Gravitational radiation from a rotating protoneutron star

M. V. Sazhin*)

P. K. Shternberg State Astronomical Institute, 119899 Moscow, Russia

S. D. Ustyugov, and V. M. Chechetkin

M. V. Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, 125047 Moscow, Russia
(Submitted 31 July 1997)
Zh. Éksp. Teor. Fiz. **113**, 1153–1166 (April 1998)

This paper discusses the gravitational radiation that accompanies the formation of a rotating protoneutron star. Mainly large-scale entropy inhomogeneities develope inside the star. As a result, bubbles of hot nuclear substance are formed, and convective motions arise: bubbles of hot neutron matter float to the surface of the star, while cold matter sinks to its center. Such large-scale motions of material give rise to an inhomogeneous mass distribution inside the star. Variable asymmetry appears in the mass distribution, and this causes gravitational radiation.
 [S1063-7761(98)00104-8]

1. INTRODUCTION

Explosions of supernovas accompanying the collapse of the core and resulting in the formation of a neutron star or a black hole are one of the main sources of gravitational radiation. Since laser detectors of gravitational waves that are sensitive enough to detect the first bursts of such radiation from space have begun to operate in the last few years, a detailed analysis of possible sources has become especially crucial, including detailed numerical modelling of the physical processes in the sources themselves. Although gravitational radiation can already be regarded as having been detected from the secular variation of the semimajor axis of the double pulsar PSR 1913+16, which gave astronomers confidence in the rapid detection of bursts of gravitational radiation from space, the direct action of gravitational pulses on detectors has not yet been observed. Besides the direct problem of detecting bursts, the problem of what information can be extracted from gravitational pulses is also important.

Coalescing double neutron stars are currently regarded as the most promising source for detecting gravitational pulses.[1,2] Such events are much rarer than are the flashes of supernovas in our galaxy and in the nearby parts of the universe. Gravitational waves are emitted far more often during the flashes of supernovas, and therefore it can be even more promising to detect them than to detect pulses from the coalescence of compact objects: neutron stars, neutron-star–black-hole pairs, or possibly black-hole–black-hole pairs. Pulses from the coalescence of compact objects will probably provide us with information on the epoch of the generation of galaxies, while pulses of gravitational radiation from supernovas make it possible for us to study the initial stage of formation and evolution of protoneutron stars, which will apparently be the most abundant source of bursts of gravitational radiation.

The main problem associated with computing the gravitational radiation accompanying a supernova explosion has for a long time[1–3] been to determine the degree of asymmetry of the explosion. Two main mechanisms for the collapse of a presupernova are known. They are distinguished by the fact that the energy of the propagating shock wave in the first mechanism is sufficient to escape to infinity. The propagation of the shock wave in the second mechanism slows down (because of the intense loss of energy) and is restored after the energy is replenished at the expense of neutron radiation. We shall work in terms of the second mechanism. We neglect the stage of compression of the iron core when a presupernova collapses and the gravitational radiation that accompanies this process. Our treatment will begin with the stage of formation of the protoneutron star. We shall investigate the gravitational radiation that appears at the stage of cooling of a protoneutron star and its evolution into an ordinary neutron star.

Reference 4 discussed a model of presupernova evolution. It consisted of the formation of a rapidly rotating protoneutron star, which, because of instabilities, decays into two components, forming a close-lying pair of protostars. Such a pair rotates around the common center of gravity and produces powerful gravitational radiation.[4,5] This model is similar to those in papers discussing the radiation of gravitational waves from rapidly rotating nuclei of stars and neutron stars.[6–8]

The mechanism that we consider is associated with convective flows of matter in a protoneutron star.[9] Any nonequilibrium processes in the central regions of a protoneutron star increase the entropy. In particular, the nonequilibrium neutronization of matter first noted by Bisnovatyĭ-Kogan and Seĭdov[10] can be such a process; this process was also subsequently considered in Ref. 11. In this case, because of the strong heating in the central region of the star, the mechanism of convective heat dissipation from the center begins to operate, accompanied by the onset of large-scale inhomogeneities of entropy. These produce rather strong asymmetry and result in strong gravitational radiation.[12]

Gravitational radiation caused by asymmetric motions of matter associated with convective processes was also consid-

ered in Ref. 13. However, the authors of that paper obtained the evolution of convection only for small-scale fluctuations of matter. We obtain large-scale convection accompanied by the ascent of large bubbles of hot matter, which is associated with the possible role of nonequilibrium neutronization when a protoneutron star is formed.

Our modelling of the source of gravitational radiation during the explosion of a supernova is associated with the idea of explosive convective instability in a protoneutron star due to nonequilibrium neutronization of matter at high densities. The nonsteady-state cooling of a protoneutron star was simulated numerically, allowing for its rotation and its influence on the development of instabilities. To solve the second part of the formulated problem—finding the shape of a gravitational pulse and the characteristics that allow observers to determine the main parameters of protoneutron stars—we used a numerical code similar to that developed and tested in Ref. 14.

According to current concepts,[15] stars with $M>10M_{\odot}$ at the end of their evolution begin to collapse as a result of the processes of thermal dissociation and electron capture by the nuclei. During the collapse of the iron core of a star about 99% of the gravitational energy is radiated in the form of neutrino radiation. Part of this energy can be imparted to the outer layers of the star and can subsequently cause its shell to be ejected. However, since the characteristic diffusion time of a neutrino is about 10 sec, some mechanism is necessary that would speed up the transport of the neutrino energy to the front of the shock wave. Convection both inside and outside the neutrinosphere has recently been offered as such a mechanism.[15–17] Calculations for the two-dimensional case showed an increase in the time it takes for matter to arrive in the region of heating (as a consequence of convection) behind the front of the shock wave and hence an increase in the radius of the wave. However, when neutrino transport in the region above the neutrinosphere was accurately taken into account, it was shown[18] that convection, having been initially generated, subsequently ceases to be developed and ceases to replenish the energy of the shock wave, and this, in the final analysis, causes it to be damped out.

The process of collapse is extremely inhomogeneous. A collapsing central core with a mass of about $1M_{\odot}$ appears, surrounded by a shell in which the main mass of the star is concentrated.[15] The collapsing core subsequently reaches nuclear densities, which causes an expanding shock wave to be formed, and the process of nonequilibrium neutronization of this core begins. Such a neutronization process causes the entropy to increase at the center of this core; i.e., hot matter is formed at the center of the core. This structure is hydrodynamically unstable.

Reference 9 discussed the hydrodynamic growth mechanism of neutrino radiation, based on Ref. 19. It discussed the development of convective instability in a gaseous sphere. Large-scale instabilities grow most rapidly in this case. Analytical estimates showed that the inner hot layers are transported to the surface of the star in a characteristic time $\tau \sim R/v_{\text{sound}}$. Reference 9 described calculations of the development of hydrodynamic instability in a protoneutron star with excess entropy at the center for the three-dimensional case. The inviscid Euler equations were solved numerically. The TVD difference scheme with coordinate splitting was used (see Ref. 9 for more detail). The characteristic distribution of rising matter in the form of large-scale entropy variations gave rise to shapes reminiscent of the mushroom clouds of a nuclear explosion in the earth's atmosphere, escaping from the center of the star. The same paper includes figures that show contours of constant entropy as a function of time. Since the calculation was done in the adiabatic approximation, increased entropy corresponds to reduced density.

In the simplest model under consideration (without rotation or a magnetic field), six bursts appear, and the gravitational radiation is small in this case. It should be emphasized that large-scale inhomogeneities grow most rapidly, as follows from the analysis carried out in Ref. 19.

The characteristic times for the inhomogeneities to develop were about 4 ms when the characteristic scale of the ascending region was about 20 km. The characteristic time for a bubble to reach the surface of the protoneutron star was 1 ms, which corresponds to a mean velocity of $c/150$. The gravitational radiation from such a process, calculated in Ref. 12, is not very great. However, breakdown of the symmetry of the picture, caused, say, by rotation or by the presence of a magnetic field in the protostar, increases the gravitational radiation.

If the rotational velocity in a protoneutron star is nonzero (this assumption is quite reasonable), the symmetry of the problem breaks down. In the present paper, we consider the hydrodynamic instability that arises in a protoneutron star for two cases. The first case involves the development of instabilities when rotation is absent, and the second case involves the development of instabilities when weak rotation is present. In both cases, the gravitational radiation from such a source was computed.

2. MODELLING OF THE HYDRODYNAMIC INSTABILITY

In calculating the density and temperature distributions inside a protoneutron star, we used the following values of the central density and central temperature: $\rho_c=2\times10^{14}$ g/cm^3 and $T=10^{11}$ K. It was assumed that relativistic degenerate electrons and an ideal nonrelativistic Fermi gas of nucleons contribute to the equation of a state of the substance inside the star. Note that excess entropy was produced close to the center of the star ($r=0$) by the process of nonequilibrium neutronization.

The equations of hydrodynamics used to model the processes that accompany the collapse of a protoneutron star in the three-dimensional case have the form

$$\rho\frac{d\mathbf{v}}{dt}=-\operatorname{grad}P-\frac{\rho GM}{r^3}\mathbf{r}, \tag{1}$$

$$\frac{d\rho}{dt}+\rho\operatorname{div}\mathbf{v}=0, \tag{2}$$

$$\frac{dE}{dt}=T\frac{dS}{dt}+\frac{P}{\rho^2}\frac{d\rho}{dt}, \tag{3}$$

$$\frac{dS}{dt}=0. \tag{4}$$

Here ρ is the mass density, $\mathbf{v}$ is the material velocity, P is the pressure inside the protoneutron star, E is the energy density, and S is the entropy. All the quantities are functions of three spatial coordinates and time.

To carry out numerical calculations, it is necessary to choose dimensionless quantities. The physical variables of the given problem were reduced to dimensionless form by dividing by the corresponding quantities: density was expressed in units of $\rho_0=2\times10^{14}$ g/cm^3, length in units of $L_0=2\times10^7$ cm, time in units of $t_0=7.74\times10^{-3}$ sec, mass in solar masses, $M_\odot=2\times10^{33}$ g, and temperature in units of $T=10^{11}$ K. These units were chosen for convenience of the computations.

An explicit Godunov-type conservative TVD difference scheme was used in the calculations. This means that the system is written in so-called divergence form, where the symbolic density vector $\mathbf{U}$ appears in a partial derivative with respect to time, while the density-flux vector $\mathbf{F}$ appears in a partial spatial derivative:[20]

$$\frac{\partial\mathbf{U}}{\partial t}+\frac{\partial\mathbf{F}}{\partial x}=0.$$

Three-dimensional space was broken up into cubic cells with a constant step of the lattice. All the variables of the density vector $(\rho,\rho\mathbf{u},\rho S)$ were referred to the centers of the cells, while the fluxes of these variables $(\rho\mathbf{u},\rho u^2,\rho Su)$ were computed on the boundaries between the cells. The variables were computed at time t' by solving the one-dimensional problem for each spatial direction. Cyclic permutation of these directions was used to maintain second-order accuracy. After this, a function consisting of the source caused by the action of the gravitational field of the star, which was taken as constant in the time of the calculations, was added to the equations. The calculations were carried out on the Convex computer of the Institute of Applied Mathematics, Russian Academy of Sciences. One simulation was carried out on a three-dimensional $51\times51\times51$ lattice with a step of $h=0.0025$, which corresponded to a maximum distance of the bubble from the center of $r_0=0.0675$. A second model was calculated on the same lattice with a step of $h=0.004$, with the bubble at $r_0=0.1$ from the center. This was done in order to estimate how the total gravitational radiation depends on the maximum distance of the bubble from the center. The time to calculate the first model was 15 h. The time to do the calculation in the second model changed slightly because of the increase in the timestep.

