose of avoiding these limitations, we apply a more general technique.

We go over from the fields $\overline{u}_1$ and $\overline{u}_2$ to normal modes R and Q:

$$\mathbf{S}(\nabla)\begin{pmatrix}\overline{u}_1\\ \overline{u}_2\end{pmatrix} = \begin{pmatrix}R\\ Q\end{pmatrix} ; \quad \begin{pmatrix}\overline{u}_1\\ \overline{u}_2\end{pmatrix} = \mathbf{S}^{-1}(\nabla)\begin{pmatrix}R\\ Q\end{pmatrix} .$$

The operator $\mathbf{S}(\nabla)$ is chosen so as to diagonalize equations (5.120), which are linearized near the ground state $\mathbf{V} = 0$:

$$\rho\partial_t^2 R + \rho\omega_1^2(\mathrm{i}\partial_\xi, \mathrm{i}\partial_\eta)R + \cdots = 0 \,, \quad \rho\partial_t^2 Q + \rho\omega_2^2(\mathrm{i}\partial_\xi, \mathrm{i}\partial_\eta) + \cdots = 0 \,. \tag{5.122}$$

The explicit form of the differential operators $\rho\omega_j^2(\mathrm{i}\partial_\xi, \mathrm{i}\partial_\eta)$, $(j = 1, 2)$ in (5.122) is dictated by the dispersion laws (5.119). We act as the operator $\mathbf{S}(\nabla)$ to equation (5.120). The linear terms of the equations take the form (5.122). In the longwave approximation, to transform the nonlinear terms (5.121), it is sufficient for us to restrict ourselves to the expressions for $\mathbf{S}(\nabla)$ and $\mathbf{S}^{-1}(\nabla)$, which do not contain derivatives:

$$\mathbf{S}(\nabla) = \begin{pmatrix} 1 & N \\ N & -1 \end{pmatrix}, \quad \mathbf{S}^{-1}(\nabla) = \frac{\mathbf{S}(\nabla)}{1+N^2}, \quad N(\varepsilon_1) = \left[\mu \sin 2\varphi\right]^{-1} (-\nu \cos 2\varphi - \varepsilon_1 \delta).$$

It should be emphasized that this approximation agrees with the reductive perturbation theory (5.121). For directions that satisfy the condition $\sin 2\varphi = 0$, it is convenient to choose $\varepsilon_1 = -\operatorname{sign}(\nu \cos 2\varphi)$. Then the expression for N permits the passage to the limit of $N \to 0$ as $\sin 2\varphi \to 0$.

When exciting one of the branches of the spectrum and $R \gg Q$, the interaction with the modes of the other branch can be neglected. Then the modes of the main branch of the spectrum evolve by the closed equation:

$$\begin{gathered} \rho\partial_t^2 R + \rho\omega_1^2(\mathrm{i}\partial_\xi, \mathrm{i}\partial_\eta)R - g_1\partial_\xi^2 R^2 = 0, \\ g_1 = \tfrac{3}{4}\left(1+N^2\right)^{-1}\left[A_1(\cos\varphi + N\sin\varphi) + \tfrac{1}{2}A_3(\sin\varphi + N\cos\varphi)\sin 2\varphi + \right. \\ \left. + A_2\cos 2\varphi(\cos\varphi - N\sin\varphi)\right]. \end{gathered} \tag{5.123}$$

In the other limit case, when $Q \gg R$, $\sin 2\varphi \neq 0$, an equation for Q comes out of (5.123) after the formal substitution $\varepsilon_1 \to -\varepsilon_1$, $R \to Q/N(\varepsilon_1)$ as $N^{-1}(-\varepsilon) = -N(\varepsilon)$. For the sake of definiteness, we will consider equation (5.123), omitting the lower index of the parameters g_1, $\alpha_1^{(ik)}$, etc.

Let us look into waves moving in one direction along the direction where $\alpha^{(11)}(\varphi) = 0$, with velocities close to the speed of sound ($s = \sqrt{\alpha^{(20)}/\rho}$). Then, assuming that $R = R(x, \eta, t)$, where $x = \xi + st$, and taking into account that $[\rho\partial_t^2 - \alpha^{(20)}\partial_\xi^2]R \cong 2\rho s\partial_t\partial_x R$, we reduce equation (5.122) to the KP model integrable by means of the inverse scattering problem [68]:

$$\partial_x\left[2\rho s\partial_t R + \alpha^{(40)}\partial_x^3 R - g\partial_x R^2\right] = \alpha^{(02)}\partial_\eta^2 R. \tag{5.124}$$

For purely one-dimensional motions (the dependence on η can be neglected), equation (5.124) becomes the Korteweg–de Vries equation (KdV). It is essential that neglecting of geometric nonlinearity in the expression for the tensor η_{ij} within the 1D model destroys cubic displacement terms in the expression for the energy [59, 61]. Consequently, the equation for one-dimensional waves in the small amplitude limit reduces a modified KdV equation with a stronger cubic nonlinearity (it is less important for low-amplitude waves), rather than to the KdV equation with a nonlinearity quadratic in displacements.

It has been shown in [66] that the low amplitude formations against the background of the 2 mm phase are described by effective equations similar to (5.123). The

only difference is to renormalize the elastic moduli and the constants of the wave interaction due to spontaneous deformations $(u_{i,j}^{(0)}) \neq 0$. This circumstance ultimately leads to a nontrivial temperature dependence of the parameters of the nonlinear excitations.

Two-Dimensional Solitons as Precursors of a Martensitic Transition

Equation (5.123) allows a Bäcklund transformation: if u_0 is a solution of equation (5.123) and the function φ satisfies the equation

$$\left[\rho D_t^2 + \rho\omega^2(\mathrm{i}D_\xi, \mathrm{i}D_\eta) - 2gu_0 D_\xi^2\right] f \cdot f = 0\,, \tag{5.125}$$

then $u_1 = -6\alpha^{(40)} g^{-1} \partial_\xi^2 \ln f + u_0$ is also a solution of equation (5.123) [65–67]. Here, $D_t f \cdot g = (\partial_t - \partial_{t'}) f(t) g(t')|_{t=t'}$, etc. The bilinear form (5.125) allows the employment of the Hirota method [69] to obtain soliton like solutions of the equation (5.123). In particular, the author of [65–67] has found an N-soliton exponential solution:

$$\begin{gathered}
R = -6\alpha^{(40)} g^{-1} \partial_\xi^2 \ln f\,, \\
f = \sum_{\mu=0,1} \exp\left[\sum_{i>j} A_{ij}\mu_i\mu_j + \sum_{i=1}^{N} \mu_i\eta_i\right]\,, \quad \kappa^2(p,q) = -\rho\omega^2(ip, iq)\,, \\
\exp\eta_i = \exp\left[\Omega_i t + p_i\xi + q_i\eta + \eta_{0i}\right]\,, \quad \Omega_i = \sigma_i\left[\rho^{-1}\kappa^2(p_i,q_i)\right]^{\frac{1}{2}}\,, \quad \sigma_i = \pm 1\,, \\
\exp A_{ij} = -\left[\rho(\Omega_i - \Omega_j)^2 - \kappa^2(p_i - p_j, q_i - q_j)\right]\left[\rho(\Omega_i + \Omega_j)^2 - \kappa^2(p_i + p_j, q_i + q_j)\right]^{-1}\,.
\end{gathered} \tag{5.126}$$

Here, $\sum_{\mu=0,1}$ means summation over all possible combinations of $\mu = 0, 1$, and $\sum_{i>j}$ means summation over all possible pairs of N elements. The parameters $\Omega_j, p_j, q_i, \eta_{0j}$ must satisfy the reductions guaranteeing the reality of f.

When $N = 2M$, and the parameters are also connected in pairs in such a way that $\Omega_s = \Omega_{s+M}^*, p_s = p_{s+M}^*$, etc., $(s = 1, 2, \ldots, M)$, expression (5.126) controls a system of M pulsating solitons. Such solitons become singular at some space time points, so they are difficult to physically interpret.

