Physics – Uspekhi **46** (5) 493 – 547 (2003)

FROM THE HISTORY OF PHYSICS

PACS numbers: **01.65.+g**, **42.50.–p**, **67.40.–w**, **74.20.–z**

On the centenary of the Nobel Prize: Russian laureates in physics

K N Mukhin, A F Sustavov, V N Tikhonov

DOI: 10.1070/PU2003v046n05ABEH001288

Contents

Abstract. The history and development of the branches of physics which profited significantly from the work of Russian Nobel laureates (P A Cherenkov, I E Tamm, I M Frank, L D Landau, N G Basov, A M Prokhorov, P L Kapitza, and Zh I Alferov) are reviewed in popular form to mark the recent Nobel Foundation centenary. Apart from the Russian prize winners' achievements, the major contributions of their colleagues — Russian and foreign, predecessors and successors — are briefly discussed. The current state of the branches of physics advanced with the participation of Russian laureates is reviewed, and the practical implications of their work for science, technology, and everyday life are discussed.

K N Mukhin, A F Sustavov, V N Tikhonov Russian Research Centre 'Kurchatov Institute', Institute of General and Nuclear Physics
pl. Kurchatova 1, 123182 Moscow, Russian Federation
Tel. (7-095) 196 75 71, 196 76 63. Fax (7-095) 196 91 33
E-mail: mukhin@chen.net.kiae.ru

Received 16 July 2002, revised 25 February 2003
Uspekhi Fizicheskikh Nauk **173** (5) 511 – 569 (2003)
Translated by A S Dobroslavskiĭ, V I Kisin, and E Yankovsky;
edited by A V Getling, S N Gorin, and A Radzig

1. Introduction

This review describes the discovery and study of several exotic phenomena in macrophysics[1]. In a sense, it is a sequel to previous articles that considered exotic processes in microphysics, namely in nuclear physics [2] and the physics of elementary particles [3]. In getting down to writing this article, we immediately ran into great difficulty due to the immense amount of a pertinent material. As a matter of fact, macrophysics as a whole is an exotic domain and is much more vast than, for example, the physics of elementary particles, the popular description of the exotic features of which took so many pages [3].

What to choose? The only thing that was clear implied that the article ought to be on as popular a level as the two previous articles, because the entire cycle of 'exotic' articles is aimed at beginner physicists (students), or practising specialists in fields other than the topic discussed. This means that the article cannot be confined to one particular matter, because this will compromise its popular character and winnow out a large portion of the audience.

To tell the truth, at one point the idea of a spacious review seemed attractive. In connection with the recent

[1] We use the terminology of the author of Ref. [1], who divides physics as a whole into macrophysics and microphysics (also specially distinguishing astrophysics). In the Russian scientific literature, the topics pertaining to macrophysics often fall under the concept of 'general physics'.

Figure 1. Russian laureates of the Nobel Prize in physics.

award of the Nobel Prize in Physics 2000 to Zh I Alferov, we were tempted to write about his achievements in the field of semiconductors. We even began writing thusly, but very soon it became clear that what results is a general survey, which was not our goal. However, the success of Alferov helped us to find the criterion for making a selection. Out of the exotic phenomena of general physics we decided to select those whose discovery, explanation, and study especially benefited from the work of the Russian Nobel Prize laureates in physics (Fig. 1).

There are now eight of them, and they all have been elected to the USSR Academy of Sciences. They are P A Cherenkov, I E Tamm, and I M Frank (1958), L D Landau (1962), N G Basov and A M Prokhorov (1964), P L Kapitza (1978), and, after a long hiatus, Zh I Alferov (2000). Their lines of inquiries were the following: Vavilov – Cherenkov radiation, superfluidity and superconductivity, quantum oscillators (masers and lasers), and the physics of semiconductors (including heterostructures). These directions (in a somewhat different order) will be the thrust of our treatise.

We shall start (in Section 2) with superfluidity and superconductivity, because historically the two fundamental results in these lines of investigation (liquefaction of helium in 1908, and the discovery of superconductivity in 1911) were obtained earlier (although not in Russia but in the Netherlands) than the rest of the fields discussed in this paper, and their author Kamerlingh Onnes had already received the Nobel Prize in Physics 1913.

Incidentally, the Nobel Prize Foundation is seldom quick to react to discoveries in general physics. All the Russian laureates, except Basov and Prokhorov, had to wait their turn for 20 – 30, or even 40 (Kapitza) years!

So we start Section 2 with the story about the discovery and popular explanation of two 'superexotic' phenomena: superfluidity and superconductivity, which are associated not only with the work of Kamerlingh Onnes, Kapitza, and Landau, but also with the studies of Meissner, the brothers Fritz and Heinz London, Ginzburg, Tisza, Bogolyubov, Bardeen, Cooper, Schrieffer (the last three also being the Nobel Prize laureates), and many others. Moreover, in a special section we discuss the discovery of high-temperature superconductivity by Bednorz and Müller (another 'superexotic' phenomenon and another Nobel Prize in physics, this time awarded just one year after the discovery), and the exceptionally fast rise in the critical temperature of high-temperature superconductors. Our description of the properties of Bose systems will close with the latest accomplishment in this domain (the Nobel Prize in Physics 2001) — the achievement of Bose – Einstein condensation in rarefied gases of alkaline metals (to be more specific, ^{87}Rb) in laboratory conditions. At the end of the section we describe the technological applications of low-temperature superconductivity and the promises of high-temperature superconductivity.

Section 3 is devoted to Vavilov – Cherenkov radiation, which was explained by Tamm and Frank. This process also seems to be associated with two superexotic phenomena: the motion of particles at *a superluminal velocity* (not in a vacuum, however, but in a medium with a refraction index greater than one) and the emission of electromagnetic radiation by such particles in the course of *uniform rectilinear motion* (also in a medium). In addition to the works of the pioneers mentioned above, we shall speak of the contributions to the studies of this effect made by E Fermi and V L Ginzburg, and of the use of Cherenkov counters in the physics of elementary particles. In Section 3.4 we discuss a related phenomenon — the transition radiation predicted in 1945 by Ginzburg and Frank and confirmed experimentally in 1959 by R Goldsmith and J Jelley.

Section 4 is devoted to the advent of quantum electronics through the efforts of C Townes, Basov and Prokhorov, who shared (in the proportions 1/2, 1/4, and 1/4) the Nobel Prize in Physics 1964. Here we discuss the origin (A Einstein, 1917), development (P Dirac, 1927 [4]), and implementation of the fundamental idea of the possible generation of high-frequency oscillations in an active medium with inversely populated energy levels. In particular, we discuss the three- and four-level method of creating inverse population, the necessity of pumping and cavities for realization of generation, the construction of the first maser — the molecular oscillator operating in the microwave range — by the physicists named above, and later on the first optical oscillator (ruby laser) by Theodore Maiman (USA). In the subsequent sections we consider various classes of existing masers (oscillators and amplifiers, including those occurring in nature), and lasers [solid-state, liquid, gas, plasma (X-ray), and those on free electrons], describe their operating regimes, and outline their fields of applications.

Finally, Section 5 is devoted to semiconductors, for the studies of which three physicists — Jack Kilby (USA), Zhores

Alferov (Russia), and Herbert Kroemer (USA) — received the Nobel Prize in physics recently, in 2000. As previously, this section begins with a popular introduction of basic concepts and model representations developed in classical and quantum physics for explaining the properties of semiconductors — in particular, the elements of band theory, the concept of the electron–hole transition, etc. Then we discuss the properties of certain particular semiconductors and the methods of production of semiconductors with desired properties. Special attention is paid to the modern branch of semiconductor physics — heterostructures (in particular, double heterostructures) that have allowed the creation of continuous-wave semiconductor lasers operating at room temperature and powerful solar batteries for space stations. The new perfectly exotic quantum physics of the low-dimensional electron gas involves such weird concepts as 'quantum well', 'quantum wire', and 'quantum point'. Of course, in this section we pay homage to the predecessors of the new laureates — John Bardeen, Walter Brattain, and William Shockley (the Nobel Prize in Physics 1956), who discovered the transistor effect. Finally, we describe the numerous and diverse semiconductor devices (including semiconductor lasers), and emphasize the outstanding capabilities of modern microelectronics based on the use of integrated circuits which are obviously exotic in and of themselves.

Some people may regard our criterion of selection of the material as artificial and not quite scientific. In our defense, we can put forward the following arguments.

(1) Beginner physicists must know their 'heroes'. This is especially true now, when the world is celebrating the centennial of the Nobel Prize.

(2) At the same time, our criterion is by no means dictated by jingoism or the desire to glorify the accomplishments of the Nobel laureates. Even the above brief synopsis of the article shows that we are going to speak not only about the work of the Russian scientists, and not only about the laureates, but also about the achievements of many foreign physicists (some of whom are also Nobel Prize winners) who work or used to work in the fields of science discussed in the article.

(3) All selected topics are not simply exotic — they have brought about tremendous breakthroughs in science, technology, and everyday life. Just recall the Cherenkov detectors and the supermagnets with superconducting windings used for modern accelerators, medical lasers and tomographs, deep space communications, cell phones and audio and video CDs, computers and the Internet, visual displays in home appliances, multicolored street signs, and so on and so forth.

(4) It is important that all these avenues of investigation continue to develop, and have the potential of bringing about results of great consequence — like superconductivity at room temperature, computers operating at the speed of light, and superexotic discoveries in the physics of elementary particles.

Of course we understand that our selection of material is not without fault. Firstly, the scope of the article is by conception confined to the description of those results that were obtained in the period of activity of the eight Russian scientists, with only brief mention of the things done before or after, trying to make up for omissions with references to the literature (for the reader's convenience, mainly to the reviews published by *Phys. Usp.*). Secondly, because of our particular selection criterion, the content of different sections of the article is rather kaleidoscopic. Indeed, in one article we had to speak not only about the relatively straightforward concepts of classical electrodynamics, but also about quantum liquids, topics of solid-state theory and the properties of semiconductors, basic concepts of quantum electronics and the principles of operation of masers and lasers, the mysteries of superfluidity and superconductivity, integrated circuits, and microelectronics. And all this had to be done in a concise and popular manner ('without equations', so to say), but at the same time in an interesting way, using correct physics without too much vulgarization. The reader will appreciate how we were able to cope with all these difficulties. We are open to criticism, but would ask the reader to view our attempts to popularize the various theories with sympathy, since we, after all, are experimental physicists.

2. Superconductivity and superfluidity

2.1 Historical background

Before embarking on the presentation of the material in this section, we have to note that in writing it we widely used a fairly recent review by Ginzburg [5], devoted to the history of studies of superconductivity and the prospects for future investigations. We hope that the reader will turn to this extremely interesting and instructive publication in order to fill in the lacunas inevitable in our popular presentation of the matter. For us, reading this review was an absolutely necessary step. We cannot recall another example of such a concentrated account of superconductivity (and superfluidity as well).

As known, superconductivity is the transition of a substance cooled below a certain critical temperature T_c (specific to this substance) from the normal state into a state of zero resistance to direct current (as well as low-frequency alternating current). The most convincing evidence for the zero resistance of a superconductor is the ability of a closed ring to carry a once-induced current for years. This allows putting an upper limit on the resistivity of a superconductor at $10^{-23}\ \Omega\ \mathrm{cm}^{-1}$, which is 10^{14} times less than that of a very-high-purity copper specimen at low temperatures (Cu, Ag, Au, Pt, as well as a number of alkaline and alkaline-earth elements, are not superconductors).

Distinguished today are two kinds of superconductivity: low-temperature (LTSC) and high-temperature superconductivity (HTSC). LTSC was discovered first. This was natural, because experiments indicated that the resistance of metals and alloys gradually decreases with decreasing temperature. The lowest temperatures in the beginning of the 20th century could be attained with liquid helium, which allows conducting experiments at temperatures close to absolute zero (the boiling point of liquid helium at atmospheric pressure is 4.2 K and even less at a reduced pressure).

Helium was first liquefied in 1908 in the laboratory of Kamerlingh Onnes in Leiden (Holland), after which measurements of the electric conductivities of metals at temperatures close to absolute zero immediately began in this laboratory. The resistance of superconductors was expected to fall gradually as the temperature approached $T = 0$ K, as in the higher-temperature range studied previously. However, the experiments brought about a totally unexpected result: the conductors passed from the normal state into the superconducting state *abruptly* at certain critical temperatures T_c, which were *different* for various particular substances. Thus,

a superconductor is not simply an ideal conductor with a vanishingly low resistance, but rather a new phase state of the given substance.

The first result was obtained by Kamerlingh Onnes in 1911 for mercury, for which $T_c = 4.15$ K [6], and in 1913 superconductivity was discovered for white tin ($T_c \approx 3.7$ K) and lead ($T_c \approx 7.2$ K). In the same year Kamerlingh Onnes found that superconductivity disappears when a sufficiently strong current $I > I_c$ (where I_c is the so-called critical current) is passed through the superconductor, which returns to the normal state[2]. In 1914, a similar effect of destruction of superconductivity was discovered in the same laboratory when a superconductor was placed into a sufficiently strong external magnetic field $H > H_c(T)$, where, to a good approximation,

$$H_c(T) \approx H_0\left[1 - \left(\frac{T}{T_c}\right)^2\right]. \quad (1)$$

It can be seen from (1) that, as T decreases, $H_c(T)$ increases from zero (at $T = T_c$) to a maximum H_0 (at $T = 0$), which is regarded as the magnetic field strength that destructs the superconductivity. For the superconductors mentioned above these strengths are $H_0(\mathrm{Hg}) \approx 400$ Oe, $H_0(\mathrm{Sn}) \approx 300$ Oe, and $H_0(\mathrm{Pb}) \approx 800$ Oe.

Today the superconductors exhibiting such a behavior in external magnetic fields are known as type-I superconductors. In type-II superconductors, the transition to the normal state with increasing external field occurs gradually rather than abruptly. It starts at the lower critical field H_{c1} and ends at the upper critical field H_{c2} (the properties of type-II superconductors will be discussed in greater detail in Section 2.3).

The study of the magnetic properties of superconductors in the same laboratory and in the same year, 1914, led to the first practical application of superconductivity — an electromagnet with a superconducting winding. As a matter of fact, until the 1920s, liquid helium was only produced in the laboratory of Kamerlingh Onnes in Leiden, and new superconductors were discovered and studied only there. The situation changed in the early 1920s, when cryogenic laboratories were established in Canada and Germany. In particular, at W Meissner's laboratory in Berlin in the late 1920s, many new superconducting alloys and compounds were found, and in 1933 the effect of expulsion of the magnetic field from the bulk of the superconductor was discovered (the Meissner effect) [7].

A theoretical description of the Meissner effect was provided in 1935 by Fritz and Heinz London, who proposed a phenomenological equation describing the distribution of the magnetic field in superconductors [8].

In particular, they found the law of the decrease of the magnetic field strength $H(z)$ with the depth z measured inward in the superconductor:

$$H(z) = H(0)\exp\left(-\frac{z}{\lambda_L}\right), \quad (2)$$

where $H(0)$ is the field on the surface of the superconductor, $\lambda_L = (mc^2/4\pi n_s e^2)^{1/2}$, m and e are the mass and the charge of the electron, and n_s is the concentration of the superconductivity electrons. From equation (2), we can see that the magnetic field penetrates into the superconductor only to a depth of order λ_L, which for metals is about 10^{-6} cm. The essence of the Meissner effect lies in the fact that, at $H < H_c$, a persistent circular current is induced in the surface layer of a massive superconducting body, whose magnetic field cancels out the external field in the bulk of superconductor.

Londons' theory has played an important role in explaining the phenomenon of superconductivity, but it holds only in the case of weak magnetic fields ($H \ll H_c$) and fails, for example, to give a correct expression for the critical field H_c in thin films. In 1950, it was superseded by the Ginzburg – Landau theory, of which we will speak a little later. Now we will continue with the history of investigations in the field of low temperatures, focusing our attention on the study of the properties of liquid helium.

The advances of Kamerlingh Onnes and his colleagues from Leiden laboratory in the field of low temperatures were no less impressive than their success in superconductivity. As early as in 1911, they discovered the anomalous dependence of the density of helium ρ on the temperature T. It was found that at $T = 2.17$ K the steady increase in density with decreasing temperature stops abruptly, and the density levels off. The temperature $T = 2.17$ K at the pressure of the saturated helium vapor, 38.8 mmHg, is called the point of λ transition. The λ transition is associated with anomalies (λ singularities) in the behavior of certain other parameters, which are attributed to the type-II phase transition of liquid helium from state HeI (at $T > T_\lambda$) to state HeII (at $T < T_\lambda$).

Let us note certain λ singularities of HeII. In 1922, Kamerlingh Onnes discovered that HeII can flow from one container into another, creeping up the walls whose edges were above the level of the fluid. HeII flows over an open wall forming a thick (about a hundred atomic layers) film. In the 1930s, a λ singularity was discovered in the behavior of the heat conductivity of HeII, and in 1936 Willem Keesom discovered the super-thermal conductivity of HeII, which is about 10^6 times higher than the heat conductivity of HeI [9].

In the late 1930s studies of the properties of liquid helium were carried out by Kapitza [10] and Allen and Missner [11], who discovered in 1938 the superfluidity of HeII, that is, its ability to flow without friction through very narrow ($d \sim 10^{-5}$ cm) capillaries and slits (nondissipative dragless flow). Further studies of superfluidity were described in the papers published in the early 1940s by Kapitza [12] and the originator of the theory of superfluidity, Landau[3] [13]. It is remarkable that, at the end of his paper, Landau pointed to the possible treatment of superconductivity as the superfluidity of the electron liquid in metals.

True, from the perspective of our knowledge today, this indication seems somewhat phenomenological, because in the construction of his theory for ^{4}He — that is, for Bose particles — Landau did not use the special properties of the Bose – Einstein statistics to which the superfluidity of ^{4}He at low temperatures can be related. Nobody knew at that time (and Landau was no exception) how to use the Bose – Einstein statistics for explaining the superfluidity of the electron liquid made up of particles obeying the Fermi – Dirac statistics.

[2] Kamerlingh Onnes won a Nobel Prize in physics in 1913 for his studies on the properties of matter at low temperatures and the production of liquid helium.

[3] L D Landau became a Nobel laureate in physics (1962) for his pioneering research in the theory of condensed matter, especially liquid helium. P L Kapitza became a Nobel laureate in physics (1978) for his discoveries in the physics of low temperatures [14].

However, Landau's theory of 1941 based on a two-liquid model was very important for the qualitative explanation of the superfluidity of ^{4}He, and became especially famous after having been augmented in 1947 by considerations regarding the shape of the energy spectrum of elementary excitations of HeII [15].

It would be wrong to say that the features of the Bose–Einstein statistics were not used for describing the superfluidity of liquid helium. The first to make this attempt were F London [16], who noted that superfluidity can be regarded as a macroscopic quantum phenomenon associated with Bose–Einstein condensation, and Tisza [17], who based his theory on the concept of ^{4}He as a degenerate ideal Bose gas. The latter, however, was justly criticized by Landau for neglecting interactions between particles. The next step was made in 1947 by N N Bogolyubov, who constructed the first microscopic and fairly realistic theory based on the model of a *nonideal* Bose–Einstein gas with weak interactions between particles [18]. (See section 2.2 for more details on the works of Kapitza, Tisza, Landau, and Bogolyubov.)

Now we go back to superconductivity, and begin with the promised discussion of the Ginzburg–Landau theory [19], which is an extension of the Londons' theory mentioned above [8]. The new theory extended the range of agreement with experiment to stronger fields and explained the behavior of thin films and other superconducting materials in a magnetic field. In our opinion, however, the biggest accomplishment of the Ginzburg–Landau theory [19] was associated with the very important step from the phenomenological analogy between superconductivity and the superfluidity of electron liquid to a microscopic theory of superconductivity. We will discuss this in greater detail in Section 2.3, and here we are only going to describe one very instructive and not widely known episode that confirms this view (see Refs [5, 20]).

Obviously, any theory of superconductivity ought to involve electric charge. In the Ginzburg–Landau theory it was denoted by the symbol e^*, but the authors of the theory disagreed about the physical meaning and value of e^*. Ginzburg suggested regarding e^* as the effective charge and considering it a free parameter of the theory. Landau did not accept this view because of the nonuniversality of the effective charge, and the value $e^* = e$ was assumed in Ref. [19], where e is the charge of the electron (which is, of course, universal).

However, Ginzburg did not abandon the idea of regarding e^* as a free parameter, and in Ref. [21] proposed a simple method of determining its value from comparisons of certain results of the theory (which include e^*) with experiment; in Ref. [22], he used this method to find that $e^* = (2 \text{ to } 3)e$, which brought the theory into better agreement with experiment than for $e^* = e$. With this result (duly quoting the opinion of Landau in this regard), Ginzburg published his paper in 1955 [22].

This controversy was finally resolved after the development of a microscopic theory of superconductivity by Bardeen, Cooper, and Schrieffer in 1957 (the BCS theory) [23], and the proof given by Gor'kov in 1959 [24] for the fact that the equations and coefficients of the Ginzburg–Landau theory [19] rigorously follow from the BCS theory and that $e^* = 2e$, which corresponds to the charge of Cooper pairs ([25], 1956). In this way, the dispute between Ginzburg and Landau ended in a draw. Each of them was right, because the value $e^* = 2e$ is, on the one hand, universal and, on the other hand, not equal to e. Moreover, the authors understood how close they had approached the BCS theory in 1950, seven years before its appearance. In Section 2.3 we will give even more convincing evidence of such proximity.

The BCS theory gave a brilliant explanation to the phenomenological analogy between superconductivity and superfluidity. The two electrons of the Cooper pair with opposite spins and equal and opposite momenta experience mutual *attraction* by virtue of the electron–phonon interaction (that is, interaction between electrons and oscillations of the crystal lattice)[4] and form a bound state with a total spin of $I = 0$ and a charge of $2e$. Such electron pairs (known as Cooper pairs) obey the Bose–Einstein statistics and can thus form Bose condensate and exhibit the property of superfluidity. It turns out that superconductivity is indeed similar to superfluidity, being the superfluidity of the electron liquid. The phenomenological conjecture that Landau made in 1941 proved to be prophetic and was confirmed 16 years later (the Ginzburg–Landau and BCS theories will be discussed in greater detail in Section 2.3).

The success of the BCS theory stimulated many studies in the theory of superconductivity and superfluidity (since the superfluidity of ^{3}He could be treated, according to the BCS scheme, as the formation of Bose pairs from ^{3}He fermions). In particular, the main results of the BCS theory were obtained by a different method and refined by Bogolybov [26] and other authors [27]. Also, many studies of this period (see Section 2.3) dealt with analysis of the Ginzburg–Landau theory [19] (developed long before the BCS theory) in the context of the fact that its equations strictly follow from the BCS theory, developed much later.

As we noted before (and will discuss in greater detail in Section 2.3), the mechanism of attraction between the electrons of the Cooper pair consists in the electron–phonon interaction, which involves them into joint coherent motion in the form of a charged Bose particle with a total charge of $2e$ and zero net spin and momentum. It ought to be added that the Cooper pair exhibits another exotic property: it involves not adjacent electrons, separated by an average distance of 10^{-8} cm, but electrons that are relatively far from each other ($\sim 10^{-4}$ cm). Therefore, the size of such a 'Bose particle' is several orders of magnitude greater than the size of the common neutral atom ^{4}He. Incidentally, it is because of such a large distance between the electrons in the Bose pair that the electrostatic repulsion between them is shielded and does not neutralize the weak Cooper attraction.

We can conclude from the BCS theory that the critical temperature of the superconductor must increase with the increasing Debye temperature of the metal θ_D. First, this explains the low values of T_c for the discovered low-temperature semiconductors (since the corresponding metals have relatively low θ_D); second, in view of this fact, it seems

[4] The classical notion of oscillations of atoms in a crystal lattice, which emit sonic (and other) waves, is replaced in quantum mechanics by the concept of quasi-particles — phonons — traveling over the lattice. A phonon is a particle that can be associated with a certain direction of motion and a certain energy $h\nu$ (where h is the Planck constant and ν is the classical oscillation frequency); however, the phonon's momentum is somewhat different from the classical concept (although close to it), which is the reason why we call it quasi-momentum and refer to a phonon as a quasiparticle. The formation of Cooper pairs can be viewed as the exchange of phonons between electrons. One electron emits a phonon, the other one absorbs it, which, under certain circumstances (to be discussed later), can give rise to attraction.

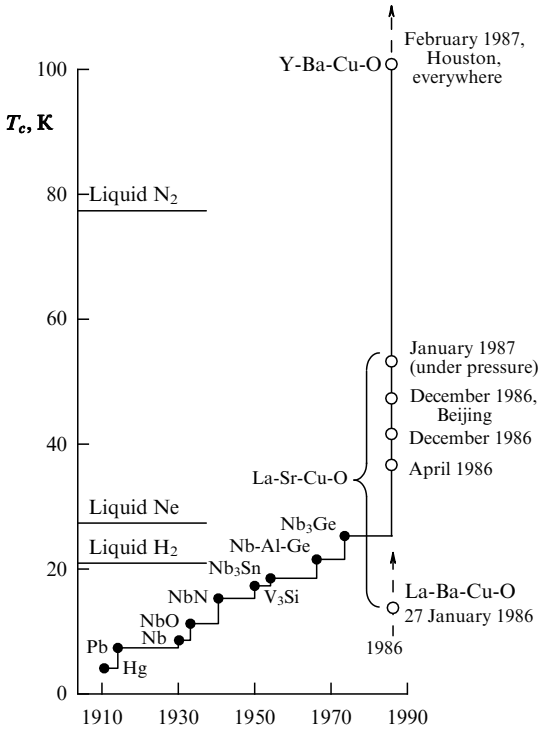

Figure 2. Growth of T_c before and after the discovery of HTSC.

obvious that the search for superconductors with higher T_c ought to be steered towards metals with higher θ_D. Under the assumption of another (exciton)[5] mechanism of Cooper pairing, high-temperature superconductors ought to be sought among quasi-one-dimensional [28] and quasi-two-dimensional [29, 30] metals (one-dimensional 'needles' and two-dimensional films).

HTSC was discovered by Bednorz and Müller in 1986 [31] and Wu with colleagues in 1987 [32] in layered compounds — cuprates (which are, incidentally, oxides)[6]. In the first experiments, they achieved critical temperatures of $T_c \simeq 30-34$ K, which was a fundamental breakthrough, because up to that point the critical temperatures of low-temperature superconductors had crawled up the scale of critical temperatures very slowly (Fig. 2) and stopped at 24 K in 1973.

The situation with HTSC was entirely different. Ten months after the discovery of HTSC, the critical temperature of the compound studied in Ref. [32] exceeded the boiling point of liquid nitrogen (77.3 K). This result was of extreme importance, since it promised possibilities of wider applications of superconductivity as compared with the use of LTSC. Naturally, such a fast growth of critical temperatures tempted physicists to dream of room-temperature superconductivity (RTSC). All over the world, various substances started being tested on an unprecedentedly wide scale in the quest for HTSC with a higher T_c [34]. As a result, by 1994, T_c was raised to 135 K [5], almost halfway to the cherished value of $T_c \simeq 300$ K.

Experts hold that there are no fundamental barriers to prevent the existence of RTSC [34], but today's theory gives no guidelines to indicate what compounds are more likely to exhibit higher values of T_c [5], because the mechanism of HTSC in cuprates has not yet been clarified [35] (for more details, see Section 2.4)

Concerning the practical use of superconductivity, LTSC is still used much more widely than HTSC. Its best known applications are the physics and technology of producing elementary particles and studying their properties, medicine (NMR tomographs), and investigations of hot plasmas (which might seem a little paradoxical for *low-temperature* superconductivity). Nevertheless, there is also some progress in the practical use of HTSC. Technical difficulties have already been overcome in the development of HTSC wires and cables, and a number of HTSC projects have been implemented and others are close to completion [36–38] (see Section 2.5).

2.2 Superfluidity: Kapitza's experiments and an elementary exposition of the theories of Tisza, Landau, and Bogolyubov

As we already said in Section 2.1, liquid helium was first produced by Kamerlingh Onnes in 1911, and it was found somewhat later that it remains in the liquid state at normal pressure all the way down to $T = 0$ K. From the standpoint of classical physics, this property of liquid helium seemed completely exotic and was explained only in quantum mechanics.

According to classical physics, the kinetic energy of thermal motion of particles of any body reduces with decreasing temperature, and at $T = 0$ the particles occupy fixed positions within the body, so that the body changes from a liquid into a solid state. According to quantum mechanics, the kinetic energy of the particles does not become zero even at $T = 0$, because they participate in so-called zero-point oscillations. The amplitude of zero-point oscillations increases as the mass of the particles and the force of their interaction decrease. If the amplitude becomes comparable to the mean distance between the particles, they will obviously be unable to occupy fixed positions, so that the body will remain in the liquid state all the way down to $T = 0$.

Of all substances, the two helium isotopes ^{4}He and ^{3}He (the latter being present in atmospheric helium in a proportion of 10^{-4}%) have the least atomic mass combined with the least force of interaction. For this reason, they remain liquid at atmospheric pressure in the immediate proximity to absolute zero $T = 0$ (^{4}He at $0 < T < 4.2$ K, ^{3}He at $0 < T < 3.35$ K), and solidify at a pressure of 25 (^{4}He) and 30 atm (^{3}He). Since such a retention of the liquid state down to $T = 0$ can only be understood in the framework of quantum mechanics, such liquids are referred to as quantum liquids.

Another feature of quantum liquids is their superfluidity, which, at the pressure of saturated vapor, is exhibited by ^{4}He at $T < 2.17$ K and by ^{3}He at $T < 0.9 \times 10^{-3}$ K. As noted in Section 2.1, superfluidity was discovered in 1938 by Kapitza [10] and Allen and Missner [11]. It is interesting that these two papers were published next to one another in the same issue of *Nature*, and Allen and Missner mentioned that they were aware of Kapitza's results. The scheme of Kapitza's experi-

[5] In this case, by exciton we mean a quasi-particle that is a hydrogen-like bound state of a conduction electron and a hole, which, however, exhibits a much lower bond energy compared to the energy of the bond between the electron and proton in the hydrogen atom (10^{-2} and 13.6 eV, respectively), and is considerably larger in size (10^{-4} vs. 10^{-8} cm). An exciton can form a bound state with another electron, and the two electrons will, as a result, experience mutual attraction.

[6] For the discovery of HTSC, J G Bednorz and K A Müller won a Nobel Prize in physics in 1987 [33].

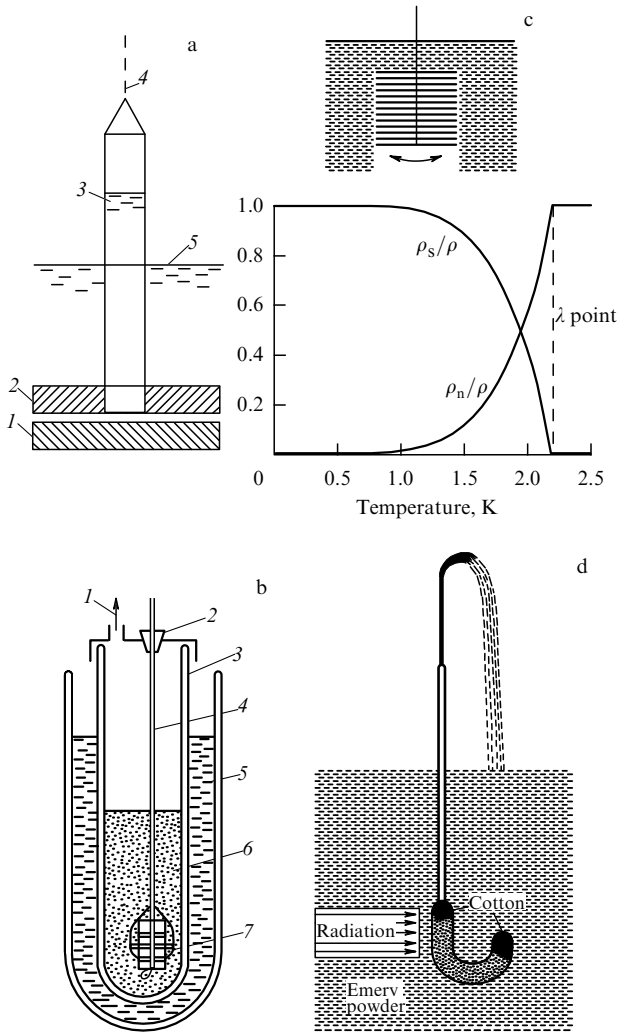

Figure 3. Properties of superfluid helium: (a) experiment of Kapitza: *1*, *2*, glass disks; *3*, tube; *4*, filament; *5*, level of fluid in Dewar flask; (b) double Dewar flask for working with liquid helium: *1*, fan or pump; *2*, rubber plug; *3*, double Dewar flask; *4*, measuring instrument; *5*, liquid nitrogen; *6*, liquid helium; *7*, specimen; (c) setup and results of the experiment of Andronikashvili; (d) the fountain effect.

ment is shown in Fig. 3a[7]. Here, *1* and *2* are two optically polished glass disks, 3 cm in diameter, the gap between which can be varied and measured. The upper disk has an opening 1.5 cm in diameter, and tube *3* is placed above this opening. The entire assembly was suspended by the filament *4* in the Dewar flask with liquid helium so that the level of fluid in the tube could be placed above or below the level of fluid *5* in the Dewar flask. The rate of fluid flow through the gap between the disks was monitored by the height of the fluid column in the tube. Measurements made with a gap thickness of ~ 0.5 μm indicated that the flow rate of HeII was higher by a factor of 1,500 than that of HeI (or even 10,000 if some other circumstances are taken into account). In this connection, Kapitza concluded that 'liquid helium below the λ point assumes a special modification that could be referred to as *superfluid*.'

[7] We note that this and subsequent diagrams are schematic. To work with liquid helium, one needs two nested Dewar flasks: the outer one filled with liquid nitrogen, and the inner one with liquid helium (Fig. 3b).

In addition to the preservation of the liquid state down to $T = 0$ and the superfluidity, liquid helium displays a number of other typically quantum properties, one of which was noted by Pitaevskiĭ [39]. Let us compare the laws of rotational motion of a normal fluid and a superfluid poured into cylinders rotating at an angular velocity Ω. As is known, the linear velocity v for a normal rotating fluid varies with the distance r from the axis of rotation as

$$v = \Omega r\,, \tag{3}$$

whereas, for a superfluid, we have

$$v = \frac{\hbar}{M}\frac{1}{r}n\,. \tag{4}$$

Here, M is the mass of the atom and n is an integer. This formula holds for $\Omega > \Omega_{cr}$, where Ω_{cr} is a certain critical angular velocity. Thus, in the latter case v decreases with r rather than increasing and is quantized. The formula for v explicitly involves the principal constant of quantum mechanics $\hbar$, which in principle can be determined from this experiment. We note the *macroscopic* nature of this *quantum* motion (r being much greater than the mean distance between the atoms). Later we shall discuss some other properties of quantum liquids.

It can be said that the theory of superfluidity evolved in a spiral. As noted above, the first theories (London [16], Tisza [17]) assumed Bose properties for liquid helium. For example, Tisza in 1938 proposed that the theory be based on the notion of liquid superfluid helium as a degenerate ideal Bose gas, which, by virtue of the Bose–Einstein statistics, can form Bose condensation with its special properties.

Later (in 1941), Landau noted that the notion of liquid helium as a degenerate ideal Bose gas is not acceptable because of the interaction between its atoms, and constructed a theory now known as two-fluid hydrodynamics [13]. Even though Landau in his theory does not use Bose properties of ^{4}He, he, in essence, outlines certain main ideas of the future microscopic theory of superfluidity. In particular, precisely in this theory Landau tackled for the first time the problem of elementary excitations of a quantum liquid (quasi-particles), which are responsible for the energy spectrum of HeII, and described this spectrum (refined in 1947 in Ref. [15]).

Finally, in the contemporary microscopic theory of superfluidity, physicists again consider liquid helium II a (nonideal) Bose system with an energy spectrum of elementary excitations similar to that predicted by Landau in 1947 and calculated in the same year by Bogolyubov in his microscopic theory [18]. Quasi-particles that form the spectrum at low temperatures do not interact with one another; that is, they behave as an ideal Bose gas. Now we will give a brief and very popular account of all these developments.

Let us start with Tisza's proposal to consider HeII a degenerate ideal Bose gas. First of all, we recall a few relevant concepts. In a broad sense, a Bose gas is a quantum gas consisting of bosons — identical particles with an integer spin 0, 1, or 2 — in units of $\hbar$, which obeys the Bose–Einstein statistics; accordingly, any number of particles can occupy the same state. The ideal Bose gas is one of noninteracting particles. A typical Bose gas of elementary particles is the photon gas, and a typical Bose gas of quasi-particles is the phonon gas. At temperature T below the so-called tempera-

ture of degeneration T_d, the Bose gas turns into the degenerate Bose gas, with most of its particles gathering in the state with zero momentum (Bose – Einstein condensate).

The temperature of degeneration T_d is for most gases well below the temperature of liquefaction and even the temperature of solidification. The only exception is helium, which goes over at $T = T_\lambda$ into the superfluid HeII modification. Tisza associated this property with the Bose condensation of the degenerate ideal Bose gas. As noted above, Landau considered this point of view incorrect, since it disregards interactions between the particles.

Thus, liquid helium is not a Bose gas. What will occur, however, if we consider, following Tisza, his model, assuming for the moment that liquid helium at low temperatures can be regarded as an ideal degenerate Bose gas? The result proves to be quite wonderful. If we substitute the parameters of liquid helium (density and atomic mass) into the expressions for a real Bose gas, we will find the range of existence of the Bose condensate $(0-T_d)$ for our pseudo-ideal 'Bose gas' to be 0 – 3.1 K, which differs little from $0-T_\lambda$, where $T_\lambda = 2.17$ K. This is definitely not so bad for a model that totally neglects the interactions. This agreement allows us to assume that the Bose – Einstein statistics is definitive for the superfluidity of ^{4}He. Another advantage of Tisza's model was that it stimulated (although in a rather special way) the advent of Landau's theory.

Landau's theory is often referred to as phenomenological[8] or quasi-microscopic. However, it was the first (and seminal, in the opinion of the Nobel committee) theory that allowed the explanation of many known features of liquid helium and the prediction of a number of new ones. This theory is based on the assumption that HeII consists of two interpenetrating components — a normal one with density ρ_n and velocity v_n, and a superfluid one with density ρ_s and velocity v_s, so that the total density is

$$\rho = \rho_n + \rho_s \,, \tag{5}$$

and the total momentum of unit volume is

$$p = \rho_n v_n + \rho_s v_s \,. \tag{6}$$

At $T \to 0$, we have $\rho_n \to 0$ and $\rho_s \to \rho$; at $T \to T_\lambda$ (from zero), we have $\rho_s \to 0$ and $\rho_n \to \rho$.

The two-component nature of HeII was experimentally confirmed by Andronikashvili [40], who measured the moment of inertia of a system consisting of thin aluminium disks performing torsional oscillations in HeII at different temperatures. Owing to the smallness of the gaps between the disks, at $T > T_\lambda$, the liquid helium was completely entrained by the motion, which was indicated by the large period of torsional oscillations. At $T < T_\lambda$, the period abruptly became much shorter, because only the normal component of liquid helium in this case took part in the motion together with the disks.

[8] We call upon young physicists not to consider 'phenomenological' a derogative term. Quite the opposite. A result derived (and subsequently validated by experiment) not from the first principles of quantum mechanics (which would also be difficult but ideologically straightforward) but from general considerations based only on certain assumptions and indirect experimental indications is a sign of outstanding scientific intuition and creative imagination. It is probably not by accident that the words 'phenomenological' and 'phenomenal' have the same root. Landau's theory is a brilliant confirmation of this opinion.

Andronikashvili's experiment and its results, expressed in the form of the ratios ρ_s/ρ and ρ_n/ρ as functions of the temperature T, are illustrated in Fig. 3c, reproduced from the book [41], which we recommend to the reader as the source closest to this section of our article in terms of content and popularity of presentation (but certainly much more informative). We can see from the diagram that, as the temperature varies from zero to T_λ, ρ_n actually grows from zero to ρ, whereas ρ_s falls from ρ to zero.

In order to explain the main feature of liquid helium — its superfluidity at $T < T_\lambda$ — Landau considered the problem of the energy spectrum of liquid helium, stressing the fact that the point was not the levels of individual atoms, but the energy states of the liquid as a whole (cf. a similar situation in solid state physics in Section 5.3). According to the original standpoint of Landau, developed in his paper of 1941 [13], the two possible forms of motion of a quantum fluid (potential and vortical) correspond to two spectra whose lowest levels are separated by a certain finite energy interval $\Delta \simeq kT_\lambda$. Landau assumed that the lowest level corresponding to potential motion lies below the lowest level for vortical motion. This is the case for the ground states (or 'normal' states, in Landau's usage). As far as the excited levels are concerned (we mean the levels very close to the ground levels), they are considered as a combination of individual elementary excitations — phonons with the energy $\varepsilon = u_1 p$ (where u_1 is the speed of sound)[9] for potential motion and rotons with the energy $\varepsilon = \Delta + p^2/2\mu$ (where μ is the effective mass of the roton) for vortical motion. In the case of weakly excited states, the number of phonons and rotons is not large, and they can be represented as a mixture of two ideal Bose gases.

In terms of a semiclassical and semiquantum treatment of the energies of individual excited atoms of liquid HeII, one can say that the process of formation of phonons and rotons leads to the deceleration of the atoms. It should be remembered, however, that, in the context of the quantum theory of liquids, one cannot distinguish atoms in the ground state from those in excited states; that is, one should speak of the deceleration of the fluid as a whole, or at least, of some macroscopic portion of the fluid.

Based on the above-described structure of the energy spectra of liquid helium and the numerical values of u_1 and Δ, Landau demonstrated that at near-zero temperatures, given that the velocity of HeII particles is small, the birth of quasi-particles is energetically not beneficial; that is, phonons and rotons, which cause the deceleration of the fluid, are not generated. And "*this means*," Landau writes, "*that the flow of liquid will not be impeded, and helium will exhibit the phenomenon of superfluidity*."

In 1947, Landau, in view of the experimental discovery of the phenomenon of a second sound in helium, which he had predicted (see details below), obtained in his two-fluid theory a very important result, which later became famous — the form of the energy spectrum of elementary excitations of HeII [15]. In contrast to his study of 1941, where the spectrum was represented by two (photon and roton) branches, in Ref. [15] Landau represented the spectrum by a common function $\varepsilon(p)$ (where ε is the energy of elementary excitations and $\mathbf{p}$ is their momentum). Landau did not speak of phonons and rotons as different types of elementary excitations, whose spectra are

[9] Unlike Landau, who denotes the velocity of sound by c, we denote it by u_1, because below we will also deal with a second sound, whose velocity we denote by u_2.