In solving the hydrodynamic problem, data on the density had to be held in the computer memory for values from ρ_0 to $10^{-3}\rho_0$. This made it possible to trace the ascent of the bubble to values of $r_0=0.1$ and no further. In order to completely trace the evolution of a bubble from the center to the neutrinosphere, it is necessary to find the law of motion of a bubble out to $r_0=0.17$. This requires either the calculational technique to be changed or the lattice step to be reduced from 0.004 to 0.0001, which makes the time of the calculation unacceptably large. We calculated the total luminosity in the form of gravitational radiation for both models and used extrapolation to estimate the total luminosity for the case in which a bubble reaches the boundaries of the neutrinosphere.

At the initial instant, the self-consistent field method was used to obtain the equilibrium configuration, taking into account rotation.[21]

The calculations in this paper were carried out for two cases. In the first case, in order to test the equilibrium reached by a star, rotation was neglected. In the second case, weak rotation was taken into account, with the ratio of the rotational kinetic energy T to the gravitational energy $|W|$ being

$$T/|W|=0.01.$$

The angular velocity of the protoneutron star was assumed to be fairly small to avoid Jacobi instabilities, which cause a star to rupture into two components. The kinetic and potential energies were computed from integral representations, taking into account the density profile obtained in the calculation. The rotational period of a star corresponding to a ratio of kinetic to potential energy of 0.01 equalled 14 ms. The coordinate system was chosen so that the plane of rotation of the star coincided with the xy plane. This means that the star's angular velocity vector has the components

$$\Omega_x=0,\quad \Omega_y=0,\quad \Omega_z=\Omega=\text{const};$$

i.e., we are considering solid-body rotation.

The entropy distributions for the second case are shown in Fig. 1. Two cross sections of the star are chosen as images. In the first cross section (Figs. 1a–1i), the angular velocity vector lies in the image plane. In other words, the horizontal axis corresponds to the x axis, while the vertical axis corresponds to the z axis of our coordinate system. Figures 1a′–1i′ correspond to a view from above onto the plane of rotation, corresponding to the plane of the equator. The initial configuration is chosen at time $t=0.075$ ms, and the final configuration is shown for time $t\approx6.31$ ms. The complete evolution lasted 20 ms.

Unlike the model considered in Refs. 9 and 12, in this case two bubbles initially appear (at about 3 ms), elongated in opposite directions along the axis of rotation. Four additional bubbles appear soon afterwards (at about 5 ms) and lie in the plane of rotation of the protoneutron star. Such an entropy distribution breaks the symmetry of the pattern considered in Ref. 12 and causes quadrupole gravitational radiation to appear. The bubbles located along the axis of rotation are broken off the hot core by the first bubbles and float to the surface. This occurs because the density varies more rapidly along the axis of rotation. In the interval between these bubbles, cooler matter sinks to the center of the protoneutron star. The bubbles located in the plane of rotation break off the hot nucleus core later and also float to the surface. Our calculations show that, following the first bubbles, additional bubbles with significantly less volume are formed, which also begin to float to the surface.

The first stage of the development of the perturbations is the growth of asymmetry along the axis of rotation of the star (Figs. 1a–1c). which lasts for 3 ms. The asymmetry is still very weakly expressed in the plane of rotation (Figs.

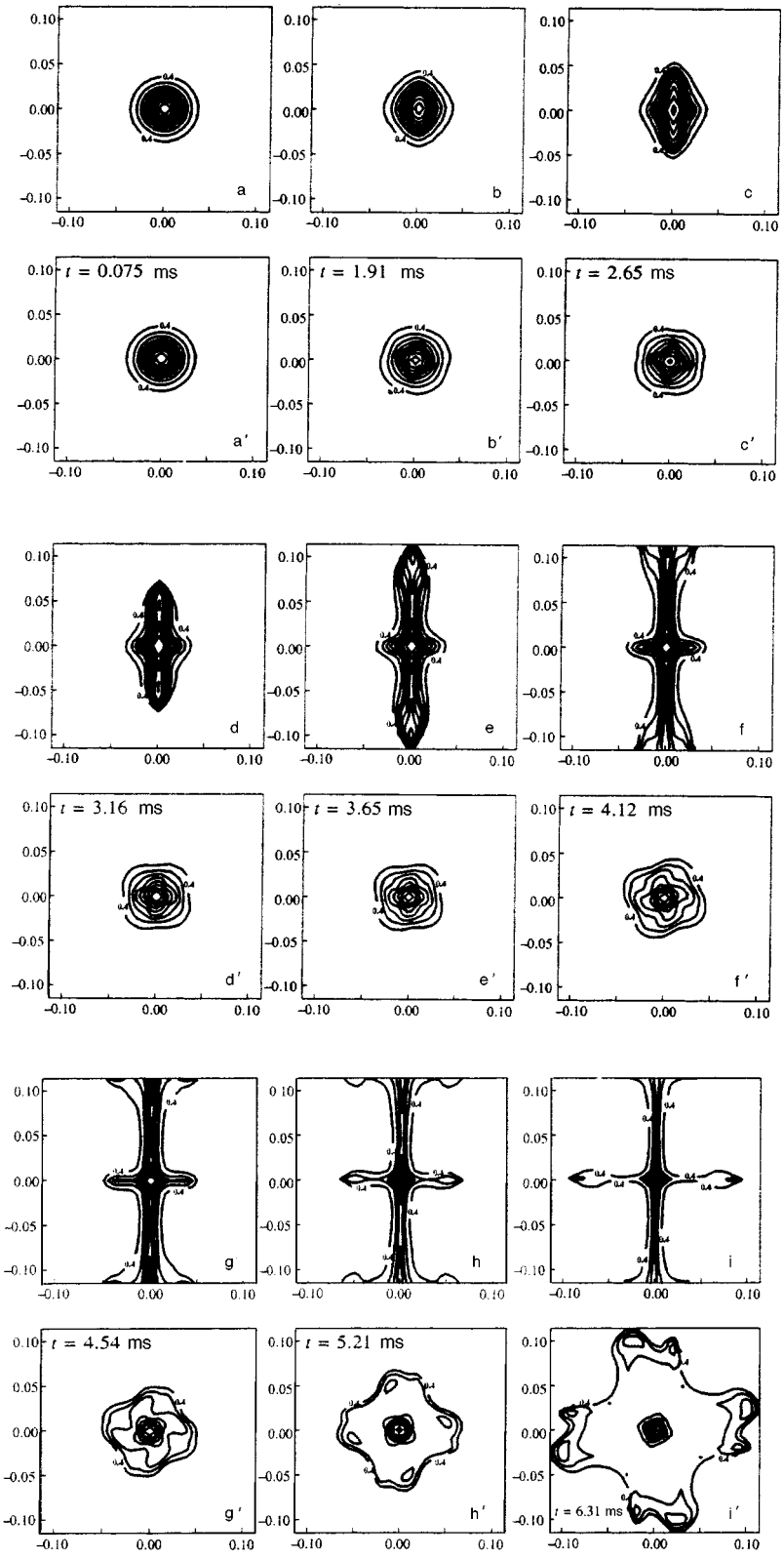


FIG. 1. Distribution of levels of constant entropy in the zx plane (a)–(i) and in the equatorial plane (a′)–(i′). The length scale along the axes shows the fraction of the characteristic length: 0.1 corresponds to 20 km in linear measure. The entropy is expressed in dimensionless units (normalized to Boltzmann's constant and the nucleon density).

1a′–1c′), and the evolution proceeds peacefully. The asymmetry of the distribution of the substance is negligible during this stage, and bubbles of hot matter have not yet been formed. Gravitational radiation of the star, as will be shown below, is also negligible at this stage. Bubbles moving along the axis of rotation have already been formed for the stages shown in Figs. 1d–1f, and bubbles in the plane of rotation (Figs. 1d′–1f′) begin to be distinguished at this stage. This instant corresponds to the most intense gravitational radiation. As can easily be seen, the main gravitational radiation occurs from bubbles rising along the axis of rotation of the protoneutron star. The stages shown in Figs. 1g′–1i′ correspond to the final formation of bubbles in the plane of the equator and the beginning of their ascent to the surface. The

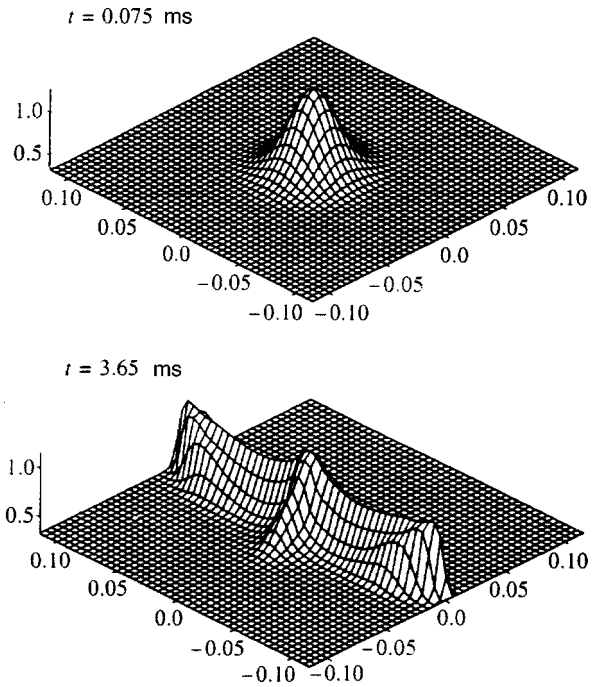


FIG. 2. Three-dimensional profile of the entropy for two characteristic times.

bubbles along the axis of rotation (Figs. 1g–1i) have already escaped from the calculated region; i.e., they are farther than 20 km from the center.

Figure 2 shows the three-dimensional profile of the entropy distribution initially ($t=0.075$ ms) and when the bubbles have already formed ($t=3.65$ ms).

On figures with the velocity-field distribution (these figures are not shown here), convective cells are clearly visible that show both the rise of hot matter to the surface and the fall of cold matter to the center of the protoneutron star.