When the parameters $\Omega_j, p_j, q_i, \eta_{0j}$ are real, solution (5.126) delineates elastic interactions of N flat solitons of the type:

$$R^{(0)} = -\frac{6\alpha^{(40)} d^2}{g\,\mathrm{ch}^2\theta}\,, \quad \theta = d(\xi + \mathrm{v}t)\,, \quad d^2 = \frac{\alpha^{(20)} - \rho\mathrm{v}^2}{4\alpha^{(40)}} > 0\,. \tag{5.127}$$

Here, v is a real parameter (a soliton velocity). The flat solitons are not singular.

It is interesting to note that, for the solitons (5.126), (5.127) associated with the branch of the spectrum $\omega_1(k)$ of the linearized problem ($\varepsilon_1 = \pm\,\mathrm{sign}\,\mu \sin 2\varphi$) and

propagating in the directions $\varphi = \pm(\pi/4)(\mathrm{mod}\,\pi)$ against the 4 mm phase backdrop, we have $N = \mp 1$. Consequently, in the main approximation, we arrive at:

$$\begin{aligned} e_1 &\approx \left[\sqrt{2}(1+N^2)\right]^{-1} \partial_\xi R^{(0)}(\cos\varphi + N\sin\varphi) = 0\,, \\ e_3 &\approx \left[\sqrt{2}(1+N^2)\right]^{-1} \partial_\xi R^{(0)}(\sin\varphi + N\cos\varphi) = 0\,, \\ e_2 &\approx \left[\sqrt{2}(1+N^2)\right]^{-1} \partial_\xi R^{(0)}(\cos\varphi - N\sin\varphi) \neq 0\,. \end{aligned} \tag{5.128}$$

Since the field e_2 is a parameter of the theory order, its deviation from zero in the soliton localization region means that the square lattice, typical of the ground state, turns into a rectangular lattice inside the soliton (Figure 5.23). Although the rectangular lattice corresponds to a new phase, in this case the phase transition does not yet occur, since the amplitude of the soliton is small. However, forming the inhomogeneous internal structure of the ground state with deformations typical of the other phase indicates the appearance of tendencies for a phase transition to take place in the system. Apparently, the solitons (5.128) can be interpreted as low amplitude twinned interlayers parallel to the (110) planes, or as residual traces of the low temperature 2 mm phase (martensite) against the background of the main 4 mm phase (austenite). Quasiperiodic finite zone solutions of the KP model (5.124) correspond to small amplitude twinned strips. Relations (5.128) meet no other directions of the wave propagation. Other solitons cannot be treated as precursors of a new phase, because their excitations are far too involved. Interestingly, near the cubic tetragonal martensitic transition, the twinned interlayers parallel to the (110) planes and tetragonal modulations against the 4 mm phase backdrop are experimentally observed [59, 70].

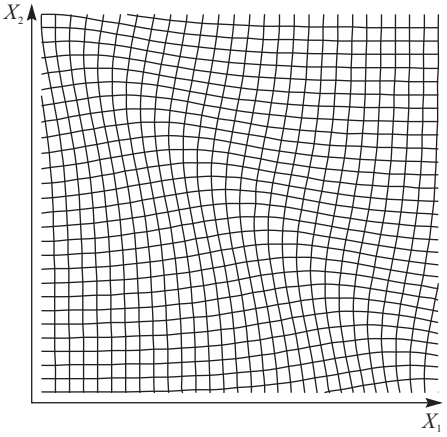


Fig. 5.23: A flat soliton (5.127) is a precursor of a martensitic transition: $\varphi = \pi/4$, $\alpha^{(40)} g > 0$.

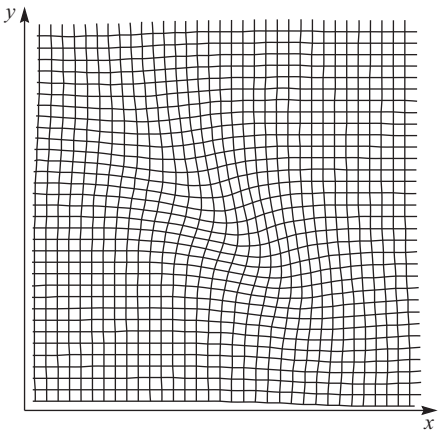


Fig. 5.24: Deformations of a square lattice, generated by a cigar shaped soliton (5.129): $\varphi = \pi/4$, $\alpha^{(40)} g > 0$.

Besides the exponential solutions, there are also polynomial solutions of the equation (5.128). The simplest of them is a two-dimensional cigar shaped soliton:

$$
\begin{aligned}
R &= \frac{12\alpha^{(40)}\left[A^2(\xi + \mathrm{v}t)^2 - (AB - 2C^2)\eta^2 + 2CA\eta(\xi + \mathrm{v}t) - A\right]}{g\left[1 + A(\xi + \mathrm{v}t)^2 + B\eta^2 + 2C\eta(\xi + \mathrm{v}t)\right]^2}\,; \\
A &= \left[4\alpha^{(02)}(\alpha^{(20)} - \rho\mathrm{v}^2) - (\alpha^{(11)})^2\right]\left[12\alpha^{(40)}\alpha^{(02)}\right]^{-1}\,, \\
B &= (\alpha^{(20)} - \rho\mathrm{v}^2)\left[\alpha^{(02)}\right]^{-1} A\,, \quad C = \alpha^{(11)} A\left[2\alpha^{(02)}\right]^{-1}\,.
\end{aligned}
\tag{5.129}
$$

The requirement of boundedness of the solution (5.129) reduces it to the condition: $A, B > 0, AB - C^2 > 0$. The two-dimensional soliton (5.129) (Figure 5.24) is spatially localized ($R = O(r^{-2})$ as $r \to \infty$); it can be motionless ($\mathrm{v} = 0$). The initial assumption of a weak dependence of the waves on the variable η implies a limitation on the applicability region for solution (5.129):

$$
\alpha^{(11)} \ll 2\alpha^{(02)}, (\alpha^{(20)} - \rho\mathrm{v}^2) \ll \alpha^{(02)}\,.
$$

The solitons may emerge under the action of impulsive forces applied to the surface of the sample. They can locally arise during its deformation, such as at grain boundaries which have suffered a local phase transition. In addition, the solitons can arise as pretransitional states as a result of energy exchange between different phonon branches, and due to progressing instability in the phonon system [71].

Stability of Solitons

For working out various soliton scenarios, the answer to the question of the stability of solitons is important. In the KP model (5.124), the stability or instability of solitons depends on the sign of the combination of $\alpha^{(40)}\alpha^{(02)}$. In [68] it is established that, for

$\alpha^{(40)}\alpha^{(02)} < 0$, flat solitons of the type (5.127) are stable, relative to two-dimensional perturbations. A nontrivial picture occurs for $\alpha^{(40)}\alpha^{(02)} > 0$. In this case, the flat solitons of the KP model are unstable [72], but cigar-shaped solitons of the type (5.129) that are stable relative to two-dimensional perturbations exist. The nonlinear stage of development of the instability of flat solitons of the KP model has been investigated by [72–76]. The author of [73], in particular, claims that an unstable flat soliton spreads as a result of the excitation of small oscillations on its front, which move faster than the soliton itself, overtake it, and take away a part of its energy. The radiated energy is further dissipated in space due to dispersion effects. An alternative has been found in [76]. Once the flat soliton decays, chains of cigar-shaped solitons and a new flat soliton are formed, with the amplitudes of these new solitons being less than that of the original soliton. Finally, there is no dissipation of the energy, but condensation into new soliton structures is produced.

As compared to the KP model, the more general equation (5.123) requires modifying the soliton stability conditions. Let us construct a criterion for the instability of the flat soliton (5.127), relative to two-dimensional perturbations. Here, it is worth noting that the solution of (5.127) exists only in a definite interval of soliton velocity values. In particular, for $\alpha^{(40)}, \alpha^{(20)} > 0$, the soliton velocity must be less than the sound velocity $0 < v^2 < \alpha^{(20)}/\rho$. For $\alpha^{(20)} > 0$, $\alpha^{(40)} < 0$, the soliton moves with a supersonic velocity.