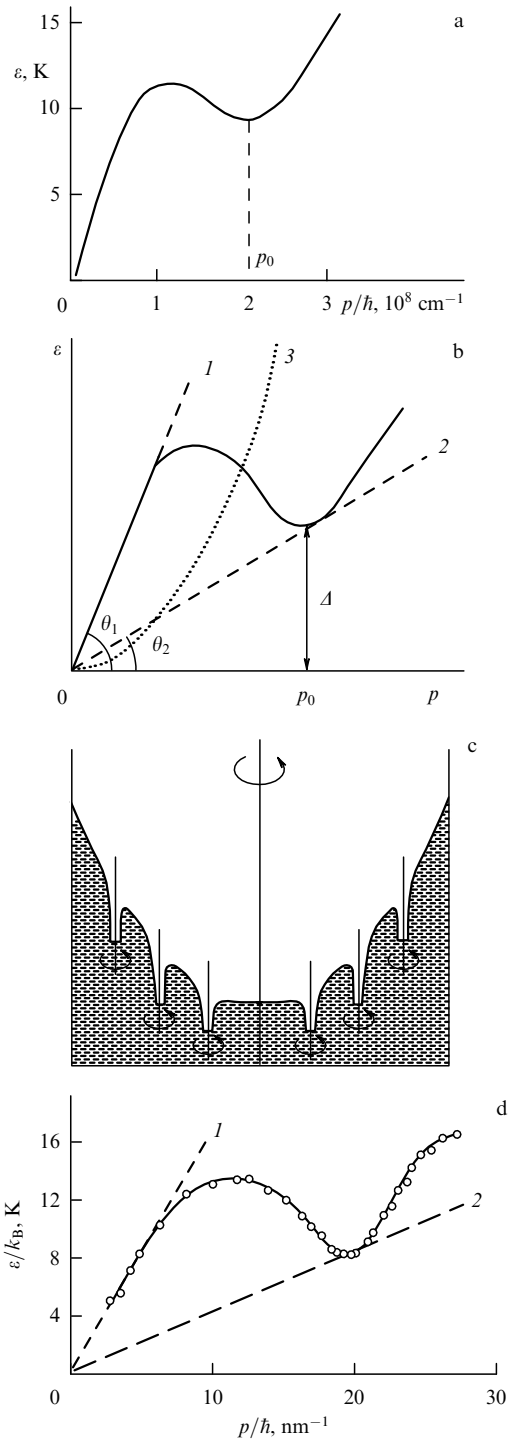

Figure 4. Spectrum of elementary excitations and critical velocity: (a) Landau's spectrum; (b) determination of critical velocity according to Landau: *1*, first solution; *2*, second solution; *3*, the case of free particles; (c) arrangement of vortex filaments in rotating HeII; (d) experimental validation of Landau's spectrum: *1*, first solution; *2*, second solution; ○, experimental results.

separated by an energy gap Δ; instead, he referred to long-wave (at small p) and short-wave (at $p \approx p_0$) excitations (Fig. 4a, b). It became possible to combine the two branches of the previously considered spectrum into one branch by assuming that the function $\varepsilon(p)$ is linear in the long-wave region and has a minimum in the short-wave region, where $\varepsilon = \Delta + (p - p_0)^2/2\mu$ (μ is a constant). In Fig. 4b, the shape of the spectrum of elementary excitations $\varepsilon(p)$ can be compared with the function $\varepsilon = p^2/2m$ (dashed line *3*), which characterizes the spectrum of free particles. We can see from the diagram that they have nothing in common. This provides convincing evidence that Landau's spectrum characterizes the motion of the entire fluid rather than the motion of its individual atoms. The spectrum shown in Fig. 4a was constructed by Landau for the values

$$\frac{\Delta}{k_B} = 9.6\ \text{K}\,, \qquad \frac{p_0}{\hbar} = 1.95 \times 10^8\ \text{cm}^{-1}, \qquad \mu = 0.77 m_{\text{He}}\,, \quad (7)$$

where k_B is the Boltzmann constant. This spectrum was confirmed by Bogolyubov somewhat later in his microscopic theory of superfluidity [18] (constructed, in essence, from the first principles of quantum mechanics) and validated experimentally in measurements of neutron scattering (see p. 504).

Let us now discuss how Landau derived the criterion of superfluidity of HeII. Following Khalatnikov (whose books [42, 43] we strongly recommend to the reader for a closer examination of the theory of superfluidity), we will not speak separately of Landau's papers of 1941 and 1947, but will discuss instead what follows from both of them. First we consider the case of $T = 0$.

Assuming that HeII flows through a capillary with a certain velocity $\mathbf{V}$, we find its critical value $\mathbf{V}_{\text{cr}}$ at which the friction resulting from the interaction with the walls of the capillary leads to thermal excitations in the fluid, which are of the type of quasi-particles with the energy $\varepsilon(p)$ and momentum $\mathbf{p}$; the fluid itself slows down to the velocity $\mathbf{V}'$.

To avoid complications due to using different reference frames (the fluid moves in the laboratory frame, while the excitations are generated in the fluid's rest frame) and going from one frame to another, we assume that the fluid is at rest, while the capillary of mass M moves with the velocity $\mathbf{V}$. Assume also that the 'drag' of the capillary reduces its velocity to $\mathbf{V}'$, and that we perform our treatment in the frame fixed in the fluid. (Clearly, such a trick will not change the value of $\mathbf{V}_{\text{cr}}$.)

In this frame the conservation of momentum has the form

$$M\mathbf{V} = M\mathbf{V}' + \mathbf{p}\,, \quad (8)$$

which, for the convenience of subsequent calculations, we rewrite as

$$M\mathbf{V}' = M\mathbf{V} - \mathbf{p}\,. \quad (9)$$

Hence from some straightforward algebra, we obtain

$$\frac{MV^2}{2} = \frac{MV'^2}{2} + \mathbf{Vp} - \frac{p^2}{2M}\,. \quad (10)$$

In addition, from the law of conservation of energy

$$\frac{MV^2}{2} = \frac{MV'^2}{2} + \varepsilon(p)\,, \quad (11)$$

we find

$$\varepsilon(p) = \mathbf{Vp} - \frac{p^2}{2M}\,, \quad (12)$$

or, disregarding the small quantity $p^2/2M$ (since M is large),

$$\varepsilon(p) \approx \mathbf{Vp} \leqslant Vp\,. \quad (13)$$

In this way, elementary excitations in HeII can arise only if it flows with a velocity $V \geqslant \varepsilon(p)/p$. At $V < \varepsilon(p)/p$, the formation of elementary excitations is not beneficial in terms of energy, and the flow of HeII with a velocity $V < \varepsilon(p)/p$ ought to proceed without energy dissipation, which means superfluidity. It is also clear that we need to take the least value of $\varepsilon(p)/p$, that is, to find $\min(\varepsilon(p)/p)$. The flow velocity of liquid HeII at this value of $\varepsilon(p)/p$ is called critical:

$$V_{\rm cr} = \min \frac{\varepsilon(p)}{p} \,. \tag{14}$$

The minimal value of the function $\varepsilon(p)/p$ can be found from the extremum condition

$$\frac{\mathrm{d}}{\mathrm{d}p} \frac{\varepsilon(p)}{p} = 0 \,, \tag{15}$$

which leads to the equation

$$\frac{\varepsilon(p)}{p} = \frac{\mathrm{d}\varepsilon(p)}{\mathrm{d}p} \,. \tag{16}$$

The left-hand side of equation (16) is the tangent of the angle θ between the axis of momentum and the straight line drawn from the origin of coordinates to the intersection with the curve $\varepsilon(p)$. The right-hand side is the slope of the tangent to this curve (Fig. 4b). The equality of the two sides means that $\min(\varepsilon(p)/p)$ is reached at the location defined by the tangent to the curve $\varepsilon(p)$ passing through the origin. This condition is satisfied by the two straight lines marked in Fig. 4b with numbers *1* and *2*. The first one is a tangent to the initial portion of the curve $\varepsilon(p)$ (and coincident with it) and gives for $V_{\rm cr}$ a value equal to the speed of sound u_1 (since $\varepsilon = u_1 p$), which is approximately 250 m s^{-1}. The second solution can be obtained if we draw a tangent from the origin to the curve in the region of its minimum [10]. It can be proved (and is more or less obvious) that the point of contact is near the minimum of the curve $\varepsilon(p)$; in this case, $V_{\rm cr} \simeq 60$ cm s^{-1}.

It is easy to see that the case of $T \neq 0$ (at $T < T_\lambda$) does not compromise the reasoning performed under the assumption of $T = 0$, because the absence of 'proper' excitations (caused by thermal oscillations rather than by the motion of the fluid) was not a prerequisite for the argumentation. In the case of $T \neq 0$ such 'ready-made' excitations exist at $V < V_{\rm cr}$ as well, but new excitations (caused by the flow of fluid) can appear only at $V \geqslant V_{\rm cr}$, so that all arguments pertaining to the case of $T = 0$, including the value of $V_{\rm cr}$, still hold. The only specific feature of the case of $T \neq 0$ is that these 'ready-made' oscillations will interfere with the motion of the part of fluid to which they pertain, slowing down this motion. (Recall that we cannot consider the motion of individual atoms of liquid helium, but can speak of the motion of some portion of the fluid.) As a result, we get (this time on a microscopic level) the two-fluid model with normal and superfluid components of liquid HeII, which we discussed earlier. The normal component is represented by the gas of quasi-particles — elementary excitations.

The clarity and transparency of this explanation of the superfluidity of HeII notwithstanding, this argument is only qualitative, because the only quantitative conclusion is the form of the equality $V_{\rm cr} = \min(\varepsilon(p)/p)$, while the function $\varepsilon(p)/p$ can have several minima, (we have already seen two of them, and we shall shortly see that even the least of them is not the absolute minimum). For this, however, we need to stop our discussion of Landau's theory temporarily and make a digression to some other results.

Subsequent experiments with HeII demonstrated that the value of $V_{\rm cr}$ shows a strong dependence on the diameter of the capillary and on the temperature, and can be two or three orders of magnitude less than the theoretical result presented above. Such huge disagreement between theory and experiment was attributed to specific features of the rotational motion of HeII. A little earlier we mentioned that a *quantum* fluid is characterized by a specific *macroscopic* quantization of the linear velocity v of the rotational motion, and an uncommon dependence of velocity on the distance to the axis of rotation. It is easy to see that formula (4) describes the quantization of the angular momentum of the superfluid component of HeII:

$$mvr = n\hbar \,. \tag{17}$$

The question of quantization of rotational motion of liquid helium was first considered in 1949 by Onsager [44] and then in 1955 by Feynman [45]. The main ideas of this treatment are as follows.

The above-described particular features of the rotational motion of the superfluid component of liquid helium were called the quantum vortex, and the axis around which it forms came to be known as the vortical filament. The quantum vortex is formed when the kinetic energy of rotation is sufficiently large, that is, when $\Omega > \Omega_{\rm cr}$, where $\Omega_{\rm cr}$ is the critical angular velocity mentioned in the discussion of formula (4). Since the kinetic energy is proportional to v^2, then, in accordance with expression (17), in which $v^2 \sim n^2$, at a given angular velocity Ω, the formation of two vortices with $n = 1$ and a smaller radius is energetically more beneficial than the formation of one vortex of a larger radius with $n = 2$. (The excitation of a vortex with $n = 2$ requires a greater kinetic energy of rotation of the layer near the larger radius, since this energy increases with increasing n. This, however, is not possible, because, according to (4), v decreases with increasing r.) As v increases (with increasing Ω), this tendency towards reduction of the radius of vortices leads to a situation where the rotating cylinder with HeII is filled with vortical filaments in the entire bulk of the fluid (Fig. 4c).

Vortex filaments arise not only in rotating HeII but also on any particles moving in it — for example, on helium ions (see below), as well as in the course of flow of HeII through the capillary, if it occurs at a sufficiently high velocity $V > V_{\rm cr}$. In this case, vortical filaments can curve up and form *vortex rings*, which travel over the entire volume of HeII, similarly to the elementary excitations considered above. The interaction between the vortex rings arising in the superfluid component at $V > V_{\rm cr}$ and the quasi-particles of the normal component results in friction, which destroys the property of superfluidity. The calculated value of $V_{\rm cr}$ depends on the diameter of the capillary d, and, at $d \simeq 0.01$ μm (as in Kapitza's experiments), is rather close to the experimental value, but still well below the value of $V_{\rm cr}$ calculated by formula (14).

The theory of formation of vortex rings, based on the quantization of the angular momentum of the rotating

[10] The use of condition (14) in the case of free particles (dashed line *3* for the function $\varepsilon = p^2/2m$ in Fig. 4b) leads to the value $V_{\rm cr} = 0$, which means that superfluidity is not compatible with their motion.

superfluid component of HeII, has been validated experimentally. The idea of the experiment consists of the deliberate formation of vortex rings on helium ions produced in HeII by exposing it to ionizing radiation. The movement of ions (together with vortex rings) can be controlled with the electric field, whose energy is expended for the production of the vortex rings. Experimental monitoring of the motion of the ions allows us to assess the energy and momentum of the attached vortex ring.

After this brief digression, we return to the description of Landau's theory, which is often called two-fluid or two-velocity hydrodynamics (because of the use of hydrodynamic variables — density ρ, flux j, entropy s, and chemical potential μ).

In addition to having provided a qualitative explanation of superfluidity, Landau's theory also explained some other known features of the behavior of liquid helium, while it predicted new phenomena (which were later discovered experimentally). For example, the extremely high thermal conductivity of HeII is attributed to the fact that the transfer of heat depends not on the common mechanism of heat conduction, but on the mechanical movement of the normal component that carries heat. The superfluid component does not carry heat, and the equalization of temperatures in HeII thus occurs through the motion of the two components of the fluid in opposite directions (without friction). The helium as a whole remains motionless, that is, $\rho_n v_n + \rho_s v_s = 0$, and v_n and v_s are oppositely directed (one part of the liquid flows through the other).

A similar explanation is given for the so-called mechanocaloric effect, which consists of the heating of HeII in a vessel when the superfluid component leaves through a capillary placed on the top without carrying away any heat (so that the amount of fluid is reduced but the amount of heat does not change). An equally simple explanation is given to the thermomechanical effect of intense fountaining of the superfluid component from the upper end of a capillary when its lower end, immersed into a tray with liquid helium, is heated. A scheme of the experiment is shown in Fig. 3d [46], from which we can see that the lower end of the 'capillary' is in fact a tube fairly large in diameter, filled with a dark emery powder, whose particles form winding channels with a clearance of about 0.1 μm; they pass (like the upper capillary) only the superfluid component, which does not carry the heat. When the tube with the powder is heated (with the beam of a pocket flashlight), the pressure in the tube increases, and the fluid spouts from the capillary. The depletion of the superfluid component in the tube is made up by its supply from the tray, so that the fountaining stops only when the flashlight is turned off.

Among the new effects predicted by the theory of Landau (and Tisza), we should mention the second sound, which is manifested in the form of a temperature wave that forms during the above-mentioned flows of the two components of HeII in opposite directions, with the helium as a whole being motionless (in other words, there are no density oscillations that result in the conventional (first) sound). The second sound was observed experimentally in 1946 by Peshkov [47], with the aid of a sensitive thermometer. Its velocity increases with decreasing temperature from $u_2 = 20$ m s^{-1} at $T = 1$ K to $u_2 = 140$ m s^{-1} at $T = 0$ (the speed of first sound is $u_1 = 230$ m s^{-1}). Thus, Landau's theory is successful not only in explaining the known effects, but also in predicting new ones.

Let us now give a brief and clear account (which is, as we will shortly see, not an easy task) of the microscopic theory of superfluidity, pioneered by Bogolyubov [18]. To give a feel of the difficulty and complexity of the problem handled and solved by Bogolyubov, we quote the opening words from his paper: "*If we set the task of developing not a phenomenological but a molecular theory, and start only from the conventional 'microscopic' equations of quantum mechanics, it becomes immediately clear that attempts at theoretical calculations of the properties of a real fluid are hopeless. From the molecular theory of superfluidity, at least at the first stage, one can demand only a basic qualitative explanation, relying on a simplified model*."

Making further mention of Tisza's model [17] (the degenerate ideal Bose–Einstein gas) and the fairness of Landau's criticism of the failure to take into account the neglecting the interactions between 'superfluid' particles (degenerate condensate) with excited particles during collisions, Bogolyubov proposes his own model, which, like Tisza's, is based on the Bose–Einstein statistics, but is more realistic. Bogolyubov stresses that his notion of collective excitations (instead of considering individual molecules) was first used by Landau in Refs [13, 15], and that his (Bogolyubov's) theory is constructed "*without making* certain assumptions regarding the nature of the energy spectrum", based "on the model of a nonideal Bose–Einstein gas *with weak interactions between particles*."

Bogolyubov's paper [18] is styled in such a way that "the words are few and the ideas are many." It is mathematically rather sophisticated, and we will only quote here those few words from this paper which can sometimes be found among the numerous formulas.

Before embarking on a popular analysis of Bogolyubov's work, let us quote another passage that can be regarded as a kind of summary:

"*Thus, the total energy of the nonideal gas under consideration is combined from the energy of the ground state H_0 and the sum of the individual energies of separate quasi-particles. The quasi-particles evidently do not interact with one another and form an ideal Bose–Einstein gas. This result only applies to weakly excited states*."

And now let us give our popular commentary (based on information from the literature cited at the end of this section). According to Bogolyubov, the ground state with an energy H_0 (at $T = 0$) is occupied by a macroscopically large number of particles with zero momentum (Bose condensate in the space of momenta), which are accelerated by an external force to the same velocity and (if this velocity is not too high) move without friction, since separation of even one particle from the entire moving mass is energetically not beneficial (cf. Landau's reasoning concerning his estimate of V_{cr}). Bose condensate at $T = 0$ K is described by the common quantum mechanical wave function

$$\psi = \sqrt{n_0}\exp(i\varphi), \tag{18}$$

where n_0 is the density of the condensate (virtually coinciding with ρ_s in Landau's theory) and φ is the phase that determines the motion of the superfluid component v_s in this theory ($v_s \propto \nabla\varphi$). The coherent motion of the superfluid component with all its particles having the same velocity means that the motion of each volume of the fluid is potential (non-vortical) and experiences no drag.

As the temperature increases, the number of particles in the condensate decreases ('depletion of condensate'), and the Bose system goes gradually into excited states, which, at very low temperatures, can be described in terms of noninteracting quasi-particles — elementary excitations. In this context, they can be regarded as an ideal Bose gas with a total energy equal to the sum of the energies of the individual particles. Thus, the complex problem of determining the energy states of the quantum fluid made up of *weakly interacting* atoms of ^{4}He reduces to a simpler problem of constructing the energy spectrum of *noninteracting* quasi-particles that exhibit properties of an ideal Bose gas. Since their energy is additive, one only needs to find the energy of one quasiparticle.

The energy ε of a quasiparticle is a function of its momentum p, and the function $\varepsilon(p)$ is known as the energy spectrum of elementary excitations. It was first derived by Landau in 1947 [15] based on his two-fluid theory of 1941 [13], under certain assumptions regarding the nature of the spectrum (Fig. 4a). In the same year, 1947, Bogolyubov obtained this spectrum in his consistent molecular theory of superfluidity without additional assumptions regarding the structure of the spectrum.

We will not write out the rather sophisticated expression for the energy of the ground state H_0 (Bose condensate), but will only reproduce Bogolyubov's formula for the energy of a quasiparticle (replacing his notation with the symbols used in this review):

$$\varepsilon(p) = \left[\frac{\rho v}{M^2}\,p^2 + \left(\frac{p^2}{2M}\right)^2\right]^{1/2}, \qquad (19)$$

where ρ is the density of HeII, M is the mass of the ^{4}He atom, and $v = \text{const}$ (the scattering length in the s-state). From formula (19), it is clear that, at small p (when the second term can be neglected), ε is a linear function of p (the phonon portion of Landau's spectrum), and it is a quadratic function as $p \to \infty$. However, the 'roton' portion of Landau's spectrum (the minimum of the curve in Fig. 4a) cannot be described by formula (19). At the same time, this formula gives a good description of experiments with gaseous condensates carried out half a century (!) later (see the end of Section 2.2).

Experimentally, the spectrum $\varepsilon(p)$ was measured in 1961 by Henshaw and Woods [48] in their experiments on the scattering of monochromatic neutrons ($\lambda = 4.04$ Å) in HeII maintained at the temperature $T = 1.12$ K. In these experiments, the energy and momentum of quasi-particles were determined from the changes in the energy and momentum of neutrons scattered by angles of 10° to 140.7° from the initial direction of the beam. These changes, as demonstrated in 1957 by Cohen and Feynman [49], must be equal to the energy and momentum of the excitation impulse generated in liquid helium. Since the experimental broadening of the spectral line of scattered neutrons is small for any angle of scattering, one can assume that this excitation corresponds to the formation of a single phonon.

The results of the experiment are represented in Fig. 4d with circles, whose sizes crudely correspond to the accuracy of the experiment (the smooth curve was drawn freehand by the experimenters). The dashed straight line *1*, which coincides with the initial linear portion of the curve $\varepsilon(p)$, has a slope corresponding to the velocity of the first sound $u_1 \simeq 237\ \text{m s}^{-1}$ [the first solution to equation (16) for V_{cr}, according to Landau]. After this first segment (at $p > 0.6\ \text{Å}^{-1}$) the curve $\varepsilon(p)$ goes below the phonon branch, reaches a maximum at $p = 1.10\ \text{Å}^{-1}$ and $\varepsilon = 13.7$ K, passes through a minimum at $p = 1.91\ \text{Å}^{-1}$ and $\varepsilon = 8.65$ K, and goes up again. (Dashed line *2* describes the second solution to equation (16) for V_{cr}, according to Landau — $V_{\text{cr}} \approx 60\ \text{m s}^{-1}$.) At $p > 2.4\ \text{Å}^{-1}$ the second derivative of the function $\varepsilon(p)$ becomes negative, and the curve exhibits the signature of a second maximum.

Comparing Fig. 4d with the above-described Fig. 4a and 4b, we can see that they are identical. The experiment of 1961 fully validated the spectrum of elementary excitations proposed by Landau in 1947 and calculated by Bogolyubov in the same year.

With this we end our exposition of superfluidity, assuming that we have completed our task (a description of the work done in this field during the period when Kapitza and Landau worked in this direction). Of course, the science of superfluidity did not stop at this point. We will give a few examples. As noted above, the formation of vortical filaments and vortical rings in HeII was studied in the late 1940s and early 1950s. Somewhat later (in 1958), Ginzburg and Pitaevskiĭ constructed a theory of superfluidity of liquid helium near the λ point [50]. Simultaneously, right after 1957, when the BCS theory of superconductivity was developed, intense theoretical and experimental studies of the possible superfluidity of ^{3}He were launched.

In principle, this possibility seemed obvious, since the formation of a Cooper pair of two Fermi atoms of ^{3}He, similar to the pairing of two electrons in the BCS theory, ought to result in the formation of a Bose pair of atoms of ^{3}He with opposite momenta on the Fermi surface. However, the realization of this possibility required a mechanism of attraction between atoms. In the BCS theory, this mechanism was the electron – phonon interaction (for more details, see Section 2.3). To explain the possible superfluidity of ^{3}He, Pitaevskiĭ [51] proposed in 1959 a mechanism of formation of Cooper pairs, based on the van der Waals attraction between two atoms of ^{3}He with an orbital moment $L \neq 0$.

Therefore, the analogy between Cooper electron pairs and Cooper pairs of ^{3}He atoms cannot be regarded as complete. Unlike the case of superconductivity, where the spin J and the orbital moment L of a pair of electrons are zero, a pair of He atoms in superfluid ^{3}He has $J = 1$ and $L = 1$. It turns out that, depending on the values of the projections of J and L onto the directions of their quantization axes, superfluid ^{3}He should have three phases, known as A, B, and B_1.

After a long search, the superfluidity of ^{3}He was discovered in 1972 by Osheroff, Halley, Richardson, and Lee [52]. It was observed at highly exotic parameters: at $T_c = 2.7 \times 10^{-3}$ K and a pressure of 34 atm, as well as at 0.9×10^{-3} K and the pressure of saturated vapor[11].

Many studies have been carried out (and are being continued) using the method of neutron spectroscopy of the fraction of liquid helium ^{4}He that is in the condensate state. However, the discussion of these and other contemporary studies of superfluidity falls beyond the scope of this review. We hope that the minimum of information that we have managed to present here will help the reader to digest more serious literature, such as, for example, the above-mentioned

[11] David M Lee, Douglas D Osheroff, and Robert C Richardson received a Nobel Prize in physics in 1996 for the discovery and study of the superfluidity of ^{3}He [53 – 55]. In his Nobel lecture [55], Richardson emphasized the importance of the Pomeranchuk effect [56], used for cooling helium by compression.

article [5] and books [41–43]. In addition, we recommend articles by Rozhkov [57], Mineev [58], and Volovik [59] dealing with the superfluidity of ^{3}He, by Ginzburg and Sobyanin [60, 61] (on the generalized theory of superfluidity of ^{4}He), and their paper [62] (on the superfluidity of HeII near the λ point), as well as the publication by Bogolyubov, Jr; and colleagues [63] (on the microscopic theory of superfluids), and an autobiographical article by Ginzburg [34] (on the research on superconductivity and superfluidity). Close to this review in scope and style are writings by Kukin [64] on superfluidity and superconductivity.

At the end of this section on superfluidity and Bose systems, let us note one of the most recent accomplishments in this field, distinguished by the 100th Nobel prize for "the achievement of Bose–Einstein condensation in dilute gases of alkali atoms and for earlier fundamental studies of the properties of condensates." The laureates were Eric Cornell, Carl Wieman, and Wolfgang Ketterle. The results of their early work on the production and study of Bose–Einstein condensates of diamagnetic atoms were published in 1995. This work used technological achievements from other domains of physics, such as magnetic traps developed for the magnetic confinement of high-temperature plasmas, the laser technique of cooling atoms (described in Section 4.4), and high-frequency magnetic fields for achieving superlow temperatures ($< 10^{-6}$ K). The first to be obtained were the condensates of atoms of ^{87}Rb, ^{23}Na, and ^{7}Li; in experiments with sodium, the number of atoms in the condensate was as large as 10^6. Note that the Bose–Einstein gaseous condensates discovered by Cornell and colleagues can well be (quantitatively) described by Bogolyubov's theory [see comments on formula (19)], because the spectra of elementary excitations of these condensates do not exhibit the roton minimum. For a detailed familiarization with the properties of the new condensates, experimental techniques, and theory, we recommend reviews by B B Kadomtsev and M B Kadomtsev [65], and Pitaevskiĭ [66].

2.3 Low-temperature superconductivity (LTSC): experimental background and elementary presentation of the Ginzburg–Landau and Bardeen–Cooper–Schrieffer theories

This section mainly contains pictorial presentation of elements of the microscopic theory of superconductivity of Bardeen, Cooper, and Schrieffer (BCS) [23], developed by physicists outside of Russia. However, keeping in mind the main purpose of our article (to honor the Russian contribution to exotic features in general physics), we will preface this presentation with a brief account of the Ginzburg–Landau theory [19], published 7 years before the advent of the BCS theory. Recall that we have already touched upon this theory in Section 2.1, where we noted that not only did it overcome certain difficulties of the theory of the London brothers [8], but also it actually predicted one of the main results of the BCS theory.

As opposed to the purely classical Londons' theory, the Ginzburg–Landau theory, for describing the behavior of superconducting electrons, uses a quantum mechanical wave function (termed 'effective' by the authors), whose squared modulus $|\psi|^2$ is set equal to the density of the superconducting electrons. The form of the function $\psi(x, y, z)$ can be found from the equations of the theory by analytical methods in the case of a weak field, and by numerical methods in the case of a strong field. Using the quantum mechanical treatment of the phenomenon of superconductivity, Ginzburg and Landau were able not only to explain numerous known experimental results, but also to find two essentially new features — the dependence of the depth of field penetration λ (that is, the thickness of the layer that carries the compensating current) on the external magnetic field (in the case of a thick plate) and on the thickness of the sample (in the case of thin films).

Unfortunately, the popular approach adopted in this review is not suitable for a more comprehensive treatment of the Ginzburg–Landau theory. This would require introducing too many concepts that might be new to our readership. Suffice it to say that the theory [19] uses the nontrivial concept of type-II phase transitions (with zero latent heat)[12], a theory of which was developed by Landau in 1937 [67], as well as such concepts as the free energy of the superconductor or the complex order parameter (whose description requires some knowledge in the group theory), and many other no less exotic (at first glance) concepts. In addition, Ref. [19] uses sophisticated mathematics.

The difficulty of explicating the Ginzburg–Landau theory [19] is also noted by the authors of a fairly sizable book on superconductivity, Rose-Innes and Roderick [68]. Nevertheless, we found a way of paying due tribute to the Ginzburg–Landau theory and emphasizing its perennial importance: we will cite the opinion of renowned foreign authors (the views of Russian specialists on the theory [19] can be found in textbooks on theoretical physics and articles in the 'Physical Encyclopedia', as well as in books by Schmidt [69, 70]).

For example, the author of one of the best known books on superconductivity, Linton, writes about the *increased* importance of the Ginzburg–Landau theory 19 (!) years after its publication (or 12 years after the development of the BCS theory) and discusses it in three chapters of the 3rd edition of his book [71]. In Ref. [41], D Tilly and J Tilly devote a whole 60-page chapter to the Ginzburg–Landau theory and write that this theory "*has launched one of the most productive approaches to superconductivity*"; they also note that even though the Ginzburg–Landau theory was developed before the advent of the BCS theory, its equations for certain ranges of temperatures and magnetic field strengths strictly follow from microscopic theory, and have been repeatedly validated by experiment. Moreover, since these equations are much simpler than the 'microscopic' equations, they are normally used in their range of applicability. Sometimes (for qualitative evaluation of the behavior of a semiconductor), they are used even outside of this area. To confirm this view, the authors of Ref. [41] at the end of this chapter give a long list of references in which the importance of the Ginzburg–Landau theory is discussed.

In Russia, the Ginzburg–Landau theory [19] was also developed by other physicists, in particular, by A A Abrikosov and L P Gor'kov. Gor'kov's work [24], in which he expressed the equation and the coefficients of the theory [19] based on the BCS theory, developed in 1957, has already been mentioned above (Section 2.1). Here, we would like to emphasize the importance of studies by A A Abrikosov, who introduced in 1957 the concept of quantized vortical filaments (rings) [72] in type-II superconductors. According

[12] Recall that common phase transitions of the first kind, like the evaporation or freezing of water (or the condensation of steam or the melting of ice) are associated with the release or absorption of the latent heat of melting or evaporation.

to Abrikosov, in the interval between the lower (H_{c1}) and the upper (H_{c2}) critical field, a type-II superconductor is in a mixed state, that is, it is made up of many small domains of superconducting and normal phases, with a developed phase interface. At $H = H_{c1}$, the superconductor is mainly in the superconducting phase, while the normal phase only starts to appear in the form of vortical filaments implanted in the bulk of the semiconductor, near which a penetrating quantized magnetic flux is concentrated (Abrikosov's vortex lattice). The number of filaments grows as H approaches H_{c2}, when the superconductor goes over into the normal state, completely permeable to the magnetic flux. The fact that the superconductivity is preserved until $H = H_{c2}$ is important because H_{c2} is several times greater than H_{c1}.

It is important to note that, according to Abrikosov's model, the properties of vortical filaments depend on the inhomogeneities in the superconductor and other defects of its structure, to which the vortices attach themselves. The strength of such attachment determines the value of the critical current and the critical field for the given semiconductor. Since the concentration of defects in the superconductor can be controlled in the manufacturing process, it proved to be possible to produce type-II superconductors that work with extremely high critical fields ($H > 10^5$ Oe) and critical currents ($I_c \approx 10^5 - 10^6$ A cm^{-2}). These are so-called hard type-II superconductors, which are sometimes referred to as type-III superconductors. They are used to manufacture powerful semiconducting magnets with high critical fields and currents.

Let us now turn to a popular exposition of the experimental background of the BCS theory and of elements of this theory. At the time of development of the BCS theory, the following experimental results on superconductivity were known.

(1) The transition of a material into the superconducting state occurs very rapidly. At $T = T_c$, the resistance of the superconductor abruptly drops down to zero, and the current increases sharply. Since the electric current both at $T > T_c$ and at $T < T_c$ is produced by the flow of the same conductivity electrons, these electrons at $T < T_c$ assume a certain new state of motion — less chaotic and more ordered, i.e., correlated. It appears that at $T < T_c$ the electrostatic repulsion between the electrons is supplemented by a certain new interaction working at a longer range, which steers the motion of the electrons and establishes a sort of long-range order. It is said in such cases that the system of particles exhibits a long-range order[13]. A correlation between electrons separated by a fairly large distance (of about 10^{-4} cm, according to certain estimates) indicates that there is some kind of attraction between electrons. The existence of a long-range order would hardly be possible if this additional interaction were of the repulsion type: after all, a (Coulomb) repulsion exists even at $T > T_c$, but there is no long-range order in this case.

(2) Low-temperature superconductivity is displayed by many metals and alloys with widely varying crystallographic and atomic characteristics. It would seem, therefore, that its existence depends on some very general properties of matter, such as the specific structure of the lattice at $T \leqslant T_c$. However, it follows from X-ray structural analyses that the parameters and properties of the superconductor lattice do not change as the superconductor is cooled below the critical temperature.

(3) At the same time, in 1950, it was established that the critical temperature T_c for a given element depends on the mass M of its isotopes approximately as

$$T_c \propto M^{-1/2} \tag{20}$$

(isotopic effect). This indicates that superconductivity has something to do with the positively charged ions of the crystal lattice, whose masses are different for different isotopes and can therefore differently affect the parameters of superconductivity, for example, because of the different frequencies of lattice oscillations in interactions with the conductivity electrons. This conjecture is confirmed by another experimental fact, which is again related to electrons.

(4) At a frequency of about 10^{11} Hz (the far infrared part of the electromagnetic spectrum) the superconductor features strong absorption of photons, similar to the well-studied absorption of radiation at about 10^{14} Hz by semiconductors (see Section 5.2), where it is explained by the presence of a forbidden band in the spectrum of electrons, which does not contain allowed energy levels (the energy gap). At $\sim 10^{14}$ Hz, the energy of a photon is large enough to carry an electron across the gap from the occupied allowed band into an empty band, which is observed as a sharp absorption edge. From the value of the absorption frequency, it is easy to estimate the width of the gap in the electron spectrum of the semiconductor: $\sim 0.1 - 1$ eV. In a similar way, the absorption of radiation by superconductors at $\sim 10^{11}$ Hz may point to the existence of an energy gap of $\sim 10^{-4}$ eV in their electron spectra. As the temperature increases from 0 K to T_c, the gap width decreases from its maximum value to zero.

Thus, from these experimental results (what we have discussed being far from the totality of experimental evidence), one can deduce that the cause of superconductivity lies in the interaction between electrons and the lattice, which can lead at sufficiently low temperatures to a weak attraction between individual electrons and to the existence of an energy gap with a width of about 10^{-4} eV in the electron spectrum.

A theoretical study of the interaction of electrons with the oscillating lattice — or, as we said in Section 2.1, of the electron – phonon interaction — was performed in 1950 (the year of the discovery of the isotopic effect) by Fröhlich [73], who not only predicted the isotopic effect before its experimental discovery, but also demonstrated that the electron – phonon interaction actually leads to a weak attraction between electrons and to an energy gap of $\sim 10^{-4}$ eV in the electron spectrum.

Fröhlich's train of thought is similar to a scheme used in quantum electrodynamics, where the interaction between two charged particles (for example, electrons) is described by the emission of a virtual photon by one of the particles with its subsequent absorption by the other. In the case of the

[13] By long-range or short-range order in a system of particles, the corresponding order in their spatial arrangement (or in the distribution of spins, magnetic moments, etc.) is meant. A classic example of long-range order with respect to coordinates in space is the regular arrangement of atoms in a crystal, which are separated by the same distance, and an example of short-range order is the mean distance between atoms in amorphous solids (which have no long-range order). The ideal gas has neither long-range nor short-range order. In our case of electrons in a superconductor, the short-range order is the mean distance between electrons, and the long-range order (at $T < T_c$) is the mutual orientation of momenta and spins of two electrons situated not too close to each other and form a Cooper's pair (see below for more details).

electron–phonon interaction, one of the phonons emits a virtual phonon, which is absorbed by another electron. In accordance with the quantum mechanical uncertainty relation

$$\Delta E \Delta t \simeq \hbar\,, \tag{21}$$

the energy of the emitted phonon ($h\nu = \Delta E$) depends on its lifetime Δt before the absorption of the phonon by another electron, and can be fairly large if Δt is small. From the standpoint of classical physics, this process is not possible, since it violates the energy conservation law for the time Δt. In quantum mechanics, it is not only possible — it can even be proved that as long as $h\nu > E_1 - E_2$ (where E_1 and E_2 are the energies of the electron before and after the emission of the phonon), the interaction between two electrons will take the form of attraction, which can be greater than Coulomb repulsion between the electrons.

Experimentally, the linkage between superconductivity and the electron–phonon interaction is confirmed by the fact that poor normal conductors (such as lead) are good superconductors (i.e., have relatively large T_c values), whereas good conductors (platinum, gold, silver, copper), as far as we know, cannot be brought into a superconducting state. According to Frohlich, this can be explained by the difference in the intensity of the electron–phonon interaction: if large, it is unfavorable for conductivity in the normal state (because of scattering of conductivity electrons by phonons), and favorable for superconductivity (because the strong electron–phonon interaction gives rise to attraction between remote electrons).

In the classical approach, the attraction between two negatively charged electrons can be described as follows. When one of the electrons interacts with the oscillating crystal lattice, the lattice shields the electron's negative charge with the positive charges of its ions, so that the electron appears to be positively charged and can attract another electron.

Fröhlich's theory allowed us to take a very important step forward in the development of the microscopic theory of superconductivity. In quantum mechanics, however, not just any kind of attraction will lead to the formation of a bound state. For that, the attraction must be strong enough. (It is said in quantum mechanics that the level corresponding to a bound state can appear only in a sufficiently deep potential well, characterizing the potential of attraction.) Fröhlich's theory, however, predicted the existence of only a very weak attraction between electrons. Therefore, of great significance was the next step made by Cooper on the way towards microscopic theory. In 1956, Cooper [25] considered the problem of two additional electrons added to the *complete* set of electrons characterizing the metal at $T = 0$ K and proved that, under certain conditions, they can form a bound state even if the attraction between them is arbitrarily weak. To better understand Cooper's reasoning, let us recall the main ideas of the solid state theory related to metals.

In contrast to our quantum Bose fluid obeying the Bose–Einstein statistics, which leads to the Bose condensation phenomenon, that is, to the condensation of particles in the lower energy state and eventually to superfluidity, the electrons in metal provide an example of a quantum Fermi system, which follows the Fermi–Dirac statistics and the Pauli principle.

A Fermi system made up of noninteracting particles is called the ideal Fermi gas, whose spectrum in the ground state at $T = 0$ K (a degenerate Fermi gas) is a sequence of filled energy levels, each of which (including the lowest one) is occupied, in compliance with Pauli's exclusion principle, by only two particles with opposite spins. The occupation of spectral levels continues up to a certain maximal energy, which depends on the density of particles in the gas (Fermi energy E_F). The energy states with $E > E_F$ are free (for more details, see Section 5).

Now let us go back to Cooper's problem of two additional electrons added, so to say, on the top of the system of energy levels of the metal filled up to E_F at $T = 0$. According to Pauli's principle, they must occupy the nearest free energy states above the Fermi surface (that is, with $E > E_F$). Based on Fröhlich's result, Cooper suggested that there is a very weak interaction between these electrons and showed that, under certain conditions (equal and oppositely directed momenta and spins), they can form a bound state (which later became known as a Cooper pair), and their total energy (kinetic plus potential) will be less than $2E_F$. In this way, these electrons will go down into the spectral region that is occupied and therefore forbidden for any new single unpaired electrons, but is available for the bound Cooper pair.

The attractive exoticness of this result calls for a more detailed treatment. To save time and space, however, we will consider not the *ad hoc* Cooper's problem of two additional electrons, but a more realistic Fermi system with many electrons, which is actually the beginning of our presentation of the BCS theory [23].

Again, we turn to the electron spectrum in a metal at $T = 0$, occupied up to the energy E_F. Let us consider what should result from the interaction of electrons with the lattice and what the conditions are for maximizing the number of such interactions and, therefore, the number of emerging Cooper pairs, and for the reduction of the total energy of electrons.

Assume that two electrons with momenta $\mathbf{p}_1$ and $\mathbf{p}_2$ are involved in electron–phonon interaction, so that their momenta assume new values $\mathbf{p}_1'$ and $\mathbf{p}_2'$; in this case, according to the conservation law,

$$\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}_1' + \mathbf{p}_2' = \mathbf{p}\,. \tag{22}$$

Since the energy of interaction is of the order of kT_c, that is, quite small (since at $T_c = 1$ K this energy is $kT_c \approx 10^{-4}$ eV) compared with the Fermi energy ($E_F \approx 10-20$ eV), the points of all four vectors corresponding to the momenta will fall within narrow spherical layers near the Fermi surface of radius p_F. The ends of the vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ should be in the layer located under the surface of the sphere, and $\mathbf{p}_1'$ and $\mathbf{p}_2'$ should be in the layer located over the surface of the sphere (that is, in the region of the unfilled energy states, where the interacting electrons are only allowed to go according to the Pauli principle).