3. GRAVITATIONAL RADIATION

We shall calculate the gravitational radiation emitted by a protoneutron protostar from the quadrupole formula of Ref. 22. The transverse-traceless part of small perturbations of the metric is determined from

$$h_{ab}^{TT}=\frac{2G}{Rc^4}\,\ddot{I}_{ab}^{TT},$$

where R is the distance from the star to the observer, and I_{ab}^{TT} is the projection of the tensor of the quadrupole moment of the mass distribution inside the star onto the operator

$$P_{ab}=\delta_{ab}-n_a n_b$$

(n_a is the unit vector directed from the star to the observer). The tensor of the quadrupole moment of the mass distribution is defined by[22]

$$I_{ab}(t)=\int \rho(\mathbf{r},t)\left(r_a r_b-\frac{1}{3}\,\delta_{ab}r^2\right)dV.$$

In these equations, the subscripts a and b take the values 1, 2, and 3, while a dot denotes the derivative with respect to time.

A gravitational wave is characterized by two degrees of freedom and accordingly by two independent amplitudes, usually denoted by h_+ and $h_\times$. Here we shall not evaluate each of these two quantities in detail, but shall restrict ourselves to a computation of the luminosity of the source of gravitational radiation and the total quantity of energy given off when a star collapses.

The luminosity of a star in the form of gravitational radiation is determined by[22]

$$\dot{E}=\frac{G}{5c^5}\,\langle \dddot{I}_{ab}\dddot{I}^{ab}\rangle,$$

while the total amount of energy given off in the form of gravitational radiation is calculated from the integral formula

$$\Delta E=\int \dot{E}dt.$$

These equations, except for the last, contain the second or third derivative with respect to time of the quadrupole tensor I_{ab}. In analytical calculations, taking the corresponding derivative with respect to time does not affect the accuracy of the computations. The situation is different in numerical simulation. Even though evaluating an integral in numerical calculations (for example, by means of Simpson's rule) is a simple problem that arouses no concern about the accumulation of numerical errors, differentiation is a nontrivial problem. This is because, when the first difference is computed and then divided by the step, small errors (for example, rounding errors) grow in magnitude. For the third derivative, the corresponding errors can already be comparable with or even greater than the quantity itself, thereby introducing unacceptable errors into the computation. Numerical differentiation is an ill-posed problem. There are several methods of overcoming this ill-posedness.

We shall use the method considered in detail in Ref. 14. The essence of this method consists of the transformation of the first derivative of I_{ab} with respect to time. In the expression under the integral, the first derivative with respect to time appears only as the partial derivative of the density of the substance. Using the continuity equation

$$\frac{\partial\rho}{\partial t}+\mathrm{div}(\rho\mathbf{v})=0,$$

we transform the time derivative to derivatives of the density of the substance and its velocity with respect to the spatial coordinates. Next the expression is integrated by parts. Assuming that there are no mass fluxes at infinity, the first derivative of the quadrupole moment tensor with respect to time can be reduced to

$$\dot{I}_{ab}=2\int \rho\left\{\frac{1}{2}\,v_a r_b+\frac{1}{2}\,v_b r_a-\frac{1}{3}\,\delta_{ab}(\mathbf{v}\cdot\mathbf{r})\right\}dV.$$

The velocities that appear under the integral sign were already computed when the self-consistent picture of motion of the substance inside the star was constructed. The accuracy of their computation was checked when the development of large-scale inhomogeneities was modelled. Such an

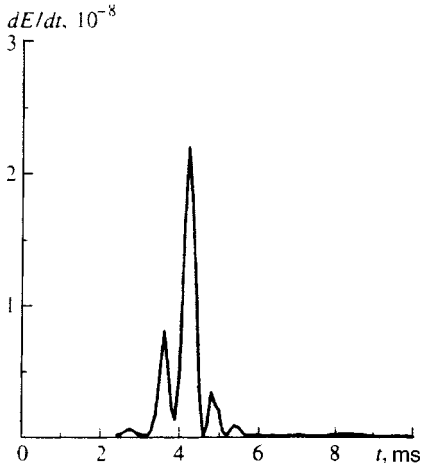


FIG. 3. Luminosity curve (in units of $M_\odot c^2$/ sec) of a protoneutron star in the form of gravitational radiation.

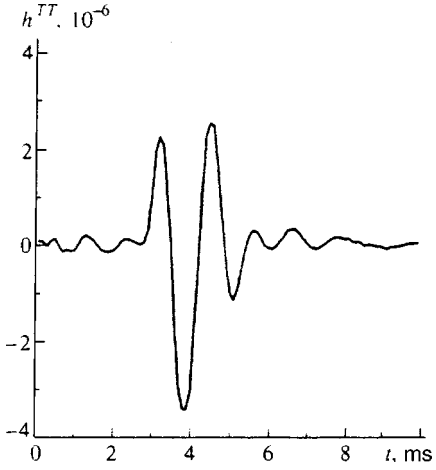


FIG. 4. Pulse profile. The dimensionless amplitude of the gravitational wave is plotted on the vertical axis in units of 5.94×10^{-17} for a distance of 10 kpc to the protoneutron star.

approach makes it possible to avoid one differentiation with respect to time and to calculate the gravitational radiation with satisfactory accuracy.

The luminosity of the star, taking into account the factors that reduce the problem to dimensionless form, is

$$\dot{E}=\frac{G\rho_0^2 L_0^{10}}{c^5 t_0^6}\,\dot{e}.$$

Here $\dot{e}$ is the dimensionless luminosity of the star in the form of gravitational radiation.

A problem arises when computing the mean values over time, denoted in the analytical formulas by angle brackets $\langle\ \rangle$. The difficulty is that the period over which the values are averaged in the theoretical formulas is the rotational period of the star, which is comparable in magnitude with the duration of the process itself or, more precisely, even somewhat exceeds it. A protostar makes one rotation around its axis in 14 ms, whereas the bubbles ascend in characteristic times of about 5 ms. In general, there are two characteristic time intervals in this process: the first equals the characteristic rise time of the bubbles, and the second is the rotational period of the protoneutron star. In order to obtain a sufficiently smooth curve describing the gravitational radiation, we chose an artificial averaging time interval. The time step in the calculations was variable; it was given automatically inside the program, where a short interval at the initial stage of the calculations changed to a rather long one at the final stage. Thus, the time step was about 200 μs by the time the maximum of the first peak occurred, at about 3 ms, whereas it was initially about 100 μs. By trial-and-error we chose an averaging time of 0.5 ms, which gave a fairly smooth luminosity curve without averaging out the main details of the process. This curve is shown in Fig. 3. This figure shows the time dependence of the luminosity $L=dE/dt$ in units of $M_\odot c^2$/sec for the case with rotation ($T/|W|=0.01$) for the second model. The luminosity at the maximum of the burst is $2.2\times10^{-8}M_\odot c^2$/sec. In the first model, this value was $3.4\times10^{-9}M_\odot c^2$/sec. The change of the luminosity is easily explained by the fact that, in the second model, the bubbles are twice as far from the center, increasing the effective arm, which increases the quadrupole moment and increases the gravitational radiation. The bubbles begin to move almost uniformly when they reach the limits of the neutrinosphere, and then the gravitational radiation from them virtually ceases. In order to estimate the total gravitational radiation in the process under consideration, we extrapolated the values found for the total luminosity to $r_0^{(\max)}=0.17$, i.e., to the limit of the neutrinosphere. The resulting luminosity is $4.5\times10^{-7}M_\odot c^2$/sec.

The total energy emitted in the form of gravitational waves in the course of the entire process can reach a value of

$$\Delta E\approx 2.5\times10^{-10}M_\odot c^2$$

taking into account rotation when the ratio of the kinetic energy to the gravitational binding energy is 0.01.

Figures 4 and 5 show, respectively, the pulse profile of h^{TT} and the spectral power distribution in a pulse of gravitational radiation.

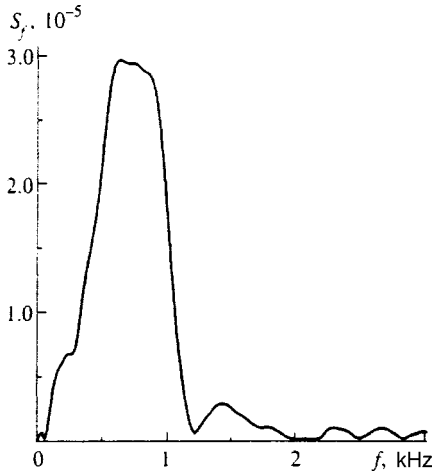


FIG. 5. Spectral power of the pulse. The quantity S_f—the square of the Fourier component h^{TT}—is plotted along the vertical axis. The maximum spectral power is reached at a frequency of 700 Hz.

4. CONCLUSION

We shall introduce additional determinations of the amplitude of gravitational radiation and estimate the amplitude observed on earth from the flashes of supernovas at the center of our galaxy and in a nearby galaxy.

We introduce the notation

$$h^2(\theta,\varphi)=\frac{1}{2}\,(h_+^2(\theta,\varphi)+h_\times^2(\theta,\varphi))$$

and we call the quantity $h(\theta,\varphi)$ the mean amplitude of the gravitational wave. The flux of gravitational radiation is expressed in terms of the introduced mean amplitude of the wave as

$$I=\frac{\pi c^3}{2G}\,\nu^2 h^2.$$

The luminosity in terms of flux is determined from

$$\dot{E}=R_0^2\int_{-1}^{1}d\mu\int_0^{2\pi}d\varphi I(\theta,\varphi).$$

The directionality diagram of the radiator of gravitational waves considered here has a complex structure. However, for our estimates, we restrict ourselves to the approximation of a quadrupole dependence of the radiation flux on angles. Then the relationship of the wave amplitude h determined here and the luminosity of the star will have the form

$$h=3.7\times10^{-19}\sqrt{\frac{\dot{E}\cdot 1\ \text{sec}}{M_\odot c^2}}\,\frac{10\ \text{kpc}}{R_0}\,\frac{1\ \text{kHz}}{\nu}.$$

We shall estimate the amplitude of the gravitational radiation when a supernova explodes at the center of our galaxy. Naturally, we shall consider the rotation to be small, i.e., about 1%.

Then

$$h=10^{-22}\sqrt{\frac{\dot{E}\cdot 1\ \text{sec}}{10^{-7}M_\odot c^2}}\,\frac{10\ \text{kpc}}{R}\,\frac{1\ \text{kHz}}{\nu},$$

where R is the distance from the center of the galaxy. However, such pulses rarely appear, about once every 30 y. The amplitude of the pulses from protostars possessing slow rotation arriving, say, once a month is

$$h=5\times10^{-26}\sqrt{\frac{\dot{E}\cdot 1\ \text{sec}}{10^{-7}M_\odot c^2}}\,\frac{20\ \text{kpc}}{R}\,\frac{1\ \text{kHz}}{\nu}.$$

Note that we have restricted ourselves in this article to demonstrating that rotation (even slow rotation) breaks the symmetry of the convective processes and increases the gravitational radiation by comparison with that of a nonrotating star. In Ref. 12, we estimated the gravitational radiation when the symmetry was broken as a consequence of rotation. An exact calculation of the energy output is given above. However, it should be mentioned that a long ''comb'' of pulses was not obtained in our calculations, as was assumed in the preceding article.