We seek the solution of the perturbed equation (5.123) in the form [77]:

$$R = R^{(0)}\left[\theta + \alpha(Y_i, T_i)\right] + \sum_{n=1}^{\infty} \varepsilon^n R^{(n)}\left[\theta + \alpha(Y_i, T_i), Y_i, T_i\right] . \tag{5.130}$$

Here ε is a small parameter characterizing the ratio of diffraction and nonlinearity, $Y_i = \varepsilon^i \eta$, $T_i = \varepsilon^i t$ are slow variables, and i are natural numbers. By substituting the expansion (5.130) into (5.123), arraying terms of the same order over ε, and by neglecting higher powers of α (linear approximation), we obtain:

$$\begin{aligned} d^4\alpha^{(40)}\left[-4\partial_\theta^2 + \partial_\theta^4\right]R^{(0)} - gd^2\partial_\theta^2\left[R^{(0)}\right]^2 = 0 , \\ \hat{L}R^{(1)} = -\left[2v\rho d\frac{\partial\alpha}{\partial T_1} + \alpha^{(11)} d\frac{\partial\alpha}{\partial Y_1}\right]\partial_\theta^2 R^{(0)} ; \end{aligned} \tag{5.131}$$

$$\begin{aligned} \hat{L}R^{(2)} = &-\rho\left[\frac{\partial^2\alpha}{\partial T_1^2}\partial_\theta R^{(0)} + 2vd\frac{\partial\alpha}{\partial T_2}\partial_\theta^2 R^{(0)} + 2vd\frac{\partial^2 R^{(1)}}{\partial\theta\partial T_1} + 2vd\frac{\partial\alpha}{\partial T_1}\partial_\theta^2 R^{(1)}\right] - \\ &- \alpha^{(11)}\left[d\frac{\partial\alpha}{\partial Y_2}\partial_\theta^2 R^{(0)} + d\frac{\partial\alpha}{\partial Y_1}\partial_\theta^2 R^{(1)} + d\frac{\partial^2 R^{(1)}}{\partial\theta\partial Y_1}\right] + \\ &+ \alpha^{(02)}\frac{\partial^2\alpha}{\partial Y_1^2}\partial_\theta R^{(0)} + gd^2\partial_\theta^2\left[R^{(1)}\right]^2 , \\ \hat{L}\varphi \equiv d^4\alpha^{(40)}&\left[-4\partial_\theta^2 + \partial_\theta^4\right]\varphi - 2gd^2\partial_\theta^2\left[R^{(0)}\varphi\right] . \end{aligned} \tag{5.132}$$

The conditions for the absence of secular terms in (5.131) and (5.132) are the essence of the orthogonality conditions for the right-hand sides of these equations to

the eigenfunction $\Psi = \int \mathrm{d}\theta R^{(0)}(\theta)$ of the operator adjoint to $\hat{\mathrm{L}}$ (the emerging arbitrary integration constants are eliminated by the requirement of decreasing $R^{(i)}(\theta)$ at infinity). For equation (5.131), the orthogonality condition is automatically satisfied. The solution has the form:

$$R^{(1)} = -\left[2\mathrm{v}\rho\frac{\partial\alpha}{\partial T_1} + \alpha^{(11)}\frac{\partial\alpha}{\partial Y_1}\right]\left[R^{(0)} + \frac{\theta}{2}\partial_\theta R^{(0)}\right]/4d^3\alpha^{(40)} .$$

The orthogonality condition for equation (5.132) gives an evolution of α in the geometrical optics approximation:

$$\begin{gathered}\gamma_1\frac{\partial^2\alpha}{\partial T_1^2} - \gamma_2\frac{\partial^2\alpha}{\partial Y_2^2} - \gamma_3\frac{\partial^2\alpha}{\partial Y_1\partial T_1} = 0 , \quad \gamma_1 = \rho(\alpha^{(20)} - 4\rho\mathrm{v}^2) , \\ \gamma_2 = \alpha^{(02)}(\alpha^{(20)} - \rho\mathrm{v}^2) + \frac{3}{4}(\alpha^{(11)})^2 , \quad \gamma_3 = 3\mathrm{v}\rho\alpha^{(11)} .\end{gathered} \tag{5.133}$$

From (5.133), the criterion of soliton instability follows:

$$\gamma_3^2 + 4\gamma_1\gamma_2 < 0 .$$

In particular, for $\alpha^{(11)} = 0$, $\alpha^{(20)} > 0$, $\alpha^{(40)} > 0$, the solution (5.127) can be stable only for $\alpha^{(02)} > 0$, with the range of the soliton motion velocities being narrower than that determined by formula (5.127): $0 < \mathrm{v}^2 < \alpha^{(20)}/4\rho$. In the region of high velocities ($\alpha^{(20)}/4\rho < \mathrm{v}^2 < \alpha^{(20)}/\rho$), such a soliton can be stable only for $\alpha^{(02)} < 0$.

The stability line can be achieved by a soliton due to temperature inhomogeneities or local stresses that change the shear moduli and, therefore, affect the soliton stability as well as the temperature. As instability is in progress, the solitons can emit phonons. Apparently, this allows us to take a fresh look at the problem of *anomalous acoustic emission* near the martensitic phase transitions [78].

Exercises

1. Electrons and Holes in Metals and Semiconductors

Exercise 1
Show that:

$$V_b = (2\pi)^3 / V_a \,.$$

Exercise 2
Construct the first, second, and third Brillouin zones for a plane square lattice with a lattice constant a. Consider cases when an elementary cell contains $p = 1, 2, 3, 4$ electrons. Depict the shape of the Fermi surface for each case in the reduced zone scheme. Depict the 4-th Brillouin zone.

Exercise 3
1. Having substituted

$$\Psi_{n\vec{k}}(\vec{r}) = u_{n\vec{k}}(\vec{r}) \exp\left(\mathrm{i}\vec{k}\cdot\vec{r}\right)$$

into the Schrödinger equation

$$\widehat{H}\Psi_{n\vec{k}} = \varepsilon_n(k)\Psi_{n\vec{k}} \,,$$

show that:

$$-\frac{\hbar^2}{2m}\Delta u_{n\vec{k}} + V(\vec{r})u_{n\vec{k}} - \frac{\mathrm{i}\hbar^2}{m}(\vec{k}\cdot\nabla)u_{n\vec{k}} + \frac{\hbar^2k^2}{2m}u_{n\vec{k}} = \varepsilon_n(\vec{k})u_{n\vec{k}} \,.$$

2. Considering the operator:

$$\widehat{H}' = -\frac{\mathrm{i}\hbar^2}{m}(\vec{k}\nabla) + \frac{\hbar^2k^2}{2m} \,,$$

as a perturbation operator ($|\vec{k}| \ll \pi/a$), calculate the corrections $O(k)$ and $O(k^2)$ to the electron energy $\varepsilon_n(\vec{k} = 0)$. Using this result, derive in terms of $u_{n\vec{0}}$ the expression for the average electron velocity $\vec{V} = (1/\hbar)(\partial\varepsilon_n/\partial\vec{k})|_{\vec{k}=0}$, as well as the tensor of its reciprocal effective mass $M_{ij}^{-1} = (1/\hbar)^2(\partial^2\varepsilon_n/\partial k_i\partial k_j)|_{\vec{k}=0}$.

Exercise 4
Show that when an electron reflects from the Bragg planes in crystals with a center of inversion, the wave functions of an electron are standing waves:

$$\Psi_+ = \sqrt{\frac{2}{V}} \exp\left\{-\frac{\mathrm{i}}{\hbar}\varepsilon_+\left(\frac{\vec{K}}{2}\right)t\right\} \sin\frac{1}{2}\vec{K}\cdot\vec{r} \,,$$

$$\Psi_- = \sqrt{\frac{2}{V}} \exp\left\{-\frac{\mathrm{i}}{\hbar}\varepsilon_-\left(\frac{\vec{K}}{2}\right)t\right\} \cos\frac{1}{2}\vec{K}\cdot\vec{r} \,,$$

where V is the volume of the crystal. The electron energy is $\varepsilon_\pm(\vec{K}/2) = \varepsilon_0(\vec{K}/2) \pm |U_{\vec{K}}|$.

https://doi.org/10.1515/9783110586183-006

Exercise 5

Show that for free electrons:

$$\frac{\partial f}{\partial T} = -\frac{\partial f}{\partial \varepsilon}\left[\frac{\varepsilon-\mu}{T} + \frac{\mathrm{d}\mu}{\mathrm{d}T}\right],$$
$$\frac{\partial f}{\partial \varepsilon} = -\frac{1}{4k_{\mathrm{B}}T}\mathrm{ch}^{-2}\left(\frac{\varepsilon-\mu}{2k_{\mathrm{B}}T}\right),$$

where f is the Fermi–Dirac distribution function.