Figure 5a shows schematically the cross section of two spherical Fermi layers of thickness Δp and mean radius p_F. Obviously, the configuration is axisymmetric with respect to the direction of the vector $\mathbf{p}$. Therefore, the hatched parts of the plot depict the cross section of the ring (formed where the spherical layers intersect), in which the points of the vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ (or $\mathbf{p}_1'$ and $\mathbf{p}_2'$) of any Cooper pair with a total momentum $\mathbf{p}$ should come together. Clearly, it is only in this case that all three of the above conditions will be satisfied (the

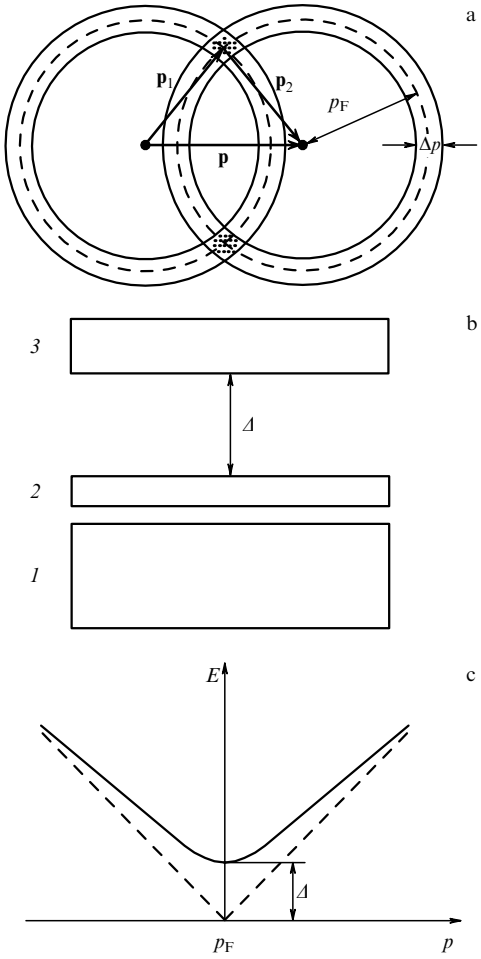

Figure 5. An illustration to the nature of superconductivity: (a) scheme of intersection of two spherical Fermi layers; (b) energy scheme of a superconductor: *1*, region of unpaired electrons with $p < p_F$; *2*, region of Cooper pairs (Δ is the energy gap); *3*, region of free electrons; (c) energy spectrum of electrons in a superconductor (solid curve) and in a normal metal (dashed line): p_F is the Fermi momentum, and Δ is the energy gap.

Pauli principle, equation (22), and $\Delta p \ll p_F$). It is also obvious that the number of interactions satisfying these conditions is proportional to the volume of the ring (in p-space). It is easy to see that, as p decreases from $p = 2p_F$ (the spherical layers are tangent to each other) to $p = 0$ (the layers coincide), the volume of the ring will initially change rather slowly, until p becomes exactly zero and the ring turns into a complete spherical layer. As this situation is approached, the volume of the ring increases sharply. The value $\mathbf{p} = 0$ corresponds to the case of $\mathbf{p}_1 = -\mathbf{p}_2$ ($\mathbf{p}'_1 = -\mathbf{p}'_2$, etc.). Thus, most Cooper pairs are formed by electrons with oppositely directed momenta. Having additionally considered the quantum mechanical problem with due account for the Pauli principle, Cooper demonstrated that the spatial component of the wave function of the electron pair is symmetric, which implies that its spin component is asymmetric — that is, the electrons forming a Cooper pair must have opposite spins (which makes the attraction stronger).

It is interesting to note (and this was also demonstrated by Cooper) that an electron finds a companion not among the nearest neighbors, but among electrons that are farther away. This is easy to understand if, in addition to the conventional form of the uncertainty relation $\Delta E \Delta t \approx \hbar$, we consider its alternative representation

$$\Delta x \Delta p \approx \hbar\,, \tag{23}$$

where Δp should correspond to the value of $\Delta E = kT_c$, equal to $0.7 \times (10^{-4} - 10^{-3})$ eV at $T_c = 1 - 10$ K. Indeed, according to our previous argumentation, we have

$$kT_c = E - E_F = \frac{p^2}{2m_e} - \frac{p_F^2}{2m_e} = \frac{(p - p_F)(p + p_F)}{2m_e} = \frac{p_F \Delta p}{m_e}\,, \tag{24}$$

where m_e is the mass of the electron, E is its energy near the Fermi surface, and p is its momentum ($p \simeq p_F$, since $E_F \simeq 10$ eV, and $kT_c \simeq 10^{-4}$ eV). Hence, $\Delta p = mkT_c/p_F$ and $\Delta x = \hbar/\Delta p \sim 1.5 \times (10^{-4} - 10^{-3})$ cm. We see that the 'size' of the Cooper pair (at $T_c \sim 1 - 10$ K) is $2 \times (10^{-4} - 10^{-3})$ cm, which is much greater than the mean distance (10^{-7} cm) between the atoms in the crystal lattice, and hence between the conductivity electrons (the 'long-range action' of the electron – phonon attraction).

We should also mention another feature of Cooper pairs. The process of electron – phonon interaction, which leads to the conversion of a pair with $\mathbf{p}_1 + \mathbf{p}_2 = 0$ into a pair with $\mathbf{p}'_1 + \mathbf{p}'_2 = 0$ occurs very rapidly (the velocity of the electron near the Fermi surface is $v_F = p_F/m_e \sim 2 \times 10^8$ cm s^{-1}), so that this new pair is converted after a short time $\Delta t \approx \Delta x/v_F \approx 10^{-12} - 10^{-11}$ s into another pair with $\mathbf{p}''_1 + \mathbf{p}''_2 = 0$, and so on, and the momenta of electrons of the newly formed pairs must lie in the spherical layer of radius p_F and thickness Δp. Therefore, particular values of momenta may reappear in the sequence of conversions of pairs. (This means that all electrons that take part in the formation of Cooper pairs are coupled.) This extension of Cooper's result from two electrons to all electrons with $p \approx p_F$ (lying in the region Δp) together with the assumption that other electrons (occupying levels with $p < p_F$) do not take part in the formation of Cooper pairs and do not hamper this process, is essentially the main content of the BCS theory, because it implies a considerable reduction in the energy of the system compared to the energy of electrons of a normal metal, as well as the stability of Cooper pairs against their breakdown, that is, the existence of an energy gap of finite width Δ. This situation is schematically represented in Fig. 5b, where the numeral *1* denotes the region of unpaired electrons with $p < p_F$, *2* is the region of Cooper pairs, *3* is the region of free electrons, and Δ is the energy gap.

The large width of the gap can be explained by the fact that the process of formation of Cooper pairs involves *many* electrons from the region Δp, so that the energy of separation of a paired electron is many times larger than the binding energy of the pair. This is not a paradox. As indicated above, in the process of continual reconversion of Cooper pairs, some particular values of momenta in a pair ($\mathbf{p}_1$ or $\mathbf{p}_2$, $\mathbf{p}'_1$ or $\mathbf{p}'_2$, etc.) may repeat again and again. This implies that the transition of an electron with a certain momentum $\mathbf{p}_i$ from the bound state into the one-electron state removes from the system of Cooper pairs not one but a large number of pairs that *could be formed with this momentum involved*. This results in the finite width Δ of the energy gap.

Because of the existence of the gap, the spectrum of electrons of a superconductor at $T < T_c$ differs dramatically from the normal spectrum of electrons at $T > T_c$. In the latter case, as indicated above, the ground state is characterized by a

progressive filling of energy levels up to the Fermi energy E_F, which lies next to the band of allowed energy states. One of these can be taken by an excited electron. In the case of a superconductor with $T < T_c$, the ground state is formed by a system of interacting electron pairs concentrated near the Fermi surface, and the excited state (characterized by one electron detached from a pair) is separated from the ground state by an energy gap Δ (*within which there are no allowed states*). The energy of electrons E as a function of momentum p in a superconductor, for the isotropic BCS model, is

$$E(p) = \left[\Delta^2 + \frac{p_F^2}{m_e^2}(p - p_F)^2\right]^{1/2}, \qquad (25)$$

where p_F is the Fermi momentum. The scheme of the energy spectrum of electrons in a metal is shown in Fig. 5c, where the dashed line corresponds to the spectrum in a normal metal, and the solid curve to the spectrum in a superconductor. At $T = 0$, the width of the gap is $\Delta > \hbar\omega$, where $\hbar\omega$ is the mean energy of lattice oscillations. Because of this, at $T = 0$ K, the described process of excitation of electrons is virtually impossible, and the metal clearly displays all features of superconductivity. As T increases (from zero to T_c), the probability of excitation of the electron increases (because of the increase in the energy of thermal oscillations), which leads to a gradual reduction in the number of pairs and a reduction in the width of the gap. At $T = T_c$ the gap disappears, and so does superconductivity.

Thus, very briefly, the essence of the microscopic theory of superconductivity consists of the following. At $T < T_c$, in a system that possesses the property of superconductivity, a macroscopically large number of strongly bound Cooper pairs of electrons form, which have equal and oppositely directed momenta, a total spin $J = 0$, and a total charge $2e$, which can be regarded as charged Bose–Einstein particles; they, similarly to liquid ^{4}He (at $T < T_c$), should feature Bose condensation and superfluidity. Since the total momentum of all electron pairs is zero, the application of an external electric field will give the same momentum $\Delta\mathbf{p}$ to each pair, which will result in a correlated motion of Cooper pairs without a drag (that is, without scattering by phonons and crystal defects, which is impossible, since $\Delta > \hbar\omega$, at $T < T_c$).

To conclude this section, let us write out two particular results of the BCS theory that are of predictive importance. The first one refers to the expression for the critical temperature T_c:

$$T_c = \theta \exp\left(-\frac{1}{\lambda_{\text{eff}}}\right), \qquad (26)$$

where, in the case of the electron–phonon mechanism of interaction, we have $\theta \simeq \theta_D$ (θ_D is the Debye temperature of the metal), and the coupling constant λ_{eff} is considered small ($\lambda_{\text{eff}} \ll 1$). Since, for most metals, $\theta_D < 500$ K, and $\lambda_{\text{eff}} < 1/3$, the BCS theory gives a fairly convincing 'validation' to the measured values of T_c, which do not exceed 24 K for LTSC (the role of formula (26) in the case of HTSC will be discussed in Section 2.4). The second experimentally validated result links the width of the gap Δ at $T = 0$ K with T_c:

$$\frac{2\Delta(0)}{kT_c} = 3.52\,. \qquad (27)$$

With this, we conclude our elementary presentation of the main principles of the microscopic theory of superconductivity. In the works of Russian scientists, this theory was further developed by Bogolyubov [26] and his school [27], in a series of articles by Ginzburg (see reviews [267, 34] and references therein), and by Gor'kov [24], Eliashberg [74], and other physicists. More detailed accounts of superconductivity can be found in books by Schmidt [69, 70], Abrikosov [75], Bogolyubov, Tolmachev, and Shirkov [76], and Vonsovskiĭ, Izyumov, and Kurmaev [77]. In addition to the foreign publications mentioned above [41, 71], we would recommend books by Schieffer [78] and Tinkham [79]. Also, we address the reader once again to articles by Kukin [64].

2.4 High-temperature superconductivity (HTSC)

As noted above, HTSC was discovered by Bednorz and Müller [31, 33] in 1986 and by Wu, Chu, and their colleagues [32] in 1987. They found that the property of high-temperature superconductivity is displayed by a class of multiphase layered ceramic materials, oxide–cuprates, which differ radically from the conventional low-temperature superconductors. The first materials of this kind were $La_{2-x}Ba_xCuO_4$ (at $x = 0.15-0.2$), in which the authors of Ref. [31] discovered superconductivity with a then-record critical temperature of $T_c = 30-34$ K. (Recall that the highest T_c for LTSC materials does not exceed 24 K even today, with the exception of the superconductivity of MgB_2 discovered in 2001 — see below.) However, a really high-temperature superconductivity is the HTSC of yttrium compounds of type $YBa_2Cu_3O_7$ (the so-called 1 : 2 : 3 crystal) with $T_c = 90$ K, which is higher than the boiling point of liquid nitrogen ($T_b = 77.3$ K). This is very important for practical applications of superconductivity, because liquid nitrogen is much cheaper than liquid helium.

In 1993 Putilin and colleagues [80] and Chu and colleagues [81] discovered HTSC in compounds of mercury of type $HgBa_2Ca_2Cu_3O_{8+x}$ with even higher critical temperatures $T_c = 130-160$ K (at normal and high pressures, respectively).

The discovery of HTSC stimulated exceptionally high interest in this area of research, with tens of thousands of publications in the past 16 years [14]. Studies of oxide materials revealed that their superconducting properties depend considerably on the oxygen content. The highest T_c corresponds to a certain optimal concentration. The superconductivity of cuprates also strongly depends on doping, i.e., adding admixtures, and also requires their optimal concentration. The structure of the elementary cells of a 1 : 2 : 3 crystal and other crystals, the nature of structural chemical Cu–O bonds, and possible effects of the polarizability of oxygen on the HTSC of cuprates is described in a review by Bobovich [82].

For illustration, we selected from this review a schematic of the elementary cell of a La_2CuO_4 crystal doped with Sr ions (Fig. 6). From this diagram, we can see that the doped elementary cell appears as almost plane layers of highly deformed octahedrons with copper–oxygen bonds, interlaid with plates of $La_{1.8}Sr_{0.2}O_2$ [83]. The state of the art in theoretical and experimental investigations of HTSC is described in reviews by Ginzburg [5] and Maksimov [84] that appeared shortly after Ref. [5].

[14] There were also some unfortunate casualties. One Russian physicist sank into major depression because, during his studies of the electric conductivity of a similar ceramic, the lack of liquid helium had prevented him from discovering HTSC before Bednorz and Müller.

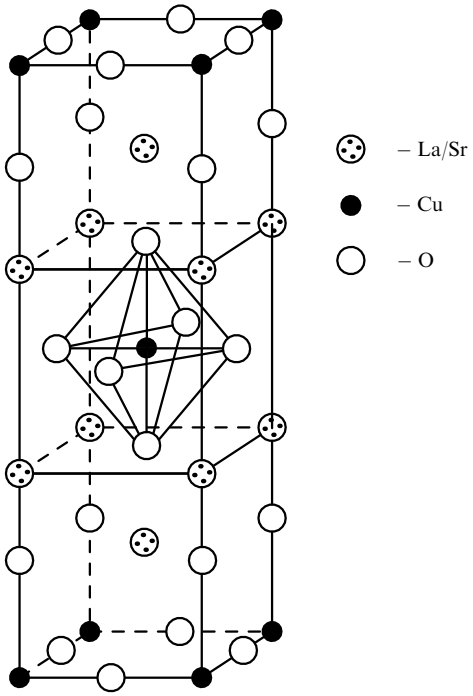

Figure 6. Elementary cell of La_2CuO_4 crystal doped with Sr ions.

One of the mechanisms of HTSC is the electron – phonon interaction (EPI), which we have discussed in detail in connection with LTSC, and which is assumed to be weak in the framework of the BCS theory. Many scientists are convinced, however, that even though EPI is important, it is not the only mechanism involved in LTSC. A number of papers in a collection [85] published in 1977, long before the discovery of LTSC, demonstrated that high values of T_c can be achieved in systems with a strong interaction. After the discovery of HTSC, such systems became the object of especially intense studies. A large number of works have been based on the Hubbard model with strong Coulomb repulsion of electrons on one center.

This model was proposed in 1963 – 1965 by Hubbard [86] for the description of various aspects of linkage between electric and magnetic properties of solids. The model is based on the concept of motion of electrons (nondegenerate with respect to their orbital state) in a crystal lattice in the form of quantum jumps from one site to another. Two electrons on the same site experience a strong local Coulomb repulsion.

For theoretical analyses of the mechanism of HTSC, the Hubbard model is used in conjunction with the ideas of Anderson regarding the superconductivity of cuprates presented by him in Ref. [87], where the La_2CuO_4 cuprate was used as an example for analyzing strongly correlated electron systems with resonant valence bonds. According to Anderson, these bonds describe the so-called spin fluid of singlet electron pairs. In other words, at low temperatures, the spin of an electron exists and is transferred separately from the charge of the electron.

No less exotic is another approach to the description of low-energy excitations in HTSC systems, proposed by Kalmeyer and Laughlin [88]. They associate these excitations with neither Fermi – Dirac nor Bose – Einstein statistics, that is, they consider them neither fermions nor bosons, but rather so-called *anions*. It may also look unusual to the layman that some models are based on the assumption of the existence of strong *repulsion* between electrons, which can lead to anisotropic p and d pairing owing to spin fluctuations. The idea of such models was first suggested by Kohn and Luttinger in 1965 [89], two decades before the discovery of HTSC. These studies attracted special attention after d-pairing in cuprates had been observed.

We realize, of course, that the previous paragraph is not highly informative and ask the reader to consider it simply as a reference to sources. In particular, we draw the reader's attention once again to two recent reviews by Ginzburg [5] and Maksimov [84]. In the same spirit (without going into detail), we mention some other publications that discuss different versions of the mechanism of the pairing of electrons — the electron – exciton mechanism discussed above [90] and the spin [91] and electron [92] mechanisms.

Thus, the situation with models of HTSC is not simple. More confusion is added by the fact that, in contrast to LTSC, which is mainly explained based on EPI, HTSC may depend on all the above mechanisms working simultaneously. This circumstance can hardly be considered an advantage of the theory of HTSC. To support this view, let us quote one of the leading experts in superconductivity [34]: "*The current state of the solid-state theory and, in particular, the theory of superconductivity does not allow us to calculate the temperature* T_c; *neither can it give any more or less precise and definite indications as to which compounds ought to be tested, especially in the case of complex materials*." And then: "*even the very mechanism of superconductivity in cuprates is not clear*." The author of Ref. [34] believes that the maximum T_c reached by the time of his writing (134 K at atmospheric pressure and 164 K at high pressures) can be explained in terms of the phonon mechanism, which, in principle, will no longer be applicable at $T_c > 200$ K. In this case, it is possible to invoke the exciton mechanism, which will be suitable even for explaining room-temperature superconductivity if it is discovered. The author of Ref. [34] is moderately optimistic: "*in principle, RTSC is feasible, but there is no guarantee of this*." Some ground for such optimism (in the sense that everything is possible) is supplied by the recent (March 2001) discovery of the superconductivity of MgB_2 by Japanese physicists. This is a very simple compound that reportedly can be bought in a pharmacy [93]. It turned out that the critical temperature for this compound is $T_c = 39$ K, which is of the same order of magnitude as the critical temperature of the first high-temperature superconductors, although the structure of MgB_2 is as simple as that of all known low-temperature semiconductors. The highest T_c, which reached 24 K in 1973, has not changed since. Thus, it is not quite clear what type of superconductivity (LTSC or HTSC) is exhibited by MgB_2. For this compound, pairing is claimed to follow the phonon mechanism [94], but at $T_c \approx 40$ K this mechanism is possible in either case. It has also been suggested [95] that such an unusually high value of T_c in a semiconductor with such a simple structure can be attributed to the existence of two energy gaps.

As to the opinion of the author of Ref. [34] regarding the status of the solid-state theory and the theory of superconductivity, expressed in 1997, this opinion had not changed by the year 2000: "*the state of the art in the solid-state theory cannot be considered satisfactory... There is no common opinion regarding the mechanism of superconductivity that could lead to a high value of* T_c *in cuprates. In real substances, of course, the electron – phonon, spin, and elec-*

tron–electron (*electron–exciton*) *interactions exist concurrently*" [5]. As far as we know, such a pessimistic view of the theory is shared by other experts in the field and probably by many readers as well. Observe, however, that the author of Ref. [5] also expressed an optimistic view about the 'golden dream' of RTSC, for which he even sets a deadline — no later than 2011 (100 years after the discovery of LTSC). Oh, V.L.! Easier said than done...[15] However, there is some ground for such hope, since HTSC was recently discovered not only in cuprates (for which no new record-breaking values of T_c have been achieved since 1994), but also in fullerites, for which critical temperatures of $T_c = 18$, 80, and 117 K were recently achieved [96, 97]. Another promising material may be carbon nanotubes [98].

2.5 Applications of LTSC and HTSC in science and technology

To end the chapter on superconductivity, let us say a few words on the use of LTSC and HTSC in science and technology. So far, LTSC has been used much more broadly than HTSC. Today, it is hard to imagine accelerators and storage rings without superconducting magnets. One of the best known installations of this type is the US Tevatron collider. Nuclotron, a superconducting accelerator of protons and atomic nuclei, is operated in Russia (Dubna). Superconducting magnets are used in various detectors of elementary particles, in medical tomographs, and in tokamaks — devices for the confinement of hot plasmas, widely used worldwide in the framework of the international project of controlled fusion. Today, Russian physicists have already manufactured a huge superconducting coil insert for the future reactor. The mere size of the coil — 2 m in diameter, 5 m in height — gives an idea of the work done. In the course of recent testing in Japan, the coil passed a current of 46040 A (higher than the design value) in a magnetic field of 13 T [99].

The supeconducting magnetometer based on the Josephson effect [100][16] is widely employed. The superconducting electric circuits are controlled with a device called a cliotron, whose operation is based on the loss of superconductivity in a supercritical external magnetic field.

Technical applications of HTSC are not as extensive because of problems with the manufacture of usable superconducting cables. The mechanical properties of ceramic material prevent the preparation of thin strands of wire fixed in a matrix of a good conductor, as is done in the case of LTSC. It was necessary to develop a sophisticated process, which only recently gave practical results. One implementation is known as 'powder in tube': a silver tube is filled with a HTSC powder and is subjected to rolling, drawing, and heat treatment to produce a HTSC ribbon. Typical dimensions of the ribbon (for a current of about 60 kA) are 0.2 mm in thickness by 3 to 8 mm in width. Another process is known as the coating process and produces a multilayered film structure made up of flexible metal strips coated with a LTSC film [104].

Many technological problems have now been overcome, and there are many applications of HTSC in power engineering. These include emergency current limiters in the existing power supply in Switzerland, a 180-meter insert of HTSC cable in Denmark, current leads for the Large Hadron Collider (LHC) for an energy of 2×7 TeV at CERN, and an experimental HTSC transformer for 0.5 MW in Japan [105]. In the US, HTSC materials are used in systems of filters and amplifiers in the base stations of the cellular network, which increases the coverage of the base station, increases the transmission rate, and improves the quality of sound of cell phones [38].

Both abroad and in Russia, there are many projects of practical applications of HTSC with varying degrees of implementation. Let us mention some of them. An HTSC current delimiter is being developed in Russia for RAO EES, an in-house HTSC power transmission line at the Noril'skiĭ Nikel Company is planned, and a completely superconducting power transmission system is proposed for the Science Center of the Russian Academy of Sciences, currently under construction near St. Petersburg. This system will consist of HTSC generators, transformers, current limiters, and transmission lines [106, 36].

Many foreign companies are developing high-voltage HTSC cables. An experimental sample of such a cable, 30 to 50 meters long, for buried routing is being developed in France. HTSC cables of different construction are designed in Denmark, the UK, Switzerland, Japan, etc. The most advanced is the US project of a HTSC cable for a power substation in Detroit. The company working on this project has already produced a flexible three-phase cable with connectors and interfaces for 2400 A at 24 kV, as well as a cooling system [37]. Industrial applications of the HTSC cables are, however, limited by their high cost compared with the cost of LT superconductors (1 US dollar per 1 kiloampere-meter (kAm) for NbTi materials). It is believed that HTSC materials will seriously compete with LT super-

[15] We allow ourselves such familiarity with reference to a very old 'half-acquaintance': one of the authors (KNM) studied with V L Ginzburg at the Department of Physics of Moscow State University at virtually the same time (1936–1941), and knew him as a senior student viewed through the eyes of a freshman; moreover, we sometimes attended the same special courses (on matrix theory by Professor Rabinovich, on group theory by Professor Delaunay, etc.) together. The main courses were also taught by the same lecturers, even though we attended these courses in different years: S É Khaĭkin (to demonstrate the outstanding capabilities of the human ear, he invited his famous brother, the conductor of the Bolshoi Theater Boris Khaĭkin to one of his lectures and made him identify sounds produced with a special multi-plate xylophone), M A Leontovich, I E Tamm, A N Tikhonov, D I Blokhintsev, G S Landsberg, A A Vlasov, and V S Fursov. In his younger years, KNM had the privilege of meeting contemporary and future big names. He wrote his diploma thesis (under the supervision of V V Migulin) while in the same office with N D Papaleksi and the young A M Prokhorov; in the early years of the war, he was assistant to L I Mandel'shtam (and often was his guest at home, so that the 'half-acquaintance' became a full acquaintance); after his transfer to the present Kurchatov Institute, he met I V Kurchatov almost every day; in 1948, he worked for some time on the team of Yu B Khariton (with G N Flerov), and often conversed with the youngest Corresponding Member of the Academy of Sciences (elected at the age of 32) Ya B Zel'dovich (with whom he lived in an adjacent room in a hotel in Arzamas); in the same year, he was transferred to the laboratory of I I Gurevich, who introduced KNM to L A Artsimovich, I K Kikoin, and A P Aleksandrov. Teaching at the same time at the Moscow Engineering Physics Institute, he knew A I Alikhanyan and V I Gol'danskiĭ well. What times, what people!

[16] B D Josephson was awarded the 1973 Nobel Prize in physics for the prediction of the effects named after him (Josephson tunneling, Josephson junction, and nonstationary Josephson effect) [101]. Some of these effects were experimentally discovered by Esaki [102] and Giaever [103], who shared the 1973 Nobel Prize in physics with Josephson for their experimental discoveries regarding tunneling phenomena in solids.

conductors if their cost is reduced to 10 dollars per 1 kAm; more realistic, however, is reaching a cost of 50 dollars per 1 kAm by 2004 (by mass production), which will make possible the use of HTSC materials only for unique projects [107].

To conclude this section, let us note one more practical application — this time not of HTSC materials, but of their cooling agent, liquid nitrogen[17]. An article published in November 2001 [108] reported that a prototype of a cryogenic automobile working on liquid nitrogen had been developed in the US. This cryogenic automobile has a 124-liter tank with liquid nitrogen, a heat exchanger in which the vapor of boiling nitrogen is heated to the ambient temperature, and a 13-kW pneumatic engine operating at an input pressure of 10 atm. The mass of the automobile is 700 kg. During tests, the automobile ran at a speed of 58 km h^{-1}, and covered 24 kilometers on one tank. The next version is expected to reach a speed of 100 km h^{-1} and cover 240 km before refueling. The important advantages of the cryogenic automobile over the ordinary car are the environmental friendliness and complete fire safety, which make it a very desirable vehicle for airfields. The drawback is some clumsiness due to the large sizes of the nitrogen tank and heat exchanger.

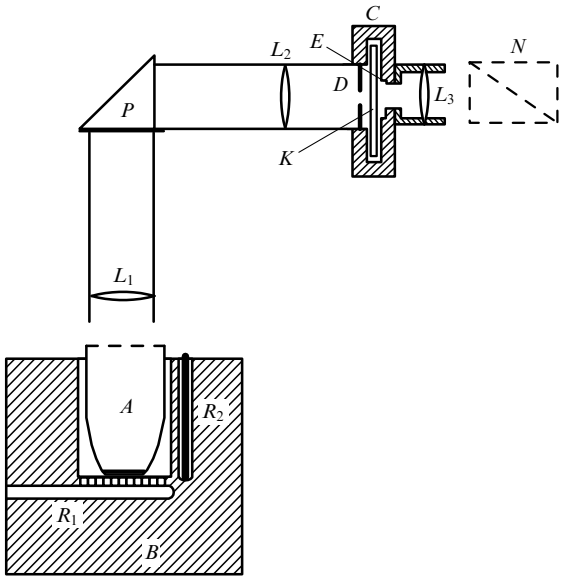

Figure 7. P A Cherenkov's installation: A — platinum vessel, B — wooden holder, R_1 and R_2 — slots which accommodate radioactive material, L_1, L_2, L_3 — lenses, P — prism, D — diaphragm, C — slit for the wedge K, E — slit for optical filters, N — Glan prism.

3. Vavilov – Cherenkov radiation and related effects

3.1 History of the discovery

The history of the discovery of Vavilov – Cherenkov radiation goes back to 1934 when P A Cherenkov, a postgraduate student of the Academician S I Vavilov, discovered an unexpected blue glow while visually studying the luminescence of solutions of uranyl salts irradiated with γ-emission of radium in absolute darkness[18], which could not be explained by the conventional mechanism that produces fluorescence [110]. Cherenkov's experiment was carried out with the installation shown in Fig. 7. It consisted of a platinum vessel A 3 cm in diameter, containing the investigated liquid whose emission was gathered by an optical system consisting of the lenses L_1 and L_2, a prism P, and a diaphragm D onto a photographic wedge K, forming a spot 3 mm in diameter. Moving the wedge in slit C, it was possible to set the brightness of the image observed through the lens L_3 on the threshold of visual reception. The auxiliary slit E held optical filters. To determine the polarization of emission, a Glan prism N was placed in front of the eye.

To maintain constancy in the observational conditions, the vessel with the liquid was placed in the wooden holder B with two slots R_1 and R_2 for the radioactive source. The liquid was irradiated with γ-emission of 103.6 mg of radium sample (α- and β-particles could not penetrate the glass walls of the radium-containing ampoule or the wall of the platinum vessel). To determine the polarization of the emitted light, the radioactive source was placed into the slot R_2 parallel to the axis of the vessel, while in all other experiments it stayed in slot R_1.

On the whole, twenty very different liquids were investigated; among these, in addition to water (tap water and triply distilled water) were such disparate liquids as sulfuric acid and glycerin, acetone and paraffin oil, and so forth. The relative luminance was found to be identical within ±18% in all the investigated liquids. A visual study of the spectral composition of the glow was conducted using optical filters; it showed that whatever the liquid, the emission was concentrated in the blue-to-violet part of the spectrum. This glow could not be quenched by applying familiar luminescence killers or by heating (which in some liquids affected not only temperature but viscosity as well). The polarization study showed that the emission from all the liquids was partially polarized.

All these observational results contradicted the properties of fluorescence that had been studied in detail by Vavilov [111 – 113]. In particular, it was known that fluorescence appears as a result of transitions of atoms or molecules between excited states and has the sign of polarization opposite to that observed in Vavilov – Cherenkov radiation. The characteristic fluorescence lifetime is $\tau \geqslant 10^{-10}$ s, and the transition probability is affected, for example, by added quencher or, conversely, by careful purification of the medium, by a change in its temperature, etc. As we outlined above, Cherenkov found it impossible to quench the glow by any of these means. In this connection, Vavilov published a paper [114] in the same issue of the journal where Cherenkov's original paper was published [110] in which he stated that the effect observed by Cherenkov is not fluorescence but is caused by fast Compton electrons as they are decelerated in a dense

[17] We would like to remind the reader who might regard this digression as strange that, without cooling with liquid nitrogen, HTSC materials lose their remarkable properties and turn into ordinary ceramics with poor mechanical properties. Only RTSC, if ever realized, will be like Kipling's cat that walked by himself. LTSC and HTSC can only walk together with their masters, cooling stations. These masters, however, not only are able to 'walk' freely in physical laboratories, attending to much other business not related to servicing LTSC and HTSC, but also can literally run on the roads (as will now become clear).

[18] I M Frank who sometimes took part in these experiments recalled [109] that before starting the count, it was necessary to first spend about an hour in darkness for the adaptation of the eyes. S I Vavilov used this time to discuss the results obtained earlier and to suggest new experiments.

medium. Even though this specific cause of the effect suggested by Vavilov (electron deceleration) proved to be erroneous, the very idea of explaining its origin in terms of *the motion of fast electrons in dense media* has played the decisive role, as we will see later, in developing a theory and in designing subsequent experiments which were suggested, as was the first one, by Vavilov. That is why the universal blue glow discovered by Cherenkov is known in Russia as Vavilov – Cherenkov radiation.

In further studies carried out already with electrons [115] Cherenkov discovered a sharp asymmetry in the distribution of the intensity of this radiation: the intensity was much higher in the direction of motion of the electron than in the opposite direction. This result played a very important role in revealing the nature of the effect.

The mechanism and the quantitative theory of Vavilov – Cherenkov radiation were worked out in 1937 by I E Tamm and I M Frank on the basis of an analysis of classical electrodynamics equations [116]. A similar result was obtained in 1940 by V L Ginzburg in quantum terms [117]. In addition, Ginzburg discussed the case of an anisotropic medium [118]. In the same year, E Fermi generalized the theory of Vavilov – Cherenkov radiation by taking into account the absorption of a particle's energy in the medium [119].

In 1958, Cherenkov, Tamm, and Frank received the Nobel Prize in physics for the discovery and interpretation of the Vavilov – Cherenkov effect [120][19].

3.2 Elements of the Tamm – Frank theory

Tamm and Frank [116] called attention to the fact that the assertion of classical electrodynamics on the impossibility of energy loss through emission of radiation by a charged particle moving uniformly along a straight line in a vacuum is invalidated if the vacuum is replaced with a medium having a refractive index $n > 1$. And they proved this new assertion, again by using classical electrodynamics.

However, the term 'classical' is not always a synonym of the word 'simple'. This only happens sometimes in classical physics, and even there a simple explanation often has to be replaced with a more complicated (quantum mechanical) one which is, however, more correct. As for classical electrodynamics, it allows the derivation of a number of exact results without resorting to a more universal theory — quantum electrodynamics. Nevertheless, even with all its classical nature, the theory in question does not simply fall within the domains of classical physics, but it belongs to classical theoretical physics, and to understand its achievements one has to know a great deal about mathematics, among other things.

This paper [116] falls precisely in this category because it requires from the reader, in addition to knowing the Maxwell equations, the ability to deal with Fourier integrals, cylindrical Bessel functions, and Hankel functions. Consequently, a popular acquaintance with the elements of the theory of the Vavilov – Cherenkov effect calls for a different method that assumes no special knowledge except the classical Huygens principle and the energy and momentum conservation laws (albeit in their relativistic forms)[20].

[19] S I Vavilov died in 1951 and therefore, according to the Statutes of the Nobel Foundation, could not be among the laureates.

[20] This method was tested by one of the authors (K N M) on several generations of students [121] and proved adequate.

We assume that the charged particle in uniform motion along a straight line can lose energy and momentum by radiation. Then the following equality must hold true:

$$\left(\frac{\mathrm{d}E}{\mathrm{d}p}\right)_{\mathrm{part}} = \left(\frac{\mathrm{d}E}{\mathrm{d}p}\right)_{\mathrm{rad}} . \tag{28}$$

It is readily seen that this condition cannot be satisfied in a vacuum but is fulfilled in a medium with $n > 1$. Indeed, in accordance with the special relativity theory, the energy E_{part} and momentum p_{part} of a particle of mass $m \neq 0$, moving freely through a vacuum at a velocity v close to the speed of light c and with kinetic energy T, are related by the following expressions

$$E_{\mathrm{part}} = mc^2\gamma = \sqrt{m^2c^4 + p^2c^2} = mc^2 + T , \tag{29}$$

$$p_{\mathrm{part}} = mv\gamma = \frac{Ev}{c^2} , \tag{30}$$

where $\gamma = 1/\sqrt{1-\beta^2}$, $\beta = v/c$, and c is the speed of light in a vacuum (3×10^{10} cm s^{-1}). Therefore, one finds

$$\left(\frac{\mathrm{d}E}{\mathrm{d}p}\right)_{\mathrm{part}} = \frac{pc^2}{E} = \beta c = v . \tag{31}$$

At the same time, we have for electromagnetic radiation:

$$E_{\mathrm{rad}} = pc , \tag{32}$$

that is

$$\left(\frac{\mathrm{d}E}{\mathrm{d}p}\right)_{\mathrm{rad}} = c , \tag{33}$$

and since $v < c$, we obtain

$$\left(\frac{\mathrm{d}E}{\mathrm{d}p}\right)_{\mathrm{part}}^{\mathrm{vac}} < \left(\frac{\mathrm{d}E}{\mathrm{d}p}\right)_{\mathrm{rad}}^{\mathrm{vac}} . \tag{34}$$

Therefore, the laws of conservation of energy and momentum forbid a charged particle in uniform motion along a straight line in a vacuum to give up its energy and momentum as the electromagnetic radiation (this radiation cannot accept the entire momentum released by the particle).

However, this prohibition is removed if a particle moves in a medium with a refractive index $n > 1$. In this case, the speed of light in the medium is given by

$$c' = \frac{c}{n} < c \tag{35}$$

and the velocity of the particle v not only may be equal to the speed of light c' in the medium but may even exceed it:

$$v \geqslant c' = \frac{c}{n} . \tag{36}$$

Obviously, if $v = c'$, condition (28) will be met for electromagnetic radiation emitted strictly in the direction of motion of the particle ($\theta = 0$). If, however, $v > c'$, condition (28) holds true for the direction θ along which $v' = c'$, where $v' = v\cos\theta$ is the projection of the particle's velocity v onto this direction. By equating v' and c', we find $v\cos\theta = c/n$,

that is

$$\cos\theta = \frac{c}{nv} = \frac{1}{\beta n}\,. \tag{37}$$

Consequently, in a medium with $n > 1$, conservation laws allow a charged particle moving uniformly along a straight line at a velocity $v \geqslant c' = c/n$ to lose such portions of its energy $\mathrm{d}E$ and momentum $\mathrm{d}p$ that can be carried away by electromagnetic radiation propagating in this media at an angle

$$\theta = \arccos\frac{1}{\beta n} \tag{38}$$

to the direction of motion of the particle.

Arguments stemming from conservation laws do not allow one to make any conclusions about the mechanism of losses by emission. It is clear that the mechanism must originate in some processes occurring in the medium because no such losses are possible in a vacuum.

Vavilov – Cherenkov radiation has common features with some other processes observed in various media through which bodies move at velocities exceeding the speed of wave propagation through these media. Familiar examples of such processes are the V-shaped wake following a steamship and the conical wave generated by a plane or missile moving at hypersonic velocities. In all these cases, the wave field of a body moving at ultra-high-speed is strongly perturbed (compared to the field of a slow moving body) and begins to decelerate that body.

In this particular case of super fast motion of a particle through a medium with the refractive index $n > 1$, the mechanism of the Vavilov – Cherenkov glow consists of the emission of coherent radiation by dipoles as a result of polarization of atoms of the medium by the charged particle moving through it. The dipoles are produced by the electric field of the moving particle, which shifts electrons of the surrounding atoms relative to their nuclei. As dipoles bring back to their normal state (after the particle leaves their vicinity), an electromagnetic pulse is generated.

If a particle moves with a relatively slow velocity, the resulting polarization should be symmetric with respect to the position of the particle (Fig. 8a) because the electric field of the particle *has enough time* to polarize all atoms in its vicinity, including those located in front of the particle along the direction of its motion. In this case, the collective field of all the dipoles far from the particle will be zero and their emissions will cancel each other out. If the particle moves through the medium at a velocity above the speed of electromagnetic wave propagation ($v > c' = c/n$), an unusual effect of retarded polarization of the medium must be observed, so that the produced dipoles will be oriented mostly in the direction of the motion of the particle (Fig. 8b). It is obvious then that in this case, firstly, the particle will be slowed down by the perturbed field of the dipoles (compare this with the examples given above of the V-shaped wake in water and conical acoustic waves) and, secondly, a direction must exist in which coherent emission of dipoles can occur because the waves emitted by dipoles at different points along the path of the particle may have the same phase. As the source of the emitted energy is the charged particle itself, the direction of emission can be found using the Huygens principle — applied to the particle acting as a source of partial waves, which it emits at each point along its path.

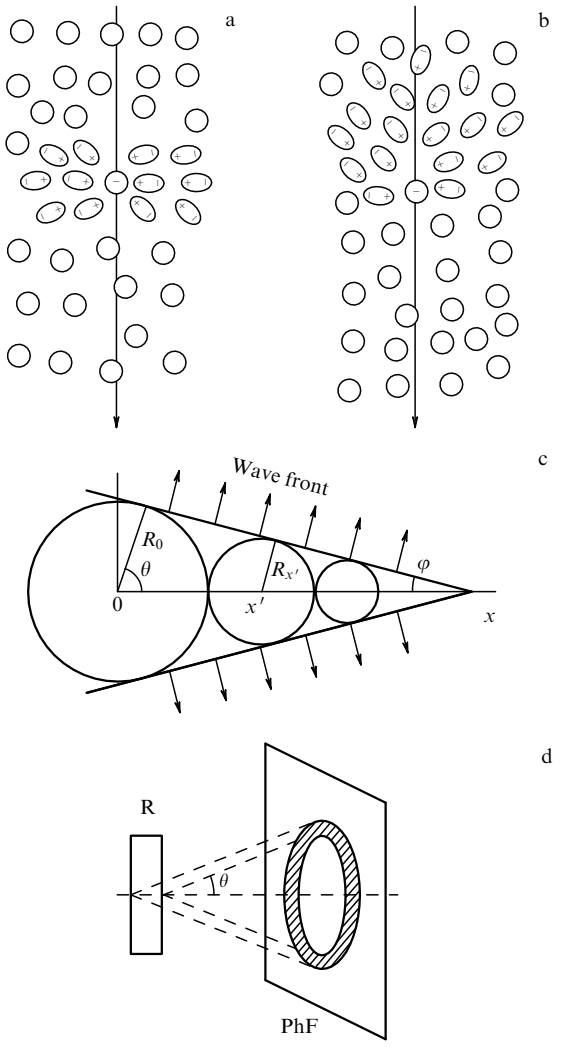

Figure 8. The nature of Vavilov – Cherenkov radiation and its special features: (a, b) excitation of atoms of the medium as a slow (a) and a fast (b) charged particle move through it; (c) determination of the direction of emission using the Huygens principle: $0x$ — direction of motion of the particle, R_0 and $R_{x'}$ — radii of the spherical waves emitted at the points $x = 0$ and $x = x'$, respectively, φ — one half of the opening angle of the conic envelope of these spherical waves, θ — the angle that determines the direction of emission, and (d) diagram of a photographic recording of the glow: R — radiator, θ — the angle between the direction of motion of the charged particle and the envelope of the emission cone, and PhF — photographic film.

The quantitative argument by which Tamm and Frank found this direction took up only one paragraph of twelve lines in their paper [116]. For a popular presentation, we have already used considerably more space and will need much more yet. Nothing can be done about it, however — this is the price of popularization. So, let the particle move from the left to the right from a point $x = 0$ at a velocity $v > c' = c/n$ (Fig. 8c). We assume that the particle is at a point $x = vt$ at some instant of time t. In accordance with the Huygens principle, the front of the emitted wave can be found by constructing the enveloping surface of the spherical waves emitted by the particle on its path from point $x = 0$ to point x. The radius of the wave at the point $x = 0$ at a moment t equals $R_0 = c't$. The radius of the wave at the point x at a moment t equals $R_x = c'(t - x/v) = 0$.