To estimate the actual gravitational radiation pulses for the flashes of supernovas, the gravitational energy radiated during faster rotations of the protostars needs to be calculated and the fraction of supernovas that possess such rotation also needs to be estimated.

The authors are grateful to A. A. Starobinsky and M. E. Prokhorov for fruitful discussions. This work was supported by the Kosmion Scientific Center and the Astronomiya Federal Program, Computational Astrophysics Section, as well as by Grant No. 97-02-16486 of the Russian Fund for Fundamental Research.

*)E-mail: sazhin@sai.msu.su

[1] B. F. Schutz, in TAUP-89, ed. Frontieres, Gif-sur-Yvette, France, 1989, p. 167.
[2] K. S. Thorne, in *Three Hundred Years of Gravitation* (Cambridge Univ. Press, Cambridge, 1987).
[3] L. M. Ozernoĭ, JETP Lett. **2**, 52 (1965).
[4] V. S. Imshennik, Pis'ma Astron. Zh. **18**, 489 (1992) [Sov. Astron. Lett. **18**, 194 (1992)].
[5] A. F. Zakharov, Astron. Zh. **73**, 605 (1996) [Astron. Rep. **40**, 552 (1996)].
[6] B. Schutz, Class. Quantum Grav. **6**, 1761 (1989).
[7] S. C. Smith, J. L. Houser, and J. M. Centrella, Astrophys. J. **458**, 236 (1996).
[8] D. Lai and S. L. Shapiro, Astrophys. J. **442**, 259 (1995).
[9] V. M. Chechetkin, S. D. Ustyugov, A. A. Gorbunov, and V. I. Polezhaev, Pis'ma Astron. Zh. **23**, 34 (1997) [Astron. Lett. **23**, 30 (1997)].
[10] G. S. Bisnovatyĭ-Kogan and Z. F. Seĭdov, Astron. Zh. **47**, 139 (1970) [Sov. Astron. **14**, 113 (1970)].
[11] V. S. Imshennik and V. M. Chechetkin, Astron. Zh. **47**, 929 (1970) [Sov. Astron. **14**, 747 (1970)].
[12] M. V. Sazhin, S. D. Ustyugov, and V. M. Chechetkin, JETP Lett. **64**, 871 (1996).
[13] E. Muller and H.-T. Janka, Astron. Astrophys. **317**, 140 (1997).
[14] S. C. Smith, J. M. Centrella, and S. P. Clancy, Astrophys. J., Suppl. Ser. **94**, 789 (1994).
[15] H. A. Bethe, Rev. Mod. Phys. **62**, 801 (1990).
[16] M. Herant, W. Benz, and S. A. Colgate, Astrophys. J. **395**, 642 (1992).
[17] A. Burrows, J. Hayes, and B. A. Fryxell, Astrophys. J. **450**, 830 (1995).
[18] A. Mezzacappa, A. C. Calder, S. W. Bruenn, J. M. Blondin, M. W. Guidry, M. R. Strayer, and A. S. Umar, E-print archive astro-ph 9612107.
[19] N. M. Zueva, M. S. Mikhaĭlova, and L. S. Solov'ev, Preprint Inst. Prik. Mat., Akad. Nauk SSSR No. 65 (1976).
[20] A. A. Samarskiĭ, Yu. P. Popov, *Various Methods of Solving Gas-Dynamics Problems* (Nauka, Moscow, 1980).
[21] K. Hachisu, Astrophys. J., Suppl. Ser. **62**, 461 (1986).
[22] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Nauka, Moscow, 1988; Pergamon Press, New York, 1971).
[23] L. Blanchet, T. Damour, and B. Iyer, Phys. Rev. Lett. **74**, 3515 (1995).
[24] K. S. Thorne, in *Particle Astrophysics* (Gif-sur-Yvette, France, 1993), p. 375.

Translated by W. J. Manthey

Forming stable nonlinear lightguides with cross-modulation self-locking of incoherent soliton modes

V. A. Vysloukh, V. Kutuzov, V. M. Petnikova, and V. V. Shuvalov*)

International Scientific Training Laser Center, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia
(Submitted 26 June 1997)
Zh. Éksp. Teor. Fiz. **113**, 1167–1180 (April 1998)

This paper discusses a new class of spatially localized soliton solutions with bounded energy in the problem of the nonlinear propagation of a light beam along a photorefractive crystal (PRC) with a drift mechanism of nonlinear response. Solitons of this class correspond to stable propagation in the PRC of two or more wave packets (components) that are matched in intensity distribution but incoherent. Their spatial distributions correspond to simultaneous cross-modulation self-locking of its first- and higher-order modes into the common nonlinear waveguide formed by them. It is shown that such multicomponent solitons can be implemented and that they are structurally stable both with respect to collisions and with respect to substantial (more than 10% in intensity) stochastic perturbations of their components as they propagate to distances of the order of several diffraction lengths. The paper discusses the possibility that conjugate polymers, antiferromagnets, and superconductors can contain unusual states corresponding to the stable propagation of two or more mutually incoherent but matched electron wave packets along conjugate chains or along isolated atomic planes.
 [S1063-7761(98)00204-2]

1. INTRODUCTION

One of the most interesting problems of modern laser physics is to investigate the self-organization processes that occur in systems consisting of nonlinear media and a light field. The stable self-consistent solutions of problems of this type (solitons) in media with spatially localized (local) nonlinear response—the so-called Kerr nonlinearity—has been thoroughly studied.[1] The concepts of one-component and two-component (vector) solitons as self-consistent spatially localized solutions (in essence, normal modes) of nonlinear problems have become firmly established in very different fields of modern physics. This includes fiber optics and the optics of supershort light pulses,[1–3] nonlinear optics and laser spectroscopy,[4–6] the physics of quasi-one-dimensional chains in ferromagnets,[7,8] high-temperature superconductors,[9] conjugate polymers,[10–12] etc. In terms of simulations, very interesting results have recently been obtained in studying solitons and multisoliton solutions, as well as stable soliton pairs in photorefractive crystals (PRCs). The importance of such studies is that the mechanism by which optical nonlinearity is formed in PRCs is one of the strongest mechanisms. It can be observed with light intensities of only a few mW/cm^2.[13] Beginning with the pioneering work of Refs. 14–16 in PRCs with a drift (local) mechanism for the formation of a nonlinear response,[17] active studies have been carried out in so-called bright,[18,19] dark,[20] gray,[21] vector,[22] and vortex[23] solitons, multisoliton solutions,[24] and questions of their propagation and interactions,[25,26] spatial dimensionality,[27,28] and stability.[29–31] It has been shown, for example, that it is possible to form soliton lightguides in the interior of a PRC, with subsequent trapping and propagation along them of relatively weak (in intensity) light beams.[24,32] Provided that the wavelength of the strong beams does not fall into the region of photorefractive sensitivity of the crystal, it is possible for them to propagate in lightguides induced in PRCs by low-intensity spatial solitons. The formation of stable pairs of two incoherent spatial solitons of any of the types enumerated above (bright–bright, bright–dark, dark–dark) has been observed.[33,34] In terms of applications, these results open up wide prospects for controlling light with light in systems for transporting and processing optical information.

The object of this paper is to analyze a new class of spatially localized stable soliton solutions with limited energy, corresponding to a bound state of two or more incoherent light beams—the components of such a soliton—in a PRC with a drift (local) nonlinear response mechanism. In terms of the character of the light-field distribution, solutions of this class correspond to the simultaneous self-locking and interlocking of several of its of zero- and higher-order modes into the common nonlinear lightguide formed by them. The interaction of the locked modes in this case has a cross-modulation (reactive) character and is not accompanied by energy-exchange processes.

2. THE STARTING MODEL

The model that we used for the nonlinear response of a PRC is based on the well-known[35] steady-state solution of the classical system of material equations for an internal electric field $E_{sc}(x,z)$, written for the two-dimensional case neglecting the photovoltaic effect,[17]

$$E_{sc}=\frac{1}{aI_0(\chi+1)(\lambda_1-\lambda_2)} \times\sum_{m=0}^{\infty}\left(\frac{\lambda_1}{\lambda_2^{m+1}}-\frac{\lambda_2}{\lambda_1^{m+1}}\right)\frac{\partial^m I(x,z)}{\partial x^m},$$
$$a=\frac{\varepsilon}{4\pi eN_a},\quad \chi=\frac{N_a}{N_d-N_a},\tag{1}$$
$$\lambda_{1,2}=\frac{eE_0}{2\theta}\pm\sqrt{\left(\frac{eE_0}{2\theta}\right)^2+\frac{e}{a\theta(\chi+1)}}.$$

Here N_a and N_d are the acceptor and donor concentrations, respectively; I_0 is a parameter that describes the dark conductivity of the PRC and determines the rate of dark ionization of the donors as sI_0, where s is the photoionization cross section; $I(x,z)$ is the radiation intensity; ε is the static permeability; e is the charge of a free carrier taking into account its sign, i.e., negative for electrons and positive for holes; and θ is the temperature of the PRC in energy units. It is assumed that an external static electric field E_0 is applied to the PRC in the transverse direction (along the x axis). It is also assumed that the optical radiation propagates along the z axis and that $N_a\gg n$, $I_0\gg I$, and $a(\partial E/\partial x)\ll 1$. Here n is the free-carrier concentration. For $E_0\sim 10$ kV/cm, $\theta\sim 300$ K, when the spatial scales of variation of the light field exceed the wavelength of the light, the terms of the series in Eq. (1) decrease so rapidly that only the first term, proportional to $I(x,z)$, need be kept in it, so that we will call this the local component of the nonlinear response. In fact, this means that we restrict ourselves in this paper to so-called Kerr nonlinearity.[1]

The propagation of a light beam with complex field amplitude $A(x,z)$ was described by the standard truncated wave equation[1]

$$i\frac{\partial A}{\partial z}=\frac{1}{2k}\frac{\partial^2 A}{\partial x^2}+k\frac{\delta\eta}{\eta}A,\tag{2}$$

written in the paraxial approximation, neglecting absorption. Here k is the wave number, $\delta\eta=-(r_{\rm eff}\eta^3/2)E_{sc}(x,z)$ is the nonlinear addition to the refractive index η, and $r_{\rm eff}$ is the effective electrooptic coefficient. Equation (2) omits the homogeneous-in-x refractive-index variation caused by E_0. Equations (1) and (2) form a self-consistent problem that takes into account the mutual influence of the redistribution processes of the light intensity and electric field in the interior of the PRC. This model gives a good description of experiments with so-called slit beams,[36] which are widely used in practice when studying soliton effects in PRCs because of the strong anisotropy of the nonlinear response of the latter.