Exercise 6

Suppose silver is a monovalent metal with a spherical Fermi surface. Calculate the following values of:

1. the Fermi energy (eV) and the Fermi temperature
2. the radius of the Fermi sphere in k-space
3. the Fermi velocity
4. the cross sectional area of the Fermi surface
5. cyclotron frequency (Hz) in a magnetic field with a voltage H = 5000 Oe
6. a mean free path of electrons at room temperature (295 K) and near absolute zero temperatures (20 K)
7. the radius of the cyclotron orbit in the field H = 5000 Oe
8. length of the edge of a cubic unit cell
9. length of the reciprocal lattice vectors
10. volume of the first Brillouin zone

There is table data for Ag: the density is $d = 10.5\mathrm{g/cm}^3$, the atomic weight is $A = 107.87$, the specific resistance is $\rho = 1.61\cdot 10^{-6}\Omega\cdot\mathrm{cm}$ at T = 295 K, and $\rho = 0.0038\cdot 10^{-6}\Omega\cdot\mathrm{cm}$ at T = 20 K.

Exercise 7

Show that the expression for the density of electron levels in a crystal can be written in two equivalent forms:

$$\nu(\varepsilon) = \frac{1}{4\pi^3}\int\limits_{\varepsilon(k)=\varepsilon}\frac{\mathrm{d}S}{\left|\nabla_{\vec{k}}\varepsilon\right|} = \frac{1}{4\pi^3}\int\limits_{V_b}\delta\left(\varepsilon-\varepsilon(\vec{k})\right)\mathrm{d}^3\vec{k}\,.$$

Exercise 8

Check directly that the function $\varphi = (Q/r)\exp(-\lambda r)$ is a solution of the differential equation:

$$-\Delta\varphi = 4\pi Q\delta(\vec{r}) - \lambda^2\varphi\,.$$

Exercise 9

Check directly that the semiclassical equations of motion of an electron:

$$\dot{\vec{r}} = \vec{V} = \frac{1}{\hbar}\frac{\partial \varepsilon(\vec{k})}{\partial \vec{k}}\,, \quad \hbar\dot{\vec{k}} = -|e|\left\{\vec{E}(\vec{r},t) + \frac{1}{c}\left[\vec{V}(\vec{k}) \times \vec{B}(\vec{r},t)\right]\right\}\,,$$

can be written in the Hamiltonian form:

$$\dot{\vec{r}} = \frac{\partial H}{\partial \vec{p}}\,, \quad \dot{\vec{p}} = -\frac{\partial H}{\partial \vec{r}}\,,$$

where $H = \varepsilon((1/\hbar)[\vec{p} + (|e|/c)\vec{A}(\vec{r},t)]) - |e|\varphi(\vec{r},t)$, wherein:

$$\vec{B} = \operatorname{rot}\vec{A}\,, \quad \vec{E} = -\nabla\varphi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t}\,, \quad \hbar\vec{k} = \vec{p} + \frac{|e|}{c}\vec{A}\,.$$

Note

In the calculations, use the abbreviated notations:

$$\left[\vec{V} \times \vec{B}\right]_s = e_{spk}V_pB_k\,, \quad \left(\operatorname{rot}\vec{A}\right)_s = e_{spk}\nabla_pA_k\,,$$

and identity:

$$e_{spk}e_{klm} = \delta_{sl}\delta_{pm} - \delta_{sm}\delta_{pl}\,,$$

where e_{spk} is an absolutely antisymmetric unit tensor ($e_{123} = 1$).

Exercise 10

Show that the dispersion law:

$$\varepsilon(\vec{k}) = \frac{\hbar^2}{2}\left[\frac{k_1^2}{m_1} + \frac{k_2^2}{m_2} + \frac{k_3^2}{m_3}\right]\,,$$

where $\vec{k} = (k_1, k_2, k_3)$, governs the period of motion of an electron along a closed orbit in a magnetic field $\vec{B} = (0, 0, B)$ by the formula $T = (2\pi c)/(|e|B)\sqrt{m_1m_2}$. The quantity $m_c = \sqrt{m_1m_2}$ is called the cyclotron mass of an electron.

Exercise 11

Show that a magnetic flux Φ enclosed by a closed orbit of an electron can only vary in a constant magnetic field by quanta:

$$\Delta\Phi = -\frac{hc}{|e|}\Delta n\,.$$

Here h is the Planck constant, c is the speed of light in a vacuum, e is the charge of an electron, and Δn is an integer.

Note
Use the Bohr–Sommerfeld formula:

$$\oint \vec{p} \cdot \mathrm{d}\vec{r} = (n + \gamma)\, 2\pi\hbar \,,$$

where n is an integer and γ is a phase correction (its typical value is $\gamma \approx 1/2$). In the given exercise

$$\vec{p} = \hbar\vec{k} - \frac{|e|}{c}\vec{A} \,.$$

Exercise 12

1. Calculate the concentration of electrons and holes, as well as an expression for the chemical potential in a strongly doped semiconductor of n-type ($N_\mathrm{d} \gg n_i$), provided that the donor levels are completely ionized.
2. Show that the chemical potential $\mu(T)$ in an n-type semiconductor in the region of ultralow temperatures has a local maximum lying in the interval $((\varepsilon_\mathrm{c} + \varepsilon_\mathrm{d})/2, \varepsilon_\mathrm{c})$. Calculate the value of the maximum and the corresponding temperature.

2. Crystal Lattice Vibrations

Exercise 13
Show that for a one-dimensional diatomic lattice we have:

$$\lim_{q\to 0}\left(\frac{A_1}{A_2}\right)\Big|_\mathrm{acoust} = 1,\ \lim_{q\to 0}\left(\frac{A_1}{A_2}\right)\Big|_\mathrm{optic} = -\frac{M_2}{M_1} \,,$$

where A_1, A_2 are the amplitudes of vibrations of lattice atoms with masses M_1, M_2, respectively, and q is a wavenumber.

Exercise 14
Show that for a one-dimensional diatomic lattice the oscillation frequencies $\omega_i(q)$ ($i = 1, 2$) as $q \to 0$ satisfy the following relations:

$$\omega_1(q) \approx c_l q \,, \quad \omega_2(q) \approx \omega_0\left(1 - \frac{\gamma^2 q^2 a^2}{8}\right) \,,$$

where

$$c_l = \sqrt{2\alpha/(M_1 + M_2)}\, a \,,$$
$$\omega_0 = \sqrt{2\alpha}\sqrt{(M_1 + M_2)/(M_1 M_2)} \,,$$
$$\gamma = 2\sqrt{M_1 M_2}/(M_1 + M_2) \,,$$

a is the distance between neighboring atoms with masses M_1, M_2, and α is the constant of quasi-elastic interaction between atoms.

These relations imply that $(\partial\omega_2/\partial q)|_{q=0} = 0$, i.e., the curve $\omega_2 = \omega_2(q)$ has a horizontal tangent at a point $q = 0$.

Exercise 15

Show that for a diatomic one-dimensional lattice we have:

$$\Delta\omega = \sqrt{2\alpha}\left(\frac{1}{\sqrt{M_2}} - \frac{1}{\sqrt{M_1}}\right),$$

and for $q = \pi/2a$ $(M_1 > M_2)$,

$$\left.\frac{\partial\omega_1}{\partial q}\right|_{q=\pi/2a} = \left.\frac{\partial\omega_2}{\partial q}\right|_{q=\pi/2a} = 0\,.$$

The last equalities mean that the curves $\omega_i = \omega_i(q)$ $(i = 1, 2)$ have horizontal tangents at the boundaries of the first Brillouin zone for $q = \pm\pi/2a$.