The radius of the wave at any intermediate point x' at a moment t equals $R_{x'} = c'(t - x'/v)$, i.e., it decreases linearly as x increases. Obviously, the enveloping surface is a cone with an opening angle 2φ, where

$$\sin\varphi = \frac{R_0}{x} = \frac{c't}{vt} = \frac{c'}{v} = \frac{c}{vn} = \frac{1}{\beta n}\,, \tag{39}$$

and the normal to the envelope, which according to the Huygens principle determines the direction of emission, lies at an angle $\theta = \pi/2 - \varphi$ to the axis x:

$$\cos\theta = \frac{1}{\beta n}\,. \tag{40}$$

Clearly, by applying the method suggested by Tamm and Frank, we obtain the same result (38) as was obtained earlier by using only the energy and momentum conservation laws. However, this is all that conservation laws can yield, so to become more familiar with the theory of the Vavilov–Cherenkov effect we have to quote further results of paper [116], albeit without derivation.

In accordance with Tamm and Frank's theory, the number of photons in the frequency range from ν to $\nu + \mathrm{d}\nu$, emitted by a particle with a charge z moving at a velocity β through a medium with the refractive index n, is given by

$$N(\nu)\,\mathrm{d}\nu = 4\pi^2\,\frac{(ze)^2}{hc^2}\left(1 - \frac{1}{\beta^2 n^2}\right)\mathrm{d}\nu\,. \tag{41}$$

This formula implies that (1) the spectra are identical for particles with equal z, for instance, for protons, electrons, muons, and π-mesons. As z changes, the number of emitted photons increases in proportion to z^2; (2) $N(\nu)$ grows with increasing β from 0 for $\beta = 1/n$ to $4\pi^2(e^2z^2/hc^2)(1 - 1/n^2)$ for $\beta = 1$, and (3) $N(\nu)$ is independent of ν. The spectrum is uniform over varying frequencies, and since $E = h\nu$, this means that the main part of the energy is concentrated in the part of the spectrum with the shortest wavelengths:

$$\mathrm{d}E_{\mathrm{VCh}} \sim \nu\,\mathrm{d}\nu\,. \tag{42}$$

This explains the blue-to-violet color of Vavilov–Cherenkov radiation.

We have already mentioned that the quantum-mechanical treatment of the Vavilov–Cherenkov effect by Ginzburg [117] gave similar results, with a small correction. Additional details on the theory of the Vavilov–Cherenkov effect can be found in the reviews by B M Bolotovskiĭ [122] and Ginzburg [123].

3.3 Practical applications

Figure 8c and formula (40) imply that Vavilov–Cherenkov radiation propagates along the elements of a cone whose axis coincides with the direction of motion of the particle and that the opening angle is 2θ. This glow can be recorded with color photographic film (PhF) placed perpendicular to the direction of motion of the particle (Fig. 8d). The emission emerging from the radiator R produces a blue ring as it intersects the photographic film.

As a result of the sharp directivity of the Vavilov–Cherenkov emission, it is possible to calculate the velocity β of the particle from the angle θ. It follows from formula (40) that the range of variation of β in which emission is possible extends over

$$\frac{1}{n} \leqslant \beta \leqslant 1\,. \tag{43}$$

If $\beta = 1/n$, the emission is observed at an angle $\theta = 0°$; if $\beta = 1$, then is appears at the maximum angle

$$\theta_{\max} = \arccos\frac{1}{n}. \tag{44}$$

For instance, in water ($n = 1.33$) $\beta_{\min} = 1/1.33 = 0.75$ and in accordance with formula (29), this condition for electrons is met already at the kinetic energy

$$T = mc^2(\gamma - 1) = mc^2\left(\frac{1}{\sqrt{1 - \beta^2}} - 1\right) \simeq 0.26\ \mathrm{MeV}\,. \tag{45}$$

Therefore, the Vavilov–Cherenkov effect in water must be observable for electrons of relatively low energy (e.g., beta-decay and Compton electrons). The maximum angle at which glow can be observed in water is found from the condition

$$\theta_{\max} = \arccos\frac{1}{1.33} = 41.5°\,. \tag{46}$$

For particles with $m > m_{\mathrm{e}}$, condition (43) is obviously satisfied at much higher energies. For instance, the threshold energies at which the glow is generated in water by the moving π-mesons and protons must, according to formula (45), be greater by the factors m_π/m_{e} and $m_{\mathrm{p}}/m_{\mathrm{e}}$, respectively, than for electrons, i.e., they must equal

$$T_\pi \simeq 70\ \mathrm{MeV} \quad \text{and} \quad T_{\mathrm{p}} \simeq 477\ \mathrm{MeV}\,, \tag{47}$$

respectively, because $m_\pi \simeq 273\,m_{\mathrm{e}}$ and $m_{\mathrm{p}} \simeq 1836\,m_{\mathrm{e}}$.

Vavilov–Cherenkov radiation is observed not only in those liquids investigated by Cherenkov but also in any other materials transparent to the visible part of the spectrum. Since the color of the glow is blue-to-violet, it is desirable that the materials retain transparency up to ultraviolet frequencies.

The Vavilov–Cherenkov effect is widely applied to designing instruments for measuring the velocity of rapidly moving charged particles — the so-called Cherenkov counters. In addition to the above-mentioned transparency to light of the maximum possible frequency, the choice of a material for designing the Cherenkov counter is determined by the range β in which measurements are to be conducted. Transparent plastics are very convenient materials. Cherenkov counters may also be filled with liquid or gas.

Depending on their destination, Cherenkov counters can be separated into threshold counters (recording particles with $\beta > 1/n$) and differential counters (making it possible to select particles with the prescribed velocity β). The schematic diagram of the simplest threshold Cherenkov counter is given in Fig. 9a. The charged particle moves along the axis of the radiator R. If the velocity of the particle $\beta \geqslant 1/n$, where n is the refractive index of the radiator's material, Vavilov–Cherenkov radiation will be emitted at an angle $\theta = \arccos(1/\beta n)$, then undergo total internal reflection (characterized by the angle $\varphi = \pi/2 - \theta$) from the walls of the cylindrical radiator, enter its conical part, get reflected by its surface and a flat mirror M (installed in order to remove the photomultiplier from the particle beam), and get collected by a lens L and a photomultiplier Ph.

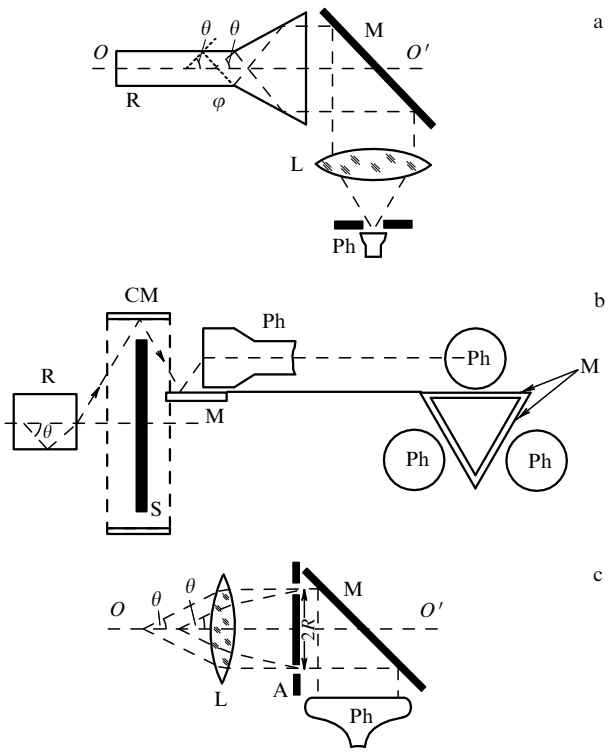

Figure 9. Schematic diagrams showing the Cherenkov counters: (a) the simplest threshold counter: OO' — direction of motion of the particle, R — radiator, θ — angle characterizing the direction of emission, φ — angle of total internal reflection, M — mirror, L — lens, Ph — photomultiplier; (b) the differential counter (in two projections): R — radiator, θ — angle determining the direction of radiation, CM — cylindrical mirror, S — screen, Ph — photomultipliers, M — flat mirrors; (c) the gas-filled Cherenkov counter: OO' — direction of motion of the particle, θ — angle determining the direction of emission, L — lens, A — ring aperture with a slit of radius $R(\theta)$, M — mirror, and Ph — photomultiplier.

Using the threshold counter, one can select beam particles of a given sort. If, for instance, the beam contains protons and π-mesons with identical momentum $p_p = p_\pi$ (selected by the magnetic system of the beam), then their velocities will also be different because their masses are different: $\beta_\pi > \beta_p$. And if the refractive index of the radiator satisfies the condition $\beta_\pi > 1/n > \beta_p$, then the photomultiplier records the light pulse only upon passage of a π-meson. Protons will go through the counter without causing any glow.

One of the possible designs of a differential counter is sketched in Fig. 9b (in two projections). The figure shows that such a device can record Vavilov – Cherenkov radiation only from particles with a specific velocity $\beta \pm \Delta\beta$ corresponding to the angle $\theta \pm \Delta\theta$. The emission produced by particles moving at a different velocity does not enter the photomultipliers.

To measure the velocities of ultrarelativistic particles ($\beta \approx 1$), gas-filled Cherenkov counters are used; the refractive index here exceeds unity only slightly (it can be controlled by changing gas pressure in the counter). The schematic diagram of a gas-filled differential counter is shown in Fig. 9c. Charged particles move at a velocity β through the gas medium along the axis OO' and emit Vavilov – Cherenkov radiation at the same angle $\theta(\beta)$ to the axis at any point of their trajectories. We know from geometric optics that such rays will be gathered by the lens L in the focal plane as a ring whose radius R is determined by the angle θ (and the focal length of the lens). If ring aperture A with an annular slit of radius $R(\theta)$ is placed in the focal plane of the lens, the radiation only from particles moving at a velocity $\beta \pm \Delta\beta$ will be allowed to pass through it. The rays of this light, reflected by the mirror M, are recorded by the photomultiplier Ph. Particles moving at a different velocity β' cause emission at a different angle θ' to the axis OO'; having passed the lens, this radiation is gathered into a ring of a different radius $R(\theta')$ and, therefore, will not pass through the aperture. A distinctive feature of gas-filled Cherenkov counters is their relatively large size (~ 10 m).

Threshold and differential counters of distinct designs were used widely in accelerating installations in which numerous elementary particles were discovered (antiproton, antineutron, τ-lepton, J/ψ-particle, b-quark, etc.) and are used at least as widely now in elementary particle physics (for details, see the review [124], a monograph by J Jelley [125], a two-volume monograph by V P Zrelov [126], and our earlier article [3]).

3.4 Ginzburg – Frank transition radiation

The preceding sections discussed the Vavilov – Cherenkov effect which, at first glance, seemed to contradict classical electrodynamics because, according to this effect, a charged particle moving uniformly along a straight line and at a velocity above the speed of light (in a medium with $n > 1$) will emit light. We were able to show ultimately that this 'doubly exotic' phenomenon does not contradict the laws of conservation of energy and momentum; we even derived from these laws one of the main characteristics of this radiation — its sharp directivity. Until this point everything was sufficiently rigorous. Now that we switch to a story about the transition radiation, which is recognized as 'related' to the Vavilov – Cherenkov effect, we will have to somewhat sacrifice rigor in favor of popularized exposition in our search for this kinship and put forward a rather unscientific and slightly artificial argument. However, that cannot be helped because "*to reach the source one has to paddle upstream*".

We remind the reader that Vavilov – Cherenkov radiation is observed if $v > c/n$, where all three quantities are constant. However, only one of them, the speed of light in vacuum c, is a world constant while the other two characterize the properties of the medium (n) and the characteristic of motion of the particle (v), i.e., both can change. Let us see if a charged particle (for the sake of simplification, we will consider electrons only) will retain the property of emitting radiation in its motion through a specific medium with a specific velocity, and if it does, what the emission will be for different values of n and v.

We begin by making electrons move under the action of a magnetic field along a circular orbit in the evacuated chamber of an accelerator — that is, we make the electron velocity constant in magnitude but varying in direction ($d\mathbf{v}/dt \neq 0$), and set the refractive index of the medium (vacuum) equal to unity ($n = 1$). Experience shows that such electrons rapidly lose energy, and it is necessary to replenish their energy constantly in order to maintain their circular motion. The question arises, where does this energy go if electrons move inside the vacuum chamber far from its walls, i.e., without colliding with anything. Obviously, only one answer is possible here: the energy is lost by emission of radiation. This emission — just like Vavilov – Cherenkov radiation — can in fact be seen by the eye if the electron

energy is on the order of 100 MeV and the magnetic field is as high as $H = 1$ T. (At higher energies of electrons, the radiation is emitted in the form of hard X-rays, but in both cases it forms thin light beams directed tangentially to the electron beam.) This is the familiar synchrotron radiation. It was theoretically predicted as a magneto-bremsstrahlung as early as 1912 but the construction of high-power synchrotrons and their utilization for practical purposes started relatively recently. The first Soviet synchrotron was constructed in 1947 under the supervision of V I Veksler, who failed to receive a Nobel Prize for it only because of an unfortunate combination of circumstances (the discovery was excessively classified).

Sometimes synchrotron radiation is still referred to as the magneto-bremsstrahlung, meaning that it is a particular case of the bremsstrahlung emitted by charged particles when they rapidly decelerate in the electric fields of the atomic nuclei and atomic electrons of sufficiently dense media. The simplest and most widely known example of bremsstrahlung is the continuous X-ray spectrum generated when relatively slow electrons decelerate by the anticathode of an X-ray tube. The intensity of this radiation is maximal in the direction perpendicular to the direction of motion of the electrons, and the spectral distribution is inversely proportional to the frequency of emitted radiation: $N(v)\,\mathrm{d}v \sim \mathrm{d}v/v$; hence, energy losses through bremsstrahlung are independent of v:

$$\left(\frac{\mathrm{d}T}{\mathrm{d}x}\right)_{\mathrm{br}} \sim N(v)hv \sim \mathrm{d}v\,. \tag{48}$$

Electrons lose the same energy by bremsstrahlung in any part of the bremsstrahlung spectrum (that is, emitting either one hard quantum hv or several soft quanta with the same total energy). The bremsstrahlung of relativistic electrons ($T > 1$ MeV), induced in their decelerating by the targets in electron accelerators, is directed forward as a narrow light beam with an average divergence angle of $\theta \approx mc^2/T$. In media with large Z, electron energy losses through bremsstrahlung become the main relaxation channel (in comparison with losses to ionization) beginning with an energy of $T \simeq 10$ MeV.

Let us formulate a preliminary summary of the above string of arguments.

(1) Vavilov–Cherenkov radiation is characterized by the condition $v > c/n$ in which all three quantities are constant.

(2) Synchrotron radiation and bremsstrahlung are emitted when v is variable either in direction or in magnitude and n is constant (even though it is different under various conditions). Therefore, we are left with one more option that we will formulate as a question:

(3) Can there be another type of radiation caused by a moving electron if its velocity remains constant (in the sense of $\mathbf{v} = \mathrm{const}$) while the refractive index of the media changes? The answer to this question is 'yes': there can and there is. This is the transition radiation predicted by V L Ginzburg and I M Frank in 1945 in works [127, 128] and analyzed in detail in the book by Ginzburg and V N Tsytovich [129]. The refractive index n of the media may be changing in space when the particle passes from one part of the optically inhomogeneous media to another, or in time when the refractive index of the homogeneous media changes abruptly (for instance, in response to a change in pressure). Velocities in this case may be high ($v > c/n$) but they can also be relatively low ($v < c/n$) — that is, the transition radiation may appear together with Vavilov–Cherenkov radiation or without it. Note also that both types of emission are related not only in that both are caused by a charged particle moving through the medium uniformly along a straight line but also because the mechanisms of emission are quite close. In both cases the sources of radiation are atoms of the media, which are excited by the passing particle, and in both cases radiation is coherent. Further details on the nature of the transition radiation and its application will be described by resorting to the articles by B M Bolotovskiĭ [130], G I Merzon [131], and V L Ginzburg [123, 132].

Transition radiation thus emerges when charged particles move at constant velocity along a straight line through a nonuniform medium, for example, when they pass through the interface of two media with different refractive indices. The cause of the transition radiation is the change in the electromagnetic field accompanying a moving particle when this particle passes from one medium into another.

As follows from calculations by Ginzburg and Frank, transition radiation must be emitted on both sides of the interface, in such a way that the backward radiation consists of visible-range electromagnetic waves regardless of the particle velocity, while the radiation emitted forward has a broad spectrum with the maximum frequency

$$\omega_{\max} = \omega_0 \frac{T}{m_0 c^2}\,, \tag{49}$$

where $\omega_0 = 4\pi n e^2/m_{\mathrm{e}}$, n is the number of electrons per unit volume of the media, m_{e} is the electron mass, e is the electron charge, m_0 is the mass of the emitting particle, and T is its kinetic energy. The forward-emitted transition radiation propagates at an angle θ to the direction of motion of the particle, which is given by the expression

$$\theta \sim \frac{1}{\gamma}\,, \tag{50}$$

where $\gamma = 1/\sqrt{1-\beta^2}$ is the Lorentz factor (or γ-factor), and $\beta = v/c$ is the velocity of the particle expressed as a fraction of the speed c of light in vacuum. If $\gamma \gg 1$, this angle is very small. Energy losses through transition radiation increase in proportion to the γ-factor (i.e., to energy) but the number of emitted photons is practically constant (roughly one photon per one hundred particles crossing the interface between the media). This is caused by the fact that as the γ-factor increases, harder photons are emitted.

The first experimental examination of backward-emitted transition radiation dates back to 1958, when a bright white luminous spot (different from the one that was caused by the bremsstrahlung) was observed on a metal surface irradiated by a particle beam in a vacuum. Today precise techniques of measuring the parameters of transition radiation in the optical range are available so that the results of such measurements allow the determination of the optical properties of irradiated surfaces.

Since the hardness of the emitted photons grows as the γ-factor of emitting particles increases, the energy of the emitted photons is a measure of the particle energy (if its mass and charge are known). If the γ-factors are large, the major part of the radiation is concentrated in the X-ray frequency range, for example, electrons with energy 10 GeV ($\gamma \approx 10^4$) emit forward photons with energies 10 keV when crossing an interface between a dense medium and a gas.

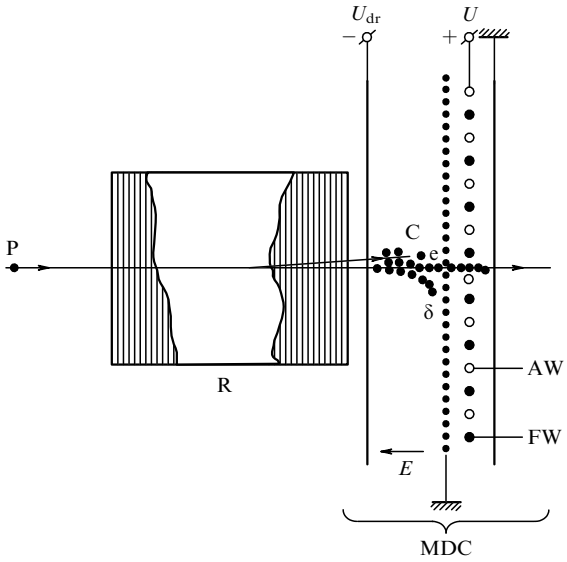

Figure 10. Schematic of the transition radiation detector: P — particle, R — radiator, MDC — multiwire drift chamber, AW — anode wires, FW — field-shaping wires, U_{dr} — drift potential, U — high voltage, e — ionization electrons, δ — delta-electrons, C — cluster formed by the X-ray emission of the radiator.

In view of the above-mentioned low intensity of transition radiation, when particles cross a single interface ($N_{ph}/N_{part} \simeq 0.01$), layered media are used for practical utilization of the transition radiation: stacks of a large number (several hundred) of thin (5 – 100 μm) plates (films or foils) that are transparent to X-ray quanta and separated by gaps of 0.1 – 2 mm. The combined emission of all layers of the stack is recorded by an X-ray detector. Therefore, a transition radiation detector consists of a radiator (a stack of emitting plates) and the detector proper — a device recording X-ray quanta (for instance, a multiwire drift chamber) and capable of measuring their energy. Sectional type detectors consisting of several sections, each containing its own radiator and its own recorder, are often used (see Fig. 10).

Transition radiation detectors allow experimenters not only to measure the energy of particles of a known kind (for example, electrons) but also to discriminate between particles having the same energy (for example, electrons and pions) because the difference in mass ($m_\pi \simeq 273 m_e$) results in different γ-factors. Note that this possibility is available in the energy range $10^2 – 10^3$ GeV in which all other methods of particle identification fail completely. In addition to their use in elementary particle physics as detectors in large accelerators, transition radiation detectors are used to study cosmic rays.

To conclude this section, we need to note that transition radiation is an all-wave phenomenon taking place not only in electrodynamics but also in acoustics and hydrodynamics (including the hydrodynamics of superfluid liquids) provided the medium is nonuniform and the emitting object (like a charged particle in electrodynamics) does not have a natural frequency and moves at constant velocity. A review of the work on transition radiation of acoustic waves in hydrodynamic media can be found in a paper by V I Pavlov and A I Sukhorukov [133].

This will conclude our discussion of the *spontaneous* emission of radiation by charged particles (mainly electrons) but we will return to this topic in Section 4.5 when discussing stimulated emission by electrons, including the stimulated Compton, Cherenkov, and undulator (i.e., magneto-bremsstrahlung) radiations and projects based on them; at least one of them (a free electron laser) is already being implemented.

4. Quantum electronics

4.1 History of the main ideas and discoveries. Masers and lasers. Work of Townes, Basov, and Prokhorov

Before describing the design and applications of lasers that have gained such popularity in our time, we will remind the reader of the main discoveries in quantum electronics, which led to the creation of quantum oscillators, first in the radio frequency band (masers) and then in the optical range (lasers)[21]. The main ideas of quantum electronics: the need for inverse population of energy states of the active medium and placing this medium in a cavity, were advanced at the beginning of the 1950s by N G Basov and A M Prokhorov (USSR) and Charles Townes (USA)[22]. In the spring of 1954, Townes, J Gordon, and H Zeiger (USA) built the first acting molecular oscillator operating on ammonium molecules (NH_3) [136]. In this case, the inverse population was created by enriching the equilibrium molecular beam with excited molecules, and positive feedback was provided by the cavity. In autumn 1954, Basov and Prokhorov suggested a similar device [137].

In 1955, Basov and Prokhorov proposed a three-level pump method with external pumping to produce inverse population [138]. In 1958, Prokhorov suggested using an open cavity of two parallel metal disks to apply positive feedback to the amplifying medium [129]. In contrast to a cavity resonator, whose dimensions are comparable with the wavelength and which thus cannot be fabricated for the optical range, the length of an open resonator must be equal to an integral number of half-wavelengths, i.e., it can have an acceptable size for designing a laser. A similar two-mirror resonator was suggested at the same time by A Schawlow and Townes [140]. In 1959, Basov, B M Vul, and Yu M Popov advanced the idea of using semiconductors to create lasers [141] and finally, in 1960, T Maiman (USA) created the first solid-state single-crystal ruby laser in which he realized all the basic ideas of quantum electronics listed above [142]. In 1961 – 1964, other solid-state, liquid, gas, and semiconductor lasers were developed [143 – 145]. In the 1970s, the first free-electron laser began operating, and in the 1980s, an X-ray laser using high-temperature plasma as the active medium.

For the fundamental research in quantum radiophysics that led to the creation of oscillators and a new types of amplifiers — masers and lasers — Townes, Basov, and Prokhorov were awarded the Nobel Prize in Physics 1964 [146 – 148].

[21] Maser is the acronym for Microwave Amplification by Stimulated Emission of Radiation; laser is the acronym for Light Amplification by Stimulated Emission of Radiation.

[22] V A Fabrikant pointed to the possibility of light amplification in the medium with inverse population in his doctorate even earlier (1939 – 1940); he proposed a specific method of generating inverse population (electrical discharge in a mixture of gases, using the resonance in collisions of the second kind) [134]. More details on the early history of the laser can be found in Charles Townes's book [135].

Here is how this section of the review is organized: first, we give a popular exposition of how quantum oscillators work, explaining as visually clearly as possible the physical foundations of the above-listed main concepts of quantum electronics. Then we discuss the design of the pioneer Townes–Basov–Prokhorov molecular oscillator and also other types of maser-oscillators and maser-amplifiers and where they are applied. Finally, in Section 4.6 we look at the design and fields of application of various types of lasers: solid-state (beginning with Maiman's first laser), liquid, gas, X-ray, and free electron lasers (because of the specific operating conditions of the semiconductor laser, it is considered in Section 5).

4.2 Quantum oscillator: principle of operation

To make our explanations specific, we will discuss the principle of operation of quantum oscillators using lasers as an example, even though the main concepts describe the maser as well (see Section 4.3). In the most general form, the principle of a laser's operation is the stimulated emission of directed coherent light radiation by the active medium which is driven by an external source of energy to an inverse-population state and where a positive feedback is applied to it at the resonance frequency. This statement is indeed sufficiently general because it works for lasers of any design; however, it cannot reveal the detailed picture of how the laser works. Therefore, in what follows we attempt to give a less rigorous interpretation of the main components of the definition.

We begin with the active (lasing) medium which can be solid, liquid, gaseous, or plasmalike with nonequilibrium distribution of active working particles (atoms, molecules, or ions) over their energy states. Assume now that active particles possess at least three energy levels with energies $E_1 < E_2 < E_3$. Of these three, E_1 corresponds to the ground state, E_2 to a metastable, sufficiently long-lived state (narrow lasing level), and E_3 to a broad level or even an entire energy band which can be populated by the action of a nonmonochromatic energy source such that spontaneous transitions from this level to the level E_2 occur sufficiently rapidly.

In conventional (equilibrium) conditions the level population[23] $N(E)$, i.e., the number of particles in a state of energy E, is determined by the Boltzmann distribution which in this particular case is of the form

$$N(E) \sim \exp\left(-\frac{E}{kT}\right), \tag{51}$$

where E is the energy in eV, $k = 0.863 \times 10^{-4}$ eV K^{-1} is the Boltzmann constant, and T is the absolute temperature. As follows from Eqn (51), one obtains

$$\frac{N_1(E)}{N_2(E)} = \exp\frac{E_2 - E_1}{kT}, \tag{52}$$

i.e., if $E_2 > E_1$, the populations of the corresponding levels satisfy the condition $N_2 < N_1$. Therefore, even though according to Einstein's theoretical findings (1917) the probabilities of elementary acts of absorption (the $E_1 \to E_2$ transition) and emission (the $E_2 \to E_1$ transition) equal one another, the total absorption by the medium in its normal equilibrium state is greater than the total emission, so that a noninverted medium cannot emit radiation.

[23] The term 'occupation' is often used interchangeably with the term 'population'. They are equivalent.

For the active medium to be able to emit photons with energy $\hbar\omega = E_2 - E_1$ that corresponds to the lasing transition $E_2 \to E_1$, the energy levels of its working particles must be possessed of an inverse population

$$N_3(E_3) > N_2(E_2) > N_1(E_1). \tag{53}$$

Obviously, relation (52) can be satisfied in this case only if $T < 0$, so that sometimes people characterize the inverted medium (not really correctly) as having negative temperature! The inverse population of the levels in the lasing medium is created by a pumping system. This system is an external source of energy (a pump lamp, discharge tube, auxiliary laser, nuclear or chemical reaction, etc.) which activates a considerable number of working particles from level E_1 to level E_3 from which they spontaneously drop back to the level E_2, thus increasing its population. The thus prepared medium (it now becomes active) can emit photons; however, these are not yet laser photons but merely luminescence photons (noncoherent and moving in all directions). Furthermore, in this medium it is also possible to observe a brighter and partially coherent superluminescence which arises from conventional luminescence as a result of quantum amplification of the latter owing to the stimulated emission triggered by luminescence photons as they propagate through the active medium. As these photons progress along their path they cause resonance emission of identical photons (identical in direction, frequency, phase, and polarization) (P A M Dirac, 1927 [4]) emitted by inverted elementary radiating elements of the active medium. The amount of this optical quantum amplification of internal radiation grows exponentially with the distance x covered by the photons:

$$I = I_0 \exp(\alpha x), \tag{54}$$

where $\alpha \sim (N_2 - N_1)$.

It is clear from the above arguments that the last step that must be accomplished to convert superluminescence into directed coherent laser radiation is to lengthen to the maximum extent the path of photons in one selected direction inside the active medium, changing its operation to the resonant mode. This is achieved by placing the active medium in an optical cavity, for which one can use a pair of strictly parallel mirrors (do you remember A M Prokhorov's metal disks?). Photons propagating perpendicular to the surface of mirrors and repeatedly reflected by them will cover a very long path through the active medium, creating newer and newer identical photons along its length (the positive feedback). Furthermore, the effect of quantum optical amplification will be especially strong if the distance L between the mirrors is equal to the whole number of half-wavelengths λ satisfying the lasing condition $E_2 \to E_1$. Obviously, in this particular case this condition is met if

$$L = n\frac{\lambda}{2} = \frac{n}{2}\frac{hc}{E_2 - E_1}, \tag{55}$$

where h is the Planck constant, and c is the speed of light in vacuum. And finally, it is necessary to extract the resulting laser radiation from the active medium, for which one of the mirrors is made partially transparent.

At the present moment an enormous number of very different quantum oscillators and amplifiers (masers and lasers) have been developed and successfully used; they differ in design, type of active medium, output power, spectral

range and type of radiation (continuous or pulsed), method of producing inverse population, method of pumping, resonator design, and so forth. Some of these will be described below.

4.3 Maser-oscillators and maser-amplifiers. Types and applications

4.3.1 Pioneer molecular oscillator in the microwave range. Basic parameters of molecular and atomic microwave oscillators. Applications. The first quantum oscillator for the microwave range ($\lambda = 1.24$ cm) was created, as we already mentioned, by Gordon, Zeiger, and Townes [136] in the spring of 1954. Somewhat later (in autumn 1954) the idea of a similar device was advanced by Basov and Prokhorov [137]. The principle of operation of a molecular oscillator consists in stimulated transitions of ammonia molecules (NH_3) from the excited states to the ground state within a closed cavity resonator. In contrast to later developments of quantum oscillators in which inverse population in the active media was created by external pumping, the first quantum oscillator used a molecular beam in an equilibrium state characterized by relation (51). We remind the reader that according to expression (51), as the excitation energy of molecules increases, their number decreases, i.e., such a beam cannot emit quanta. And since the three-level pump method was not known at the time, the problem of inverting populations of energy levels was solved by an elegant and, we can say, 'converse' method of 'pumping out' — that is, by sucking off less excited molecules from the equilibrium beam; this enriched the beam in more excited molecules and thereby made it capable of emitting light [149]. We will now describe this mode of operation in more detail.

The four-atomic ammonia molecule (NH_3) is arranged as a pyramid with three hydrogen atoms H at its base and one nitrogen atom N at the apex. The molecule can exist in two inverted (mirror-reflected) states (N 'above' three H atoms or N 'below' them). These states correspond to two values of the electric dipole moment (EDM) of opposite sign and an inverted doublet of levels E_1 (we call it the lower level) and E_2 (the upper level); the energy difference between the two is $E_2 - E_1 = h\nu \simeq 10^{-4}$ eV, and thus it lies in the microwave range ($\lambda = 1.24$ cm). However, as we already mentioned, molecules with a higher population of the lower level dominate an equilibrium beam of ammonia, and such a beam cannot emit radiation. For it to be able to do so, it is necessary to remove some molecules residing at the lower energy level.

It was possible to achieve this operation mode owing to the above-mentioned difference between EDMs of mirror configurations of the ammonia molecule. We know that a particle with nonzero EDM is oriented in a uniform electric field (but is not displaced), while in a nonuniform field it moves in a certain direction which depends on the sign of the EDM.

In this experiment, molecules were separated by the value of their EDM (and therefore, by the values of excitation energies E_1 and E_2) by using a quadrupole high-voltage capacitor with a nonuniform electric field perpendicular to the axis of the molecular beam, in such a way that grad $E \sim \mathbf{r}$, where $\mathbf{r}$ is the radius vector on the plane perpendicular to the axis of the capacitor (Fig. 11). As follows from the theory of a molecular oscillator [149], the interaction of the EDM with a non-uniform electric field deflects less excited molecules (E_1) of the beam towards the periphery of the capacitor, and molecules excited to a higher energy (E_2) towards its axis,

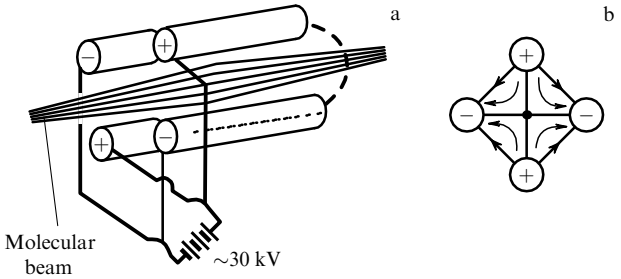

Figure 11. The sorting system of the first molecular oscillator: (a) quadrupole capacitor, (b) configuration of the electric field (arrows trace the lines of force).

from which they enter the cavity resonator built in such a way that one of its natural frequencies was close to the frequency of the transition $E_2 \to E_1$. Then the thermal electromagnetic field of the cavity on this frequency will cause stimulated emission of resonance photons with the frequency ν, which, accumulating in the cavity, stimulate emission from fresh molecules entering it and thereby placing positive feedback loop around active medium.

As a result, the electromagnetic field in the cavity quickly grows in intensity, and more and more molecules entering it drop back from the upper energy state to the lower one. Obviously, this growth stops when 50% of molecules are able to emit photons over the time of flight of molecules in the beam through the cavity — that is, when the populations at the upper and the lower energy levels become equal and the absorption probability becomes equal to the emission probability.

The power of the molecular oscillator was not high ($\sim 10^{-8}$ W) but the frequency of the radiation it emitted was quite stable ($\Delta\omega/\omega \simeq 10^{-11} - 10^{-7}$). In addition to ammonia molecules, other molecular and atomic beams were used to fabricate quantum oscillators, for example, atomic hydrogen beams. The principle of operation of the hydrogen oscillator in similar to the one discussed above but the beam is enriched in excited hydrogen atoms not by the electric field but by the nonuniform magnetic field that affects the magnetic dipole moment of the hydrogen atom.

The output power of the hydrogen oscillator is even lower than that of the ammonia maser ($\sim 10^{-12} - 10^{-10}$ W) but the stability of its frequency reaches $\Delta\omega/\omega \simeq 10^{-13}$. In view of the high frequency stability, molecular and atomic oscillators in the radio frequency band (masers) are used as quantum frequency standards, the foundation of the modern standards of time and length.

4.3.2 Quantum amplifiers for the microwave range. In addition to maser-oscillators for the microwave range, maser-amplifiers are widely applied to increase the sensitivity of radio devices required to conduct long-range cosmic radio communications with automatic probes sent to remote planets of the solar system, and in radioastronomy studies. We will describe the principle of operation of quantum amplifiers for the microwave range on the basis of the review by L S Kornienko and V B Shteĭnshleĭger [150].

The work of microwave quantum amplifiers (maser-amplifiers) is based on using stimulated emission in active quantum systems with an inverse population of energy levels. The population inversion is achieved by the above-described

methods of pumping a system of particles possessing three energy levels. The active media for the amplifiers are dielectric crystals with a small admixture ($\sim 0.01\%$) of paramagnetic ions possessing a suitable system of three (or more) levels whose energies can be controlled by applying an external magnetic field. In the microwave range (i.e., for decimeter, centimeter, and millimeter wavelengths), the level populations differ insignificantly and the efficiency χ'' of the paramagnetic crystal in the amplification mode can be approximately characterized by the quantity

$$\chi'' \sim |J| \frac{Nf}{kT}, \tag{56}$$

where $J = (N_1 - N_2)/(N_1 - N_2)_B$ is the inversion ratio, $N_1 - N_2$ is the population difference in the active crystal, $(N_1 - N_2)_B$ is the population difference at Boltzmann's equilibrium (51), N is the total number of paramagnetic ions per cm^3, f is the frequency of the amplified wave, k is the Boltzmann constant, and T is the absolute temperature. As follows from formula (56), the crystal will operate with sufficient efficiency only if it is kept at a sufficiently low temperature, usually provided by liquid helium. The inversion ratio J (its sufficiently high values are desirable) is usually several unities. The direct proportionality between χ'' and N holds only up to a certain limit, so that in reality the quantity χ'' reaches a maximum at a certain optimum value of N (for details see the paper [150]).

In the quantum amplifiers described in the literature, ions belonging to the transition groups of iron and rare earths (mostly ^{31}Cr and ^{31}Fe) are of use as active dopants in crystals. Most often ruby is employed for the active crystal; ruby is one of the crystalline modifications of aluminium oxide, Al_2O_3 (also known as α-corundum) in which a small part of Al ions is replaced isomorphically[24] by ^{31}Cr ions. Ruby is very convenient in that it can be used practically in the entire microwave range with different field strengths and different directions of the external magnetic field. Among other active crystals we will mention rutile (a crystalline modification of titanium oxide, TiO_2) doped with Cr^{3+} or Fe^{3+} ions, whose advantage lies in the possibility of achieving a high inversion ratio ($|J| > 10$).

The active crystal (i.e., one possessing inverted level population) can in principle amplify an electromagnetic signal if the wavelength corresponding to the signal equals

$$\lambda = \frac{hc}{E_2 - E_1}, \tag{57}$$

where $(E_2 - E_1)$ is the energy difference between the working levels of the lasing medium. However, the crystal would have to be very long (several meters), which is not realistic. Therefore, a quantum amplifier includes, in addition to the crystal, a cavity resonator or waveguide (which makes it possible to reduce the length of the crystal) and also, as should be obvious from the preceding discussion, a number of auxiliary devices (for pumping, for inducing the external magnetic field, for cooling the crystal to helium temperatures, etc.).

[24] Isomorphism is the ability of atoms, ions, or molecules to substitute for one another in crystalline structures, thereby forming substitution solid solutions. Isomorphic mixtures are formed only if the effective sizes and some other parameters of the substituting particles are not very different from those of the host particles.

In the first quantum amplifiers, the crystal was placed in a cavity resonator in which the amplified electromagnetic wave was multiply reflected from its walls and could interact with the crystal for a long time, providing high amplification for a small-sized crystal. The design of quantum amplifiers was later improved and they were built around several coupled cavities — active (enclosing crystals) and passive (empty).

The main type of quantum amplifier for the centimeter and millimeter wave ranges is the so-called quantum traveling-wave amplifier in which the signal is amplified in the waveguide filled with the active paramagnetic crystal. An important role in obtaining a high amplification coefficient in a small device is played then by the slowing down of the velocity of the traveling wave to the group velocity v_g, which is less than the speed of light in vacuum by a factor of 50 to 200.

The radio receivers in which quantum amplifiers are used have a sensitivity that is greater by a factor of 10^3 and a frequency stability that is approximately $10^4 - 10^5$ times more than in radio equipment used in the past. The typical output power of a quantum amplifier is 10^{-6} W which is greater by several orders of magnitude than the output power of atomic and molecular oscillators. Therefore, the application of quantum amplifiers has greatly increased the possibilities for long-range cosmic communications through systems in which quantum amplifiers are used in conjunction with specially adjusted large double-reflector low-noise antennas (several dozen meters in diameter).

Quantum amplifiers work successfully in the largest radio telescopes and planet radars. We can recall that V A Kotel'nikov and his team carried out radiolocation of such a very remote planet as Jupiter as early as 1962 [151]. Other well-known achievements in this field are the reception on the Earth of the far-side lunar image and of the Venus-surface image, and the reception of data from automatic interplanetary stations traveling towards the remotest planets of the solar system.

In radio astronomy, quantum amplifiers have made it possible to magnify by several orders of magnitude the dimensions of that part of the universe in which cosmic radiation at the 21-cm wavelength can be observed; this radiation was discovered in the 1950s and is caused by quantum transitions between the hyperfine structure components of the levels of neutral hydrogen atoms. This radiation carries very useful information on the presence of hydrogen in various ever more remote regions of the universe. Moreover, as a result of using quantum amplifiers, weak cosmic background radiation at the wavelength $\lambda \simeq 3$ cm was discovered; it became also feasible to observe very weak radio-emission spectral lines from excited and ionized hydrogen (at $\lambda = 5-6$ cm, and even $\lambda = 8$ mm) that carry information on stars in the Milky Way Galaxy.

4.3.3 Exotic masers. To conclude this section, we will describe two exotic objects in the world of masers. One of them was created by human hands, the other is the creation of nature.

One of the fundamental problems of modern astrophysics is the study of quasars and nuclei of galaxies whose angular dimensions are extremely small — less than one thousandth of one second of arc. To investigate such objects, it is necessary to have very-long-base radio interferometers because the angular resolution of the radio interferometer (in radians) is approximately equal to the ratio of the radiation wavelength to the baseline distance. We immediately see that to solve this problem, we need radio inter-

ferometers with bases of 8 to 10 thousand km and the use of the short wave part in the centimetric wave band, plus quantum traveling-wave amplifiers. Such an exotic radio interferometer with the USSR – USA baseline of about 10,000 km in length was created by the joint effort of Soviet and American physicists. Observations at the wavelength of 3.5 cm allowed astrophysicists to study the complex structure of a number of quasars, galaxies, and other compact sources and to detect in their structure regions whose size is less than 3×10^{-4} of one second of arc. To investigate these regions, similar experiments were run in 1976 with three telescopes situated in the USSR, the USA, and Australia, in which radiation at the 1.35-cm wavelength was recorded and the achieved resolving power was approximately 10^{-4} of one second of arc.

The second exotic object in the world of masers relates to a natural phenomenon. This is the so-called maser effect discovered in the cosmos in 1965 by H Weaver et al. [152] from the Radio Astronomy Laboratory of the University of California. A specific feature of the effect is that even though the observed wavelengths ($\sim$ 18 cm) could originate from the transitions between hyperfine structure components of the levels of the diatomic radical OH, their relative intensities completely contradicted this hypothesis. And to such an extent that the discoverers began to doubt the correctness of their interpretation and assigned part of the discovered spectral lines not to OH radicals but to some other radiation source which they named, because it was so quizzical, 'mysterium' (physicists love such exotic names — remember 'strange particles', about which we wrote in the review [3]). However, the initial hypothesis that all OH molecules are the source of this radiation was confirmed before long. In fact, this only deepened the mystery of the phenomenon: the radiation not only showed unusually high intensity but was also characterized by strong polarization and variability in time. Several years later a similar phenomenon was discovered by Charles Townes's group (the same Charles Townes of whom we wrote in Section 4.1) in the 1.35-cm line that corresponds to the rotational transition in H_2O molecules [153]. By 1990, several hundred such sources of unusual radiation were discovered.