3. TWO-COMPONENT SOLITONS

We shall attempt to find two-component solitonlike solutions of the system formed by Eqs. (1) and (2), corresponding to separation of variables,

$$A(x,z)=Y_1(x)\exp(-i\nu_1 z)+Y_2(x)\exp(-i\nu_2 z),\tag{3}$$

where the real functions $Y_{1,2}(x)$ give the spatial distribution in x of two mutually incoherent components of the light field, while the positive constants $\nu_{1,2}$ determine their nonlinear phase damping as a function of z. We are thus dealing with the search for intensity distributions that are stationary along z. It is easy to see, taking into account the mutual incoherence of the $Y_{1,2}$ components, that substituting Eq. (3) into Eqs. (1) and (2) results in a system of equations for the spatial profiles of their amplitudes in the form

$$\frac{d^2Y_{1,2}}{dx^2}+2k\left[\frac{\alpha_0E_0}{I_0}(Y_1^2+Y_2^2)-\nu_{1,2}\right]Y_{1,2}=0,\tag{4}$$

where $\alpha_0=(1/2)kr_{\rm eff}\eta^2$, and only the first term of the expansion is retained in Eq. (1). It is convenient to introduce into Eq. (4) the dimensionless coordinates $\xi=x/x_0$ and $\zeta=z/L_d$ and the amplitudes of the field components $\rho_{1,2}(\xi)=Y_{1,2}(\xi)\sqrt{R/I_0}$, where x_0 is determined by the characteristic transverse scale of the problem, for example, by the width of one of the light beams; $L_d=kx_0^2$ is the diffraction length corresponding to x_0; $R=L_d/L_r$; and $L_r=|1/\alpha_0E_0|$ is the nonlinear refraction length. In this case, Eq. (4) transforms into a system of equations for the dimensionless amplitudes:

$$\frac{d^2\rho_{1,2}}{d\xi^2}\pm 2[\rho_1^2+\rho_2^2-\beta_{1,2}]\rho_{1,2}=0,\tag{5}$$

where we have written $\beta_{1,2}=L_d\nu_{1,2}$, and the $+$ and $-$ signs correspond to the cases of focusing ($\delta\eta>0$) and defocusing ($\delta\eta<0$) nonlinearity. Both these cases can be implemented by appropriately choosing the orientation of the PRC and the direction of E_0.[17] It is easy to see that the system of Eqs. (5) retains the same form for oppositely propagating light beams.

Depending on the sign of the nonlinearity, the system of Eqs. (5) has well-known particular solutions, having the character of single-component bright solitons,[18,19]

$$\rho_1=\pm\frac{\sqrt{2\beta_1}}{\cosh(\sqrt{2\beta_1}\xi)},\quad \rho_2=0,$$
$$\rho_1=0,\quad \rho_2=\pm\frac{\sqrt{2\beta_2}}{\cosh(\sqrt{2\beta_2}\xi)},\tag{6}$$

or dark solitons,[20]

$$\rho_1=\pm\sqrt{\beta_1}\tanh(\sqrt{\beta_1}\xi),\quad \rho_2=0.$$
$$\rho_1=0,\quad \rho_2=\pm\sqrt{\beta_2}\tanh(\sqrt{\beta_2}\xi).\tag{7}$$

Such solutions are stable in principle, even though the necessary one-dimensionality of the diffraction and self-action processes is attained in real experiments only by using slit beams having a homogeneous field distribution in the direction orthogonal to ξ and ζ.[36] In this case, one has to consider the possibility that modulational instability will develop along the PRC,[37] leading to filamentation of the beam in this direction. The stability of the two-dimensional solitons is due to the nonlinearity-saturation effect.

In analyzing two-component self-consistent solutions with a form more general than given by Eqs. (6) and (7), it is convenient to use the mathematical analogy between the

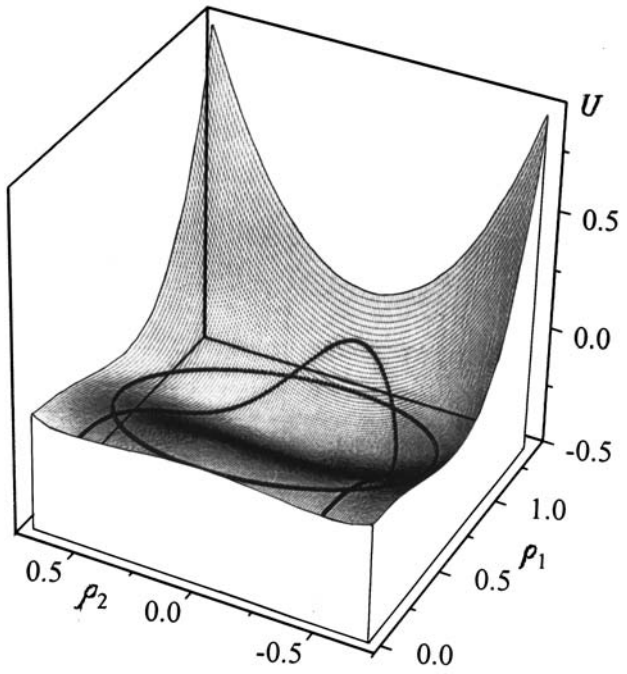


FIG. 1. The trajectory corresponding to the soliton of Eqs. (11) known for PRCs[33] and formed by two mutually incoherent (''bright'' and ''dark'') components and the trajectory corresponding to the two-component soliton of Eqs. (13) (open and closed curves, respectively). $U(\rho_1,\rho_2)$ is the potential energy, $\beta_1=1$, $\beta_2=0.25$.

character of the $\rho_{1,2}(\xi)$ dependences determined by the system of Eqs. (5) and the nonlinear oscillations of two oscillators in the common potential

$$U(\rho_1,\rho_2)=\pm\frac{1}{2}(\rho_1^2+\rho_2^2)^2\mp(\beta_1\rho_1^2+\beta_2\rho_2^2). \tag{8}$$

Here, as earlier, the + and − signs correspond to focusing and defocusing, respectively. The potential energy of the equivalent mechanical system on the $\rho_1\rho_2$ plane has several local extrema, which correspond to singular points of its phase portrait. However, by assuming $\beta_1>\beta_2$ for definiteness and considering the case of self-focusing (Fig. 1), it is easy to convince oneself that only the two absolute minima of $U(\rho_1,\rho_2)$, with coordinates $\rho_1=\pm\sqrt{\beta_1}$ and $\rho_2=0$ (one minimum with coordinates $\rho_1=\rho_2=0$ for defocusing) are points of stable equilibrium, or foci, while the only local maximum $\rho_1=\rho_2=0$ (two absolute maxima with coordinates $\rho_1=\pm\sqrt{\beta_1}$ and $\rho_2=0$ for defocusing) is a node. All the other singular points have the character of saddles. In the degenerate case of $\beta_1=\beta_2=\beta$, all the points with coordinates $\rho_1^2+\rho_2^2=\beta$, located at the bottom of the potential valley of Eq. (8), correspond to neutral equilibrium.

Besides the one-component bright and dark solitons of Eqs. (6) and (7), there are also two-component solutions of Eq. (5), which can be written in terms of elementary functions. Thus, when $\beta_1=\beta_2=\beta$, the solutions of Eq. (5) are also known pairs formed by two incoherent bright solitons (the self-focusing case),

$$\rho_1=\frac{\sqrt{2\beta}}{\cosh(\sqrt{2\beta}\xi)}\cos\alpha,\quad \rho_2=\frac{\sqrt{2\beta}}{\cosh(\sqrt{2\beta}\xi)}\sin\alpha, \tag{9}$$

or dark solitons (defocusing),

$$\rho_1=\sqrt{\beta}\tanh(\sqrt{\beta}\xi)\cos\alpha,\quad \rho_2=\sqrt{\beta}\tanh(\sqrt{\beta}\xi)\sin\alpha, \tag{10}$$

as given by Ref. 33. In (ρ_1,ρ_2,U) coordinates, both these solutions correspond to the projection onto the $\rho_{1,2}$ axes of the corresponding single-soliton solutions of Eqs. (6) and (7), the plane of whose trajectories is rotated around the axis of symmetry U through angle α.

When $\beta_1>\beta_2$ holds, stable pairs—''gray'' solitons[21]—can be formed from incoherent bright and dark solitons both with self-focusing,

$$\rho_1=\pm\frac{\sqrt{2\beta_1-\beta_2}}{\cosh[\sqrt{2(\beta_1-\beta_2)}\xi]},$$
$$\rho_2=\pm\sqrt{\beta_2}\tanh[\sqrt{2(\beta_1-\beta_2)}\xi], \tag{11}$$

and with defocusing,

$$\rho_1=\pm\sqrt{\beta_1}\tanh[\sqrt{2(\beta_1-\beta_2)}\xi],$$
$$\rho_2=\pm\frac{\sqrt{2\beta_2-\beta_1}}{\cosh[\sqrt{2(\beta_1-\beta_2)}\xi]}. \tag{12}$$

The latter solution exists only when $\beta_1<2\beta_2$. In the case of self-focusing, the trajectory of such a soliton begins ($\xi\to-\infty$) at a point of unstable equilibrium $(0,\pm\sqrt{\beta_2})$ [the local maximum at the bottom of the potential valley of Eq. (8)] and ends ($\xi\to+\infty$) at the symmetrically located point $(0,\mp\sqrt{\beta_2})$ (Fig. 1, open curve). In the $\rho_1\rho_2$ plane, this trajectory is a semiellipse with semiaxes $\sqrt{2\beta_1-\beta_2}$ and $\sqrt{\beta_2}$ and, depending on the sign, the asymptotic limits lie either in the half-plane $\rho_1\geqslant 0$ or $\rho_1\leqslant 0$. With defocusing, the trajectory of such a two-component soliton is also a semi-ellipse (but now with semi-axes $\sqrt{\beta_1}$ and $\sqrt{2\beta_2-\beta_1}$), beginning at the point of unstable equilibrium $(\mp\sqrt{\beta_1},0)$ (the absolute maximum of potential energy) and ending at the point $(\pm\sqrt{\beta_1},0)$, symmetrically located on the $\rho_1\rho_2$ plane. As in the preceding case, depending on the sign of the asymptotics, this trajectory is located either in $\rho_2\geqslant 0$ or $\rho_2\leqslant 0$.

All the two-component solutions of Eqs. (5) enumerated above are definite pairwise combinations of bright and dark solitons. At the same time, if at least one dark soliton enters into such a pair, its energy is infinitely large. Therefore, all combinations of this type known until now[33] for PRCs most likely had purely methodological interest.