Exercise 16

Show that the dynamic matrix of crystals is real in crystals with a center of inversion:

$$\varphi^*_{ss'}(\vec{q}) = \varphi_{ss'}(\vec{q})\,.$$

Exercise 17

Show that the equation of lattice dynamics can be rewritten in the form:

$$M_\kappa \ddot{U}_\alpha\begin{pmatrix} l \\ \kappa \end{pmatrix} = -\sum_{l',\kappa',\alpha'} \Phi_{\alpha\alpha'}\begin{pmatrix} l & l' \\ \kappa & \kappa' \end{pmatrix}\left[U_{\alpha'}\begin{pmatrix} l \\ \kappa \end{pmatrix} - U_{\alpha'}\begin{pmatrix} l' \\ \kappa' \end{pmatrix}\right].$$

Exercise 18

Show that if $R_m = ma$ $(1 \le m \le N)$ and $k = 2\pi n/Na$ $(m, n, N$ are integers), the following identity holds:

$$\frac{1}{N}\sum_{m=1}^{N} \exp(\mathrm{i}kR_m) = \begin{cases} 0\,, & \text{when} \quad k \neq \frac{2\pi}{a}s \equiv K \quad (s \text{ is integer})\,; \\ 1\,, & \text{when} \quad k = \frac{2\pi}{a}s \equiv K\,. \end{cases}$$

3. Superconductivity

Exercise 19

Estimate the electron-phonon interaction constant:

$$g_{s'\vec{k}',s\vec{k}}\begin{pmatrix} \vec{q} \\ j \end{pmatrix} \sim \mathrm{i}\sqrt{\frac{\hbar N}{2\omega_j(\vec{q})}}\frac{(\vec{e}\cdot\tilde{\vec{q}})}{V\sqrt{M}}e^2a^2\delta_{\vec{k}'-\vec{k},\tilde{\vec{q}}}\,,$$

where $\vec{e} \equiv \vec{e}\left(\kappa \,\middle|\, \begin{matrix} \vec{q} \\ j \end{matrix}\right)$ is the polarization vector of a phonon; $\tilde{\vec{q}} = \vec{q} + \vec{K}$, $\vec{q}$ is the wave vector of a phonon, and $\vec{K}$ is the reciprocal-lattice vector.

When $\vec{K} = 0$, the quantity $|(\vec{e}\cdot\vec{q})|$ is maximum when $\vec{e} \parallel \vec{q}$, i.e., the interaction of electrons with longitudinal lattice vibrations is the most important.

Exercise 20

Show that, in calculating the correction to the energy of two electrons through the exchange of a virtual phonon, the matrix elements of the interaction operator have the form:

$$\left|\left\langle \vec{k}_1, \vec{k}_2, 0 \left| \widehat{H}_{\text{int}} \right| \vec{k}_1 - \vec{q}, \vec{k}_2, 1 \right\rangle\right|^2 = \left|\left\langle \vec{k}_1, \vec{k}_2, 0 \left| \widehat{H}_{\text{int}} \right| \vec{k}_1, \vec{k}_2 + \vec{q}, 1 \right\rangle\right|^2 = \left|g_{\vec{k}_1, \vec{k}_1 - \vec{q}}\right|^2 .$$

Exercise 21

Consider the case of remaining the modulus of a wave function $\Psi(\vec{r})$ of the condensate of Cooper pairs as a constant in the Ginzburg–Landau theory, but only the phase $\varphi(\vec{r})$ changes.

Show that the expression for the current density in a superconductor takes the form:

$$\vec{j} = -\left[\frac{2e^2}{mc}\vec{A} + \frac{|e|\hbar}{m}\vec{\nabla}\varphi\right] n_s ,$$

where $n_s = |\Psi|^2$ is the density of superconducting electrons (the wave function is properly normalized).

4. Quantum Optics – Interaction of Radiation with Matter

Exercise 22

Show that the energy of an electromagnetic field in a cavity can be written in the form:

$$W = \frac{1}{2}\int \mathrm{d}^3\vec{r}\left[\varepsilon_0 \vec{E}^2 + \mu_0 \vec{H}^2\right] = \sum_{\vec{k},\lambda} \hbar\omega_k a^*_{\vec{k}\lambda}(t) a_{\vec{k}\lambda}(t) .$$

Note

Use the identities:

$$\frac{1}{V}\int_V \exp\left[\mathrm{i}\left(\vec{k} \pm \vec{k}'\right)\cdot\vec{r}\right]\mathrm{d}^3\vec{r} = \delta_{\vec{k},\mp\vec{k}'} ;$$

$$\left(\vec{e}_{\vec{k}\lambda}\cdot\vec{e}_{\vec{k}\lambda'}\right) = \delta_{\lambda\lambda'} ; \quad \left(\left[\vec{n}_{\vec{k}}\times\vec{e}_{\vec{k}\lambda}\right]\cdot\left[\vec{n}_{\pm\vec{k}}\times\vec{e}_{\pm\vec{k}\lambda'}\right]\right) = \pm\left(\vec{e}_{\vec{k}\lambda}\cdot\vec{e}_{\pm\vec{k}\lambda'}\right) .$$

Exercise 23

Show that eigenvectors corresponding to a given number of photons can be written in the form:

$$|n_{\vec{k}\lambda}\rangle = \frac{(\widehat{a}_{\vec{k}\lambda})^{n_{\vec{k}\lambda}}}{\sqrt{n_{\vec{k}\lambda}!}}\,|0\rangle , \quad \langle n_{\vec{k}'\lambda'}|n_{\vec{k}\lambda}\rangle = \delta_{\vec{k}\vec{k}'}\delta_{\lambda\lambda'} .$$

Exercise 24
Show that the evolution of operators of creation and annihilation of photons is determined in the Heisenberg picture by the equations:

$$\hat{a}_{\vec{k}\lambda}(t) = \hat{a}_{\vec{k}\lambda}(0)\exp(-\mathrm{i}\omega_k t)\ , \quad \hat{a}_{\vec{k}\lambda}(t) = \hat{a}_{\vec{k}\lambda}(0)\exp(\mathrm{i}\omega_k t)\ .$$

Exercise 25
Show that:

$$[\cos\hat{\varphi}, \sin\hat{\varphi}] = \frac{1}{2\mathrm{i}}\{\hat{a}^+(\hat{n}+\hat{I})^{-1}\hat{a} - \hat{I}\}\ .$$

Using the result, make sure that all diagonal elements of this commutator are zero except for one:

$$\langle 0|\ [\cos\hat{\varphi}, \sin\hat{\varphi}]\ |0\rangle = -1/2\mathrm{i}\ .$$

Exercise 26
Show that:

$$[\hat{n}, \cos\hat{\varphi}] = -\mathrm{i}\sin\hat{\varphi}\ , \quad [\hat{n}, \sin\hat{\varphi}] = \mathrm{i}\cos\hat{\varphi}\ ,$$

Where:

$$\hat{n} = \hat{a}^+\hat{a}\ , \quad \exp(\mathrm{i}\hat{\varphi}) = (\hat{n}+1)^{-1/2}\hat{a}\ , \quad \exp(-\mathrm{i}\hat{\varphi}) = \hat{a}^+(\hat{n}+1)^{-1/2}\ .$$

Exercise 27
Show that:

$$|\alpha\rangle = \exp\left(\alpha\hat{a}^+ - \frac{1}{2}|\alpha|^2\right)|0\rangle\ ,$$

and

$$|\alpha\rangle = \exp(\alpha\hat{a}^+ - \alpha^*\hat{a})\ |0\rangle\ .$$

Note
Use the Baker–Hausdorff identity. If $[\hat{A}, \hat{B}] \neq 0$, $[\hat{A}, [\hat{A}, \hat{B}]] = [\hat{B}, [\hat{A}, \hat{B}]] = 0$, then the identity is valid:

$$\exp(\hat{A}+\hat{B}) = \exp\hat{A}\cdot\exp\hat{B}\cdot\exp\left(-\frac{1}{2}[\hat{A}, \hat{B}]\right)\ .$$

Here, you should note that $\hat{A} = \alpha\hat{a}^+$, $\hat{B} = -\alpha^*\hat{a}$.