As we mentioned at the beginning of this subsection, the cause of this phenomenon is the maser effect in space, namely, the amplification of radio waves passing through a cosmic medium due to the stimulated emission of photons by excited atoms and molecules of the medium. The required inverse population of atomic and molecular energy levels is provided by a permanently acting pumping which, according to suggested theoretical models, may be of radiational or collisional nature. High-power infrared radiation from the nascent massive stars may serve as a source of pumping in the case of the observed maser effect. On details of this extremely interesting phenomenon, you can read reviews by V S Strel'nitskiĭ [154] and V V Burdyuzha [155], and N G Bochkarev's book [156].

4.4 Lasers (solid state, liquid, gas, and plasma). Energy level diagrams and operation modes. Applications

4.4.1 The first ruby laser and other solid-state lasers. Pulsed and continuous operation modes. Three- and four-level diagrams. The pioneer laser was created, as we already mentioned, in 1960 by the American physicist T Maiman [142]. In developing it, Maiman applied all the fundamental ideas of quantum electronics that were formulated by that time by N G Basov, A M Prokhorov, and C Townes, namely, the need for inverse population of certain levels of the active medium, the three-level pump method with external power supply, application of open cavities to heighten the effect of stimulated emission of radiation, and placing positive feedback path around the active medium.

For the active material, Maiman's laser used a ruby crystal (Al_2O_3) in whose crystalline lattice Cr^{3+} ions substitute for about 0.05% atoms of Al. A simplified energy level diagram of Cr^{3+} in the ruby crystal is shown in Fig. 12a. Here, E_1 is the ground energy level: transitions from this level to two broad energy bands E_3 and E_3' are produced by green and blue light, respectively, emitted by a pulsed xenon flash lamp (pumping lamp). The duration of the pumping pulse is $\tau \simeq 10^{-3}$ s. Atoms spend a very short time at the levels E_3 and E_3' ($\tau \simeq 10^{-7}$ s) and rapidly drop back to the less excited but longer-lived ($\tau \simeq 10^{-3}$ s) metastable state E_2 in a radiationless manner, i.e., they give up their excessive energy to the crystalline lattice (thereby heating it). If pumping is sufficiently powerful, the number of particles at the metastable level may be higher than at the ground level, namely, inverse populations of the levels E_1 and E_2 are produced. When population inversion reaches a sufficiently high (threshold) value, the lasing transition $E_2 \rightarrow E_1$ at the wavelength $\lambda = 6943$ Å (red light) becomes possible.

Maiman's laser used a ruby crystal 5 mm in diameter. The roles of the open cavity and 'cavity' resonator (the former by the principle of operation, and the latter by design) were played by two strictly parallel polished and silver-coated end faces of the crystal, one of which was partially transparent in order to provide an exit for the generated emission. Pumping was achieved by a pulsed lamp fed from a battery of capacitors of 400-μF capacity at a voltage of 4 kV. Here are more precise temporal and other characteristics of this first laser: pumping pulse duration $\tau \simeq 5 \times 10^{-4}$ s; coherent radiation consisted of individual flashes of $\tau \simeq 10^{-7}$ s duration with intervals of several microseconds between them and a total duration of about 10^{-3} s. The peak emission power was about 10 kW. The efficiency of the laser was approximately 1%. (A more detailed description of the first laser and other oscillators for the radio frequency band and optical range, developed in this period, can be found in the review by G M Zverev et al. [157].)

The first ruby laser belongs to the class of solid-state lasers. In this class, in addition to ruby, other dielectric crystals and glasses containing rare-earth ions are used as the lasing medium. We already mentioned one of them, rutile, when discussing masers. Among others, we will mention yttrium-aluminium garnet $Y_3Al_5O_{12}(Nd^{3+})$, nickel niobate $NiNbO_3(Nd^{3+})$, and lithium-yttrium fluoride $LiYF_4(Nd^{3+})$, all activated with neodymium Nd^{3+} ions. On the whole, the laser effect has by now been discovered in several hundred dielectric crystals alloyed with ion dopants or with active ions implanted in the crystalline lattice. The second method is preferable because in this case active ions are distributed through the crystal more uniformly, which makes it possible to increase their concentration up to $\sim 20\%$ (in the ion dopant method the limit is only $\sim 5\%$), so that the output power can be increased without risking overheating the crystal. By the way, most solid-state lasers operate in the pulsed mode precisely to avoid overheating the crystal (and also to avoid the difficulties involved in creating a high-power optical pumping system working in continuous mode); still, some continuously operating systems are also known.

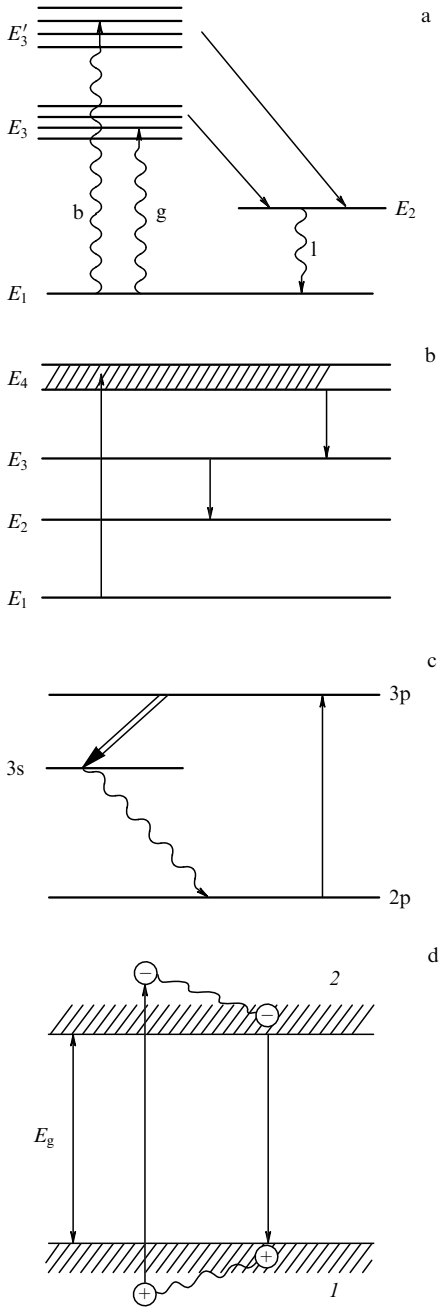

Figure 12. Energy level diagrams for laser radiation: (a) three-level scheme of a ruby's laser radiation: E_1 — ground level, E_2 — metastable level, wavy lines — pumping of the E_3 and E_3' bands by green (g) and blue (b) light and the laser transition (l), straight arrows — radiationless transitions; (b) four-level scheme: E_1 — ground level, E_2 — lower operating level, E_3 — metastable level, E_4 — absorption band; (c) diagram of transitions yielding the X-ray emission of Se^{24+} ions: 2p — ground level, 3s — lower operating level, 3p — upper operating level, straight arrow — collisional excitation, wavy arrow — radiation transition, double-line arrow — lasing transition; (d) scheme of pumping and emission in a semiconductor laser: *1* — valence band, *2* — conduction band, circles with minus inside — electrons, circles with plus inside — holes, and E_g — forbidden band width.

The pulsed operation mode (described above with the ruby laser as an example) in which the total duration of coherent emission equals the pumping duration ($\sim 10^{-3}$ s) is known as the free running mode. We recall that in this case the laser pulse consists of a series of consecutive short oscillation spikes lasting $\sim 10^{-7}$ s each and separated by gaps of a microsecond or so, and that the total pulse duration is 10^{-3} s; the peak power in this mode is approximately 10 kW. However, there exists another pulsed mode for solid-state lasers in which the duration of a total oscillation pulse can be made equal to that of one oscillation spike in the case of free running mode (and even shorter); this leads to a dramatic increase in the laser's peak power.

This mode of operation is achieved by the so-called cavity Q-switching method[25], which essentially consists of a forced delay of the triggering of the lasing process. To do this, a special fast electronic shutter is added to the design of the cavity; it can open or shut the entrance of radiation into the cavity. In this operating conditions the pumping proceeds with the shutter closed, which allows the cavity to accumulate the maximum possible number of excited ions at the metastable level (if the cavity is not operating, oscillation cannot start even when the inverse population of levels is achieved). Then the shutter is rapidly opened and the cavity is switched on, so that the energy stored in the lasing medium is released instantaneously (for $10^{-7}-10^{-8}$ s) as a high-power light pulse.

There also exist other techniques for obtaining still shorter laser pulses and enormous peak powers corresponding to this shortening. Among them is the method based on synchronization, i.e., on constructive interference of several resonance modes in the cavity. This technique also uses a shutter but it is a particular one, of, let us say, self-controlled design. This is a specific dye cell placed in the cavity; the dye is such that its opacity disrupts the positive feedback, which results in a steady growth of the inverse population in the lasing medium. This process continues until superluminescence is achieved (see Section 4.2). The most intense superluminescence pulse bleaches the liquid in the shutter and thereby switches on the cavity which provides oscillation. In order to ensure synchronized work of all cavity modes, the amplitude or phase modulation of oscillations in the cavity is employed; it consists either of the modulation of transmittance of the output mirror or of the modulation of the distance L between mirrors on the frequency of intermode beats $\Omega = c/2L$. The duration of the resulting light pulse is $10^{-9}-10^{-10}$ s, and its power is very high: $\sim 10^{10}$ W. Still higher peak power can be brought about with solid-state lasers using a glass lasing medium with an admixture of Nd^{3+} ions; these lasers generated what was at that time ultra-short pulses of $10^{-11}-10^{-12}$ s duration. (Modern ultra-short pulse-generating lasers are described in the concluding section.)

Notice that despite the enormous peak power, the total energy transferred by these pulses is relatively moderate. However, if necessary, this energy can be increased by adding one or more amplifying lasers to the laser-oscillator. The peak power is then also increased, reaching values on the order of $10^{13}-10^{14}$ W.

Continuous-wave operation is also possible for certain laser crystals (e.g., for the garnet mentioned above and also for glass with Nd^{3+}). The necessary condition for this is the presence of four-level scheme of levels in the lasing medium, as shown in Fig. 12b. The advantage of this arrangement in comparison with the three-level one is that the lasing transition occurs between the third and the second levels

[25] The Q factor is the quantity that characterizes the resonance properties of an oscillatory (in this case emitting) system. Numerically, the Q factor is equal to the ratio of the resonance frequency to the width $\Delta\omega$ of the resonance curve: $Q = \omega/\Delta\omega$.

($E_3 \to E_2$), with E_2 not being the ground level, i.e., according to formula (51) its population is lower than that of E_1. As a result, the inverse population of the metastable level E_3 is much easer to obtain in the four-level scheme (at lower pumping) than in the three-level one (there is no need then to 'raise' more than 50% of particles from the level E_1). Both continuous-wave lasers (garnet- and glass-based) belong to the same type of neodymium lasers because the generation of the laser radiation occurs in them due to quantum transitions between the energy states of trivalent Nd^{3+} ions. Lasers with Nd^{3+} ions added to a different condensed medium, for instance, a solid semiconductor or an organometallic or organic liquid, also belong to this type.

The actual energy level diagram of Nd^{3+} ions is more complicated than the one shown in Fig. 12b (it contains several pumping bands and lower intermediate levels) but in principle it works like a four-level configuration and generates, with maximum probability, infrared radiation at the wavelength $\lambda = 1.06$ μm. Note that neodymium lasers work not only in the continuous-wave mode but also in all the above-mentioned pulsed modes (free running mode, Q-switching mode, and mode synchronization). In reality, the cw mode of laser operation is essentially a pulsed mode because it yields an uninterrupted train of pulses $\sim 10^{-9}$ s long each, following one another at the frequency Ω mentioned above. For further reading on solid-state lasers (including those with tunable emission frequency) we can recommend the review written by Prokhorov [158] on the occasion of the 25th anniversary of laser development (see also the laser handbook edited by Prokhorov [159]). Note that, as we know from experts, the efficiency of modern solid-state diode-pumped lasers reaches 30%, their output power may be as high as 10 kW, and they possess high frequency stability (comparable with the stability of microwave oscillators) and a narrow emission line (on the order of 1 Hz).

4.4.2 Liquid lasers. Lasers in which the active medium is a liquid (inorganic or a dye) are known as liquid lasers. The active particles in inorganic liquid lasers (ILLs) are usually the same neodymium Nd^{3+} ions (we discussed them in Section 4.4.1) included in the composition of the liquid phosphor. Therefore, inorganic liquid lasers possess all the advantages of the solid-state neodymium lasers described above that originate from using four-level system pumping. Furthermore, they possess an additional advantage in comparison with solid-state lasers — the possibility of cooling the active medium as it is pumped through the cavity and then heat exchanger, so that the output energy and power can be further increased. Inorganic liquid lasers work both in the pulsed and continuous modes of operation. The energy ILLs generate may amount to $\geqslant 1$ kJ, and an output power to $\geqslant 1$ kW. One drawback of ILLs is the large divergence of the output beam; however, methods for overcoming this deficiency do exist (for more details see the book by Yu G Anikiev, M E Zhabotinskiĭ, and V B Kravchenko [160]).

Typically, dye lasers (DL) use aqueous or alcohol (and some other) solutions of dyes which act as the lasing medium. For dyes, complex organic compounds are applied; alternating them in a dye laser makes it possible to tune the oscillation frequency, covering the wavelength range from 330 nm to 1.8 μm. The laser effect has been achieved in almost 1000 different dyes. The most frequently used are oxazoles and oxadiazoles (the violet and ultraviolet ranges of the spectrum), coumarins (blue to green), rhodamines (yellow to red), and polymethine dyes (near infrared range). The greatest asset of the DL is the possibility of tuning the oscillation frequency. The drawback of the DL is the considerable width of the lasing spectrum; to reduce it, special optical devices are introduced into the DL cavity (e.g., dispersion prisms, interferometers, etc.). It is possible, using the same devices, to change the laser wavelength generated by a given dye in a gradual manner.

Depending on the type of pumping, a DL may work both in the pulsed and in the cw modes. To realize the pulsed mode, flashlamp pumping is typically used; to realize the continuous-wave operation, pumping by gas lasers is employed (see Section 4.4.3). However, in the case of laser pumping (e.g., by a garnet laser), pulsed mode can also be achieved with a DL. DLs are very efficient for generating ultrashort radiation pulses. Using the mode synchronization method mentioned above (Section 4.4.1), dye lasers could generate pulses of $\sim 10^{-14}$ s duration. For additional details on DLs, read the paper by A N Rubinov and V I Tomin [161]. Reference data on the parameters of various liquid lasers are given in the book [159].

4.4.3 Gas lasers. In a broadly interpreted sense, a gas laser (GL) is a laser with gases, vapors or their mixtures, or weakly ionized plasma used as a lasing medium. The most common feature of gas lasers, separating them from solid-state, liquid, and semiconductor lasers, is the substantial homogeneity of the lasing medium, its low density (even if a pressure of several dozen atmospheres is maintained in the GL cavity), and as a result, a high degree of transparency, narrow emission and absorption lines, a very small divergence of laser radiation and a highly stable frequency. Furthermore, because of the technical feasibility of high-rate circulation of the lasing medium through the cavity, gas lasers make it possible to reach very high mean radiated power without overheating the active medium. Another specific feature of the gas laser is the possibility of producing the laser emission in a very wide range of wavelengths (from ultraviolet to submillimeter).

Specific types of gas lasers vary greatly. Their designs and operation conditions are determined by numerous factors, including the choice of an active medium, type of laser transition (atomic, ion, molecular, eximer, etc.), and the type of pumping (optical, electric discharge, nuclear or chemical reaction, electron beam). Some of these factors may be interrelated. In what follows we describe several particular gas lasers and briefly characterize their features. For details of physical processes in gas lasers, see the book by A V Eletskiĭ and B M Smirnov [162] and the review by N G Basov and V A Danilychev [163].

Just like solid-state and liquid lasers, gas lasers can be classified as operating in the pulsed or cw mode. The choice of a mode is determined by the type of pumping. Most gas lasers work in the cw mode; a considerable part of them are gas-discharge lasers. In these lasers, the level population inversion is provided by electric discharge that produces electrons exciting the working particles of the gas (by electron collisions). The working particles may be neutral atoms (e.g., He, Ne), ions (e.g., Ar^{2+}, Ar^{3+}, Kr^{2+}, Kr^{3+}, Ne^{2+}), atoms and ions of vaporized metal (e.g., Cu, Cd), molecules (e.g., vibrational and rotational levels of CO_2 and N_2), or excimer (unstable) molecules (Ar_2, ArF, Kr_2, KrCl, Xe_2, XeBr, and so forth).

The first to be built from the list above was the helium–neon laser (in 1961) in which neon is the lasing medium, while helium, via resonant transfer of energy from its metastable level to neon, helps the neon to stay in the upper excited state for a longer time than is possible for pure neon (in which sufficiently high population inversion cannot be achieved).

Because the lasing levels of neon are split to a number of sublevels, the helium–neon laser can work on 30 wavelengths in the visible (red spectrum) and infrared ranges. The selection of a particular resonance frequency is provided by a specific tuning of the cavity mirrors. The power and efficiency of the He–Ne laser are not high (~ 0.1 W and $\sim 0.1\%$, respectively) but this laser provides continuous operation, high monochromaticity and directivity of radiation, and simplicity of design (a gas-discharge tube with two electrodes and two mirrors). In addition to a helium–neon mixture, lasing has been obtained with several dozen neutral atomic species both in the continuous-wave and pulsed modes (in some cases achieving very high peak power).

In 1964, the first laser working on ionized gases was built, in which inverse population was created with energy levels of atomic ions[26]. All in all, several dozen ion lasers are known, using several hundred working sublevels. Ion lasers are characterized by high current density and high ion concentration, which allows achievement of higher output power than with neutral atom lasers. However, the efficiency of ion lasers is quite low ($\sim 0.1\%$). Efficiencies greater by an order of magnitude ($\sim 1\%$) were obtained on atoms and ions of metal vapors which were found to have a more efficient (collisional) type of depopulation of the lower lasing level than the conventional spontaneous transition. Lasing has been achieved in vapors of several dozen metals with the mean power output of about 40 W (for Cu) and peak power of 200 W.

Molecular lasers that work on the molecular vibrational levels characterized by small energy separations (0.01–0.1 eV), high excitation efficiency, high quantum yield, and good energy selectivity for radiation transitions possess highest power outputs as well as high efficiency (about 25%). One of the examples of molecular gas lasers is the laser working on CO_2 molecules with an admixture of He, and N_2 and H_2O molecules that help to achieve high population inversion by effectively populating the upper laser level (N_2) and depopulating the lower level (He and H_2O). The average power of these lasers is about 1 kW. To increase this figure, a fast-flow version of a molecular laser at gas pressure up to 50 atm is utilized. Such lasers, working continuously, make it possible to achieve the power output of several dozen kilowatts; in pulsed mode, an energy output per pulse of about $\sim 10^4$ J was achieved. Details on high-power molecular lasers can be found in the book by N V Karlov and Yu B Konev [164].

The most exotic among all the listed gas lasers are probably excimer lasers in which lasing is produced by molecular transitions from the upper *bound* but very short-lived ($\sim 10^{-8}$ s) excited state to the lower *unstable* state. The exotic feature of excimer molecules is that they can be formed of two atoms of inert gases (or of an inert gas atom and a halogen atom) only in the excited state because free atoms repel one another and cannot form a molecule in its lowest (ground) energy state. This salient feature of excimer molecules leads to fast depopulation of the lower (unstable) state, facilitating the achievement of inverse level population.

Excimer lasers operate on transitions between electronic states of molecules. The emerging radiation is in the visible or ultraviolet range of the optical spectrum, and is characterized by considerable line width, which facilitates tuning the oscillation frequency. The active medium of excimer lasers is an inert gas at a pressure of about 1 atm with about 1% admixture of halogen-containing molecules. Because of a very short lifetime of the excited states of excimer molecules, high-power electric discharge pulses or a high-intensity electron beam are used for the pumping of excimer lasers. The most efficient oscillation was reported for a laser based on ArF, KrF, and XeF molecules (output energy of about 100 J, efficiency of about 10%, pulse duration $\tau \sim 10^{-8}$ s). Details revealing the mechanism of operation of excimer lasers can be found in the review by A V Eletskiĭ [165].

Also known, in addition to gas lasers pumped by an electric discharge or an electron beam, are lasers with nuclear or chemical pumping, in which the working gas medium is excited by nuclear or chemical reactions. In the case of nuclear pumping, gas atoms (e.g., Ar or Xe) are excited by the products of nuclear reactions produced in interactions of thermal neutrons ($E = kT = 0.025$ eV) with ^{10}B, ^{3}He, or ^{235}U nuclei, producing as a result ^{4}He and ^{7}Li ions, ^{1}H and ^{3}H ions, or fission fragments, respectively. The laser design is a tube filled with gas, with a thin layer of ^{10}B or ^{235}U coating the inner surface of the tube (^{3}He is introduced as an admixture to the working gas). A pulsed, moderated reactor is usually a source of thermal neutrons. Power output on infrared transitions of inert gas atoms is on the order of 10 kW, with an efficiency of about 1%.

This mechanism of pumping is not very different from that of plasma lasers in which the excitation of the lasing medium (a weakly ionized plasma) is also produced by charged particles (an electron beam). An important feature of plasma oscillators of coherent radiation is the possibility of smoothly varying the emission frequency by controlling plasma density. For details see the book by L I Gudzenko and S I Yakovlenko [166].

In the case of chemical pumping (chemical lasers), the excitation of the working medium occurs because products of many exothermal reactions are formed in excited states. For instance, in diatomic molecules this effect manifests itself as excitation of vibrational–rotational levels. The result is the inverse population of these levels that can be used to achieve lasing (if the energy separation between them is of suitable value).

Typical reactions of these types used in chemical lasers are found in substitution reactions

$$F + H_2 \rightarrow HF + H \tag{58}$$

or

$$F + CH_4 \rightarrow HF + CH_3\,, \tag{59}$$

which lead to an efficiency (the ratio of laser emission energy to energy released by the reaction) of about 10%. The energy emitted by the chemical laser using HF in the pulsed mode with a pulse duration of about 10 ns can reach 10 J. The maximum power in continuous operation is achieved when the active lasing medium is circulated through the cavity at supersonic speeds. In this case, the output power reaches

[26] Sometimes these lasers are referred to as plasma lasers, which points to cold weakly ionized plasma [not to be confused with X-ray lasers that work with high-temperature multiply ionized plasmas (see Section 4.4.4)].

several kW, with an efficiency up to 2–4%. For details on chemical lasers, see the review by A V Eletskiĭ [167] and the book edited by N G Basov [168]. Detailed data on various types of such lasers can be found in the already mentioned handbook edited by A M Prokhorov [159].

4.4.4 The X-ray laser (raser) and the problem of the feasibility of creating a gamma laser (gaser). The history of the development of X-ray lasers (rasers[27]) and the principles on which they operate are described in the review by E G Bessonov and A V Vinogradov [169] and in a paper by A V Andreev [170]; we will briefly summarize the contents of the two. (The reader can find a popularized description of new possibilities for materials research that the invention of the X-ray laser has brought out in the review by D Attwood, K Halbach, and K-J Kim [171].)

It is mentioned in Ref. [169] and in Ref. [170] that the idea of building the X-ray laser emerged immediately after the invention of the first optical laser (1960). The main concepts framing the principles of its operation and the ways of implementing them made their appearance in 1970s; the first laboratory pilot model of the X-ray laser was built in the E Lawrence Livermore National Laboratory (USA) in 1985. Lasing was obtained on a series of lines of the Ne-like selenium ion in the wavelength range of 182–263 Å (the brightest line was 206.3 Å).

The active medium of the X-ray laser is the small volume of high-temperature plasma about 0.01–0.1 cm in cross section and several centimeters long. We note that when lasers whose operation is based on transitions between discrete energy levels of atoms and molecules are created, all four aggregate states of matter are utilized: solid, liquid, gaseous (including weakly ionized, namely, cold plasma), and the state of high-temperature multiply ionized plasma.

The high-temperature plasma (at electron energy of 10^2-10^3 eV, which corresponds to the temperature of 10^6-10^7 K) is generated for its short survival time ($10^{-10}-10^{-8}$ s) by focusing the radiation of a high-power optical or infrared laser onto the surface of a target whose atoms are multiply ionized. As plasma cools down, nonequilibrium collisional and recombination processes proceed in it, which lead to the creation of inverse level population of multiply ionized ions (e.g., C^{5+}, Al^{11+}, Se^{24+}). These are the levels on which the laser effect shows itself. The lasing wavelength depends on the material of which the target is made and the type of transition. As a rule, lasing occurs in the single-pass super-radiation regime (recall the superluminescence described above), i.e., without the cavity making its contribution, or in two- or three-pass mode (if multilayer X-ray mirrors are utilized). The laser gain per passage is between 3 and 16, so that the maximum amplification relative to the level of spontaneous emission is $e^{16} \simeq 10^7$. The conversion factor of pumping laser radiation into X-ray output emission is very low ($\sim 10^{-5}$) but is nevertheless sufficient for running physical and biological experiments in which a resolving power better than 10^{-5} cm was obtained on the wavelength $\lambda = 182$ Å with the laser illuminating a carbon target. For further progress towards shorter wavelengths, the pumping laser power has to be raised (by using, for instance, a neodymium glass laser) which increases the multiplicity of ionization and ultimately shortens the laser wavelength ($\lambda = 46$ Å was reported with an aluminium target). The X-ray laser possesses the maximum pulse brightness compared to all other sources of X-ray radiation (see Section 4.5).

About a dozen various theoretical mechanisms were put forward to explain the establishment of inverse population of levels in the active zone of the X-ray laser. Two of them are most realistic: collisional excitation (with high electron density in the plasma) and recombination pumping (in the case of high-temperature plasma with low electron density). Figure 12c gives the transition pattern for the case of achieving inverse population with Se^{24+} ions by collisional pumping. The upper working level 3p is populated from the ground state as a result of collisions of plasma ions with electrons; the lower working level 3s is rapidly depopulated if the radiation transition $3s \rightarrow 2p$ is allowed; the radiation transition $3p \rightarrow 2p$ is forbidden by selection rules, and the laser transition is realized between the upper and lower working levels ($3p \rightarrow 3s$). The current state-of-the-art is well described in the review by P D Gasparyan, F A Starikov, and A N Starostin [172]. The latest achievements in the field can be found in Ref. [258] (see the concluding section).

To conclude this section, several words are in order concerning the problem of creating a gamma laser (gaser or graser[28]) — that is, the possibility of generating coherent electromagnetic radiation in the γ-ray wave band (by using transitions between levels not of atoms, molecules, or ions, but of atomic nuclei).

Researchers started to think about a gamma laser (and X-ray laser as well) immediately after the creation of the first optical laser in 1960. The problem of generating coherent radiation in the γ-ray band seemed to be solvable, in principle, in view of the discovery, made not long before that (1958), of 'recoilless' γ-radiation (the Mössbauer effect)[29]. It was assumed that the problem of creating a γ-laser could be solved by implementing stimulated emission of radiation in a system of excited isomeric nuclei[30]. It is obvious that an

[27] Raser is the acronym for Roentgen Ray Amplification by Stimulated Emission of Radiation.

[28] Graser, or gaser, is an acronym for Gamma Ray Amplification by Stimulated Emission of Radiation.

[29] The Mössbauer effect is that, whereby under certain conditions (sufficiently low transition energy and temperature that is low compared with the Debye temperature), the momentum and energy of recoil that occur in emission, scattering, or absorption of γ quanta are elastically transferred not to one atom but to the entire crystal (or more precisely, to a very large group of atoms on the order of $N \simeq 10^8$), so that (because of the large mass of this group) the recoil energy per one nucleus is practically zero (much less than the natural width Γ of the level). Since low recoil energy implies very small Doppler broadening D ($D < \Gamma$), the Mössbauer effect allows one to measure with relative ease the energy E with a relative accuracy $\Gamma/E \simeq 10^{-12}$. Under special experimental conditions it was even possible to measure the famous red shift of the energy of γ quanta in the gravitational field of the Earth, whose scale (for the 20-m path length) was $\Delta E/E \simeq 2 \times 10^{-15}$ (which is 10^9 times smaller than the familiar solar effect measured by astrophysical methods).

For the discovery of recoilless emission, scattering and absorption of γ quanta, R Mössbauer received the Nobel Prize in Physics 1961. For details of the Mössbauer effect and its applications, see, e.g., Ref. [121].

[30] Nuclear isomers are atomic nuclei with metastable, i.e., sufficiently long-lived excited states (with a lifetime from a small fraction of one second to thousands of years). An isomeric nucleus with a sufficiently long-lived excited state manifests the properties of two nuclei; for example, a β-radioactive isomeric nucleus may emit β particles with two different half-lives. It was owing to this property that nuclear isomerism was discovered in 1921 by the German scientist O Hahn in the naturally radioactive $^{234}_{91}$Pa nucleus and was investigated in detail in 1935 by Russian physicists and chemists headed by I V Kurchatov using an artificial radioactive $^{80}_{35}$Br nucleus (for details, see, e.g., Ref. [2]).

analysis of this problem is very important because a successful solution would lead to new possibilities in conducting structural analysis of matter and for solving a number of problems in nuclear physics. Therefore, many leading physicists have joined the work on the theoretical analysis of the gamma laser problem. In Russia, these were V I Gol'danskiĭ, Yu M Kagan, R V Khokhlov, and others.

Gol'danskiĭ and Kagan considered gamma laser schemes using short-lived isomeric nuclei ($\tau < 10^{-5}$ s), while Khokhlov thought of long-lived nuclei ($\tau \gg 10^{-5}$ s). The energy range of suitable radiation transitions lies within $10\ \text{keV} < \hbar\omega < 150\ \text{keV}$. The lower bound is imposed by sharply intensified competition from the photoeffect when energy is reduced further, and the upper bound comes from the reduced probability of recoilless emission as energy increases above 150 keV. Details of the theoretical models of the gamma laser and the difficulties encountered in analyzing them can be found in the paper by A V Andreev [173].

As far as we are aware, none of these models has ever been implemented even though we cannot know, for obvious reasons, in what state the American developments in the Star Wars project are. The state-of-the-art in this field by 1987 can be found in a report to the American Physical Society [174].

4.5 Free transition lasers

So far we have been describing various types of lasers: solid-state, liquid, gas, and plasma (low-temperature as well as high-temperature X-ray lasers) that have one important common feature: for generating monochromatic and coherent radiation, these lasers use transitions between the levels of a discrete energy spectrum of atoms, molecules, or ions, and also nuclei (when the feasibility of a gamma laser was discussed). The case of the semiconductor laser (to be discussed in Section 5.8) is somewhat different in that it involves transitions not between discrete energy levels but between allowed energy bands of a solid, but these are also discrete. Therefore, the monochromaticity of laser emission in all laser types listed above should not be surprising.

The situation is very different if the source of coherent and monochromatic radiation is a beam of relativistic electrons which, as we mentioned in Section 3, is characterized by continuous spectrum in all the cases discussed there (the Vavilov–Cherenkov, transition, bremsstrahlung, and synchrotron radiations). This is not surprising because the electron beam energy in all these processes changes not by identical discrete amounts but *freely*, i.e., within certain, rather arbitrary, limits. Nevertheless, lasers on such free transitions (sometimes referred to as free–free transitions) can indeed be created. This statement carries no paradox because in Section 3 we were discussing *spontaneous emission* by electrons, which indeed proceeds freely to a certain extent; but now we will address the *stimulated emission* whose freedom is very much restricted.

It appears that the characteristics of all the above-considered processes of spontaneous emission by beam electrons may change sharply if electrons move in an external electromagnetic field that transforms the spontaneous electron emission into a stimulated one. The main feature of the stimulated electron emission is the possibility of exchanging energy between the electron beam and the field. In doing so, electrons can both transfer a part of their energy to the field, undergoing additional deceleration (stimulated emission), and be involved in an opposite process — they can receive energy from the field and accelerate (stimulated absorption).

An electric and/or magnetic field of special configuration, complemented with laser radiation, is typically used nowadays as an external electromagnetic field. The field is modulated onto the electron beam density, which may result under certain conditions in the transformation of continuous-spectrum emission by electrons into monochromatic and coherent, namely, in producing the laser effect *on free transitions*.

The theory of the free transition laser was developed for several versions of such devices (undulator radiation, the Compton laser, the free electron laser, the Cherenkov laser, etc.) and some of them were implemented experimentally. A spectacular advantage of these devices is the possibility of generating coherent radiation in the ultraviolet and soft X-ray ranges, and also of frequency tuning in a broad frequency band. These properties are obviously very alluring since they may solve certain problems in the soft X-ray band more easily than the X-ray laser allows (see above).

Historically, the first paper on stimulated light scattering or, more precisely, the stimulated Compton effect[31], was written in 1927 by E Schrödinger [175] — the famous physics theoretician, one of the creators of quantum mechanics (the Nobel Prize in Physics 1933). In 1933, a similar problem was discussed by two no less famous physicists — P L Kapitza and P A M Dirac [176]. We wrote much about the former in Section 2, and about the latter (who received the Nobel Prize together with E Schrödinger, also for the creation of quantum mechanics) in our earlier paper [3]. It goes without saying that neither the first nor the second paper dealt with the Compton laser because lasers were not invented yet; nevertheless, the ideas advanced in these papers were very close to the Compton laser.

The idea of using the stimulated Compton effect for creating a Compton laser was advanced 'in these very words' by R Pantell et al. [177] 40 years later (!) when lasers were already a reality. The idea of the Compton laser was that a relativistic electron and an amplified wave interact with the high-power pumping wave (e.g., from a microwave oscillator or a CO_2 laser) that propagates towards the electron beam.

Notice that for the geometry described above the spontaneous Compton effect leading to a sharp increase in the frequency of the scattered wave is well known (radiation scattering by a relativistic electron moving towards the incident optical wave). This is the so-called inverse Compton effect which is used to *increase* radiation frequency via scattering (in the conventional Compton effect, the radiation frequency is reduced). As follows from the theory of inverse Compton effect, the energy E of a photon scattered backwards by an electron with energy E_e is related to the energy E_0

[31] The ordinary (spontaneous) Compton effect is a feature of scattering of hard X-ray radiation by electrons of matter. In 1923, A Compton, when studying how the radiation from an X-ray tube was scattered by graphite and other targets, showed that the spectrum of the scattered radiation contained, in addition to the original wavelength λ_0, a shifted line with $\lambda' > \lambda_0$, and that $\Delta\lambda = \lambda' - \lambda_0$ increases with an increasing scattering angle θ but that at a given scattering angle $(\Delta\lambda)_\theta$ is independent of either λ (i.e., $\Delta\lambda$ is appreciable only at short wavelengths) or the material of the scatterer. Compton interpreted this phenomenon using quantum theory (as scattering of a photon by a free electron). For this discovery, A Compton was awarded the Nobel Prize in Physics 1927.

of the incident photon by the expression

$$E = \frac{E_0}{(m_e c^2/2E_e)^2 + E_0/E_e}\,, \tag{60}$$

which gives for $E_0 \ll E_e$:

$$E \simeq 4E_0\left(\frac{E_e}{m_e c^2}\right)^2 = 4E_0\gamma_e^2\,, \tag{61}$$

where γ_e is the electron Lorentz factor. We see from formula (61) that for high electron Lorentz factors we have $E \gg E_0$, so that the inverse Compton effect is employed to generate X-rays and γ-emission when laser light is scattered by a beam of oppositely moving relativistic electrons. Note also that the hypothesis of the existence of the inverse Compton effect in nature is used by astrophysicists to explain the isotropic X-ray and γ-backgrounds observed in the universe; perhaps their origin is connected with the inverse Compton scattering of thermal background radiation ($T = 2.7$ K) by high-energy cosmic electrons. More details on inverse Compton scattering of blackbody radiation can be found in the paper by D Fargion and A Salis [178].

Pantell's stimulated inverse Compton effect differs from the spontaneous Compton effect in the additional amplified wave that moves parallel to the electron beam and whose frequency coincides with that of the light wave reflected in the inverse Compton scattering. The theory of the Compton laser was considered, for example, in a paper by D F Zaretskiĭ, É A Nersesov, and M V Fedorov [179]. As far as we are aware, the Compton laser has not been implemented experimentally yet. (For details, see the reviews by V L Kuznetsov [180] and M V Fedorov [181].)

The history of the emergence of another laser utilizing free–free transitions, which is called just that — the free electron laser (FEL) — also goes back some years but, in contrast to the Compton laser, working pilot devices have already been produced.

At present, several detailed theoretical FEL schemes have been developed. We will describe one of them; its idea lies in using the stimulated undulator emission of relativistic electrons. The history of this effort goes back to 1947 when V L Ginzburg [182] and later on H Motz [183] treated the spontaneous undulator emission of electrons in a periodic magnetic field, constant in time but variable in space, and showed theoretically that undulator emission must possess essential advantages compared to synchrotron radiation: monochromaticity in a given direction, and higher spectral radiation flux density[32]. In 1951–1953, Motz first built sources for spontaneous incoherent and spontaneous coherent undulator radiation and showed that the properties of the emitted radiation agree with the theoretical expectations for both. The idea of a source of stimulated undulator emission was advanced and substantiated in 1958–1959 by H Motz, R Pantell (USA), A V Gaponov-Grekhov (USSR), and some others. The first sources of stimulated undulator radiation were designed and studied at the wavelength of $\lambda \simeq 10$ cm in 1960 by the American physicist R M Phillips. A review by D F Alferov, Yu A Bashmakov, and E G Bessonov [186] discusses further theoretical and experimental studies of the undulator radiation.

New possibilities were created for increasing the frequency of undulator emission when specialized storage electron rings were put in operation with undulators inserted into their rectilinear segments. The electron energy on these storage rings is from several hundred MeV to several GeV, which promises generation of undulator radiation at a wavelength down to 10^{-9} cm, which roughly corresponds to the energy of hard X-ray quanta (~ 100 keV). This wavelength is suitable for studying biological objects and artificial microstructures in microelectronics, and much more. For details see the above-mentioned reviews by Attwood, Halbach, and K-J Kim [171], Bessonov and Vinogradov [169], and P Sprangle and T Coffey [187].

The stimulated undulator emission (or absorption) arises if an external electromagnetic wave propagates along the undulator axis towards the electron beam. The corresponding setups for the nonrelativistic energy range are known as ubitrons; ubitrons are powerful sources of radiation in the centimetric and millimetric wave bands. Oscillation frequency rises considerably ($\sim \gamma^2$) when passing to relativistic electron energies. It seems that J Madey [188] was the first to suggest in 1971 the idea of building a free electron laser (FEL) using an undulator; in 1976–1977, this suggestion was implemented under his guidance at the Stanford linear accelerator [189].

A schematic of the first FEL is shown in Fig. 13. The beam electron energy was 43 MeV. The electron beam and the wave experiencing amplification were fed into the setup as short (~ 1 mm) trains each consisting of an electron bunch and an electromagnetic pulse approaching simultaneously the entrance into the undulator. As both trains pass through the undulator, their interaction results in amplification of the electromagnetic pulse which is retained within the resonator by two mirrors until another pair of trains arrives. Both components of the pair (the electron bunch and the electromagnetic pulse) again approach the entrance of the undulator simultaneously, and the amplification cycle is repeated, and so forth.

Here are some of the parameters of the first free electron laser: electron beam diameter $\simeq 0.3$ cm, current 2.6 A, current density $N_e = 5 \times 10^{10}$ cm^{-3}, pitch of the helical magnet $l = 3.2$ cm, FEL length $L = 5$ m, magnetic field intensity $B = 2.4 \times 10^3$ G, and oscillation frequency $\omega = 5.5 \times 10^{14}$ s^{-1} ($\lambda = 3.4$ μm); the maximum radiation power outside the resonator was 7 kW.

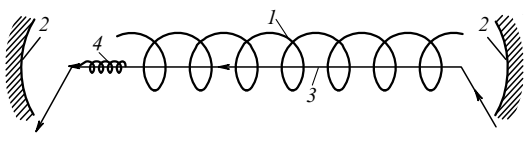

Figure 13. Schematic of the first free electron laser: *1* — helical magnet, *2* — mirrors, *3* — electron beam, *4* — wave being amplified.

[32] Undulator (from French 'onde' — wave) is a device in which periodic (in space or time) fields are produced and they act on charged particles (e.g., electrons) traveling through it. As a result, a charged particle moving through the undulator performs a periodic translatory–oscillatory motion — that is, becomes a moving oscillator emitting electromagnetic radiation referred to as undulator emission.

The simplest undulator is a periodic sequence of sign-changing magnets between whose poles a particle moves (in a parallel plane), being alternately deflected in opposite directions (e.g., along a sinusoidal trajectory), i.e., undergoing acceleration and hence emitting quanta (compare with the synchrotron radiation mentioned in Section 3.4). Bessonov's papers [184, 185] provide additional details on undulators and undulator radiation.

A considerably higher output radiation power (~ 1 MW) was obtained already the next year at a wavelength 0.5 mm [190]. In this experiment, an electron beam of lower energy (1.2 MeV) was used but the current (25 kA) was much higher. The pitch of the periodic field in this case was 8 mm.

As we have mentioned once, storage rings with undulators placed in their linear segments were put into operation in the 1990s as electron sources, and new possibilities arose for generating higher-frequency spontaneous undulator radiation. Obviously, this inevitably led to improvements in FELs.

Indeed, we learned fairly recently (in March 2002) from a publication [191] that the most powerful frequency-tuned free electron laser in the world was developed at the Thomas Jefferson National Accelerator Facility (USA). The principle of operation of the FEL is to pass electrons accelerated to 42 MeV through a wiggler creating a periodic magnetic field and exciting their emission, which is then amplified in an optical cavity with mirrors whose reflection coefficient is 100% and 90%. The initial radiation power of the FEL (which started operations in 1999) reached 1.72 kW (at the wavelength $\lambda = 3.1$ μm), which already allowed researchers to carry out fundamental and applied investigations in chemistry, biology, nanotechnology, and materials science. After planned modernization of the laser facility, the output radiation power of FEL will grow to 10 kW in the infrared range, and to 1 kW in the visible and ultraviolet spectral ranges (down to $\lambda = 0.3$ μm). The same communication [191] describes the Argonne *mirrorless single-pass FEL* which made it possible to generate laser radiation at considerably shorter wavelengths (including the ultraviolet region at $\lambda = 385$ nm) than was possible earlier on mirrorless free electron lasers. This result can be regarded as an essential step forward to creating an X-ray free electron laser with exceptionally high radiance in the ultrashort pulse mode, which will allow, for instance, the monitoring of chemical and biological reactions and other very fast processes by using a sequence of instantaneous snapshots (a sort of 'filming'). For additional information on undulator radiation, the Compton laser, and FEL we recommend turning to the already mentioned reviews [169, 171, 180, 181, 186, 187] and also to a book by T Marshall [192].