We have succeeded in finding a new two-component soliton solution of Eqs. (5) for PRCs with bounded energy for the case of nonlinearity of the focusing type. A similar solution was obtained earlier, for example, in the problem of the propagation of two ultrashort light pulses with mutually orthogonal polarizations along a lightguide with Kerr nonlinearity.[38,39] Even though its trajectory on the $\rho_1\rho_2$ plane, like the self-matched pair formed by two bright incoherent solitons, begins and ends at the point (0,0), the $\rho_2(\xi)$ dependence is odd in this solution (Fig. 1, closed curve). At the same time, unlike the conventional dark soliton of Eq. (10), the function $\rho_2(\xi)$ decreases exponentially as $\xi\to\pm\infty$. For $\beta_1=4\beta_2$, this new two-component solution for PRCs can be written explicitly

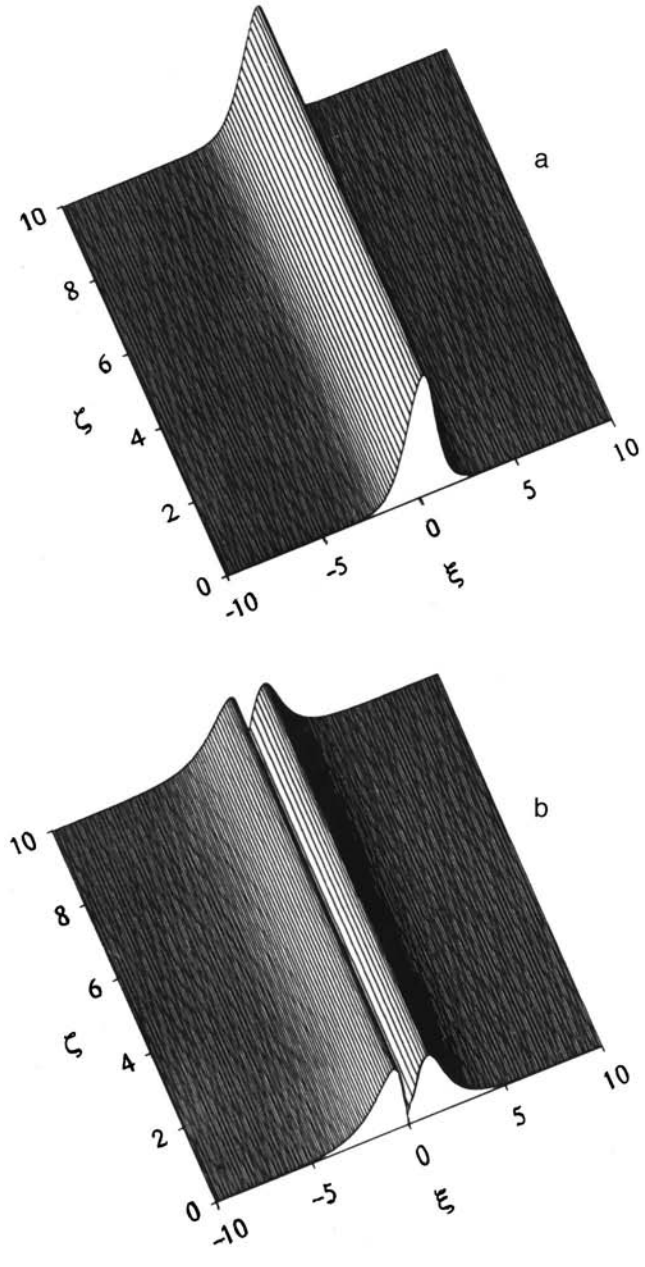


FIG. 2. Distributions of $|\rho_1(\xi)|$ (a) and $|\rho_2(\xi)|$ (b) for the components of the paired soliton of Eqs. (13) and their stable propagation along the ζ axis: $\beta_1=1$, $\beta_2=0.25$.

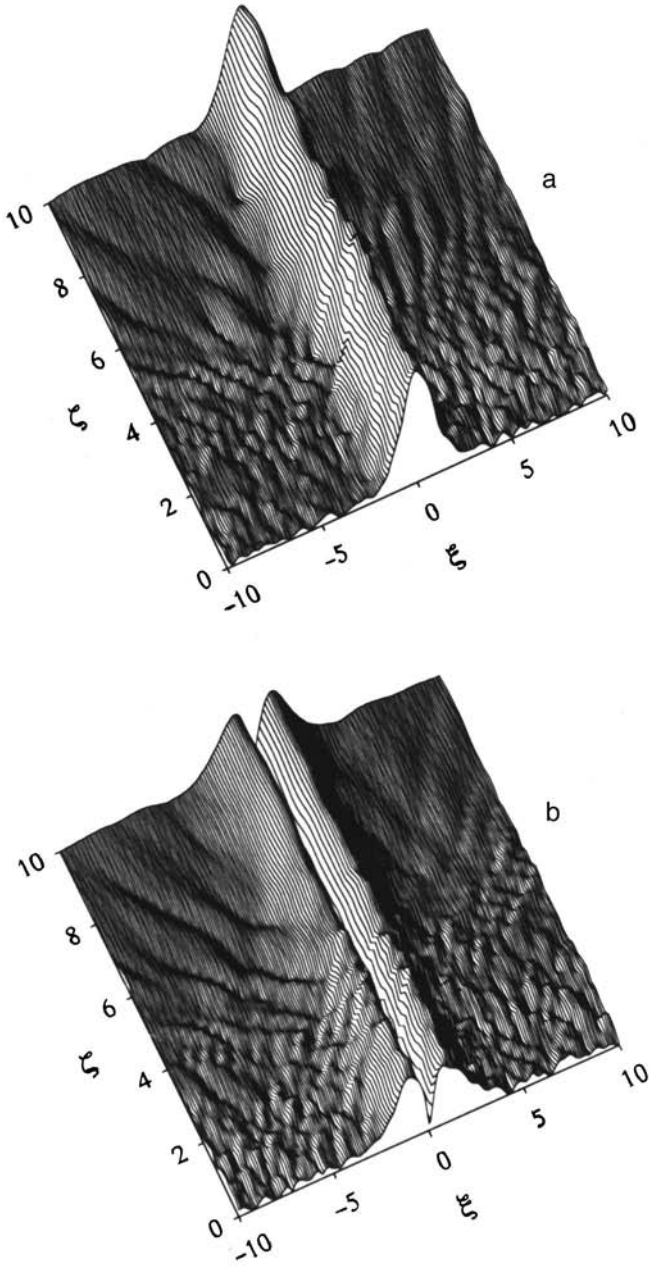


FIG. 3. Structural stability of the distributions of $|\rho_1(\xi)|$ (a) and $|\rho_2(\xi)|$ (b) perturbed by Gaussian noise (10% in intensity) for the components of the paired soliton of Eqs. (13) when they propagate along the ζ axis: $\beta_1=1$, $\beta_2=0.25$.

$$\rho_1=\pm\frac{\sqrt{6\beta_2}}{\cosh^2(\sqrt{2\beta_2}\xi)},\quad \rho_2=\pm\frac{\sqrt{6\beta_2}\,\sinh(\sqrt{2\beta_2}\xi)}{\cosh^2(\sqrt{2\beta_2}\xi)}. \tag{13}$$

Figure 2 shows the spatial distributions of the amplitude of the light field in the two mutually incoherent components of the solution of Eqs. (13) [the symmetric (Fig. 2a) and antisymmetric (Fig. 2b) components] and their stable propagation to a distance of $\zeta=10$, which on the scale of an actual experiment would correspond to a length of about 5 cm for the PRC. A numerical calculation shows that, unlike the antisymmetric component ρ_2, the symmetric component ρ_1 is transformed as it propagates along the PRC into an ordinary one-component bright soliton of the form of Eq. (6). At the same time, the antisymmetric component ρ_2 in the absence of a ρ_1 component gradually transforms into a pair of bright small-amplitude solitons with opposite phase that diverge along the transverse coordinate ξ.

Figure 3 illustrates the structural stability of the two-component soliton of Eqs. (13) against substantial (10% in intensity) perturbations of the input amplitude profiles of both field components $\rho_{1,2}$ by Gaussian noise. With an increase of the noise level (20% in intensity), such a soliton of Eqs. (13) was observed to decay in a computer experiment to ordinary bright solitons, which consequently have a somewhat large margin of stability against perturbations. A numerical calculation also proved that the two-component soli-

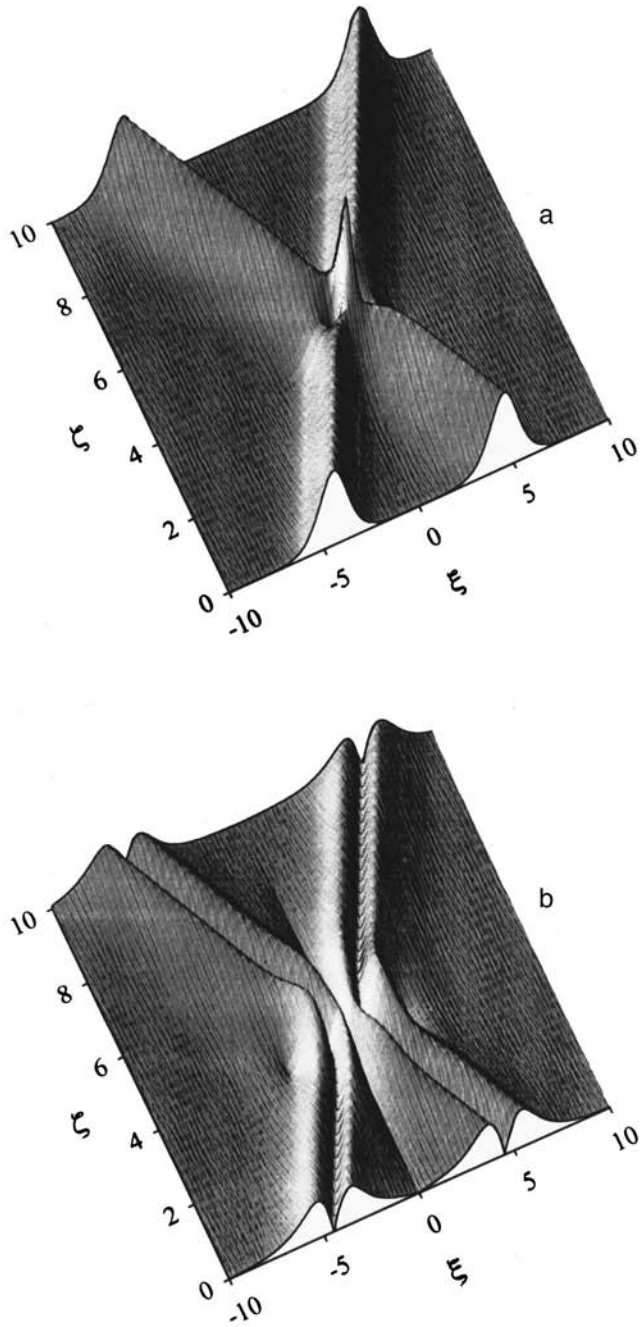


FIG. 4. Stability of the distributions of $|\rho_1(\xi)|$ (a) and $|\rho_2(\xi)|$ (b) for the components of the paired soliton of Eqs. (13) against intersecting with the same paired soliton in a PRC: $\beta_1=1$, $\beta_2=0.25$.

ton solution of Eqs. (13) has structural stability with respect to collisions (mutual intersections) with the same two-component solitons (Fig. 4) or with ordinary one-component bright solitons.