Exercise 28
Show that:

$$\langle\alpha|\ \cos^2\hat{\varphi}\ |\alpha\rangle = \frac{1}{2} - \frac{1}{4}\exp\left(-|\alpha|^2\right) + \frac{|\alpha|^2}{2}\cos 2\theta\exp\left(-|\alpha|^2\right)\sum_{k=0}^{\infty}\frac{|\alpha|^{2k}}{k!\sqrt{(k+1)(k+2)}}\ ,$$

where $\alpha = |\alpha|\exp\mathrm{i}\theta$.

Exercise 29
Show that:

$$\langle\alpha|\ \widehat{E}_z^2\ |\alpha\rangle = \left\langle E_z^2\right\rangle = \frac{\hbar\omega}{2\varepsilon_0 V}\left\{1 + 4\left|\alpha^2\right|\sin 2\left(\vec{k}\cdot\vec{r} - \omega t + \theta\right)\right\}\ ,$$

where $\alpha = |\alpha|\exp \mathrm{i}\theta$. Also, check that:

$$\langle\alpha|\ \widehat{E}_z^2\ |\alpha\rangle - \langle\alpha|\ \widehat{E}_z\ |\alpha\rangle^2 = \frac{\hbar\omega_k}{2\varepsilon_0 V}\ .$$

Exercise 30
Suppose two statistically independent mechanisms for line broadening existed:

$$\varphi_i(\varepsilon) = \frac{1}{\sqrt{\pi\delta_i^2}}\exp\left[-\frac{(\varepsilon - \varepsilon_0)^2}{\delta_i^2}\right]\ , \quad i = 1, 2\ .$$

Find the function of the compound line shape.

Exercise 31
Show that the average of the product of any two functions of the type:

$$A(t) = A(\omega)\exp(-\mathrm{i}\omega t) + \text{c.c.}\ , \quad B(t) = B(\omega)\exp(-\mathrm{i}\omega t) + \text{c.c.}\ ,$$

over the period $T = 2\pi/\omega$ is determined by the relation:

$$\langle A(t)\ B(t)\rangle\ = \frac{1}{T}\int_0^T A(t)B(t)\mathrm{d}t = 2\mathrm{Re}\left\{A(\omega)B^*(\omega)\right\}\ .$$

Exercise 32
Show that coherent states $|\alpha\rangle$ satisfy the completeness condition:

$$\frac{1}{\pi}\int |\alpha\rangle\,\langle\alpha|\ \mathrm{d}^2\alpha = \widehat{I}\ ,$$

where $\mathrm{d}^2\alpha = \mathrm{d}(\mathrm{Re}\alpha)\mathrm{d}(\mathrm{Im}\alpha)$.

Note
Use the completeness condition for vectors $|n\rangle$:

$$\sum_n |n\rangle\,\langle n|\ = \widehat{I}\ ,$$

as well as the value of the integral:

$$\int \alpha^{*n}\alpha^m \exp\left(-|\alpha|^2\right)\mathrm{d}^2\alpha = \pi n!\delta_{nm}\ .$$

Exercise 33

Argue that the identity:

$$\frac{1}{\pi}\int \exp\left(\beta^*\alpha - |\alpha|^2\right)(\alpha^*)^n \mathrm{d}^2\alpha = (\beta^*)^n\,,$$

is valid. Using it, show that for any state of an electromagnetic field of the type:

$$|f\rangle = f(\hat{a}^+)\,|0\rangle \equiv \frac{1}{\pi}\int |\alpha\rangle\, f(\alpha^*)\exp\left(-\frac{|\alpha|^2}{2}\right)\mathrm{d}^2\alpha\,,$$

where $f(x)$ is an arbitrary function expanded in a power series, the equality:

$$\langle\beta|f\rangle = \exp\left(-|\beta|^2/2\right) f(\beta^*)\,,$$

holds. Here $|\alpha\rangle$, $|\beta\rangle$ are coherent states.

Exercise 34

Show that the Poisson distribution:

$$p(n) = \frac{\langle n\rangle^n \exp\left(-\langle n\rangle\right)}{n!}\,,$$

determines the probability of locating n photons in a coherent state $|\alpha\rangle$.

Exercise 35

Show that any coherent state can be expressed through other coherent states:

$$|\alpha\rangle = \frac{1}{\pi}\int \mathrm{d}^2\alpha'\;|\alpha'\rangle \exp\left[-\frac{1}{2}\left(|\alpha|^2 + \left|\alpha'\right|^2\right) + \alpha\alpha'^*\right]\,.$$

This is a consequence of the nonorthogonality and overfilling of coherent states.

Exercise 36

Show that, for the quantity $\hbar g^2_{\vec{k}} n_{\vec{k}\lambda}$ that characterizes the probability of transitioning an atom from the ground state to an excited state under the action of an electromagnetic wave, the following estimate ($n_{\vec{k}\lambda} \gg 1$) holds:

$$\hbar g^2_{\vec{k}} n_{\vec{k}\lambda} \sim |e|^2 \left\langle n_{\vec{k}\lambda}\left|\widehat{E}^2\right|n_{\vec{k}\lambda}\right\rangle a_{\mathrm{B}}^2\hbar^{-1}\,.$$

Note

Accept the estimate $|\vec{D}_{12}| \sim a_{\mathrm{B}}$ for the matrix element $|\vec{D}_{12}|$, where a_{B} is the Bohr radius of an atom.

Exercise 37

The operators $\exp(\mathrm{i}\widehat{\varphi})$ and $\exp(-\mathrm{i}\widehat{\varphi})$ introduced in considering coherent states do not possess the exponential properties of the operators $\mathrm{i}\widehat{\varphi}$ and $-\mathrm{i}\widehat{\varphi}$. Verify that:

$$\exp\left(-\mathrm{i}\widehat{\varphi}\right)\cdot\exp\left(\mathrm{i}\widehat{\varphi}\right) \neq 1\,.$$

Note
Consider the action of the operator $\exp(-\mathrm{i}\widehat{\varphi})\cdot\exp(\mathrm{i}\widehat{\varphi})$ on a complete set of vectors $\{|n\rangle\}$.

5. Dislocations and Martensitic Transitions

Exercise 38
Find the explicit form of the dynamic Green's function, using the Fourier transform method.

Exercise 39
Derive the static Green's function by integrating the expression for the dynamic Green's function over time.

Exercise 40
Calculate the displacement fields of motionless rectilinear screw and edge dislocations using the Green's function. Select the *Oz*-axis along the dislocation line. The Burgers vectors of the dislocations are $\vec{b} = (0, 0, b)$ and $\vec{b} = (b, 0, 0)$,respectively.

Exercise 41
Using the results of the previous exercise, calculate the components of the stress tensor of rectilinear screw and edge dislocations.

Exercise 42
Calculate the elastic stresses of walls of rectilinear screw and edge dislocations.

Note
Use the formulas:

$$\frac{y}{x^2+y^2} = \frac{1}{2}\left(\frac{1}{y+\mathrm{i}x} + \frac{1}{y-\mathrm{i}x}\right); \quad \frac{x}{x^2+y^2} = \frac{1}{2\mathrm{i}}\left(\frac{1}{y-\mathrm{i}x} - \frac{1}{y+\mathrm{i}x}\right);$$

$$\operatorname{ctg}\xi = \sum_{n=-\infty}^{\infty} \frac{1}{\xi + n\pi}.$$

Exercise 43
Using the results of the previous exercise, show that the shear stresses decrease exponentially with distancing away from the boundary of subgrains of a slightly misoriented crystal:

$$\sigma_{12} \approx 4\pi^2 D \frac{x}{w^2} \exp\left(-\frac{2\pi x}{w}\right) \cos\frac{2\pi y}{w}, \quad x \gg w.$$

Note

The grain boundary is treated as a wall of rectilinear edge dislocations. The dislocation lines are parallel to the *Oz*-axis and lie in the *Oyz*-plane. The distance w between the dislocations in the wall is much larger than the length of their Burgers vector: $w \gg b$.