To conclude this section, we will briefly outline the stimulated Cherenkov effect (SCE) whose theory is discussed in the review by V M Arutyunyan and S G Oganesyan [193] (see also Ref. [194]). If the project is successful, it may lead to developing a Cherenkov laser. The main feature of SCE showing itself in a field of an external electromagnetic wave is (as in all the other above cases of free – free transitions) the energy exchange between a moving particle and an external field. As this takes place, the particle can either transfer its energy to the wave and be decelerated (stimulated Cherenkov emission) or, conversely, receive energy from the field and accelerate (stimulated Cherenkov absorption). Both these processes typically occur at the same time; it is not yet possible to separate them completely.

The authors of Ref. [193] conducted their theoretical analysis of the SCE in terms of both classical and quantum approaches. In the quantum theory of SCE, they were able to analyze separately the stimulated emission and the stimulated absorption. This allowed them to derive conditions under which the emission processes dominate absorption, and to develop a theory of the electromagnetic wave amplification at the expense of the electron kinetic energy undergoing transformation in the SCE — that is, they were able to consider the possibility of building a Cherenkov klystron and a Cherenkov laser. In their Conclusions, the authors of Ref. [193] discussed the feasibility of experimental examination of the SCE and the difficulties encountered on the way.

4.6 Laser applications

Applications of laser-oscillators and laser-amplifiers are so widespread and diverse that it would be impossible to write about this in any detail without expanding the size of this article. Therefore, we only describe certain fields of laser application, albeit very briefly, while signposting other applications only by headings and, where possible, indicating the accessible literature.

4.6.1 Laser (thermonuclear) fusion. Laser fusion — one of the proposed methods of implementing inertial controlled thermonuclear fusion — was proposed in 1962 by N G Basov and O N Krokhin (in a report to the Academy of Sciences of the USSR, published as an article in 1964 [195]). Some physicists remember that a similar proposal was also made at security-restricted Arzamas-16 (and probably somewhat earlier) by A D Sakharov. The principle of this method is the superfast ($\sim 10^{-9}$ s) pulsed heating of thermonuclear fuel, condensed and uniformly compressed to a density of $10^2 - 10^3$ g cm^{-3}, to a temperature of $10^7 - 10^8$ K (fuel prepared as a superminiaturized spherical target of volume $\sim 10^{-6}$ cm^3, made of a mixture of ^{2}H and ^{3}H as ice or gas under high pressure and coated with a multilayer shell). Basically, it is possible to try to implement all the above-listed conditions using a powerful multichannel (hundreds of beams) 4π-laser facility with an energy per pulse of 1 MJ or higher. If this is achieved then the thermonuclear fuel which is evaporating under heating does not have time to fly away before fusion starts. Hence the name of the method: inertial fusion. Laboratories in many countries have tackled the problem of laser fusion and promising results have been reported in all parameters, including a considerable yield of thermonuclear neutrons generated in the reaction

$$^2\mathrm{H} + {}^3\mathrm{H} \rightarrow \mathrm{n} + {}^4\mathrm{He}\,. \qquad (62)$$

However, the required amplification coefficient (the ratio of the energy released in the fusion reaction to the laser energy) which has to be on the order of $10^2 - 10^3$ has not yet been achieved. For details, see the earlier journal publication by N G Basov, V B Rozanov, and N M Sobolevskiĭ [196], and a later paper by N G Basov et al. in the book [197]. In the concluding section, we will outline another recently reported project covering inertial fusion.

4.6.2 Laser isotope separation. The laser isotope separation method is based on using the isotope shift of spectral absorption lines of electromagnetic radiation. When the wavelength λ of the incident light coincides with an absorption line of one of the isotopes in the mixture, its atoms are raised to a strongly excited state (up to ionization). After this they can be separated from nonexcited atoms by photochemical or physical techniques (e.g., by deflecting the resulting ions in an electric field).

In addition to selective excitation of electronic levels in atoms, the laser method of isotope separation also uses the excitation of molecular vibrational levels. As radiation sources, CO_2 lasers or dye lasers are employed (to excite atoms of vaporized metal). For details, see the reviews by

N V Karlov and A M Prokhorov [198], N G Basov et al. [199], N V Karlov et al. [200], and V S Letokhov [201].

4.6.3 Laser chemistry. It appears that the laser isotope separation described above can be regarded as one of the lines of inquiry of laser chemistry. The other two are connected with the chemical transformation of materials via photochemical and thermochemical processes. A process is known as photochemical if chemical reactions in this process are initiated by a resonance excitation of electronic levels in atoms and molecules by a laser electromagnetic radiation. The thermochemical branch of laser chemistry is concerned with the chemical transformation of materials by the purely thermal action of laser radiation. More details of the photochemical field of laser chemistry can be found in the book by I M Dunskaya [202], and of the thermochemical branch in the review by F V Bunkin, N A Kirichenko, and B S Luk'yanchuk [203].

4.6.4 Laser spectroscopy. Optical spectroscopy in its conventional interpretation is the field of physics in which energy levels of atoms and molecules and quantum transitions between them are found from the experimental data on the intensity of electromagnetic emission and absorption as a function of frequency; in the long run, this makes it possible to obtain the macroscopic characteristics (density, temperature, and velocity of motion) of various objects (including astrophysical ones) residing in various aggregate states of matter (solid, liquid, gaseous, and plasma). The most widespread instruments of conventional optical spectroscopy that uses continuous-spectrum light sources simultaneously exciting a large number of spectral lines are various types of monochromators: prisms, interferometers, spectrophotometers, and other dispersing devices. These instruments allow the researcher to single out the desired region of a spectrum.

The resolving power and sensitivity of optical spectroscopy grew abruptly after the advent of laser radiation sources with their uniquely high intensity, monochromaticity and directivity of outgoing beams. The monochromaticity of laser radiation makes it possible to activate a specific quantum transition, thus making unnecessary the above-listed dispersing instruments operating in conventional optical spectroscopy. The resulting shape of the investigated spectral lines is the true one (not distorted by the instrument). With the arrival of laser sources of radiation in optical spectroscopy, it became possible to study microscopic amounts of materials in small volumes ($\sim 10^{-10}$ cm^3) and remote macroscopic objects. For example, the distance to remotely studied astrophysical objects by the method of absorption of the laser signal sent to the object was considerably increased. Laser spectroscopy uses frequency-tuned laser monochromators with the frequency recorded by a special device (λ-meter), which allows quantitative spectroscopic studies in the ultraviolet, visible, and infrared parts of the spectrum. With new achievements in the technology of generating ultrashort nano-, pico-, and then femto- and even attosecond-pulses (1 as $= 10^{-18}$ s), the temporal resolving power of laser spectroscopy dramatically increased (see earlier descriptions and the concluding section). As a result, very fast relaxation processes in condensed media can now be studied; for example, it is possible to measure lifetimes of excited states and decay channels, to investigate internal motion within molecules and elementary stages of chemical reactions, and many other phenomena. Moreover, owing to the unique sensitivity of laser spectroscopy, the possibility opened of detecting single atoms and nuclei. The reader will find additional information on this topic in two reviews [204, 205] and in abstracts of V S Letokhov's report [206], in the review by V S Antonov, V S Letokhov, and A N Shibanov [207], in the book by S A Akhmanov, V A Vysloukh, and A S Chirkin [208], and in a paper by V F Kamalov [209].

4.6.5 Laser technology. The intense thermal effect of the laser beam, with its short duration and localization of the affected area, makes lasers extremely convenient for a wide range of methods of technological processing of materials and finished products (cutting, drilling, engraving, welding, hardening, annealing, and so on); the results are often impossible to achieve with any other techniques. For instance, laser cutting can produce cuts 5 – 30 μm wide, while laser drilling enables the processing of ruby bearings for watches and diamond dies for wire drawing, parts made of superhard alloys and ceramics, and so forth. Laser welding has a unique feature: welding can be carried out inside a closed vessel (through a window transparent for laser radiation). The thermochemical effect of laser radiation is used for profiling surfaces of materials used in microelectronics. (See Conclusions for news on novel applications.)

4.6.6 Laser physics of ultrahigh pressures. High-power pulsed lasers with energy per pulse of 10^4 J, designed at the beginning of the 1990s to study controlled thermonuclear fusion, can be applied to producing and studying plasmas with extremely high pressure (hundreds of megabar) and temperature ($\sim 10^6$ K). Some of the experiments in this field and the theory of the effects are outlined in the review by S I Anisimov, A M Prokhorov, and V E Fortov [210].

4.6.7 Laser cooling of atoms. This heading may sound paradoxical. Indeed, welding, hot plasma, in other words, any laser heating — this is quite clear, but cooling is strange! However, as was shown by Ya B Zel'dovich [211] in 1974, there is no contradiction here. The law of conservation of energy allows a material interacting with radiation to be cooled, provided the radiation is 'heated' to the same extent (i.e., provided the frequency of radiation increases as a result of scattering[33]). The following simple arguments illustrate how this can be implemented. Assume that we irradiate a material with the laser radiation whose frequency is slightly lower than the resonance frequency of absorption by atoms of this material. Obviously, in this case resonant absorption is possible only for atoms that move towards the laser beam at a velocity at which an atom, owing to the Doppler effect, responds to the laser frequency coinciding with its resonance frequency. Because the atom receives momentum from the photon absorbed, it is slightly decelerated ('cooled'). At the same time, the quantum of radiation emitted then by the excited atom will have the absorption resonance frequency of

[33] There is no conflict here with the second law of thermodynamics because entropy of a closed system (material plus radiation) still increases. This occurs because the increase in the radiation's entropy during the scattering appreciably exceeds the decrease in the material's entropy as it cools down [mostly because of the sharp rise in the solid angle (up to 4π) for the isotropically scattered radiation]. Steven Chu, Claude Cohen-Tannoudji, and William Phillips received the Nobel Prize in Physics 1997 for developing the methods for cooling and trapping atoms using laser radiation.

the given material, and it is higher than the laser frequency; as a result of interaction with the material, the radiation 'heats up'.

Obviously, since it is unknown in what direction an atom will move after its deexcitation (photons are emitted isotropically and therefore atoms can receive recoil momenta in arbitrary directions), it is not possible to re-apply the cooling operation to the same atom using a single laser. However, this can be performed multiply if, instead of a unilateral irradiation, laser illumination of atoms from all sides is arranged. It was approximately in this way, together with additional cooling by evaporation, that a temperature below 10^{-6} K was obtained, which was necessary to study the Bose–Einstein condensation of rarefied gases of alkali metals, discussed in Section 2.2.

For more details on the above-described effect and on the methods of laser cooling of atomic beams, see the review by V I Balykin, V S Letokhov, and V G Minogin [212].

4.6.8 Lasers and fiber optics. A review by E M Dianov and A M Prokhorov [213] with the same title as that of this subsection describes the success achieved by combining these two spectacular inventions 25 years after the first laser was built. Of course, now that 40 years have passed, we have a very good idea of the possibilities offered by fiber optics communications — suffice it to refer to the example of cable television. However, for those who wish to learn the history of the birth and subsequent evolution of fiber optics, including the technology of manufacturing glass fibers, the review is still quite topical. It demonstrates very clearly the equal necessity for progress in fiber optics communications both of lasers as sources of monochromatic radiation in the optical range and of fibers that provide transmission of the optical signal without losses (which are inevitable in the free atmosphere). Furthermore, the article especially emphasizes the importance for further progress in fiber optics communications of the semiconductor laser created by Zh I Alferov's group for continuous-wave operation at room temperature [214].

4.6.9 Fields of application of semiconductor lasers. Owing to their advantages — some already mentioned and some not, such as high efficiency, low-voltage power supply, rapid response, and simplicity of design — semiconductor lasers find application not only in fiber optics communications but also in a number of other regions among which we find the spectroscopic diagnostics of various media and materials, metrological calibration, high-resolution spectroscopy, optical memory systems, video disk technology, laser printing, etc. For details, see the review by N G Basov, P G Eliseev, and Yu M Popov [215]; on the principles of operation and design of the semiconductor lasers, see Section 5.8.

4.6.10 Fields of application of the X-ray laser and FEL. This subsection is a case in which giving the heading suffices because very much has already been written about applications of these lasers in the appropriate sections, where literature was also cited.

This completes our list of applications of lasers to science, technology, domestic devices, etc., even though we also planned to write about laser medicine, the laser microprojector, the laser gyroscope, laser frequency standards, and so on and so forth. However, the saying goes that no one can embrace the boundless. We also complete our story of quantum electronics as a field, again leaving out a great deal without even listing the topics, in the hope that the literature cited in the subsections will help the interested reader fill in the gaps.

5. Semiconductor physics

The 20th century is often called the atomic and space century, which is probably true since the most striking achievements of that century are related to these areas of science and technology. However, as the century drew to its close, an ever increasing role came to be played by an area of physics that at first played a very minor role but whose advances are used today not only in science and technology (including the nuclear and space industries) but also in everyday life, giving life its modern civilized hallmark and making our lives more comfortable.

We are speaking of semiconductor physics, of course. In 2002 three scientists, Jack S Kilby (USA), Zhores I Alferov (Russia), and Herbert Kroemer (USA), were awarded the Nobel Prize in physics for their outstanding discoveries and achievements in this field of science [216–218]. The achievements are certainly enormous and, notwithstanding the smaller scale, are as outstanding as those of nuclear power engineering and space technology. In the life of an ordinary person they play a greater role, although many people do not realize that semiconductor physics is behind all these achievements. Let us list some of them.

In engineering, these are high-current rectifiers and voltage stabilizers, solar batteries for space stations and communication satellites, and devices for measuring temperature, pressure, radioactivity, magnetic field strength, and other parameters in nuclear power plants and particle accelerators; in physics, these are integrated circuits, which contain millions of transistors and thyristors in a very small chip; in instrumentation, these are photodiodes and photoelectric elements, the semiconductor laser and the Peltier cooler, detectors of elementary particles and piezoelectric transducers, thermistors, and magnetoresistors. Finally, in household appliances, these are the well-known radio and TV sets, in which semiconductor devices long ago replaced electron tubes; computers with laser printers; cell phones, cable TV, audio CD and DVD players, effective lighting in pop music performances, calculators with solar batteries; and scales that show the price, weight, and cost of the produce in the store, etc.

In this section, after a brief historical survey (Section 5.1), we attempt to tell as simply as possible (for nonspecialists in this field of science) the story about just what a semiconductor is (Section 5.2), the elementary model concepts (Section 5.3), the parameters of specific semiconductors (Section 5.4), the electron–hole junction (Section 5.5), and the heterojunction (Section 5.6). We also discuss the design principles and the areas of application of some semiconductor devices (Section 5.7), and go into greater detail about integrated circuits (Section 5.8) and semiconductor lasers (Section 5.9). Finally, in Section 5.10 we return to our initial idea of discussing the role that Alferov and his colleagues played in creating all this semiconductor exotica and in developing the new quantum physics, for which semiconductor physics was such good 'soil.'

But first we would like to briefly tell the story of the Russian predecessors of Alferov, scientists who studied the physics of semiconductors before him and then together with him.

5.1 Scientific schools of A F Ioffe and B M Vul

The pioneer in this area of physical science in Russia was of course A F Ioffe, who in his time (just as Alferov in his) headed the famous Leningrad (nowadays St. Petersburg) Physicotechnical Institute (later named after him) and founded a well-known scientific school of physicists and chemists, who later also became well-known and founded their own scientific schools. Here are some of these scientists. Physicists: the Nobel Prize winner in physics P L Kapitza; the outstanding atomic physicists I V Kurchatov, A P Aleksandrov, A I Alikhanov, I K Kikoin, and Yu B Khariton; the equally well-known L A Artsimovich (electromagnetic isotope separation and, later, fusion plasma), A I Leĭpunskiĭ (nuclear physics and reactor physics), K D Sinel'nikov (nuclear physics and accelerator physics), V P Zhuze (semiconductor physics), and Ya I Frenkel (a universal theorist — from solid state physics to nuclear physics). Chemists: the Nobel Prize winner in chemistry N N Semenov (chain reactions) and B V Kurchatov (semiconductor physics, nuclear isomerism, the first Russian plutonium).

Ioffe began his research into semiconductor physics in the early 1930s, and in 1931 he analyzed (with Ya I Frenkel) the tunnel effect in semiconductors. Other work of Ioffe was also important, in which he discovered the effect of impurities on the electrical properties of semiconductors (e.g., on conduction and the sign of charge carriers). This finally enabled him to plan a method of fabricating semiconducting materials with variable properties. By the late 1930s, Ioffe had formulated the mechanism of current rectification via semiconductors, and his ideas are still valid. Other important work was carried out during the same period by V P Zhuze and B V Kurchatov, pupils of Ioffe. In it they studied the intrinsic and impurity conduction of semiconductors.

The second area of research in semiconductor physics emerged in the same 1930s in B M Vul's laboratory at the P N Lebedev Physics Institute (FIAN) of the USSR Academy of Sciences, where the physics of insulators was the main topic of research. Vul began his intensive research in semiconductor physics in 1948, when the first semiconductor diodes, transistors, and solar cell elements were being developed under his leadership. In 1959, N G Basov, Yu M Popov, and Vul published a paper [141] mentioned in Section 4.1, devoted to study of lasing in the optical range via semiconductors. In 1961 – 1965, Basov, Vul, O N Krokhin, V S Bagaev, O V Bogdankevich, A Z Grasyuk, and others proposed and developed various types of semiconductor lasers [143 – 145, 219]. But the first practical results were achieved not at FIAN but at the Physicotechnical Institute in 1962, where D N Nasledov, A A Rogachev, S M Ryvkin, and B V Tsarenkov observed stimulated emission of light from a gallium arsenide (GaAs) diode [220][34]. However, it was the American physicist R N Hall and coworkers who built the first semiconductor laser (in the same year and based on the same materials) [221]. In Russia, semiconductor lasers were built a year later at FIAN, and Vul actively participated in this work [144].

But let us return to Leningrad, where in 1950 the young student Alferov was beginning his research in semiconductor physics. In 1953 he is already a scientific worker at the Physicotechnical Institute in V M Tuchkevich's laboratory (in 1967 Tuchkevich became Director of the Institute) and is actively participating in the fabrication of the first Russian semiconductor devices. In 1963 he defends his thesis on semiconductor rectifiers and becomes a Ph.D. The double heterostructure was invented by Alferov in 1963, while in 1968 he builds, together with his colleagues, the first heterolaser [214]. We discuss this new advance in semiconductor physics in greater detail in Section 5.10. And so, after this brief historical survey of Russian semiconductor physics, we proceed with our outline.

[34] In 1964, Vul, Krokhin, Popov, A N Shotov, Nasledov, Ryvkin, Rogachev, and B V Starenkov were awarded the Lenin Prize for their fundamental studies that led to the creation of semiconductor lasers.

5.2 What are semiconductors?

Everybody knows what conductors and insulators (dielectrics) are. The first have high conductivity, i.e., a low but finite (at any temperature) resistance to electric current (we do not discuss superconductors in this section; see Sections 2.3 and 2.4). Solid and liquid (mercury) metals, electrolytes, and plasma are all conductors. The second do not conduct electrical current (they have very high ohmic resistance). Some solids (not metals), glass, resins, some liquids, e.g., water (distilled water, but not sea water), and gases (not ionized) belong to this group of substances. The difference in electrical conductivity of conductors and insulators is enormous. In this parameter they may differ by a factor of $10^{12} - 10^{23}$.

There are also two groups of substances that occupy an intermediate position between conductors and insulators. These are semimetals and semiconductors. The first are closer to metals in their conductivity. Just as metals, they are conductors even at $T = 0$ K, but, in contrast to metals, which have a high and weakly temperature-dependent conductivity (which decreases with increasing T), their conductivity is moderate and rapidly increases with temperature. Semiconductors are closer in conductivity to insulators; i.e., at $T = 0$ K their conductivity is zero, but as T increases the conductivity becomes finite and increases exponentially. A semiconductor may also become a conductor when there is some external agent acting on it, such as an electric field, light, or pressure, or when impurities are injected into it. Below we will discuss in greater detail only semiconductors, but first we will briefly characterize all four groups of substances mentioned above and will try to understand why they differ so much in their electrical conductivity, although on the whole they are neutral, i.e., have an equal number of positive and negative charges that ensure interatomic interaction (oppositely charged ions in the case of the ionic bond, electrons and positively charged metal atoms in the case of the metallic bond, etc.).

5.3 Model concepts

The answer to the above question in classical physics is as follows. Let us begin with metals and insulators. The number of positive and negative charges in both is indeed the same. But in insulators all these charges (including electrons) are bound, so that an electric field can only slightly shift these charges but not free them. On the other hand, in conductors (metals), even at $T = 0$ K there is a large number of free mobile electrons (conduction electrons), which are usually valence electrons. The number of free electrons in semimetals is very small, but still there are enough of them to ensure small electrical conductivity even at absolute zero. In semiconductors, all electrons at $T = 0$ K are bound, but this binding can be disrupted by heating or other external effects, with the

results that the conductivity of semiconductors can be controlled.

Of course, this classical explanation is not very persuasive, since it does not answer the question of why in some substances the electrons are bound while in other substances they are partially free.

A more meaningful explanation is provided by the quantum theory of the energy spectrum of electrons in solids, what is known as the band theory of solids, which was developed by F Bloch in 1928[35] and L M Brillouin in 1930. According to this theory, due to the proximity of atoms in a crystal, the wave functions of electrons belonging to neighboring atoms overlap and the discrete levels, a characteristic feature of free atoms, merge into energy bands. Here is a very elementary picture of the band theory.

A crystal with N atoms is considered a single quantum mechanical system with NZ collectivized electrons (Z is the nuclear charge) described in what is known as the one-electron approximation. According to this description, an electron (more precisely, a quasiparticle with a mass m^*, which may differ substantially from the electron mass m_e ($10^{-2}m_e < m^* < 10^2 m_e$) moves in a periodic potential generated by the nuclei and the other electrons. Its energy, in contrast to the energy of a free electron (which is $E = p^2/2m_e$, where p is the electron momentum), is expressed by a multivalued function $E(k)$, where k is the wave vector. Each branch of the function $E(k)$ corresponds to a band of allowed energies with a different function $E_l(k)$, where l is the order number of the band. The electron energy spectrum is obtained by solving the time-independent Schrödinger equation for a wave function that satisfies the periodicity condition (determined by the lattice parameter of the crystal) and certain boundary conditions.

In contrast to the discrete electron spectrum of a free atom (consisting of narrow levels of width Γ with big gaps between the levels, $\Delta E \gg \Gamma$), the spectrum of electrons in a crystal consists of alternating bands of allowed and forbidden energies. Each level of the atomic spectrum (*of all the atoms in the crystal*) is transformed into a corresponding allowed band, while the gap between this level and the adjoining level is transformed into an adjoining forbidden band, often called the energy gap. Since the crystal has N atoms, to each level of a free atom there correspond $2N$ levels in the allowed band that this level produced (the factor 2 accounts for the two values of the projection of electron spin). The same is true for all allowed bands: each has a discrete structure (quasi-discrete, to be exact). The filling of the bands by electrons proceeds according to Pauli's exclusion principle; i.e., the first levels to be occupied are those in the lowest band (the most tightly bound electrons of the atoms of the crystal), and then the remaining allowed bands are filled one after another (the forbidden bands remain unoccupied). The process continues up to the value $E(k) = E_F$, where E_F is the Fermi energy determining the upper limit of filling of the energy levels at absolute zero. In a real solid, bands are produced only for the least tightly bound electrons in an atom (e.g., valence electrons), while for the most tightly bound electrons (those of the ion core), whose wave functions do not overlap, the structure of the energy state remains practically the same as for a free atom (Fig. 14a).

[35] In 1952, Bloch and E Purcell were awarded the Nobel prize in physics for their discovery of nuclear magnetic resonance.

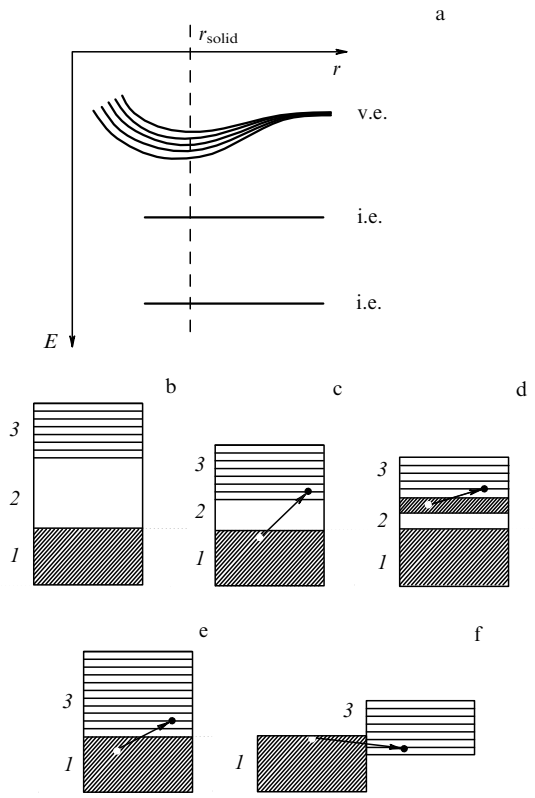

Figure 14. Structure of allowed and forbidden bands for different substances: (a) formation of an allowed band: E is the level energy, r is the distance between the atoms, r_{solid} is the average distance between atoms in the solid, v.e. is a valence electron, and i.e. is an inner electron; (b) insulator: *1*, valence band; *2*, forbidden band; and *3*, conduction band; (c) semiconductor (the notations are the same); (d) and (e) metal; and (f) semimetal.

The physical properties of a crystal, e.g., its electrical conductivity, are determined by the way in which the two upper allowed bands, the valence band and the conduction band, are filled, the distance between these bands (the gap width in energy units), and the arrangement of these bands relative to the Fermi level. For instance, the case of an insulator (Fig. 14b) corresponds to a system consisting of several filled allowed bands, with the last (valence) band being completely filled and separated from the next allowed band (not occupied by electrons) by a fairly wide (~ 7 eV) gap (forbidden band) that electrons cannot overcome even under external forces (provided that these forces are not too strong; in a very strong electric field there may be a breakdown of the insulator accompanied by the formation of a conducting channel, or a discharge filament).

A semiconductor has a similar band structure, but, in contrast to an insulator, the width of the forbidden band is much smaller ($\leqslant 3$eV; see Fig. 14c). An electron can overcome such a band in the presence of the external forces mentioned earlier, including the injection of impurities into the crystal, which results in the appearance of new (allowed) states in the forbidden band (see below).

In metals two means of filling allowed bands are realized. In the first case, the valence band becomes completely filled while the conduction band is filled only partially, so that unoccupied levels are left in it. Conduction is ensured by the transition of electrons to these unoccupied levels (Fig. 14d).

In the second case, there are enough electrons to fill only the valence band, while the conduction band remains unoccupied, but the two bands merge (there is no gap between them; see Fig. 14e). Here, the electrons have the opportunity to go over to the free levels, thus ensuring conduction.

A characteristic feature of semimetals is the partial overlapping of the upper two allowed bands (Fig. 14f), which explains the moderate conductivity even at $T = 0$ K (and this conductivity increases with temperature). The conduction of semimetals is explained by the fact that in the case of overlapping bands it may be energetically preferable for an electron not to complete the filling of the lower band but instead to go to the upper band, where there are many unoccupied levels, which ensures conductivity. Several words should be said about amorphous substances, which have no crystal structure (e.g., liquids). Obviously, there can be no periodic potential in them, but because of the proximity of the neighboring atoms there is also an overlap of the wave functions of the weakly bound electrons, so that instead of the discrete levels of separate atoms a pattern of allowed and forbidden bands emerges for the sample as a whole (see Section 2.2).

5.4 Properties of some semiconductors

The elements silicon (Si) and germanium (Ge) of Group IV of the Periodic Table, and selenium (Se) of Group VI, are typical homogeneous semiconductors, and their forbidden bands at room temperature (~ 300 K) are 1.1, 0.65, and 1.89 eV, respectively. Homogeneous, absolutely pure (without impurities) semiconductors not conducting electric current at $T = 0$ K acquire *intrinsic* conduction as the temperature increases, and also as other external factors come into play. Intrinsic conduction amounts to some of the electrons from the occupied valence band becoming free and then going over to the conduction band and becoming the carriers of electron (n-type) conduction. Simultaneously with electrons from the valence band becoming free, there appear vacant places on the levels of the valence band that are not occupied by electrons, and these vacant places are known as *holes*. Since the filled state of the valence band of a semiconductor is the normal state, a hole rapidly captures an electron, but the new vacant place is also a hole, etc. Thus, just like a conduction electron, a hole moves inside the semiconductor, thus being the carrier of hole (p-type) conduction. Obviously, when there is intrinsic conduction, the number of free electrons and holes is the same.

If the semiconductor is not absolutely pure and homogeneous (contains impurities), additional (impurity) energy levels appear in its forbidden band, whose levels are very close either to the valence band (acceptor levels) or to the conduction band (donor levels).

For instance, donor levels in tetravalent silicon appear when pentavalent elements such as phosphorus (P), arsenic (As), and antimony (Sb) are added to it as impurities. Introduced into the crystal lattice of silicon with four valence electrons at each atom, these impurities fairly easily transfer their 'redundant' (for this lattice) fifth valence electron from the donor level to the closest conduction band, with the result that electron (n-type) conduction emerges. Obviously, there is no place for hole formation in this mechanism.

For acceptor levels to appear in silicon, trivalent elements such as boron (B), aluminium (Al), or indium (In) must be added as impurities. These elements, injected into the silicon lattice, take from the near valence band of the semiconductor the fourth valence electrons needed in the new state of the element, so that holes form in the valence band of the semiconductor, which create hole conduction.

Controlling the type and number of impurities (doping the semiconductor), we can easily create a semiconductor with the desired properties. Other elements, in addition to those mentioned above, are used as dopants and in different combinations, depending on the desired properties. Semiconductors are very sensitive to the type and number of impurities. At high impurity concentrations, a semiconductor may become similar to a metal in its properties, i.e., acquire residual conductivity at $T = 0$ K that is weakly temperature-dependent.

Looking slightly ahead, we note that semiconductors with given properties are produced not only by doping but also by forming binary, ternary, and more complicated compositions of different elements, what is known as heterostructures (see Section 5.6).

5.5 The electron – hole junction

If a semiconductor is doped in one of its parts by a donor impurity and in the other by an acceptor impurity, a region is formed in which electron (n-type) conduction is replaced with hole (p-type) conduction, and this region is known as an electron – hole, or p – n, junction. Since in the n region the concentration of the majority charge carriers, or electrons, is much higher than that of the minority carriers, or holes, while in the p region there are many more holes than electrons, there is mutual diffusion of electrons into the p region and holes into the n region. As a result, near the electron – hole junction there forms a sort of double electrical layer consisting of positively charged donor and negatively charged acceptor atoms with a contact electric field E between them and a diffusion potential difference U of about 1 V (Fig. 15a). The field generated hinders further diffusion of the majority charge carriers, which can now only proceed via tunneling through the potential barrier [36]. Here the charge carriers that are the majority carriers in their region become the minority carriers in the 'alien' region. Obviously, these minority carriers flow in the alien region in the opposite direction in relation to the direction of diffusion of the carriers that are the majority carriers in this region, i.e., in the direction determined by the sign of the contact electric field (the drift current). Under conditions of thermal equilibrium, both currents, the diffusion and the drift, are equal, so that the net current flowing through the electron – hole junction is zero.

This equilibrium can be violated by an external field. If the field is applied in such a way as shown in Fig. 15a, i.e., with its plus to the p region (forward-bias, or transmission, regime), the potential barrier lowers, which leads to an increase in diffusion of the majority charge carriers through the barrier and, as mentioned before, these carriers become, after passing the barrier, minority carriers in the 'alien' region. This increase in the concentration of the minority carriers is called injection. The charges of the injected minority carriers recombine with the opposite charges of the majority carriers of this medium supplied by the impurity atoms. The result is an exponentially increasing current flowing through the electron – hole junction.

[36] The tunnel effect, or tunneling, is a quantum mechanical process, forbidden by classical physics, in which a microparticle 'tunnels' through a potential barrier whose height is greater than the particle's energy.

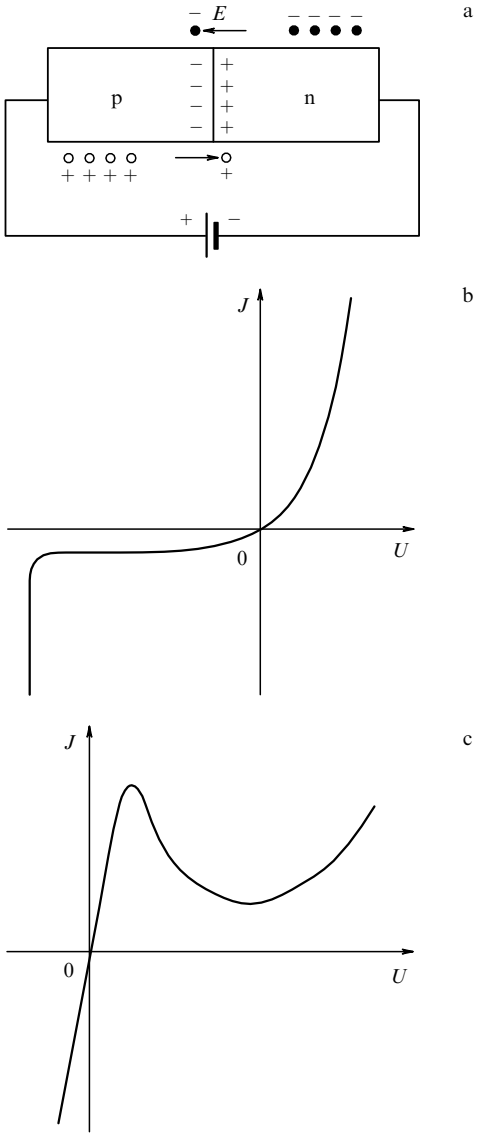

Figure 15. Electron–hole junction: (a) circuit diagram of the p–n junction; (b) current–voltage characteristic of the p–n junction; and (c) current–voltage characteristic of the tunnel diode.

If the external field is applied to the p region by its minus (reverse-bias, or cutoff, regime), the barrier becomes higher and the diffusion of the majority carriers rapidly decreases, while the current of the minority carriers remains the same — in fact, its value is generally independent of the strength of the external field (saturation current). This feature of the dependence of the current J flowing through the electron–hole junction on the applied voltage U (the current–voltage characteristic in Fig. 15b) is used in semiconductor diodes, extremely effective rectifiers of ac current (the current in the forward-bias regime is 10^5-10^6 times greater than in the reverse-bias regime).

The contact potential difference and the saturation current depend on various external conditions (heat, mechanical force, and optical and other types of electromagnetic radiation), which makes it possible to use the electron–hole junction as a sensor, in converters of light into electricity, in photomultipliers, and in solar cells. The injection of majority carriers that causes electroluminescence is used in constructing light-emitting diodes and injection lasers (see below).

One of the most important variants of the electron–hole junction is the three-layer structure of the p–n–p or n–p–n type, which served as the basis for the bipolar transistor invented by J Bardeen, W H Brattain, and W B Shockley in 1948, capable of amplifying electrical signals just as an electron tube is (the Nobel Prize in physics for 1956)[37]. In greater detail, the bipolar transistor, as well as one more semiconductor transistor (the field-effect transistor) and other semiconductor devices mentioned earlier, will be discussed in Section 5.7.

The electron–hole junction created in a single-crystal semiconductor by doping with donor and acceptor impurities in various combinations is known as a monojunction (also called a homojunction). But the electron–hole junction can be implemented in another form known as a heterojunction, which has certain advantages over a homojunction. The study of the properties of heterostructures (especially binary and quaternary) is one of the main areas of research of Alferov. But what are heterojunctions and heterostructures?

5.6 Heterojunction, heterostructures, and their advantages. Alferov's work

A heterojunction is the contact of two semiconductors that differ in chemical composition but have (ideally) the same types, orientations, and periods of crystal lattices without any defects. In 1963, Alferov chose as such an ideal pair gallium arsenide (GaAs) and aluminium arsenide (AlAs), with a lattice mismatch not exceeding 0.1%. These semiconductors were used to fabricate a defectless mixed crystal AlGaAs used to build the first heterolaser. Soon it was found that a heterojunction could be fabricated without the stringent restrictions concerning the degree of matching of the crystal lattices and even using amorphous semiconductors. The semiconductors that comprise a heterojunction may have the same type of conduction (p–p or n–n heterojunctions) or different types (p–n heterojunction). In the first case, the heterojunction is called isotype and in the second, anisotype. At the interface of the heterojunction, the properties of the semiconducting material, such as the energy band structure, the band gap, and the effective masses of the charge carriers, undergo a change. A composition of various hetero- and homojunctions is called a heterostructure. By varying the composition of the heterostructure, one can obtain properties that differ favorably from those of homostructures.

The main advantage of heterostructures is the presence of what is known as the superinjection effect, due to which practically any injected charge carrier concentrations can pass through the heterojunction. The double AlGaAs heterostructure proposed by Zh I Alferov and the fabricated low-threshold semiconductor lasers that are based on this heterostructure and operate in the cw mode at room temperature — all this played an enormous role (together with low-loss optical fibers manufactured in the same 1960s) in the process of the creation and rapid development of optical fiber communications.

[37] It is interesting that Bardeen won the Nobel prize *in physics* twice (as mentioned earlier, in 1972 he shared it with L N Cooper and J R Schrieffer for the development of a theory to explain superconductivity). This is certainly unusual. There were only two other scientists who became Nobel prize laureates twice, M Sklodowska-Curie and L C Pauling. But Sklodowska-Curie won the 1902 Nobel Prize in physics and the 1911 Nobel Prize in chemistry, while Pauling won the 1954 Nobel Prize in chemistry and the 1962 Nobel Peace Prize.

In 1970 the first heterostructure-based solar batteries were built. The commercial production of these batteries made it possible to equip artificial satellites and the orbital space station Mir with them. Heterostructure-based semiconductors are used to read information from laser disks; heterostructure transistors are inside every cell phone — examples abound. To find out more about the heterojunction, heterostructures, and their applications and about the heterostructure laser, the reader should turn to the popular articles by Alferov and his coworkers [222 – 225], Alferov's report at an international conference held in 1970 in Budapest [226], his 1971 article in Soviet Science Reviews [227], and essays in popular science in various collections of articles (see Refs [228 – 231]). The most complete list of scientific publications by Alferov and his coworkers that could interest specialists in this area of research can be found in his Nobel lecture [217].

In addition to effectively solving many technical problems that play an important role in practical life, scientific investigations of the properties of double heterostructures have led to the discovery of a new and interesting field of physics. A double heterostructure with a layer separation comparable to the de Broglie wavelength for the electron makes it possible to experimentally analyze quantum mechanical problems dealing with the properties of a low-dimensional electron gas (for more details see Section 5.10).

5.7 Design principles of some semiconductor devices

In this section, to economize on space, we proceed in the same way as we did in Section 4.6, where we described the various applications of lasers; i.e., we limit ourselves to a brief summary of the data on the design and operation principles of semiconductor devices and describe only some of these.

5.7.1 Semiconductor diodes. These are devices with a single p – n junction and two terminals. There are several types of semiconductor diodes: rectifiers, pulse diodes (used to detect high-frequency currents), stabistors and crystal stabilizer (Zener) diodes (used to stabilize voltage), variable-capacitance diodes, and tunnel diodes (used in amplifiers and oscillators of the microwave frequency range).

All these applications of semiconductor diodes can easily be explained by the typical features of the corresponding current – voltage characteristics. For diodes of the first four types this characteristic is similar to that in Fig. 15b for a p – n junction. When $U < 0$, the reverse-bias, or cutoff, regime is operating, i.e., the 'plus' is on the n region and the 'minus' is on the p region (Fig. 15a), and there is practically no current through the diode (with the exception of a small reverse current generated by the minority carriers). When $U > 0$, the forward-bias, or transmission, regime is operating, i.e., the 'plus' is on the p region and the 'minus' is on the n region, and the current increases very rapidly as soon as the applied potential difference exceeds the diffusion potential difference across the p – n junction (~ 0.2 V for Ge and ~ 0.7 V for Si). This ensures a string rectifying (and detection) effect.

Stabistors, which are used to stabilize small voltages (up to 1 V), use the right-hand side of the current–voltage characteristic, where a large variation in current corresponds to a very small variation in voltage. Crystal stabilizer diodes are used to stabilize higher voltages (from several volts to several dozen volts). These diodes are operated in the reverse direction; when the stabilization voltage is reached, the current flowing through the diode rapidly increases. After the voltage is removed, the junction is restored. A specific feature of variable-capacitance diodes (varicaps, or varactors) is the dependence of their capacitance on the applied voltage, in view of which they are used as variable capacitors (0.01 – 100 pF). Tunnel diodes have an extremely thin barrier (or blocking) layer (5 – 15 nm) through which the electrons travel because of the tunnel effect. The current – voltage characteristic of the tunnel diode is shaped like an 'N' (Fig. 15c), with a region of negative slope under forward bias (the negative resistance region). This character of the current – voltage curves makes it possible to use tunnel diodes in amplifiers and oscillators of the microwave frequency range.

5.7.2 Light-emitting diodes (LEDs). These devices emit light under the effect of a passing electric current. The mechanism of light generation is injection electroluminescence, which amounts to the current exciting the atoms of the semiconductor to metastable states, with light emitted when these atoms go back to the ground state. Ordinarily, light-emitting diodes are fabricated from gallium arsenide and gallium phosphide (GaAs, GaP) and also from other semiconductors of a more complex composition. Depending on the composition of the semiconductor, LEDs may emit red, yellow, green, or blue light. By design, LEDs are either point-like, which are widely used in technical and consumer equipment, or mosaic-like, consisting of several dozen to hundreds of LEDs of different shapes with each element switched on or off independently or in various combinations. In this way, almost any luminous image can be 'drawn' (alphanumeric texts, luminous adverticements, etc.). For instance, it is clear that only seven elongated LEDs forming a stylized '8' pattern are needed to produce the image of any digit from 1 to 9 and 0. Such light-emitting digits are used in the displays of calculators, electronic watches, and digital scales.