Our computer modeling allowed us to track the transformation of the structure of the new two-component solitons as the ratio of parameters $\beta_{1,2}$ varied. It turned out that the limiting case of a solution of the form of Eqs. (13) as $\beta_1/\beta_2 \to 1$ is the pair $[\rho_1^{(1,2)}, \rho_2^{(1,2)}]$ of two-component photorefractive bright solitons of the form given by Eqs. (9), which are infinitely remote from each other on the ξ axis and consequently do not interact. Each of the solitons of this pair consists of two mutually incoherent components $\rho_1^{(1,2)}$ and $\rho_2^{(1,2)}$. However, the corresponding components of these $\rho_{1,2}^{(1)}$ and $\rho_{1,2}^{(2)}$ solitons moving apart along ξ are pairwise coherent. Moreover, if the coherent components $\rho_1^{(1)}$ and $\rho_1^{(2)}$ in these solitons are in phase, the components $\rho_2^{(1)}$ and $\rho_2^{(2)}$ have opposite phase. As the ratio β_1/β_2 increases, the solitons of a pair gradually approach each other along ξ. When they spatially coincide completely (the limiting transition $\beta_1/\beta_2 \to \infty$), the out-of-phase components undergo complete interference quenching, and a one-component bright soliton of the form of Eqs. (6) develops. Naturally, such an analogy is extremely tentative, since the shape and parameters of the solitons of a pair vary because of cross-modulation interaction as they approach along ξ.

Thus, our solution of Eqs. (13) is a new type of stable and physically implementable two-component soliton for PRCs, belonging to the class of solutions with separable variables.

4. SELF-LOCKING OF HIGHER-ORDER SOLITON MODES

In essence, the two-component solution of Eqs. (13) that we considered above describes cross-modulation self-locking and propagation along the resulting nonlinear waveguide of its zeroth- and first-order modes. We shall show in this section that there also exist more complicated stable solutions of this type, i.e., multicomponent solitons that include more than two incoherent self-consistent components of the light field with finite energy and spatially limited distributions corresponding to the structure of the higher-order normal modes of their common nonlinear lightguide. As far as we know, multicomponent solutions of this type have not yet been discussed in the literature. This is probably because in order to put them into practice, the most important thing is to eliminate interference between their components. When there are only two such components, their polarization can be chosen to be orthogonal.[38,39] However, if there are more such components, a different method has to be used. To do this, one can, for example, use components with different carrier frequencies and a nonlinear medium with a comparatively slow response, which, as a consequence of its inertia, is incapable of tracking the interference beats. It is this case, considered earlier for the description of spatially localized paired electronic states (the so-called bisoliton model of high-temperature superconductivity[40]) that can be put into practice in PRCs.

It is easy to convince oneself that, as a result of nonlinear interaction of the two components of the soliton of Eqs. (13), a common nonlinear lightguide is formed in a PRC with a transverse distribution of the refractive index of

$$\Delta n \propto \Delta n_{\max}/\cosh^2(\xi/\xi_0), \tag{14}$$

where the parameters $\Delta n_{\max}=6\beta_2$ and $\xi_0=1/\sqrt{2\beta_2}$ describe the maximum refractive-index change and the width of this lightguide. We shall assume that all such solutions form a common nonlinear lightguide in a PRC, with a refractive-index profile similar to Eq. (14). We shall regard $\Delta n_{\max}$ and ξ_0 as variable parameters. Then, at the first stage, our problem actually reduces to one of two already rather well-

studied problems: the calculation of localized states in a potential well corresponding to Eq. (14) of the form treated in Ref. 41, or the calculation of the normal modes of an optical gradient lightguide with a known refractive-index distribution profile (a particular case of the so-called Epstein profile) given by Eq. (14).[42] Next, we also require that the solutions found at the first stage form a common lightguide of the required profile in the PRC because of Kerr nonlinearity, and we determine the coefficients of the corresponding expansions over the normal modes from this condition.

In the general case, the problem under discussion has no elementary analytical solutions, and its eigenfunctions are expressed in terms of hypergeometrical functions. However, the character of the solutions is substantially simplified when quite definite discrete relationships exist between $\Delta n_{\max}$ and ξ_0,[41,42] for which the eigenfunctions can be expressed in terms of the so-called associated Legendre functions.[43] In this case, the system of equations that we need to solve for an arbitrary whole number $n=1, 2, \cdots$ is written in the form

$$\frac{d^2\rho_i^{(n)}}{d\xi^2}+2\left[\frac{\beta_0 n(n+1)}{\cosh^2(\sqrt{2\beta_0}\xi)}-\beta_i\right]\rho_i^{(n)}=0, \tag{15}$$

where $i=1, 2, \cdots n$. The condition of self-consistency of the multicomponent solutions of soliton type is defined in this case as

$$\sum_{i=1}^{n}[\rho_i^{(n)}]^2=\frac{\beta_0 n(n+1)}{\cosh^2(\sqrt{2\beta_0}\xi)}. \tag{16}$$

Without dwelling here on the procedure of routine calculation, we proceed to only the first three multicomponent solutions of this class for $n=1, 2, 3$:

$$\rho_1^{(1)}=\frac{\sqrt{2\beta_0}}{\cosh(\sqrt{2\beta_0}\xi)}, \tag{17a}$$

$$\begin{cases}\rho_1^{(2)}=\dfrac{\sqrt{6\beta_0}}{\cosh^2(\sqrt{2\beta_0}\xi)},\\ \rho_2^{(2)}=\dfrac{\sqrt{6\beta_0}\,\sinh(\sqrt{2\beta_0}\xi)}{\cosh^2(\sqrt{2\beta_0}\xi)},\end{cases} \tag{17b}$$

$$\begin{cases}\rho_1^{(3)}=\dfrac{(3/2)\sqrt{5\beta_0}}{\cosh^3(\sqrt{2\beta_0}\xi)},\\ \rho_2^{(3)}=\dfrac{\sqrt{30\beta_0}\,\sinh(\sqrt{2\beta_0}\xi)}{\cosh^3(\sqrt{2\beta_0}\xi)},\\ \rho_3^{(3)}=\dfrac{(1/2)\sqrt{3\beta_0}[4-5/\cosh^2(\sqrt{2\beta_0}\xi)]}{\cosh(\sqrt{2\beta_0}\xi)}.\end{cases} \tag{17c}$$

It is easy to convince oneself that the first solution, Eq. (17a), is none other than the usual bright soliton of the form given by Eqs. (6), that the second solution, Eq. (17b), corresponds to the two-component soliton of Eqs. (13), described above and well known in a number of other problems,[38,39] and that the third solution, Eq. (17c), is an additional, new solution, including a third incoherent component of the light field, with a spatial distribution corresponding to a second symmetrical mode of a nonlinear lightguide written in the PRC.

Figure 5 illustrates the spatial distributions of the amplitude of the light field for all three incoherent components of the new solution given by Eq. (17c) (Figs. 5a, 5b, and 5c, respectively) and their stable propagation to a distance of $\zeta=5$, which, on the scale of an actual experiment, corresponds to a length of the PRC of a few centimeters.

In conclusion, one more feature of the class of multicomponent solitons that we have found should be pointed out. Besides the limitation of their energy, it is in our opinion extremely important for applications that, unlike the two-component solitons of Eqs. (9) and (13) known earlier, the modulation depth of the refractive index, $\Delta n_{\max}$, can be varied in solitons of the new type—i.e., the parameters of the light-induced lightguide can be controlled by varying the number of components in the soliton.

5. CONCLUSION

To summarize this paper, we shall briefly enumerate its main results and formulate certain conclusions. Starting from a completely realistic model of the formation of the nonlinear response of a PRC,[35] we have shown for the first time that it is possible to implement in a PRC the two-component soliton solution already known for other physical problems,[38,39] corresponding to the stable propagation of two self-consistent incoherent spatially localized beams with limited energy and nontrivial spatial structure in a PRC with a drift mechanism of the nonlinear response. The light-field distributions in this solution correspond to simultaneous cross-modulation self-locking and to the propagation of its zeroth and first-order modes in the nonlinear lightguide formed by the pair of components itself. Further analysis has shown that there also exist more complicated multicomponent solitons of a similar type, which now include more than two self-consistent incoherent light-field components—the normal modes of their common nonlinear lightguide. Our computer modelling has shown that such solutions not only can be physically implemented but also are structurally stable.

In our opinion, the class of multicomponent solitons found and analyzed here is important for applications, since they exhibit a definite freedom of choice of the relationship between the maximum refractive-index change $\Delta n_{\max}$ in the resulting lightguide channel and its spatial scale ξ_0.

Such multicomponent solutions of the nonlinear Schrödinger equation can also be of interest from a methodological viewpoint, especially in connection with problems of forming nonlinear collective modes in waveguide lattices,[45] as well as problems of the physics of quasi-one-dimensional chains in ferromagnets,[7,8] high-temperature superconductors,[9,40] and conjugate polymers.[10–12] It seems to us that the last group of problems can provide fruitful ideas concerning the possibility of the formation and subsequent stable propagation of several (two or more) mutually incoherent (because of rapid phase relaxation or different eigenfrequencies) but stable electronic wave packets (the components of multicomponent solitons) with unusual mutually orthogonal spatial distributions along one-dimensional conjugate chains or isolated atomic planes. Moreover, the for-

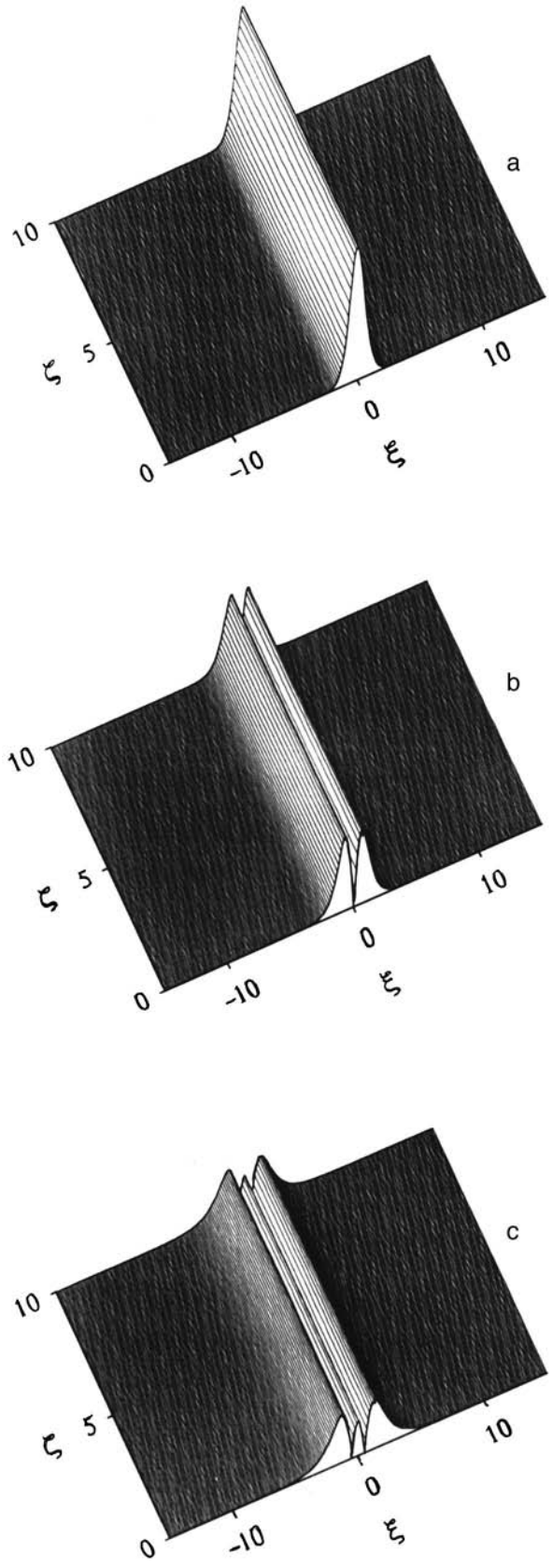


FIG. 5. Distributions of $|\rho_1^{(3)}(\xi)|$ (a), $|\rho_2^{(3)}(\xi)|$ (b), and $|\rho_3^{(3)}(\xi)|$ (c) for the components of the soliton of Eq. (17c) and their stable propagation along the ζ axis: $\beta_0=1$.