Exercise 44

Show that for a system of parallel screw dislocations, the relations

$$i\omega \int \rho(\vec{r}, \omega)\, \vec{r}\, d^2\vec{r} = -\int \vec{j}(\vec{r}, \omega)\, d^2\vec{r};$$

$$i\omega \int \rho(\vec{r}, \omega) \left(r_s r_p - \frac{\delta_{sp}}{2} r^2 \right) d^2\vec{r} = -\int \left[r_p j_s(\vec{r}, \omega) + r_s j_p(\vec{r}, \omega) - \delta_{sp} \left(\vec{r} \cdot \vec{j}(\vec{r}, \omega) \right) \right] d^2\vec{r}.$$

are valid.

Note

Use the law of conservation of the Burgers vector as dislocations move.

Exercise 45

Using multipole expansions, obtain the differential intensity of dipole radiation in a wave zone for:

1. screw dislocations
2. edge dislocations
3. dislocation loops

Exercise 46

Using Cauchy's theorem on residues, verify the validity of formulas:

1. $$\widehat{H}\left[\frac{1}{x - v}\right] = -\frac{i \operatorname{sign} \operatorname{Im} v}{x - v};$$

2. $$\widehat{T}\left[\frac{1}{\operatorname{sh} \frac{\pi}{d}(x - v)}\right] = -\frac{i}{\operatorname{sh} \frac{\pi}{d}(x - v)} \operatorname{sign}\left[\sin\left(\frac{2\pi}{d} \operatorname{Im} v\right)\right].$$

Bibliography

[1] Ashcroft NW, Mermin DN. Solid State Physics. Saunders College Publishers, Fort Worth et al., 1976, 826.
[2] Kittel C. Quantum Theory of Solids. John Wiley & Sons Inc., New York, 1963, 435.
[3] Kittel C. Introduction to Solid State Physics. John Wiley & Sons Inc., New York, Chichester, Brisbane, Toronto, Singapore , 1996, 674.
[4] Abrikosov AA. Fundamentals of the Theory of Metals. North-Holland, Amsterdam, New York, 1988, 630.
[5] Lifshits IM. Electron Theory of Metals. Nauka, Moscow, 1971, 415.
[6] Zaiman J. Principles of the Theory of Solids. Cambridge University Press, Cambridge, 1979, 448.
[7] March N, Parinello M. Collective Effects in Solids and Liquids. Adam Hilger Ltd., Bristol, 1982, 274.
[8] Voronov VK, Podoplelov AV. Modern Physics: A Condensed State. Textbook. Publishing House LCI, Moscow, 2008, 336.
[9] Maradudin AA, Montroll EW, Weiss OH. Theory of Lattice Dynamics in the Harmonic Approximations. Academic Press, New York and London, 1963, 319.
[10] Lobzovsky LN (ed). Physics on the Threshold of New Discoveries. Leningrad State University, Leningrad, 1990, 320.
[11] Schrieffer J. The Theory of Superconductivity. Benjamin, New York, 1964, 282.
[12] Bednorz JG, Müller KA. Perovskite-Type Oxide – The New Approach to High-T_c Superconductivity. Rev. Mod. Phys. 60(3), 1988, 585–600.
[13] Loundon R. Quantum Theory of Light. Clarendon Press, Oxford, 2000, 448.
[14] Scully MO, Zubairy MS. Quantum Optics. Cambridge University Press, Cambridge, New York, 1997, 630.
[15] Yariv A. An Introduction to Theory and Applications of Quantum Mechanics. Dover Publications, Inc., Mineola New York, 2013, 300.
[16] Voronov VK, Podoplelov AV. Modern Physics: Textbook. Comkniga, Moscow, 2005, 512.
[17] Lonngren K, Scott A (eds). Solitons in Action. Academic Press, New York, 1978, 300.
[18] Dodd RK, Eilbeck JC, Gibbon JD, Morris HC. Solitons and Nonlinear Wave Equation. Academic Press, London et al., 1984, 630.
[19] Kiselev VV, Dolgikh DV. Nonlinear-Elastic Dent Patterns on the Surfaces of Loaded Plates and Shells. Physmatlit, Moscow, 2012, 164 (In Russian).
[20] Borisov AB, Kiselev VV. Nonlinear Waves, Solitons and Localized Structures in Magnets. Vol.1. Quasi-One-Dimensional Magnetic Solitons. UB Press of RAS, Yekaterinburg, 2009, 512 (In Russian).
[21] Borisov AB, Kiselev VV. Nonlinear Waves, Solitons and Localized Structures in Magnets. Vol.2. Topological Solitons, Two- and Three-Dimensional Patterns. UB Press of RAS, Yekaterinburg, 2011, 416 (In Russian).
[22] Borisov AB, Kiselev VV. Quasi-One-Dimensional Magnetic Solitons. Phizmathlit, Moscow, 2014, 520 c.
[23] Umezava H, Matsumoto H, Tachiki M. Thermo Field Dynamics and Condensed States. North-Holland Publishing Company, Amsterdam, 1982, 591.
[24] Truesdell C. A First Course in Rational Continuum Mechanics. Academic Press, New York, 1977, 304.
[25] Murnaghan FD. Finite Deformation of an Elastic Solid. John Wiley & Sons, Inc., New York; Chapman & Hall Ltd., London, 1951, 140.

https://doi.org/10.1515/9783110586183-007

[26] Goldenblat II. Some Problems of Mechanics of Deformable Media. Gostechizdat, Moscow, 1955, 271.
[27] Frantsevich IN. Elastic Constants and Elastic Moduli of Metals and Nonmetals. Naukova Dumka, Kiev, 1982, 288.
[28] Landau LD, Lifshitz EM, Kosevich AM, Pitaevskii LP. Theory of Elasticity. Vol. 7., Pergamon Press, New York, 1986, 187.
[29] Kosevich AM. Dislocations in the Theory of Elasticity. Naukova Dumka, Kiev, 1978, 219 (in Russian).
[30] Stroh AN. Theory of Fracture of Metals. Advan. Phys. 6(24), 1957, 418–428.
[31] Eshelby JD. The Equation of Motion of Dislocation. Phys. Rev. 90, 1953, 248–255.
[32] Boiko VS, Garber RI, Kivshik VF, Krivenko LF. Experimental Investigation of the Transition Sound Radiation Emitted by Dislocations Emerging to the Surface. Soviet Physics JETP 44(2), 1976, 372–375.
[33] Kosevich AM. Equation of Motion of a Dislocation. Soviet Physics JETP. 16, 1962, 455–462.
[34] Kosevich AM. Dynamical Theory of Dislocations. Soviet Physics Uspekhi 7(6), 1965, 837–854.
[35] Kosevich AM. The Deformation Field in an Isotropic Elastic Medium Containing Moving Dislocations. Soviet Physics JETP 15(1), 1962, 108–115.
[36] Mura T. Continuous Distribution of Moving Dislocations. Phil. Mag. 8(89), 1963, 843–857.
[37] Stenzel G. Lineare Kontinuumstheorie Bewegter Versetzungen. Phys. Stat. Sol. 34, 1969, 351–364.
[38] Shmatov VT. Method of the Green Functions in the Continuum Theory of Dislocations. Physics of Metals and Metallography 42(1), 1976, 7–18.
[39] Peach MO, Koehler JS. The Forces Exerted on Dislocations and the Stress Field Produced by Them. Phys. Rev. 50(4), 1950, 436–439.
[40] Weertman J. The Peach–Koehler Equation for the Force on a Dislocation Modified for Hydrostatic Pressure. Phil. Mag. 11(114), 1965, 1217–1223.
[41] Li JCM. The Stress Field of an Infinite Edge Dislocation Wall. Acta Met. 9(4), 1961, 384–385.
[42] Kiusalaas J, Mura T. On the Elastic Field around Edge Dislocation with Application to Dislocation Vibration. Phil. Mag. 9(97), 1964, 1–7.
[43] Laub T, Eshelby JD. The Velocity of a Wave along a Dislocation. Phil. Mag. 14(132), 1966, 1285–1293.
[44] Natsik VD. Acoustic Emission Associated with the Annihilation of Dislocations. Soviet Physics Solid State. 14(11), 1973, 2678–2682.
[45] Chishko KA. Acoustic Emission during Annihilation of Prismatic Dislocation Loops and Kinks in Line Dislocations. Ukrainian Phys. Journ. 19(8), 1974, 1264–1271.
[46] Jackson JD. Classical Electrodynamics. John Wiley & Sons Inc., New York, London, Sydney, 1962, 641.
[47] Kiselev VV, Shmatov VT. Multipole Moment Expansion of Fields Generated by Dislocations Moving in an Unbounded Elasto-Isotropic Medium. Physics of Metals and Metallography 50(1), 1980, 20–30.
[48] Bateman H, Erdélyi A. Higher Transcendental Functions. Vol. 2. McGraw–Hill Book Company, Inc., New York, Toronto, London, 1953, 396.
[49] Love AEH. A Treatise on the Mathematical Theory of Elasticity. Cambridge University Press, Cambridge, 1927, 643.
[50] Landau LD, Lifshitz EM. Statistical Physics. Vol. 5. Butterworth–Heinemann, Oxford, 2013, 544.
[51] Eshelby JD. Dislocations as a Cause of Mechanical Damping in Metals. Proc. Roy. Soc. A. 197, 1949, 396–416.
[52] Gradshteyn IS. and Ryzhik IM. Table of Integrals, Series, and Products. Academic Press, Elsevier, Amsterdam et al., 2013, 1162.
[53] Peierls RE. The Size of a Dislocation. Proc. Phys. Soc. London A 52, 1940, 34–37.