5.7.3 Bipolar transistors. These are three-layer semiconductor structures of the p – n – p type (if the middle region, the base B, is an n-type semiconductor) or the n – p – n type (if the base is a p-type semiconductor). The side regions are the emitter E and the collector C (Fig. 16). The majority carriers of the emitter p – n junction in transistors of the n – p – n type are electrons and in transistors of the p – n – p type, holes. In both cases they are emitted by the emitter, pass through the base, and are collected by the collector.

Figures 16a and 16b show the circuit and current diagrams and the polarities of various elements for bipolar p – n – p and n – p – n transistors, respectively. We see that the emitter in p – n – p transistors is under positive potential and that in n – p – n transistors, under negative potential; such polarity ensures the passage of the majority charge carriers from emitter to base (holes in the p – n – p case and electrons in the n – p – n case). In both cases the collector is under the potential opposite to that of the emitter and several times higher, which is the cutoff potential for the p – n junction of the collector proper. However, for the majority carriers of the emitter's p – n junction that pass from it into the base, this reverse bias is a forward bias. The base, which separates the emitter and the collector, has a thickness of a few tenths of a micrometer. Here, the collector potential is much higher than the base potential. As a result, an overwhelming fraction of the emitter current ($\sim 99\%$) flows directly into the collector, while only a small fraction ($\sim 1\%$) flows in the emitter – base circuit. Note that only the presence of a small base current

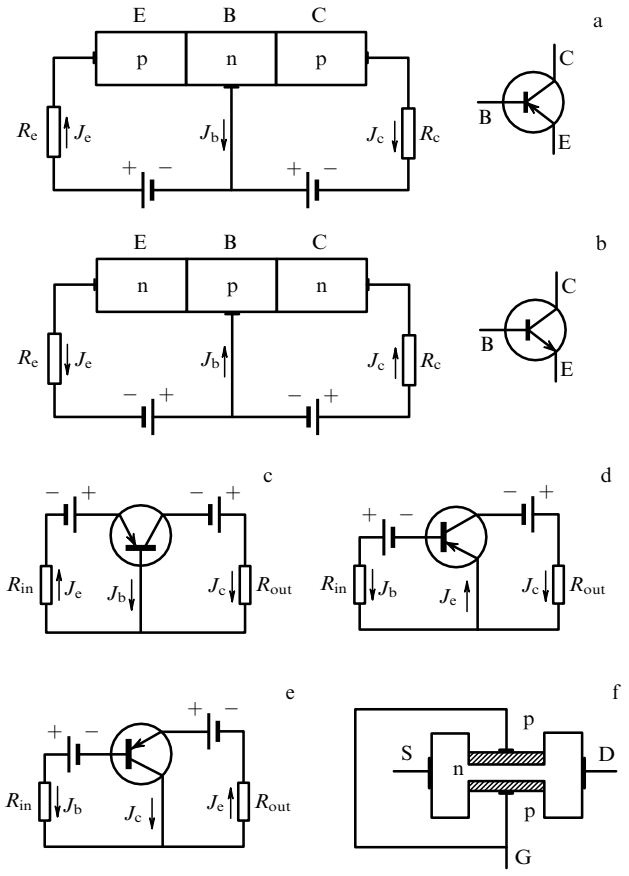

Figure 16. The current and circuit diagrams for transistors: (a) bipolar transistor of the p–n–p type: E, emitter; B, base; C, collector; and R_e and R_c resistors; (b) bipolar transistor of the n–p–n type; (c–e) circuit diagrams for a p–n–p bipolar transistor with a common base (CB), common emitter (CE), and common collector (CC), respectively: R_{in} and R_{out} are the input and output resistances; and (f) a unipolar transistor: S, source; n, an n-type plate; p, a p-type plate; D, drain; and G, gate.

(the emitter–base junction is open) enables the collector current to flow: if the emitter and the base are connected, there will be neither an emitter current nor a collector current (except for a small, $\sim 1-10$ μA, leakage reverse collector current). This uncontrollable current is very detrimental to the operation of transistors because it is highly dependent on the temperature of the p–n junction.

The resistor R_e in the emitter circuit serves as a current limiter. The output voltage can be tapped off from the resistor R_c in the collector circuit. The connections of transistors shown in Figs 16a and 16b are called common-base (CB) connections. In addition to this type of connection, there are common-emitter (CE) and common-collector (CC) connections. In Figs 16c–16e, all three types are shown schematically for a p–n–p transistor. (Obviously, for an n–p–n transistor the polarity of the potentials in the circuit must be reversed.)

The basic characteristic of all types of connections of transistors is the ratio of the increments of the output and input currents, which is known as the current amplification factor (current gain). Figure 16a shows that the current gain for the CB connection is smaller than unity, since the emitter current is larger than the collector current due to the small leakage of the emitter current flowing through the emitter–base circuit. In CE and CC connections, for the same reason (the smallness of the base current, which in these circuits is the input current) the current gain is greater than unity and may reach values of several dozen to several hundred. Thus, the CB connection cannot be used to amplify currents, but it can be used to raise voltages if the output resistance (in the collector circuit) is made higher than the input resistance (in the emitter circuit). The CB connection is used chiefly in oscillators. As for circuits with CE and CC connections, in the first usually $R_{in} \sim R_{out}$, which makes it possible to amplify both current and voltage, while in the second $R_{in} \gg R_{out}$, which allows for only current gain. The advantage of a CB circuit, which is known as the emitter follower, is its high input resistance and low output resistance. This makes it possible to use it as the coupling stage between devices with high input and low output resistances.

5.7.4 Unipolar or field-effect transistors. In contrast to the bipolar transistor (the name indicates that there are two p–n junctions with two different types of conduction), in the unipolar transistor there is only one p–n junction (the origin of the second name is discussed below). There are two types of unipolar transistor: with a controlled p–n junction and with an isolated gate. Figure 16f depicts a schematic of a unipolar transistor with a controlled p–n junction. In this case the base of the device is an n-type semiconductor plate with two terminals: the source (S), and the drain (D). Both the upper and lower sides of this plate have a p-type region, and because of the p–n junctions at the interfaces with the n-plate layers depleted of majority carriers (electrons in an n-region and holes in a p-region) form. If the 'minus' of the power source is connected to the source and the 'plus' to the 'drain', the electron will move in a channel between the depleted layers from source to drain.

The electron flux can be controlled by a gate (G) connected to the p regions, with the gate being under negative potential relative to the source. Obviously, as the negative potential on the gate grows, the depleted layers grow, which leads to a narrowing of the channel and reduction in the current. At a certain sufficiently high negative potential on the gate (the cutoff potential), the current ceases to flow through the channel. Thus, the current through a unipolar (channel) transistor is controlled by an electric field. Hence the second name, the field-effect transistor (in this respect the device is similar to an electron tube).

Note that in addition to the field-effect transistor depicted in Fig. 16f and based on the use of an n-type semiconductor plate, there are symmetric (in design) field-effect transistors with p-type semiconductor plates. Naturally, in such a transistor the p–n junctions exist as a result of formation of n regions on the surface of the p plate, and all potentials must be of opposite signs. The main advantage of field-effect transistors with a controlled p–n junction is the large input resistance, i.e., a very small input current.

The other type of field-effect transistor (with an isolated gate) differs from the field-effect transistor with a controlled p–n junction by the presence of a thin layer of insulator between the gate and the semiconducting channel. In view of this, the channel is controlled only by the potential on the gate (with a zero gate current). In this case, obviously, the input resistance is even higher than it is in a field-effect transistor with a controlled p–n junction. It can be said that it is infinitely large, i.e., simply equal to the resistance of the insulator. In view of this, the electrical properties of a field-

effect transistor with an insulator gate are better than those of a field-effect transistor with a controlled p – n junction.

The use of the miniature transistors that replaced the cumbersome, fragile, vacuumized, and complex (in design) electron tubes in various commercial radio and TV appliances, electronic computers, and other devices led not only to the miniaturization of mass and dimensions but also to a considerable reduction in the power, current, and voltage consumption.

In many cases the invention of the field-effect transistor, which consumes power only during operation (an 'infinite' input resistance) made it possible to do away with the power switch (the reader will recall the remote control units accompanying TV sets, DVD players, etc.). They can always be in a stand-by mode, ready for operation. Further progress in constructing miniature transistors finally led to the invention of miniature integrated circuits (see Section 5.7.7), containing up to several tens of thousands of transistors in a single silicon wafer. At the same time, transistors are used in high-current devices, for instance for switching on and of high currents (instead of knife switches and electromagnetic relays).

5.7.5 Thyristors. It is more convenient to control high currents ($\geqslant 10$ A) with special multilayer semiconductor devices with alternating layers of electron and hole conductors. Such devices are known as thyristors. Figure 17 shows a schematic of a four-layer thyristor with three p – n junctions J_1, J_2, and J_3, and three terminals, the anode (A), the cathode (C), and the gate electrode (G), which is the control electrode. (Thyristors with a control electrode are sometimes called trinistors and those without such an electrode, dynistors.) The external voltage is applied to the anode A ('plus') and cathode C ('minus'). As a result, the J_1 and J_3 junctions are in the forward-biased (on) state and the J_2 junction, in the reverse-biased (off) state, and only a small reverse current flows through the off-state thyristor. As the external voltage is increased, the current at first increases only slightly, but when the voltage reaches what is known as the forward-breakage value, the junction J_2 is 'broken' and the current through the thyristor increases dramatically. The device is in its on state. As the voltage is decreased, the J_2 junction is restored, and the thyristor returns to its off, weakly conducting, state. The forward-breakage voltage can be lowered by applying a positive voltage to gate G in relation to the cathode.

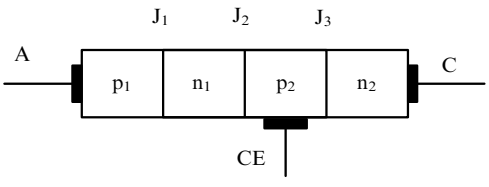

Figure 17. A schematic of a four-layer, three-terminal thyristor: A, anode; p_1, n_1, p_2, and n_2, layers with alternating hole (p) and electron (n) conduction; J_1, J_2, and J_3, p – n junctions; C, cathode; and CE, control electrode.

There are also bidirectional thyristors, known as triacs. These devices can switch the current in either direction by applying a small current between the gate and one of the two main terminals.

5.7.6 Photoelectronic devices and optrons. Semiconductor devices whose operation is controlled by light are called photoelectronic. Their operation is based on the photoelectric effect (internal and external). The internal photoelectric effect amounts to the electrons of the semiconductor being transferred to higher levels (into the conduction band), while the external photoelectric effect amounts to photoemission, i.e., the ejection of electrons from the surface of the semiconductor.

There are many semiconductor photoelectronic devices, which differ in the ways they are used. Among these are photoresistors, whose resistance changes under light; photodiodes, which generate an electromotive force under light (solar batteries are the best-known application of such devices); phototransistors, whose construction differs from that of an ordinary transistor with the CB connection in that it has a window for light to illuminate the base; and photothyristors, in which the gate electrode G (see Fig. 17) is a window fabricated in the same place.

A combination of any of these photoelectronic devices controlled by light and an LED (which generates the light) is called an optron. In such a combined device the light emitted by the LED is directed into the window of the photoelectronic device and controls its operation. Thus, an optron makes it possible to transmit current solely by optical means, without using galvanic coupling (i.e., without wires).

5.8 Integrated circuits

Earlier we noted that the invention of the transistor, which replaced the electron tube, led to considerable reduction in the size and weight of electronic equipment and in power consumption. However, the ever growing complexity of the problems that science, industry, nuclear power engineering, and the defence, rocket, and space industries began to resolve in the second half of the 20th century, and the swift progress in optoelectronic communications, the Internet, consumer electronics, etc., required the development of electronic equipment of considerable sophistication and, in particular, new-generation electronic computers. For these complex electronics not to become too unacceptably cumbersome and uneconomical, further microminiaturization of the various components was needed, which finally led in the late 1950s to the invention of integrated circuits (ICs), also called microcircuits, and to the birth of microelectronics.

Integrated circuits are devices in the form of small semiconductor plates (most often silicon-crystal plates)[38] with a large number of electronic microdevices. The invention of ICs and their mass production that followed shortly afterwards (in the early 1960s) became possible as a result of advances in the technology of fabricating subminiature transistors. Such microtransistors of about 1 μm in size are fabricated by depositing tiny droplets of donor or acceptor impurities onto the semiconductor plate. The circuits connecting the transistors are fabricated by sputtering very thin and narrow (~ 1 μm) metallic stripes along which microcurrents flow. Even narrower or thinner places in these conductors serve as natural resistors. Microscopic capacitors, diodes, and other microdevices are implanted by one or another method in the proper positions. This system of internal communication lines is said to be electric, while

[38] To increase the operation speed of integrated circuits they are sometimes fabricated, not on the basis of silicon, but on the basis of other semiconductors, e.g., GaAs, which has a higher charge carrier mobility.

other types of communication, such as optical or magnetic, also exist.

The number of microdevices used in integrated circuits determines the names of the ICs: small-scale integration circuits (SSI; the late 1950s), medium-scale integration circuits (MSI; the 1960s); large-scale integration circuits (LSI; the 1970s), and very large-scale integration circuits (VLSI; the second half of the 1970s). SSIs contain several dozen microdevices, MSIs about 10^3, LSIs 3×10^4, and VLSIs nearly 10^6. (For the current state of affairs with ICs see Section 6.) In view of the very small size of electronic microdevices, their concentration may reach 10^4 per cubic millimeter.

An integrated circuit without a casing and leads, which is called a chip (literally, a piece of crystal), has a very small volume ($0.2-50$ mm^3) and mass ($0.5-50$ mg) and can be directly installed into equipment or used as a base element for automatic formation of VLSIs (see below)[39]. A chip inside a casing and having several leads is used to solve a specific functional problem for which the integrated circuit was designed. To solve a different functional problem, a new IC is built, and so on. The general problem of building a certain piece of equipment is solved by combining different functional circuits.

Naturally, the fabrication of integrated circuits is a very complex technological process. It includes microlithography, growing thin films of metals and semiconductors, etching, introduction of impurities, and other physicochemical processes. All these processes are done automatically on special assembly lines. Moreover, the very designing of integrated circuits is performed by special automatic systems, which make it possible to form VLSIs for a specific problem from standard building blocks, i.e., chips connected by special internal communication lines.

For more details on the construction and use of integrated circuits and on the problems that emerge when these circuits are modernized, see Meindl's review [232] and two articles written by Dorfman [233, 234].

5.9 The semiconductor lasers

Before we begin to discuss the design, operation principles, and the pluses of the semiconductor laser, we remind the reader (see Section 5.1) about the pioneering work done in this field by Russian scientists (see Refs [141, 143–145, 219]). As for the specific features of the operation of the semiconductor laser, the easiest way to understand them is to recall the general operation principle of the lasers described earlier, which use the quantum transitions between the electronic levels of atoms, molecules, and ions.

As discussed in Section 4.2, this principle amounts to stimulated emission of directional, coherent optical radiation (including the IR and UV regions of the spectrum) by an active medium placed inside an optical cavity and transformed by pumping into the inverted state. Here, since the quantum optical gain of the lasers described earlier is moderate ($0.5-0.7$ cm^{-1}), the size of their active medium and the cavity must be substantial to ensure a long path for the resonance wave initiating the stimulated emission of light by inverted atoms (up to ~ 20 cm).

In contrast to these lasers, the semiconductor laser uses the quantum transitions, not between discrete energy levels, but between the allowed energy bands, and population inversion near the bottom of the conduction band and the top of the valence band, so that the wavelength of the laser light is

$$\lambda = \frac{hc}{E_g}, \tag{63}$$

where E_g is the band gap (see Section 5.3). Since E_g for different semiconductors varies considerably, the spectral range of semiconductor lasers is also very broad. It extends from the UV range to the far IR ($0.3 \leqslant \lambda \leqslant 45$ μm). And this is one of the main advantages of semiconductor lasers. The second advantage is related to the very large value ($\sim 10^4$ cm^{-1}) of quantum optical gain, in view of which the cavity length may amount only to values of several micrometers to several hundred micrometers (usually 200–300 μm), which means that semiconductor lasers are really very small.

The other advantages of semiconductor lasers are their high efficiency (up to 90%), very small response times ($\sim 10^{-9}$ s), the possibility of using low-voltage power sources (1–3 V), and long service lives (~ 10 years). Finally, one more advantage of the semiconductor laser is the possibility of a gradual spectral tuning of the emitted light. Here, by modulating the pump current, one can achieve frequency modulation of the emitted light. Such possibilities stem from the fact that, according to equation (63), the wavelength of the laser light is uniquely determined by the band gap, which depends on temperature, pressure, magnetic field strength, and other factors (see Section 5.3).

In its design, a semiconductor laser is an active element fabricated from a semiconductor single crystal whose active part proper is very small ($10^{-11}-10^{-10}$ cm^3). The mirror end (lateral) faces of the crystal or special external reflectors serve as the optical resonator. Pumping is achieved via bombardment with high-energy electrons or light or via electrical breakdown. However, the most common mechanism of pumping uses injection of excess charge carriers through the p–n junction of the heterojunction (injection laser; see Fig. 18).

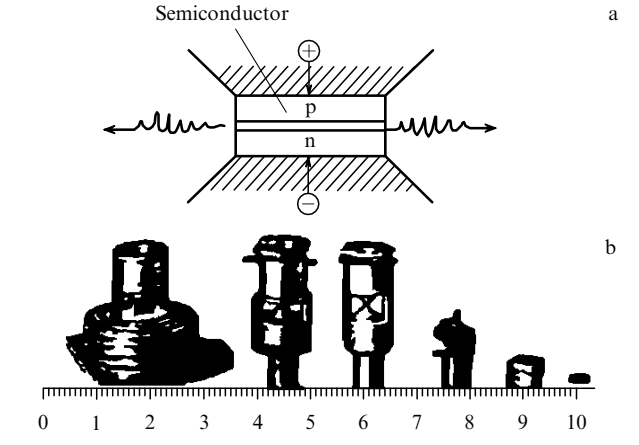

Figure 18. Injection semiconductor lasers: (a) schematic cross section: p, p region of the semiconductor; n, n region of the semiconductor; minus inside a circle — an electron; plus inside a circle, a hole; (b) photographs of specimens.

[39] According to Moore's empirical law, the number of microdevices (e.g., transistors) that can be placed within a single chip doubles every 18 months.

The operation of an injection laser is clear from Fig. 18. Injection is initiated by an electric field, the reverse bias on the p–n junction. Here, excess electrons go to the conduction band and excess holes to the valence band. As a result of relaxation (electron–electron or electron–phonon collisions), the electrons rapidly 'lower' to the bottom of the conduction band, while the holes 'rise' to the top of the valence band (because the relaxation of a lower hole amounts to the hole being occupied by an upper electron, in whose place another hole 'opens', and so on). Since relaxation occurs very rapidly, emission begins only after an electron in the conduction band and a hole in the valence band take their places at the boundaries of these bands, with the result that the emitted photon acquires an amount of energy exactly equal to the difference of the energies corresponding to these boundaries (edges), i.e., the band gap (Fig. 12d).

The most important type of injection laser is the injection heterolaser, which operates on heterojunctions between semiconducting materials with different electrical and optical properties. The characteristic features of heterolasers are low threshold lasing currents at room temperature (~ 300 K), high efficiencies, and the possibility of generating light within a broad frequency range by using different semiconducting materials. These properties, together with the high speed of response, very small size, low power consumption, and a long service life, have made injection heterolasers absolutely irreplaceable in optical fiber systems used in optical communication. Another area of application of injection heterolasers is laser memory systems (audio and video compact discs for various systems with an 'optical needle'), spectroscopy, laser projection TV, various tracking and guarding systems, etc. Figure 18b presents photographs of specimens of injection lasers, and the ruler under the photographs gives an idea of the size of such devices.

5.10 The new quantum physics of semiconductors

At the end of our article we would like to briefly highlight the role in the development of modern semiconductor physics played by the person to whom this article was initially intended to be dedicated in full, Zhores I Alferov [235].

Alferov was born on March 15, 1930. Already in 1950, when he was a third-year student at the Leningrad Electrical-Engineering Institute, Alferov began to study the properties of semiconductors. After graduation he was invited by V M Tuchkevich to work in a laboratory at the famous Physicotechnical Institute (Tuchkevich was then head of that laboratory), where he immediately got involved in the process of creating modern semiconductor electronics, namely, the first Soviet transistors, photodiodes, and high-current germanium and silicon rectifiers. His skills greatly increased in that period. In 1960 he was already a participant in an international conference in semiconductor physics, and in 1961 he defended his thesis devoted to high-current semiconductor rectifiers. In these same years he participated in the fabrication of semiconductor devices for nuclear submarines.

The year 1962 can be marked as the year when Alferov began his scientific work in a new area of research, which soon determined the entire development of modern semiconductor physics. Immediately after the publication of the paper by R M Hall and his collaborators [221] concerning a semiconductor laser based on a p–n homostructure of gallium arsenide, Alferov suggested using double heterostructures in lasers and began to investigate the possibility of fabricating such a heterostructure. This proved to be a highly nontrivial problem, since what was needed was an ideal heterojunction without defects at the interface. By 1967, Alferov and his coworkers were able, employing the liquid epitaxy method[40], to fabricate an AlGaAs heterojunction with properties close to those of an ideal heterojunction.

As noted in Section 5.6, two substances with very similar properties were used as the basis for fabricating this extremely complex heterostructure, gallium arsenide (GaAs) and aluminium arsenide (AlAs), which have nearly equal crystal lattice parameters and thermal expansion coefficients.

The result of the studies of the electrical and optical properties of heterostructures was the discovery by Alferov and his coworkers of the superinjection effect, which amounts to the creation of an enhanced density of carriers in the narrow bandgap of the active substance located between two heterojunctions. Also, Alferov's group first predicted and then discovered two other effects, which ensure localization of carriers in a layer of a narrow-gap semiconductor (electron confinement) and localization of electromagnetic waves (optical confinement). These findings have made it possible to effectively control light fluxes and the movements of charge carriers.

The fabrication of ideal heterostructures and the investigation of their properties made it possible to build, in the late 1960s and the early 1970s, a wide range of highly effective semiconductor devices (photodiodes, solar elements, light-emitting diodes, etc.), including heterolasers with low lasing thresholds and cw operation at room temperature.

In the same period (the early 1970s), Alferov and his collaborators designed and fabricated quaternary heterostructures (e.g., InGaAsP), which made it possible to broaden the frequency range of the light emitted by heterolasers and widely use such lasers in optical fiber communication lines. At the same time work was underway to build highly effective and reliable cells for solar batteries. Actually, Alferov and his team created a new field of research in semiconductor physics. Their work provided the basis for the development of optical fiber communications, the laser-disc industry (CD, DVD, etc.), solar batteries, cell phones, and heterolasers.

The work of Alferov in the realm of heterostructures was not only of great practical importance, it also provided a new path for the solution of some fairly exotic quantum mechanical problems. As already noted in Section 5.6, a double heterostructure with a very small layer separation δ ($\delta \simeq \lambda_e$, where λ_e is the de Broglie wavelength for the electron) makes it possible to formulate and experimentally solve new quantum mechanical problems dealing with the properties of a low-dimensional electron gas. In contrast to an ordinary electron gas, which moves in metals and semiconductors in all directions of three-dimensional space, by using specially selected heterostructures the electron gas can be partially or fully surrounded by a kind of energy barrier, which lowers the number of dimensions in which the gas can move from three to two (quantum well), to one (quantum wire), or even to zero (quantum dot).

A graphic image of a two-dimensional region in which an electron gas can exist (quantum well) is a thin flat plate of a

[40] Epitaxy in chemistry is the oriented growth of one crystalline substance on a substrate of a different crystalline substance. Epitaxy is widely used in quantum electronics (multilayer heterostructures, injection lasers) and microelectronics (integrated circuits, transistors, light-emitting diodes, etc.).

narrow-gap semiconductor (i.e., a semiconductor with a narrow band gap) surrounded on both sides by a wide-gap semiconductor. Earlier we noted that the band gap determines the probability of an electron overcoming a barrier: the greater the gap, the lower the probability. Hence, this construction is a peculiar two-dimensional flat barrier that limits the movement of electrons to two dimensions (in the plane)[41].

Obviously, if we cut a very narrow strip out of this plane and surround the edges of the strip by a wide-gap semiconductor, we have a quantum wire. And if we are able to inject a very small fraction ($\sim 10^{-7}$ cm across) of a narrow-band semiconductor into a wide-gap semiconductor, we obtain a quantum dot. In view of these restrictions imposed on the number of dimensions of the electron gas, the latter acquires new quantum properties.

According to Alferov, a quantum dot is a cluster consisting of $10-10^3$ 'narrow-gap' atoms inside a wide-gap semiconducting matrix, and this cluster has a discrete system of energy levels, as a free atom does (so we have an 'artificial atom'). The technology of growing these minute clusters inside another semiconductor was developed by Alferov and his coworkers in the late 1990s. Such microclusters match very well the crystal lattice of the main semiconductor even when the lattice parameters of the two semiconductors are different.

For more details about the modern methods of quantum-dot formation in the epitaxy process the reader should read the note [236] and the article by Sutter and Lagally [237]. The use of heterostructures with quantum dots in fabricating vertical-cavity emitting lasers is discussed in the report of Ustinov et al. [238] delivered at a scientific session of the General Physics and Astronomy Division of RAS. At the same session, Krestnikov et al. [239] presented a report concerning heterostructures based on nitrides of Group III elements (growing techniques and properties), which discussed the use of such heterostructures in light-emitting devices. Alferov contributed to each report.

The results of studies of quantum dots created in heterostructures in the growing process hold out hope of a dramatic increase in laser gain and a lowering of the lasing threshold. Quantum dots will possibly find their use in modernizing other semiconductor devices.

The successful research of Alferov in semiconductor physics has been fittingly acknowledged by the scientific community. In 1972 he was elected Corresponding Member of the USSR Academy of Sciences and became a professor at the Leningrad Electrical-Engineering Institute, in 1973 he became the head of a laboratory at the Physicotechnical Institute, in 1979 he was elected Full Member of the Academy of Sciences, in 1987 he became the Director of the Physicotechnical Institute, and in 1990 he was elected Vice President of the Academy of Sciences.

Alferov is a member of several foreign academies (including two academies in the United States) and an honorary doctor of many universities, a laureate of several international, domestic, and foreign prizes, including the State Prize of the Russian Federation for 2001. His candidacy for the Nobel Prize in Physics was chosen by the respective Nobel committee from a list of 300 [240].

[41] Apparently, this plane is called a well because electrons cannot leave it.

6. Conclusions

In our article we discussed five areas of general physics selected according to the topics studied by Russian physicists, laureates of the Nobel prize (and, of course, not only laureates), with Alferov still an active researcher. The five areas are superfluidity, superconductivity, Cherenkov radiation and related processes (in the Russian literature this radiation is called Vavilov–Cherenkov radiation), quantum electronics (masers and lasers), and semiconductor physics.

In these concluding remarks, we will try to list the most outstanding results obtained in these areas quite recently and tell about the hopes that physicists have when they look into the future (we call these 'blue dreamboats' the ultimate goals). Note that it would be much simpler not to write this section at all and instead recommend that the reader look through the list of pressing problems in physics given in a recent article by Ginzburg (just in time for us) and devoted to advances in all physics and astronomy from 1999 to 2001 [241]. However, since the number of problems covered by Ginzburg's article is 30, while we consider only 5, we will select these five problems from the list for our readers, add some comments, and name some of the ultimate goals (if they exist).

Here are the most striking (in our opinion) achievements in the five areas of general physics we selected. They are given in the same order as they appear in the present article.

(1) In the section devoted to superfluidity and the properties of Bose systems — the Bose condensation of rarefied gases of alkali metals described in the papers of B B Kadomtsev and M B Kadomtsev [65] and L P Pitaevskiĭ [66], and also a method for manipulating the movements of the condensate along a surface (with the coherence of the atomic states conserved) via electric fields, developed by J Reichel and others [242].

(2) In superconductivity — the continuing use of low-T_c superconductors (LTSCs) in colossal accelerators with superconducting magnetic systems. For instance, in the coming years LTSC will be used in the 7×2-TeV Large Hadron Collider (LHC) that is being built at CERN. Note that the number of unique facilities that employ or will employ LTSC, e.g., the unprecedented international collaboration on the next major step for the development of fusion — International Thermonuclear Experimental Reactor (ITER) — is steadily growing. Incidentally, extremely instructive stories about the initiation of research into controlled fusion (and also uncontrolled fusion) can be found in a series of reviews by V D Shafranov [243], B D Bondarenko [244], and G A Goncharov [245]. Furthermore, Shafranov's review provides a brief report on the current state of research into controlled fusion at the I V Kurchatov Institute of Atomic Energy. For instance, he mentions the development of the T15-M tokamak for modeling the operational modes of ITER–FIAT (a new version of ITER). And, speaking of controlled fusion, we should probably mention one more idea concerning its implementation, an idea proposed by Academician R I Nigmatulin (Russia) and Prof. R T Lahey, Jr. (USA). Their paper has been published in *Science*, with a popular version in Russian given in the newspaper *Poisk* [246].

The idea amounts to using the effect of strong compression, heating, and consecutive collapse of the microbubbles present in cold ($0-2\,^\circ$C) deuterated acetone C_3D_6O that is made to vibrate through excitation by an ultrasonic wave (vapor cavitation). The experimenters at Oak Ridge National

Laboratory, Tenn., USA, were able to create such strong compression and intense heating of acetone vapor bubbles that the two well-known reactions of the fusion of deuterium atoms became possible,

$$^2\mathrm{H} + {}^2\mathrm{H} \rightarrow \mathrm{n} + {}^3\mathrm{He} \quad (64)$$

and

$$^2\mathrm{H} + {}^2\mathrm{H} \rightarrow {}^1\mathrm{H} + {}^3\mathrm{H} \quad (65)$$

accompanied by the formation of thermonuclear neutrons in reaction (64) and tritons in reaction (65) (bubble fusion).

Both reactions in this setting produce very few nuclei (about 10^5 every second), and some physicists are sceptical about this idea, even more so because they still remember the ignominious end of the 'cold fusion' idea that occurred fairly recently (more about this see in Ref. [247]). Moreover, in July 2002 it was rumored that the results of the experiments carried out by Lahey and Nigmatulin had not been confirmed. On the other hand, the first experiments of L A Artsimovich and others produced very few neutrons, and it was rumored that the neutrons originated not in the thermonuclear reaction but in the accelerator. However, today nobody doubts the feasibility of the ITER project[42].

Going back to superconductivity, we remind the reader of recent achievements in the field of high-T_c superconductivity. What is most striking here is the rate with which the transition temperature T_c of the materials discovered earlier increases (such an increase in the transition temperature is achieved through impurity and field doping and by applying pressure) and the discovery of new superconducting materials, including fullerenes and fullerites [96, 249], nanotubes [250], and MgB_2 [93, 94] (whose properties are almost the same as those of the known low-T_c materials, although it has $T_c = 39$ K, which is roughly the same as for the first high-T_c materials). We also note the ever increasing practical use of high-T_c materials (e.g., see Refs [36–38, 251], which describe the progress made in producing HTSC cables). Lately, reports have appeared about the manufacture of whole HTSC assemblies. For instance, on June 15, 2002, there was a report about the first successful tests of a 100-kW HTSC electric motor conducted at the Moscow Aviation Institute (MAI) [252]. This fairly complex project was implemented by Russian–German collaboration in which, in addition to the MAI, four other Russian and five German institutes participated.

The ultimate goal here is room-temperature superconductivity.

(3) In the physics of electron beams (Cherenkov radiation and related processes) — the experimental proof of the existence of neutrino oscillations (what is known as the nonvanishing neutrino mass), which was obtained by using an enormous number of Cherenkov counters that covered a colossal amount (1000 tons) of heavy water (for more details see Refs [241, 253]). Also impressive is the new tunable free-electron laser [191]. Some physicists hope to build Compton and Cherenkov lasers based, like the free-electron laser, on relativistic electron beams [193].

[42] It must be noted, however, that some reports about discoveries have truly been premature. In particular, this is true of the recent report [248] about the discovery of the 118th element in the United States, about which we wrote in Ref. [2]. The authors of Ref. [248] have recently reported that their results have not been confirmed.

There is also news about another related process in the article by Platonov and Fleĭshman [254]. The researchers analyze the radiation produced by randomly inhomogeneous media excited by fast particles, i.e., polarization bremsstrahlung for inhomogeneities in thermodynamic equilibrium or transition radiation for nonthermal inhomogeneities.

(4) In quantum electronics — the transition to investigating ultrashort pulses, whose unit pulse duration is 1 ac $= 10^{-18}$ s [255]. In particular, Refs [256, 257] report about the building of a device that studies individual atoms. Another important advance is the building of a more powerful X-ray laser with hard radiation, whose active medium is high-temperature plasma created by focusing the beam of a 500-W neodymium laser on a jet of pulverized liquid xenon. It is assumed that in the future a 15-kW laser will be used to pump the plasma, which should increase the power of the X-ray laser considerably. For details see Ref. [258].

The ultimate goals here are the gamma-ray laser and laser-controlled fusion (the latter project is being implemented by building a large multibeam laser) [259]. Scientists also hope that in the near future gravitational waves will be detected. If the reader wishes to know more about this and, in particular, about the LIGO (Laser Interferometer Gravitational Wave Observatory) project, he or she should turn to the review by V B Braginskiĭ [260] and his recent article [261], where he talks about another bright prospect in using lasers. Since the electromagnetic field in picosecond laser pulses is several orders of magnitude higher than in ordinary accelerators, we can soon expect new (laser) techniques in accelerating elementary particles to emerge [252, 263].

(5) In semiconductor physics — the devices fabricated at the Physicotechnical Institute in St. Petersburg, namely, light-emitting diodes with optical microcavities and vertical-cavity emitting lasers based on structures with quantum dots, which are devices with narrow directional diagrams of the optical beam and an enhanced temperature stability of the wavelength [239], and also light-emitting devices that use heterostructures based on nitrides of Group III elements [239].

There are two ultimate goals here. One requires a bright blue sky for its implementation, and thus can be called a sky-high goal. What we mean is that in the future all the energy civilization needs can be generated solely by solar batteries, i.e., without burning organic fuel or uranium. The calculations done in an article in the journal *Ekologiya i Zhizn'* (Ecology and Life; No. 6 for 2001), reproduced later in Ref. [264], show that theoretically this is quite feasible, since 30% of the Earth's crust is silicon (needed to produce solar batteries), while the area that must be allocated to the solar batteries is not very large, 0.025% (for Russia this is approximately 4000 km^2, or about the area of three cities like Moscow, and then roofs can also be used).

The advantages are obvious: the technology is pollution-free and the supply of energy is inexhaustible, since solar batteries can work practically for ever (today the service life of solar batteries has been brought up to 30 years). Unfortunately, the disadvantages are obvious, too: the cost of producing silicon of the necessary purity (99.99%) is extremely high (and this is where the term sky-high comes into play for the second time), comparable with the cost of producing nuclear fuel (although silicon is 100,000 times more abundant in the Earth's crust than uranium).

The second ultimate goal is to fabricate very, very large-scale integration circuits (super-VLSIs, so to say), which would make it possible to achieve even higher computer

speeds in even smaller computers. Such hopes are justified by the results of a successful experiment with lithography that uses a very short wavelength of light emitted by an excimer laser, $\lambda = 157$ nm (instead of $\lambda = 248$ nm used now) and, what is more important, by the results of the first experiments with extreme ultraviolet ($\lambda = 13$ nm) lithography, in which light is emitted by a dense cloud of hot plasma (the X-ray laser mentioned earlier; see Ref. [258]). However, mastering this wavelength range will require an entirely new technology, since modern refracting optics cannot be used.

We believe that the new way of fabricating chips described in Ref. [268] presents a simpler approach. The idea is to imprint on the surface of silicon via laser stamping under pressure the image of a relief pattern that has been prestamped on a quartz matrix by an excimer microsurgical laser. A short ($\sim 10^{-8}$ s) laser pulse instantaneously melts the silicon under the matrix, and the silicon solidifies without sticking to the matrix, thus reproducing the pattern on the matrix. The process of fabricating chips by this new imprinting technique increases the number density of transistors by a factor of approximately 100 and takes less time (also by a factor of 100) than the photolithography technique.

The examples of the use of the achievements of two different areas of general physics (semiconductor physics and laser physics in the case above) out of the five discussed in this article are not unique. Just recall the free-electron laser, where the physics of relativistic electron beams is used to create a laser effect. Or Bose condensation via laser cooling and a superconducting magnetic trap (three areas of physics out of the five!). Incidentally, B B Kadomtsev and M B Kadomtsev [65], on the one hand, and A N Oraevskiĭ [265], on the other, seem to have established an analogy between Bose condensation and laser radiation.

And now we present an example in which all five areas of general physics discussed in this article are used.

A short communication published on April 30, 2002 [226] tells about a new large international project, the building of the unique ALPHA Magnetic Spectrometer (AMS). The spectrometer will be used to study the composition and energy spectrum of primary cosmic rays and to search for antihelium nuclei (the Alpha in the name of the spectrometer stands for the alpha particle, the nucleus of ^{4}He).

More than 200 physicists from 33 scientific centers from 11 countries (including three scientific centers from Russia) are participating in the project. In November 2004 the spectrometer will be launched to the International Space Station (ISS), where it will operate for three years. And now we will briefly describe the design of AMS. We would like the reader to focus on the italicized words.

The Alpha Magnetic Spectrometer consists of a *superconducting* magnet (450 A, 0.85 T) cooled by *superfluid* helium down to 1.8 K; two refrigerators; a detector of cosmic synchrotron radiation; a *silicon microstripe* detector for precise measurements of particle energies and the sign of particle charges; a system for processing the data, which is impossible to build without using *integrated circuits* and other *semiconductor* elements of *microelectronics*; a *Cherenkov* counter; and an electromagnetic calorimeter. The latest technology is being used to build the AMS, including *laser* technology, due to which it was possible to guarantee the acceptable parameters of the spectrometer so it could be fitted into the ISS (the mass of the magnetic system is 3 t, the overall mass is 6 t, and the power consumption is 2 kW). Comparing the italicized words with the names of the areas of general physics discussed in the present article, we can easily see that all these areas were indeed used in building the AMS. It appears that the words used in the Introduction concerning the possible different nature of these areas were unjustified. Physics is indeed indivisible.

In conclusion we would like to thank Yu G Abov and the members of the editorial board of *Physics Uspekhi*, headed by V L Ginzburg, who carefully read the manuscript and made many valuable remarks. We are also grateful to P A Alekseev for his recommendations concerning several sections in the article.

This work was supported by the Russian Foundation for Basic Research (grant No. 00-02-17582).

References

1. Ginzburg V L *Usp. Fiz. Nauk* **169** 419 (1999) [*Phys. Usp.* **42** 353 (1999)]
2. Mukhin K N, Patarakin O O *Usp. Fiz. Nauk* **170** 855 (2000) [*Phys. Usp.* **43** 799 (2000)]
3. Mukhin K N, Tikhonov V N *Usp. Fiz. Nauk* **171** 1201 (2001) [*Phys. Usp.* **44** 1141 (2001)]
4. Dirac P A M *Proc. R. Soc. London Ser. A* **114** 243 (1927)
5. Ginzburg V L *Usp. Fiz. Nauk* **170** 619 (2000) [*Phys. Usp.* **43** 573 (2000)]
6. Kamerlingh Onnes H *Commun. Phys. Lab. Univ. Leiden* **120B, 122B, 124C (1911)**[43]
7. Meissner W, Ochsenfeld R *Naturwissenschaften* **21** 787 (1933)
8. London F, London H *Proc. R. Soc. London Ser. A* **149** 71 (1935); *Physica* **2** 341 (1935)
9. Keesom W H *Physica* **3** 359 (1936)
10. Kapitza P L *Dokl. Akad. Nauk SSSR* **18** 21 (1938) [*CR Acad. Sci. USSR* **18** 21 (1938)]; *Nature* **141** 74 (1938); *Usp. Fiz. Nauk* **93** 481 (1967)
11. Allen F, Missner A D *Nature* **141** 75 (1938)
12. Kapitza P L *Zh. Eksp. Teor. Fiz.* **11** 581 (1941); *J. Phys. USSR* **4** 181, 59 (1941)
13. Landau L D *Zh. Eksp. Teor. Fiz.* **11** 592 (1941); *J. Phys. USSR* **5** 71 (1941); *Usp. Fiz. Nauk* **93** 495 (1967)
14. Kapitza P L *Rev. Mod. Phys.* **51** 417 (1979); *Usp. Fiz. Nauk* **129** 569 (1979)
15. Landau L D *J. Phys. USSR* **11** 91 (1947); *Usp. Fiz. Nauk* **93** 519 (1967)
16. London F *Superfluids* Vol. II (New York: Wiley, 1954)
17. Tisza J V *Nature* **141** 913 (1938)
18. Bogolyubov N N *Izv. Akad. Nauk SSSR* **11** (1) 77 (1947); *Usp. Fiz. Nauk* **93** 552 (1967)
19. Ginzburg V L, Landau L D *Zh. Eksp. Teor. Fiz.* **20** 1064 (1950); reprinted in: Landau L D *Sobranie Trudov* (Collected Papers) Vol. 2 (Moscow: Nauka, 1969) [Translated into English (London: Pergamon Press, 1965) p. 546]
20. Ginzburg V L *O Nauke, o Sebe i o Drugikh* (About Science, Myself, and Others) 2nd ed. (Moscow: Nauka, 2001) p. 199
21. Ginzburg V L *Usp. Fiz. Nauk* **48** 26 (1952); *Fortschr. Phys.* **1** 101 (1953)
22. Ginzburg V L *Zh. Eksp. Teor. Fiz.* **29** 748 (1955) [*Sov. Phys. JETP* **2** 589 (1956)]
23. Bardeen J, Cooper L N, Schrieffer J R *Phys. Rev.* **108** 1175 (1957)
24. Gor'kov L P *Zh. Eksp. Teor. Fiz.* **36** 1918; **37** 1407 (1959) [*Sov. Phys. JETP* **9** 1364 (1959); **10** 998 (1960)]
25. Cooper L N *Phys. Rev.* **104** 1189 (1956)
26. Bogolyubov N N *Zh. Eksp. Teor. Fiz.* **34** 58 (1958) [*Sov. Phys. JETP* **7** 41 (1958)]
27. Bogolyubov N N (Ed.) *Teoriya Sverkhprovodimosti* (Theory of Superconductivity) (Moscow: IL, 1960) [a collection of translations of articles from various sources in English and German]
28. Little W A *Phys. Rev.* **134** A1416 (1964)
29. Ginzburg V L, Kirzhnits D A *Zh. Eksp. Teor. Fiz.* **46** 397 (1964) [*Sov. Phys. JETP* **19** 269 (1964)]

[43] This article has been reprinted as an Appendix in Ref. [267].