[54] Stenzel G. Zum Peierls-Modell bewegter Versetzungen. Phys. Stat. Sol. 34, 1969, 495–500.
[55] Seeger A. Theorie der Gitterfehlstellen. In: Handbuch der Physik, Bd. 7, Flügge S (ed). Springer, Berlin, 1955, s. 383–665.
[56] Borisov AB, Kiselev VV. The Dynamics of Peierls Dislocations. Physics of Metals and Metallography 71(3), 1991, 17–25.
[57] Nabarro FRN. Dislocations in a Simple Cubic Lattice. Proc. Phys. Soc. London A 59(332), 1947, 256–272.
[58] Jacobs AE. Solitons of the Square-Rectangular Martensitic Transformation. Phys. Rev. B 31(9), 1985, 5984–5989.
[59] Barsh GR, Krumhansl JA. Nonlinear and Nonlocal Continuum Model of Transformation Precursors in Martensites. Metall. Trans. A 19, 1988, 761–775.
[60] Boiko VS, Garber, RI, Kosevich AM. Reversible Crystal Plasticity. American Institute of Physics, New York, 1994, 224.
[61] Falk F. Ginzburg–Landau Theory of Static Domain Walls in Shape-Memory Alloys. Z. Phys. B, Condensed Matter 51(2), 1983, 177–187.
[62] Falk F. Ginzburg–Landau Theory and Solitary Waves in Shape-Memory Alloys. Z. Phys. B, Condensed Matter 54(2), 1984, 159–167.
[63] Falk F. Stability of Solitary-Wave Pulses in Shape-Memory Alloys. Phys. Rev. B 36(6), 1987, 3031–3041.
[64] Barsh GR, Krumhansl JA. Twin Boundaries in Ferroelastic Media without Interface Dislocations. Physical Review Letters 53(11), 1984, 1059–1072.
[65] Kiseliev VV. Weakly-Nonlinear Dynamics and Solitons in the Vicinity of a Martensitic Phase Transition. Physics Letters A 196, 1994, 97–100.
[66] Kiselev VV. Weakly Nonlinear Solitonlike Excitations in a Two-Dimensional Model of a Martensitic Transformation. Physics of the Solid State 36(11), 1994, 1765–1775.
[67] Kiseliev VV. Solitons in the vicinity of a martensitic phase transition. In: Sakuma T, Sheppard LM, Ikuhara Y (eds). Grain Boundary Engineering in Ceramics – From Grain Boundary Phenomena to Grain Boundary Quantum Structures, Ceramics Transaction, Vol. 118. The American Ceramic Society, Westerville, 2000, 601–608.
[68] Kadomtsev BB. On the Stability of Solitary Waves in Weakly Dispersing Media. Soviet Physics Doklady 15(6), 1970, 539–541.
[69] Hirota R. Direct Methods in Soliton Theory. In: Bullough RK, Caudry PJ (eds). Solitons, Chapter 5. Springer-Verlag, Berlin, 1980, 389.
[70] Oshima R, Sugiyama M, Fujita FE. Tweed Structures Associated with Fcc-Fct Transformations in Fe-Pd Alloys. Metall. Trans. A 19, 1988, 803–810.
[71] Katsnelson MI. Phonon Frequency Stabilization and Quasi-Static Shifts of Atoms in Crystals. JETP 70(6), 1990.
[72] Kuznetsov EA, Turitsyn SK. Two- and Three-Dimensional Solitons in Weakly Dispersive Media. JETP 55(5), 1982, 844–847.
[73] Zakharov VE. Instability and Nonlinear Oscillations of Solitons. JETP Letters 22(7), 1975, 172–173.
[74] Laedke EM, Spatschek KH. On the Applicability of the Variation of Action Method to Some One-Field Solutions. Journal of Mathematical Physics 20, 1979, 1838–1841.
[75] Burtsev SP. Instability of a Periodic Chain of Two-Dimensional Solitons. JETP 61(5), 1985, 959.
[76] Pelinovsky DE, Stepanyants YA. Self-Focusing Instability of Plane Solitons and Chains of Two-Dimensional Solitons in Positive-Dispersion Media. JETP 77(4), 1993, 602–608.
[77] Pesenson MZ. Multidimensional Solitary Waves in Condensed Media. Soviet Physics, Solid State 32(5), 1990, 855–859.
[78] Vaks VG, Katsnelson MI, et al. An Experimental and Theoretical Study of Martensitic Phase Transitions in Li and Na under Pressure. J. Phys.: Condens. Matter 1(2), 1979, 5319–5335.

Index

https://doi.org/10.1515/9783110586183-008

Also of Interest

Volume 43
Robert F. Snider
Irreducible Cartesian Tensors, 2017
ISBN 978-3-11-056363-4, e-ISBN (PDF) 978-3-11-056486-0,
e-ISBN (EPUB) 978-3-11-056373-3

Volume 42
Javier Roa
Regularization in Orbital Mechanics: Theory and Practice, 2017
ISBN 978-3-11-055855-5, e-ISBN (PDF) 978-3-11-055912-5,
e-ISBN (EPUB) 978-3-11-055862-3

Volume 41
Esra Russell, Oktay K. Pashaev
Oscillatory Models in General Relativity, 2017
ISBN 978-3-11-051495-7, e-ISBN (PDF) 978-3-11-051536-7,
e-ISBN (EPUB) 978-3-11-051522-0

Volume 40
Joachim Schröter
Minkowski Space: The Spacetime of Special Relativity, 2017
ISBN 978-3-11-048457-1, e-ISBN (PDF) 978-3-11-048573-8,
e-ISBN (EPUB) 978-3-11-048461-8

Volume 39
Vladimir K. Dobrev
Invariant Differential Operators: Quantum Groups, 2017
ISBN 978-3-11-043543-6, e-ISBN (PDF) 978-3-11-042770-7,
e-ISBN (EPUB) 978-3-11-042778-3

Volume 38
Alexander N. Petrov, Sergei M. Kopeikin, Robert R. Lompay, Bayram Tekin
Metric theories of gravity: Perturbations and conservation laws, 2017
ISBN 978-3-11-035173-6, e-ISBN (PDF) 978-3-11-035178-1,
e-ISBN (EPUB) 978-3-11-038340-9