30. Ginzburg V L *Zh. Eksp. Teor. Fiz.* **47** 2318 (1964) [*Sov. Phys. JETP* **20** 1549 (1965)]
31. Bednorz J G, Müller K A *Z. Phys. B: Cond. Mat.* **64** 189 (1986)
32. Wu M K, Ashburn J R, Torng C J, Hor P H, Meng R L, Gao L, Huang Z J, Wang Y Q, Chu C W *Phys. Rev. Lett.* **58** 908 (1987)
33. Bednorz J G, Müller K A *Rev. Mod. Phys.* **60** 585 (1988); *Usp. Fiz. Nauk* **156** 323 (1988)
34. Ginzburg V L *Usp. Fiz. Nauk* **167** 429 (1997) [*Phys. Usp.* **40** 407 (1997)]
35. Kopaev Yu V *Usp. Fiz. Nauk* **172** 712 (2002) [*Phys. Usp.* **45** 655 (2002)]
36. "Elektroenergetika gotova prinyat' VTSP ustroĭstva. Chto mogut razrabotchiki?" ("Electronics is ready to adopt HTSC devices. What can the developers do?") *PersT* [44] **10** (1/2) 2–4 (2003)
37. "Izgotoviteli VTSP-kabeleĭ gotovy k demonstratsii uspekhov" ("Manufacturers of HTSC cables are ready to demonstrate their achievements") *PersT* **8** (8) 3–5 (2001)
38. "Sverkhprovodyashchie fil'try v mobil'noĭ svyazi" ("Superconductor filters in mobile networks") *PersT* **8** (20) 1 (2001)
39. Pitaevskiĭ L P *Usp. Fiz. Nauk* **90** 623 (1966) [*Sov. Phys. Usp.* **9** 888 (1967)]
40. Andronikashvili E L *J. Phys. USSR* **10** 201 (1946); *Zh. Eksp. Teor. Fiz.* **18** 424 (1948)
41. Tilley D R, Tilley J *Superfluidity and Superconductivity* (New York: Van Nostrand, 1974) [Translated into Russian (Moscow: Mir, 1977)]
42. Khalatnikov I M *Vvedenie v Teoriyu Sverkhtekuchesti* (An Introduction to the Theory of Superfluidity) (Moscow: Nauka, 1965) [Translated into English (New York: W.A. Benjamin, 1965)]
43. Khalatnikov I M *Teoriya Sverkhtekuchesti* (Theory of Superfluidity) (Moscow: Fizmatgiz, 1971)
44. Onsager L *Nuovo Cimento* **6** (Suppl. 2) 249 (1949)
45. Feynman R P "Applications of quantum mechanics to liquid helium", in *Progress in Low Temperature Physics* Vol. 1 (Ed. C J Gorter) (Amsterdam: North-Holland, 1955) p. 17
46. Allen J F, Jones H *Nature* **141** 243 (1938)
47. Peshkov V P *J. Phys. USSR* **10** 389 (1946)
48. Henshaw D G, Woods A D B *Phys. Rev.* **121** 1266 (1961)
49. Cohen M, Feynman R P *Phys. Rev.* **107** 13 (1957)
50. Ginzburg V L, Pitaevskiĭ L P *Zh. Eksp. Teor. Fiz.* **34** 1240 (1958) [*Sov. Phys. JETP* **7** 858 (1958)]
51. Pitaevskiĭ L P *Zh. Eksp. Teor. Fiz.* **37** 1794 (1959) [*Sov. Phys. JETP* **10** 1267 (1960)]
52. Osheroff D D, Gully W J, Richardson R C, Lee D M *Phys. Rev. Lett.* **29** 920 (1972)
53. Lee D M *Rev. Mod. Phys.* **69** 645 (1997); *Usp. Fiz. Nauk* **167** 1307 (1997)
54. Osheroff D D *Rev. Mod. Phys.* **69** 667 (1997); *Usp. Fiz. Nauk* **167** 1327 (1997)
55. Richardson R C *Rev. Mod. Phys.* **69** 683 (1997); *Usp. Fiz. Nauk* **167** 1340 (1997)
56. Pomeranchuk I Ya *Zh. Eksp. Teor. Fiz.* **20** 919 (1950)
57. Rozhkov S S *Usp. Fiz. Nauk* **148** 325 (1986) [*Sov. Phys. Usp.* **29** 186 (1986)]
58. Mineev V P *Usp. Fiz. Nauk* **139** 303 (1983) [*Sov. Phys. Usp.* **26** 160 (1983)]
59. Volovik G E *Usp. Fiz. Nauk* **143** 73 (1984) [*Sov. Phys. Usp.* **27** 363 (1984)]
60. Ginzburg V L, Sobyanin A A *Usp. Fiz. Nauk* **120** 153 (1976) [*Sov. Phys. Usp.* **19** 773 (1976)]
61. Ginzburg V L, Sobyanin A A *J. Low Temp. Phys.* **49** 507 (1982)
62. Ginzburg V L, Sobyanin A A *Usp. Fiz. Nauk* **154** 545 (1988) [*Sov. Phys. Usp.* **31** 289 (1988)]
63. Bogolyubov N N (Jr) et al. *Usp. Fiz. Nauk* **159** 585 (1989) [*Sov. Phys. Usp.* **32** 1041 (1989)]
64. Kukin V D, in *Fizika Mikromira: Malen'kaya Entsiklopediya* (Physics of the Microworld: a Small Encyclopedia) (Ed.-in-Chief D V Shirkov) (Moscow: Sovetskaya Entsiklopediya, 1980) pp. 352, 335
65. Kadomtsev B B, Kadomtsev M B *Usp. Fiz. Nauk* **167** 649 (1997) [*Phys. Usp.* **40** 623 (1997)]
66. Pitaevskiĭ L P *Usp. Fiz. Nauk* **168** 641 (1998) [*Phys. Usp.* **41** 569 (1998)]
67. Landau L D *Zh. Eksp. Teor. Fiz.* **7** 19, 627 (1937); *Phys. Z. Sowjetunion* **11** 26, 545 (1937)
68. Rose-Innes A C, Rhoderick E H *Introduction to Superconductivity* (Oxford: Pergamon Press, 1969) [Translated into Russian (Moscow: Mir, 1972)]
69. Schmidt V V *Vvedenie v Fiziku Sverkhprovodnikov* (Introduction to the Physics of Superconductors) (Moscow: Nauka, 1982)
70. Schmidt V V *The Physics of Superconductors: Introduction to Fundamentals and Applications* (Eds P Müller, A V Ustinov) (New York: Springer, 1997)
71. Lynton E A *Superconductivity* 3rd ed. (London: Methuen, 1969) [Translated into Russian (Moscow: Mir, 1971)]
72. Abrikosov A A *Zh. Eksp. Teor. Fiz.* **32** 1442 (1957) [*Sov. Phys. JETP* **5** 1174 (1957)]
73. Fröhlich H *Phys. Rev.* **79** 845 (1950)
74. Eliashberg G M *Zh. Eksp. Teor. Fiz.* **38** 966; **39** 1437 (1960) [*Sov. Phys. JETP* **11** 696 (1960); **12** 1000 (1961)]
75. Abrikosov A A *Osnovy Teorii Metallov* (Fundamentals of the Theory of Metals) (Moscow: Nauka, 1987) [Translated into English (New York: Elsevier, 1988)]
76. Bogolyubov N N, Tolmachev V V, Shirkov D V *Novyĭ Metod v Teorii Sverkhprovodimosti* (A New Method in the Theory of Superconductivity) (Moscow: Nauka, 1958) [Translated into English (New York: Consultants Bureau, 1959)]
77. Vonsovsky S V, Izyumov Yu A, Kurmaev É Z *Sverkhprovodimost' Perekhodnykh Metallov, Ikh Splavov i Soedineniĭ* (Superconductivity of Transition Metals, Their Alloys and Compounds) (Moscow: Nauka, 1977) [Translated into English (Berlin: Springer, 1982)]
78. Schrieffer J R *Theory of Superconductivity* (New York: W.A. Benjamin, 1964) [Translated into Russian (Moscow: Nauka, 1970)]
79. Tinkham M *Introduction to Superconductivity* (New York: McGraw-Hill, 1975) [Translated into Russian (Moscow: Atomizdat, 1980)
80. Putilin S N et al. *Nature* **362** 226 (1993)
81. Chu C W et al. *Nature* **365** 323 (1993)
82. Bobovich Ya S *Usp. Fiz. Nauk* **167** 973 (1997) [*Phys. Usp.* **40** 925 (1997)]
83. Frank-Kamenetskiĭ O V, in *Vysokotemperaturnaya Sverkhprovodimost'* (High-Temperature Superconductivity) Issue 2 (Leningrad: Izd. LGU, 1990) p. 191
84. Maksimov E G *Usp. Fiz. Nauk* **170** 1033 (2000) [*Phys. Usp.* **43** 965 (2000)]
85. Ginzburg V L, Kirzhnits D A (Eds) *Problema Vysokotemperaturnoĭ Sverkhprovodimosti* (High-Temperature Superconductivity) (Moscow: Nauka, 1977) [Translated into English (New York: Consultants Bureau, 1982)]
86. Hubbard J *Proc. R. Soc. London Ser. A* **276** 238 (1963); **281** 401 (1964)
87. Anderson P W *Science* **235** 1196 (1987)
88. Kalmeyer V, Laughlin R B *Phys. Rev. Lett.* **59** 2095 (1987)
89. Kohn W, Luttinger J M *Phys. Rev. Lett.* **15** 524 (1965)
90. Ginzburg V L, in *Progress in Low Temperature Physics* Vol. 12 (Ed. C J Gorter) (Amsterdam: North-Holland Publ. Co., 1989) p. 1
91. Izyumov Yu A *Usp. Fiz. Nauk* **169** 225 (1999) [*Phys. Usp.* **42** 215 (1999)]
92. Kocharovsky V V, Kocharovsky Vl V *Physica C* **200** 385 (1992)
93. Nagamatsu J et al. *Nature* **410** 63 (2001)
94. Eroshenko Yu N (Compiler) "Physics news on the Internet" *Usp. Fiz. Nauk* **171** 306, 648, 796 (2001) [*Phys. Usp.* **44** 330, 654, 759 (2001)]; "All-metal superconductivity at 40 K" *Phys. News Update* (526) (2001); An J M, Pickett W E *Phys. Rev. Lett.* **86** 4366 (2001); Eom C B et al. *Nature* **411** 558 (2001)
95. Eroshenko Yu N (Compiler) "Physics news on the Internet" *Usp. Fiz. Nauk* **172** 1110 (2002) [*Phys. Usp.* **45** 998 (2002)]; Choi H J et al. *Nature* **418** 758 (2002)
96. "Novyĭ rekord T_c (117 K) dlya fulleritov" ("A new record-high T_c (117 K) for fullerites") *PersT* **8** (17) 1 (2001)
97. Goss Levi B *Phys. Today* **54** (10) 19 (2001)

[44] The name *PersT* stands (in Russian) for *Promising Technologies — Nanostructures, Superconductors, and Fullerenes*. If the reader is unable to get the printed copy, its contents can be viewed on the Internet at http://perst.isssph.kiae.ru/inform/perst.htm

98. Tsebro V I, Omel'yanovskiĭ O E, Moravskiĭ A P *Pis'ma Zh. Eksp. Teor. Fiz.* **70** 457 (1999) [*JETP Lett.* **70** 462 (1999)]
99. "Bol'shoĭ uspekh rossiĭskikh uchenykh" ("A great achievement of Russian science") *PersT* **8** (19) 1 (2001)
100. Josephson B D *Phys. Lett.* **1** 251 (1962)
101. Josephson B D *Science* **184** 527 (1974); *Usp. Fiz. Nauk* **116** 597 (1975)
102. Esaki L *Rev. Mod. Phys.* **46** 237 (1974); *Usp. Fiz. Nauk* **116** 569 (1975)
103. Giaver I *Science* **183** 1253 (1974); *Usp. Fiz. Nauk* **116** 585 (1975)
104. "Vnov' o pervom VTSP ustroĭstve dlya RAO 'EÉS'" ("Once again about the first HTSC device for the Russian United Energy System, Inc.") *PersT* **8** (13/14) 1–3 (2001)
105. "Ya khochu razrabatyvat' kompleks — lokal'nuyu tselikom sverkhprovodyashchuyu energosistemu" ("I want to develop a local, entirely superconducting energy system") *PersT* **8** (17) 3 (2001)
106. "Ya khochu razrabatyvat' kompleks — lokal'nuyu tselikom sverkhprovodyashchuyu energosistemu" ("I want to develop a local, entirely superconducting energy system") *PersT* **8** (17) 2 (2001)
107. "Etyud o stoimosti izdeliĭ iz VTSP" ("An etude about the cost of devices fabricated from HTSC") *PersT* **9** (5) 1 (2002)
108. "Kriogennyĭ avtomobil'? Est'..." ("A cryogenic automobile? We have it!") *PersT* **8** (21) 4 (2001)
109. Frank I M *Usp. Fiz. Nauk* **143** 111 (1984) [*Sov. Phys. Usp.* **27** 382 (1984)]
110. Cherenkov P A *Dokl. Akad. Nauk SSSR* **2** 451 (1934) [*CR Acad. Sci. USSR* **2** 451 (1934)]; *Usp. Fiz. Nauk* **93** 385 (1967)
111. Vavilov S I, Tumermann L A *Z. Phys.* **54** 270 (1929)
112. Frank I M, Vavilov S I *Z. Phys.* **69** 100 (1931)
113. Vavilov S I *Phys. Z. Sowjetunion* **5** 369 (1934)
114. Vavilov S I *Dokl. Akad. Nauk SSSR* **2** 457 (1934) [*CR Acad. Sci. USSR* **2** 457 (1934)]; *Usp. Fiz. Nauk* **93** 383 (1967)
115. Cherenkov P A *Dokl. Akad. Nauk SSSR* **3** 414 (1936)
116. Tamm I E, Frank I M *Dokl. Akad. Nauk SSSR* **14** 107 (1937) [*CR Acad. Sci. USSR* **14** 107 (1937)]; *Usp. Fiz. Nauk* **93** 388 (1967)
117. Ginzburg V L *Zh. Eksp. Teor. Fiz.* **10** 589 (1940); *J. Phys. USSR* **2** 441 (1940)
118. Ginzburg V L *Zh. Eksp. Teor. Fiz.* **10** 608 (1940); *J. Phys. USSR* **3** 101 (1940)
119. Fermi E *Phys. Rev.* **57** 485 (1940)
120. Cherenkov P A, Tamm I E, Frank I M *Nobelevskie Lektsii* (Nobel Lectures) (Moscow, 1960); *Nobel Lectures*: *Physics, 1942–1962* (Amsterdam: Elsevier, 1964)
121. Mukhin K N *Eksperimental'naya Yadernaya Fizika* (Experimental Nuclear Physics) Book 1 *Fizika Atomnogo Yadra* (Physics of the Atomic Nucleus) Pt. 1 *Svoĭstva Nuklonov, Yader i Radioaktivnykh Izlucheniĭ* (Properties of Nucleons, Nuclei, and Radioactive Radiation) 5th ed. (Moscow: Energoatomizdat, 1993)
122. Bolotovskiĭ B M *Usp. Fiz. Nauk* **62** 201 (1957); **75** 295 (1961) [*Sov. Phys. Usp.* **4** 781 (1962)]
123. Ginzburg V L *Usp. Fiz. Nauk* **166** 1033 (1996) [*Phys. Usp.* **39** 973 (1996)]
124. Jelley J *Prog. Nucl. Phys.* **3** 84 (1953); *Usp. Fiz. Nauk* **58** 231 (1956)
125. Jelley J V *Čerenkov Radiation, and Its Applications* (New York: Pergamon Press, 1958) [Translated into Russian (Moscow: IL, 1960)]
126. Zrelov V P *Izluchenie Vavilova–Cherenkova i Ego Primenenie v Fizike Vysokikh Energiĭ* (Cherenkov Radiation and Its Application in High-Energy Physics) Vols 1, 2 (Moscow: Atomizdat, 1968) [Translated into English: *Cherenkov Radiation in High-Energy Physics* (Jerusalem: Israel Program for Scientific Translations, 1970)]
127. Ginzburg V L, Frank I M *J. Phys. USSR* **9** 353 (1945)
128. Ginzburg V L, Frank I M *Zh. Eksp. Teor. Fiz.* **16** 15 (1946)
129. Ginzburg V L, Tsytovich V N *Perekhodnoe Izluchenie i Perekhodnoe Rasseyanie* (Transition Radiation and Transition Scattering) (Moscow: Nauka, 1984) [Translated into English (London: Gordon & Breach, 1990)]
130. Bolotovskiĭ B M "Perekhodnoe izluchenie" ("Transition radiation"), in *Fizicheskaya Entsiklopediya* (Physics Encyclopedia) Vol. 3 (Ed.-in-Chief A M Prokhorov) (Moscow: Bol'shaya Rossiĭskaya Entsiklopediya, 1992) p. 578
131. Merzon G I "Perekhodnogo izlucheniya detektor" ("Transition-radiation detector"), in *Fizicheskaya Entsiklopediya* (Physics Encyclopedia) Vol. 3 (Ed.-in-Chief A M Prokhorov) (Moscow: Bol'shaya Rossiĭskaya Entsiklopediya, 1992) p. 577
132. Ginzburg V L *Usp. Fiz. Nauk* **172** 373 (2002) [*Phys. Usp.* **45** 341 (2002)]
133. Pavlov V I, Sukhorukov A I *Usp. Fiz. Nauk* **147** 83 (1985) [*Sov. Phys. Usp.* **28** 784 (1985)]
134. Biberman L M "Valentin Aleksandrovich Fabrikant (on his sixtieth birthday)" *Usp. Fiz. Nauk* **93** 757 (1967) [*Sov. Phys. Usp.* **10** 798 (1968)]
135. Townes C H *How the Laser Happened*: *Adventures of a Scientist* (New York: Oxford Univ. Press, 1999)
136. Gordon J P, Zeiger H J, Townes C H *Phys. Rev.* **95** 282 (1954)
137. Basov N G, Prokhorov A M *Zh. Eksp. Teor. Fiz.* **27** 431 (1954); *Dokl. Akad. Nauk SSSR* **101** 47 (1955)
138. Basov N G, Prokhorov A M *Zh. Eksp. Teor. Fiz.* **28** 249 (1955) [*Sov. Phys. JETP* **1** 184 (1955)]
139. Prokhorov A M *Zh. Eksp. Teor. Fiz.* **34** 1658 (1958) [*Sov. Phys. JETP* **7** 1140 (1958)]
140. Schawlow A L, Townes C H *Phys. Rev.* **112** 1940 (1958)
141. Basov N G, Vul B M, Popov Yu M *Zh. Eksp. Teor. Fiz.* **37** 587 (1959) [*Sov. Phys. JETP* **10** 416 (1960)]
142. Maiman T H *Nature* **187** 493 (1960)
143. Basov N G, Krokhin O N, Popov Yu M *Zh. Eksp. Teor. Fiz.* **40** 1879 (1961) [*Sov. Phys. JETP* **13** 1320 (1961)]; *Vestn. Akad. Nauk SSSR* **31** (3) 61 (1961)
144. Bagaev V S, Basov N G, Vul B M et al. *Dokl. Akad. Nauk SSSR* **150** 275 (1963)
145. Basov N G, Bogdankevich O V, Devyatov A G *Dokl. Akad. Nauk SSSR* **155** 783 (1964)
146. Basov N G, in *Nobel Lectures in Physics* Vol. 4 *1963–1970* (Amsterdam: Elsevier Publ. Co., 1972) p. 89; *Usp. Fiz. Nauk* **85** 585 (1965)
147. Prokhorov A M *Science* **149** 828 (1965); *Usp. Fiz. Nauk* **85** 599 (1965)
148. Townes Ch H *IEEE Spectrum* **2** (2) 30 (1965); *Usp. Fiz. Nauk* **88** 461 (1966)
149. Basov N G, Prokhorov A M *Usp. Fiz. Nauk* **57** 485 (1955); **93** 572 (1967)
150. Kornienko L S, Shteĭnshleĭger V B *Usp. Fiz. Nauk* **126** 287 (1978) [*Sov. Phys. Usp.* **21** 852 (1978)]
151. Kotel'nikov V A et al. *Dokl. Akad. Nauk SSSR* **147** 1320 (1962); **155** 1037 (1964); **163** 50 (1965)
152. Weaver H et al. *Nature* **208** 29 (1965)
153. Cheung A C, Rank D M, Townes C H et al. *Nature* **221** 626 (1969)
154. Strel'nitskiĭ V S *Usp. Fiz. Nauk* **113** 463 (1974) [*Sov. Phys. Usp.* **17** 507 (1975)]
155. Burdyuzha V V *Usp. Fiz. Nauk* **155** 703 (1988) [*Sov. Phys. Usp.* **31** 763 (1988)]
156. Bochkarev N G *Osnovy Fiziki Mezhzvezdnoĭ Sredy* (Fundamentals of the Physics of Interstellar Medium) (Moscow: Izd. MGU, 1992)
157. Zverev G M, Karlov N V, Kornienko L S et al. *Usp. Fiz. Nauk* **77** 61 (1962) [*Sov. Phys. Usp.* **5** 401 (1962)]
158. Prokhorov A M *Usp. Fiz. Nauk* **148** 7 (1986) [*Sov. Phys. Usp.* **29** 3 (1986)]
159. Pressley R J (Ed.) *CRC Handbook of Lasers, with Selected Data on Optical Technology* (Cleveland: Chemical Rubber Co., 1971) [Translated into Russian: Prokhorov A M (Ed.) *Spravochnik po Lazeram* (Laser Handbook) Vols 1, 2 (Moscow: Sovetskoe Radio, 1978)]
160. Anikiev Yu G, Zhabotinskiĭ M E, Kravchenko V B *Lazery na Neorganicheskikh Zhidkostyakh* (Inorganic-Liquid Lasers) (Moscow: Nauka, 1986)
161. Rubinov A N, Tomin V I "Opticheskie kvantovye generatory na krasitelyakh i ikh primenenie" ("Dye lasers and their applications"), in *Itogi Nauki i Tekhniki. Radiotekhnika* (Results in Science and Technology. Radio Engineering) Vol. 9 (Moscow: Izd. VINITI, 1976)
162. Eletskiĭ A V, Smirnov B M *Fizicheskie Protsessy v Gazovykh Lazerakh* (Physical Processes in Gas Lasers) (Moscow: Energoatomizdat, 1985)

163. Basov N G, Danilychev V A *Usp. Fiz. Nauk* **148** 55 (1986) [*Sov. Phys. Usp.* **29** 31 (1986)]
164. Karlov N V, Konev Yu B *Moshchnye Molekulyarnye Lazery* (High-Power Molecular Lasers) (Novoe v Zhizni, Nauke, Tekhnike. Ser. Fiz., Issue 1) (Moscow: Znanie, 1976)
165. Eletskiĭ A V *Usp. Fiz. Nauk* **125** 279 (1978) [*Sov. Phys. Usp.* **21** 502 (1978)]
166. Gudzenko L I, Yakovlenko S I *Plazmennye Lazery* (Plasma Lasers) (Moscow: Atomizdat, 1978)
167. Eletskiĭ A V *Usp. Fiz. Nauk* **134** 237 (1981) [*Sov. Phys. Usp.* **24** 475 (1981)]
168. Basov N G, Bashkin A S, Igoshin V I, Oraevsky A N, Shcheglov V A *Khimicheskie Lazery* (Chemical Lasers) (Ed. N G Basov) (Moscow: Nauka, 1982) [Translated into English (Berlin: Springer, 1990)]
169. Bessonov E G, Vinogradov A V *Usp. Fiz. Nauk* **159** 143 (1989) [*Sov. Phys. Usp.* **32** 806 (1989)]
170. Andreev A V "Rentgenovskiĭ lazer" ("X-ray laser"), in *Fizicheskaya Entsiklopediya* (Physics Encyclopedia) Vol. 4 (Ed.-in-Chief A M Prokhorov) (Moscow: Bol'shaya Rossiĭskaya Entsiklopediya, 1994) p. 365
171. Attwood D, Halbach K, Kim K-J *Science* **228** 1265 (1985); *Usp. Fiz. Nauk* **159** 125 (1989)
172. Gasparyan P D, Starikov F A, Starostin A N *Usp. Fiz. Nauk* **168** 843 (1998) [*Phys. Usp.* **41** 761 (1998)]
173. Andreev A V "Gamma-lazer" ("Gamma-ray laser"), in *Fizicheskaya Entsiklopediya* (Physics Encyclopedia) Vol. 1 (Ed.-in-Chief A M Prokhorov) (Moscow: Bol'shaya Sovetskaya Entsiklopediya, 1988) p. 411
174. "Report to the APS of the Study Group on Science and Technology of Directed Energy Weapons (DEW)" *Rev. Mod. Phys.* **59** (3, Pt. II) S1 (1987); *Usp. Fiz. Nauk* **155** 661 (1988)
175. Schrödinger E *Ann. Phys.* (Leipzig) **82** 257 (1927)
176. Kapitza P L, Dirac P A M *Proc. Camb. Philos. Soc.* **29** 297 (1933)
177. Pantell R H, Soncini G, Puthoff H E *IEEE J. Quantum Electron.* **QE-4** 905 (1968)
178. Fargion D, Salis A *Usp. Fiz. Nauk* **168** 909 (1998) [*Phys. Usp.* **41** 823 (1998)]
179. Zaretskiĭ D F, Nersesov É A, Fedorov M V *Zh. Eksp. Teor. Fiz.* **80** 999 (1981) [*Sov. Phys. JETP* **53** 508 (1981)]; Fedorov M V, Nersesov E A, Zaretsky D F *Phys. Lett. A* **82** 227 (1981)
180. Kuznetsov V L *Usp. Fiz. Nauk* **129** 541 (1979) [*Sov. Phys. Usp.* **22** 934 (1979)]
181. Fedorov M V *Usp. Fiz. Nauk* **135** 213 (1981) [*Sov. Phys. Usp.* **24** 801 (1981)]
182. Ginzburg V L *Izv. Akad. Nauk SSSR Ser. Fiz.* **11** 165 (1947)
183. Motz H *J. Appl. Phys.* **22** 527 (1951)
184. Bessonov E G "Ondulyator" ("Undulator"), in *Fizicheskaya Entsiklopediya* (Physics Encyclopedia) Vol. 3 (Ed.-in-Chief A M Prokhorov) (Moscow: Bol'shaya Rossiĭskaya Entsiklopediya, 1992) p. 406
185. Bessonov E G "Ondulyatornoe izluchenie" ("Undulator radiation"), in *Fizicheskaya Entsiklopediya* (Physics Encyclopedia) Vol. 3 (Ed.-in-Chief A M Prokhorov) (Moscow: Bol'shaya Rossiĭskaya Entsiklopediya, 1992) p. 407
186. Alferov D F, Bashmakov Yu A, Bessonov E G *Tr. Fiz. Inst. Akad. Nauk* **80** 100 (1975)
187. Sprangle P, Coffey T *Phys. Today* **37** (3) 44 (1984); *Usp. Fiz. Nauk* **146** 303 (1985)
188. Madey J M J *J. Appl. Phys.* **42** 1906 (1971)
189. Elias L R et al. *Phys. Rev. Lett.* **36** 717 (1976); Deacon D A G et al. *Phys. Rev. Lett.* **38** 892 (1977)
190. McDermott D B et al. *Phys. Rev. Lett.* **41** 1368 (1978)
191. DOR, report No. 10691, March 2002 (Moscow: TsNIIAtominform, 2002)
192. Marshall T C *Free-Electron Lasers* (New York: Macmillan, 1985) [Translated into Russian (Moscow: Mir, 1988)
193. Arutyunyan V M, Oganesyan S G *Usp. Fiz. Nauk* **164** 1089 (1994) [*Phys. Usp.* **37** 1005 (1994)]
194. Avetisyan G K *Usp. Fiz. Nauk* **167** 793 (1997) [*Phys. Usp.* **40** 755 (1997)]
195. Basov N G, Krokhin O N *Zh. Eksp. Teor. Fiz.* **46** 171 (1964) [*Sov. Phys. JETP* **19** 123 (1964)]
196. Basov N G, Rozanov V B, Sobolevskiĭ N M *Izv. Akad. Nauk SSSR Ser. Energ. Transport* (6) 3 (1975)
197. Basov N G et al. "Nagrev i szhatie termoyadernykh misheneĭ, obluchaemykh lazerom" ("Heating and compression of thermonuclear targets irradiated by laser light"), in *Itogi Nauki i Tekhniki. Radiotekhnika* (Results in Science and Technology. Radio Engineering) Vol. 26, Pt. 1, 2 (Moscow: Izd. VINITI, 1982)
198. Karlov N V, Prokhorov A M *Usp. Fiz. Nauk* **118** 583 (1976) [*Sov. Phys. Usp.* **19** 285 (1976)]
199. Basov N G, Belenov É M, Isakov V A et al. *Usp. Fiz. Nauk* **121** 427 (1977) [*Sov. Phys. Usp.* **20** 209 (1977)]
200. Karlov N V, Krynetskiĭ B B, Mishin V A, Prokhorov A M *Usp. Fiz. Nauk* **127** 593 (1979) [*Sov. Phys. Usp.* **22** 213 (1979)]
201. Letokhov V S *Usp. Fiz. Nauk* **148** 123 (1986) [*Sov. Phys. Usp.* **29** 70 (1986)]
202. Dunskaya I M *Lazery i Khimiya* (Lasers and Chemistry) (Moscow: Nauka, 1979)
203. Bunkin F V, Kirichenko N A, Luk'yanchuk B S *Usp. Fiz. Nauk* **138** 45 (1982) [*Sov. Phys. Usp.* **25** 662 (1982)]
204. Letokhov V S *Usp. Fiz. Nauk* **118** 199 (1976) [*Sov. Phys. Usp.* **19** 109 (1976)]
205. Letokhov V S *Usp. Fiz. Nauk* **125** 57 (1978) [*Sov. Phys. Usp.* **21** 405 (1978)]
206. Letokhov V S *Usp. Fiz. Nauk* **127** 729 (1979) [*Sov. Phys. Usp.* **22** 262 (1979)]
207. Antonov V S, Letokhov V S, Shibanov A N *Usp. Fiz. Nauk* **142** 177 (1984) [*Sov. Phys. Usp.* **27** 79 (1984)]
208. Akhmanov S A, Vysloukh V A, Chirkin A S *Optika Femtosekundnykh Lazernykh Impul'sov* (Optics of Femtosecond Laser Pulses) (Moscow: Nauka, 1988) [Translated into English (New York: AIP Press, 1992)]
209. Kamalov V F "Femtosekundnaya spektroskopiya" ("Femtosecond spectroscopy"), in *Fizicheskaya Entsiklopediya* (Physics Encyclopedia) Vol. 5 (Ed.-in-Chief A M Prokhorov) (Moscow: Bol'shaya Rossiĭskaya Entsiklopediya, 1998) p. 279
210. Anisimov S I, Prokhorov A M, Fortov V E *Usp. Fiz. Nauk* **142** 395 (1984) [*Sov. Phys. Usp.* **27** 181 (1984)]
211. Zel'dovich Ya B *Pis'ma Zh. Eksp. Teor. Fiz.* **19** 120 (1974) [*JETP Lett.* **19** 68 (1974)]
212. Balykin V I, Letokhov V S, Minogin V G *Usp. Fiz. Nauk* **147** 117 (1985) [*Sov. Phys. Usp.* **28** 803 (1985)]
213. Dianov E M, Prokhorov A M *Usp. Fiz. Nauk* **148** 289 (1986) [*Sov. Phys. Usp.* **29** 166 (1986)]
214. Alferov Zh I et al. *Fiz. Tekh. Poluprovodn.* **4** 1826 (1970) [*Sov. Phys. Semicond.* **4** 1573 (1971)]
215. Basov N G, Eliseev P G, Popov Yu M *Usp. Fiz. Nauk* **148** 35 (1986) [*Sov. Phys. Usp.* **29** 20 (1986)]
216. Kilby J S, in *Nobel Prizes in Physics 1996 – 2000* (Ed. G Ekspong) (Singapore: World Scientific, 2002); in *Les Prix Nobel. The Nobel Prizes 2000: Nobel Prizes, Presentations, Biographies, and Lectures* (Ed. T Frängsmyr) (Stockholm: Almqvist & Wiksell Intern., 2001) p. 126; *Usp. Fiz. Nauk* **172** 1102 (2002)
217. Alferov Zh I, in *Nobel Prizes in Physics 1996 – 2000* (Ed. G Ekspong) (Singapore: World Scientific, 2002); *Rev. Mod. Phys.* **73** 767 (2001); *Usp. Fiz. Nauk* **172** 1068 (2002)
218. Kroemer H, in *Nobel Prizes in Physics 1996 – 2000* (Ed. G Ekspong) (Singapore: World Scientific, 2002); *Rev. Mod. Phys.* **73** 783 (2001); *Usp. Fiz. Nauk* **172** 1087 (2002)
219. Basov N G, Grasyuk A Z, Katulyus V F *Dokl. Akad. Nauk SSSR* **161** 1306 (1965)
220. Nasledov D N, Rogachev A A, Ryvkin S M, Tsarenkov B V *Fiz. Tverd. Tela* **4** 1062 (1962) [*Sov. Phys. Solid State* **4** 782 (1962)]
221. Hall R N, Fenner G E, Kingsley J D et al. *Phys. Rev. Lett.* **9** 366 (1962)
222. Alferov Zh I, Gurevich S A "Geteroperekhod" ("Heterojunction"), in *Fizicheskaya Entsiklopediya* (Physics Encyclopedia) Vol. 1 (Ed.-in-Chief A M Prokhorov) (Moscow: Sovetskaya Entsiklopediya, 1988) p. 446
223. Alferov Zh I, Gurevich S A, Korol'kov V I "Geterostruktura" ("Heterostructure"), in *Fizicheskaya Entsiklopediya* (Physics Encyclopedia) Vol. 1 (Ed.-in-Chief A M Prokhorov) (Moscow: Sovetskaya Entsiklopediya, 1988) p. 448

224. Alferov Zh I, Portnoĭ E L "Geterolazer" ("Heterolaser"), in *Fizicheskaya Entsiklopediya* (Physics Encyclopedia) Vol. 1 (Ed.-in-Chief A M Prokhorov) (Moscow: Sovetskaya Entsiklopediya, 1988) p. 445
225. Alferov Zh I et al. *Pis'ma Zh. Tekh. Fiz.* **14** 1803 (1988) [*Sov. Tech. Phys. Lett.* **14** 782 (1988)]
226. Alferov Zh I, in *Proc. of the Intern. Conf. on the Physics and Chemistry of Semiconductor Heterojunctions, Budapest, Oct. 11–17, 1970* Vol. II (Budapest: Académiai Kiadó, 1971) p. 7
227. Alferov Zh I *Sov. Sci. Reviews* (May) 147 (1971)
228. Alferov Zh I "Geteroperekhody v poluprovodnikakh i pribory na ikh osnove" ("Heterojunctions in semiconductors and devices based on them"), in *Nauka i Chelovechestvo* (Science and Mankind) (Moscow: Znanie, 1975)
229. Alferov Zh I "Geteroperekhody v poluprovodnikovoĭ elektronike blizkogo budushchego" ("Heterojunctions in semiconductor electronics of the near future"), in *Fizika Segodnya i Zavtra* (Physics Today and Tomorrow) (Ed. V M Tuchkevich) (Leningrad: Nauka, 1973) p. 61
230. Alferov Zh I, in *Fizika p–n perekhodov* (Physics of *p–n* Junctions) (Ed. A Krogeris) (Riga: Zinatne, 1966) p. 203
231. Alferov Zh I, in *Poluprovodnikovye Pribory i Ikh Primenenie* (Semiconductor Devices and Their Application) Issue 25 (Moscow: Sovetskoe Radio, 1971) p. 204
232. Meindl J D *Sci. Am.* **237** (3) 70 (1977); *Usp. Fiz. Nauk* **127** 297 (1979)
233. Dorfman V F "Integral'naya skhema" ("Integrated circuit"), in *Fizicheskaya Entsiklopediya* (Physics Encyclopedia) Vol. 2 (Ed.-in-Chief A M Prokhorov) (Moscow: Sovetskaya Entsiklopediya, 1990) p. 154
234. Dorfman V F "Mikroelektronika" ("Microelectronics"), in *Fizicheskaya Entsiklopediya* (Physics Encyclopedia) Vol. 3 (Ed.-in-Chief A M Prokhorov) (Moscow: Bol'shaya Rossiĭskaya Entsiklopediya, 1992) p. 152
doi 235. Andreev A F, Velikhov E P, Golant V E et al. "Zhores Ivanovich Alferov (on his seventieth birthday)" *Usp. Fiz. Nauk* **170** 349 (2000) [*Phys. Usp.* **43** 307 (2000)]
236. "Udivitel'noe formirovanie kvantovykh tochek bez zarodysheĭ pri geteroepitaksii SiGe na Si(100)" ("Remarkable nucleationless quantum-dot formation in SiGe/Si(100) heteroepitaxy") *PersT* **9** (6) 3 (2002)
doi 237. Sutter P, Lagally M G *Mater. Sci. Eng. B* **89** 45 (2002)
doi 238. Ustinov V M et al. *Usp. Fiz. Nauk* **171** 855 (2001) [*Phys. Usp.* **44** 813 (2001)]
doi 239. Krestnikov I L et al. *Usp. Fiz. Nauk* **171** 857 (2001) [*Phys. Usp.* **44** 815 (2001)]
240. Rodgers P *Phys. World* **13** (10) 10 (2000)
doi 241. Ginzburg V L *Usp. Fiz. Nauk* **172** 213 (2002) [*Phys. Usp.* **45** 205 (2002)]
doi 242. Hänsel W, Hommelhoff P, Hänsch T W, Reichel J *Nature* **413** 498 (2001)
doi 243. Shafranov V D *Usp. Fiz. Nauk* **171** 877 (2001) [*Phys. Usp.* **44** 835 (2001)]
doi 244. Bondarenko B D *Usp. Fiz. Nauk* **171** 886 (2001) [*Phys. Usp.* **44** 844 (2001)]
doi 245. Goncharov G A *Usp. Fiz. Nauk* **171** 894 (2001) [*Phys. Usp.* **44** 851 (2001)]
doi 246. "Ataka na termoyad" ("Attacking the problem of thermonuclear power"), in *Poisk* (18–19) May 8 (2002) p. 11; Taleyarkhan R P, West C D, Cho J S, Lahey R T (Jr.), Nigmatulin R I, Block R C *Science* **295** 1868 (2002)
247. Morrison D R O, A report given at plenary session of World Hydrogen Energy Conf., Honolulu, Haw., USA, July 24, 1990; in *Sakharov Memorial Lectures in Physics: Proc. of the 1st Intern. Sakharov Conf. on Physics, Moscow, USSR, May 27–31, 1991* (Eds L V Keldysh, V Ya Fainberg) (Commack, N.J.: Nova Science, 1992); *Usp. Fiz. Nauk* **161** (12) 129 (1991)
doi 248. Ninov V et al. *Phys. Rev. Lett.* **83** 1104 (1999)
doi 249. Eroshenko Yu M (Compiler) "Physics news on the Internet" *Usp. Fiz. Nauk* **171** 118, 1116 (2001) [*Phys. Usp.* **44** 106, 1088 (2001)]; http://physicsweb.org/article/news/3/10/16/1; Schön J H, Kloc Ch, Batlogg B *Science* **293** 2432 (2001)
doi 250. Eroshenko Yu M (Compiler) "Physics news on the Internet" *Usp. Fiz. Nauk* **171** 912 (2001) [*Phys. Usp.* **44** 812 (2001)]; Tang Z K et al. *Science* **292** 2462 (2001)
doi 251. "Magnit iz VTSP: svyshe 17 T pri 29 K" ("A HTSC magnet: above 17 T at 29 K") *PersT* **10** (4) 1 (2003); Tomita M, Murakami M *Nature* **421** 517 (2003)
252. "100-kW VTSP elektrodvigatel' uspeshno proshel pervye ispytaniya v MAI" ("The first tests a 100-kW HTSC electric motor have been successfully completed at Moscow Aviation Institute") *PersT* **9** (11) 1–2 (2002)
253. Turner M S *Phys. Today* **54** (12) 10 (2001)
doi 254. Platonov K Yu, Fleĭshman G D *Usp. Fiz. Nauk* **172** 241 (2002) [*Phys. Usp.* **45** 235 (2002)]
doi 255. Hentschel M et al. *Nature* **414** 509 (2001); *Phys. World* **15** (1) 25 (2002)
doi 256. Guthöhrlein G R et al. *Nature* **414** 49 (2001)
257. Steane A *Nature* **414** 24 (2001)
258. "Mart 2002 — data rozhdeniya ekstremal'noĭ litografii" ("March 2002 — birth month of extreme ultraviolet (EUV) lithography") *PersT* **9** (7) 5–6 (2002)
doi 259. Lawler A *Science* **275** 1253 (1997)
doi 260. Braginskiĭ V B *Usp. Fiz. Nauk* **170** 743 (2000) [*Phys. Usp.* **43** 691 (2000)]
doi 261. Braginskiĭ V B *Usp. Fiz. Nauk* **173** 89 (2003) [*Phys. Usp.* **46** 81 (2003)]
doi 262. Pukhov A *J. Plasma Phys.* **61** 425 (1999)
doi 263. Gahn C et al. *Phys. Rev. Lett.* **83** 4772 (1999)
264. DOR, report No. 10738, May 2002 (Moscow: TsNIIAtominform, 2002)
doi 265. Oraevskiĭ A N *Usp. Fiz. Nauk* **171** 681 (2001) [*Phys. Usp.* **44** 647 (2001)]
266. "ALPHA Magnetic Spectrometer (AMS) — novyĭ bol'shoĭ mezhdunarodnyĭ proekt" ("ALPHA Magnetic Spectrometer (AMS): a new large international project") *PersT* **9** (8) 1–2 (2002)
267. Ginzburg V L *Sverkhprovodimost': Fiz. Khim. Tekhnol.* **5** (1) 1 (1992) [*Superconductivity: Phys. Chem. Technol.* **5** 1 (1992)]
doi 268. Pease R F *Nature* **417** 802 (2002)