mation of precisely such incoherent but coupled packets from coherent electron pairs when the latter are excited by picosecond pulses with a quantum energy of about 2 eV can explain the unexpected experimental result that we obtained in Y–Ba–Cu–O films, in which an energy gap in the spectrum of the electronic states of superconducting samples was observed for more than 3 ns after impact excitation.[44]

This work was carried out with the financial support of the Russian Fund for Fundamental Research (Grant No. 96-02-16238).

*)E-mail: vsh@vsh.ilc.msu.su

[1] S. A. Akhmanov, V. A. Vysloukh, and A. S. Chirkin, *Optics of Femtosecond Laser Pulses* (AIP, New York, 1992; Nauka, Moscow, 1988).
[2] A. S. Shcherbakov and E. I. Andreeva, Pis'ma Zh. Tekh. Fiz. **21**(7), 6 (1995) [Tech. Phys. Lett. **21**, 241 (1995)].
[3] A. Hook and V. N. Serkin, IEEE J. Quantum Electron. **30**, 148 (1994).
[4] M. Shalaby and A. C. Barthelemy, IEEE J. Quantum Electron. **28**, 2736 (1992).
[5] A. F. Kaplan and P. L. Shkolnikov, J. Opt. Soc. Am. B **13**, 347 (1996).
[6] M. Logdlund, P. Dannetun, S. Stafstrom *et al.*, Phys. Rev. Lett. **70**, 970 (1993).
[7] S. N. Martynov, Teor. Mat. Fiz. **91**, 112 (1992).
[8] J. P. Goff, D. A. Tennant, and S. E. Nagler, Phys. Rev. B **52**, 5992 (1995).
[9] D. B. Haviland and P. Deising, Phys. Rev. B **54**, R6857 (1996).
[10] S. Takeuchi, M. Yoshizawa, T. Masuda *et al.*, IEEE J. Quantum Electron. **28**, 2508 (1992).
[11] A. Takahashi and S. Mukamel, J. Chem. Phys. **100**, 2366 (1994).
[12] Xu Jizhong and Huang Jingning, Phys. Lett. A **197**, 127 (1995).
[13] P. Gunter and J.-P. Huignard, *Photorefractive Materials and Applications* (Springer Verlag, Heidelberg, 1988).
[14] M. Segev, B. Crosignani, and A. Yariv, Phys. Rev. Lett. **68**, 923 (1992).
[15] M. D. Iturbe-Castillo, P. A. Marquez-Aguilar, J. J. Sanchez-Mondragon *et al.*, Appl. Phys. Lett. **64**, 408 (1994).
[16] G. C. Duree, J. L. Schultz, G. J. Salamo *et al.*, Phys. Rev. Lett. **71**, 533 (1993).
[17] N. Kukhtarev, V. Markov, S. Odulov *et al.*, Ferroelectrics **22**, 949 (1979).
[18] B. Crosignani, M. Segev, D. Engin *et al.*, J. Opt. Soc. Am. B **10**, 446 (1993).
[19] M. Shamonin, Appl. Phys. A **56**, 467 (1993).
[20] G. C. Valley, M. Segev, B. Crosignani *et al.*, Phys. Rev. A **50**, R4457 (1994).
[21] D. N. Christodoulides and M. I. Carvalho, J. Opt. Soc. Am. B **12**, 1628 (1995).
[22] M. Segev, G. C. Valley, S. R. Singh *et al.*, Opt. Lett. **20**, 1764 (1995).
[23] G. Duree, M. Morin, G. Salamo *et al.*, Phys. Rev. Lett. **74**, 1978 (1995).
[24] Z. Chen, M. Mitchell, and M. Segev, Opt. Lett. **21**, 716 (1996).
[25] C. M. Gomez, J. J. Sanchez-Mondragon, S. Stepanov, and V. A. Vysloukh, J. Mod. Opt. **43**, 1253 (1996).
[26] A. V. Mamaev, M. Saffman, and A. A. Zozulya, Phys. Rev. Lett. **77**, 4544 (1996).
[27] M.-F. Shih, M. Segev, G. C. Valley *et al.*, Electron. Lett. **31**, 826 (1995).
[28] M.-F. Shih, P. Leach, M. Segev *et al.*, Opt. Lett. **21**, 324 (1996).
[29] M. Segev, B. Crosignani, P. Di Porto *et al.*, Opt. Lett. **19**, 1296 (1994).
[30] A. V. Mamaev, M. Saffman, and A. A. Zozulya, Europhys. Lett. **35**, 25 (1996).
[31] A. V. Mamaev, M. Saffman, and A. A. Zozulya, Phys. Rev. Lett. **76**, 2262 (1996).
[32] M. Morin, G. Duree, G. Salamo, and M. Segev, Opt. Lett. **20**, 2066 (1995).
[33] Z. Chen, M. Segev, T. H. Coskun, and D. N. Christodoulides, Opt. Lett. **21**, 1436 (1996).
[34] D. N. Christodoulides, S. R. Singh, M. I. Carvalho, and M. Segev, Appl. Phys. Lett. **68**, 1763 (1996).
[35] V. A. Vysloukh, V. Kutuzov, V. M. Petnikova, and V. V. Shuvalov, Zh. Eksp. Teor. Fiz. **111**, 705 (1997) [JETP **84**, 388 (1997)].
[36] G. Duree, G. Salamo, M. Segev *et al.*, Opt. Lett. **19**, 1195 (1994).

[37] M. D. Iturbe-Castillo, M. Torres-Cisneros, J. J. Sanchez-Mondragon *et al.*, Opt. Lett. **20**, 1 (1995).
[38] M. V. Tratnik and J. E. Sipe, Phys. Rev. A **38**, 2011 (1988).
[39] M. Florjanczyk and R. Tremblay, Phys. Lett. A **141**, 34 (1989).
[40] A. S. Davydov, *High-Temperature Superconductivity* (Naukova Dumka, Kiev, 1990), chap. 7.
[41] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics–Non-Relativistic Theory* (Pergamon Press, New York, 1977; Nauka, Moscow, 1989).
[42] M. Adams, *Introduction to Optical Waveguides* (Wiley, New York, 1981; Mir, Moscow, 1984), chap. 4.
[43] I. S. Gradstein and I. M. Ryzhik, *Tables of Integrals, Sums, Series, and Products* (Academic Press, New York, 1980; Moscow, 1971), chap. 8.
[44] A. N. Zherikhin, V. A. Lobastov, V. M. Petnikova, and V. V. Shuvalov, Physica C **221**, 311 (1994).
[45] S. Darmanyan, I. Relke, and F. Lederer, Phys. Rev. E **55**, 7662 (1997).

Translated by W. J. Manthey

Coherent repopulation of the components of a three-level quantum system in the field of a pulsed bichromatic radio frequency wave

D. F. Zaretskiĭ and S. B. Sazonov

Kurchatov Institute Russian Scientific Center, 123182 Moscow, Russia
(Submitted 4 July 1997)
Zh. Éksp. Teor. Fiz. **113**, 1181–1192 (April 1998)

The coherent repopulation of a quantum system consisting of three nonequidistant levels in the field of a resonant bichromatic rf wave is studied. The atoms are assumed to have an impulsive interaction with the rf wave in which the pulse duration is less than any of the relaxation times. The hyperfine structure of gas atoms and a system of atomic oscillator levels in a magnetic trap are considered as examples of such a quantum system. It is shown that in the second case, the coherent repopulation effect can be used to cool neutral atoms in magnetic traps.
 [S1063-7761(98)00304-7]

1. INTRODUCTION

The coherent redistribution of populations (in particular, coherent population capture) is well known in laser physics.[1] This phenomenon involves the suppression of resonance fluorescence from the upper level to the two lower levels of a three-level atomic system interacting with the field of a resonant bichromatic wave. In particular, the two lower levels can be hyperfine structure components.[2,3] In this case, the upper level is separated from the lower levels by an amount equal to the energy of a laser photon and is, therefore, not populated at the initial time. When the necessary conditions for coherent population capture are satisfied, the upper level remains unpopulated in the field of the bichromatic resonant laser wave.

Coherent population capture has been observed in a stationary laser field.[2] In this case, the effect depends strongly on the relationship between the resonance fluorescence time and the relaxation time of the lower levels. We have previously examined[4,5] the coherent redistribution of populations in a pulsed bichromatic laser pulse whose duration is shorter than the resonance fluorescence time. It was shown that coherent population capture also occurs in this case, while the population of the upper level and, therefore, the resonance fluorescence intensity depend strongly on the relative constant phase of the components of the bichromatic wave.

In this paper we examine the coherent redistribution of the populations in a three-level system interacting resonantly with a bichromatic radio frequency wave. This case differs from that of a laser wave in the following ways:

(1) All three components of the three-level system can be populated initially and the problem must be solved with different boundary conditions.

(2) Coherent repopulation in this case necessarily requires a pulsed interaction with the field in which the pulse duration is less than the longitudinal and transverse relaxation times. In the case of a stationary radio frequency wave, coherent repopulation does not take place.

We have investigated[6] the polarization of impurity center atoms in a magnetic matrix using the interaction of a pulsed resonant bichromatic radio frequency wave. The repopulation of levels owing to interactions with radio frequency waves may, however, have wider applications. Thus, using a bichromatic rf wave opens up the possibility of polarizing atoms in a gas target even when the gas atoms do not have optical transitions (atomic hydrogen, noble gases). This method can also be used to repopulate the oscillator levels of neutral atoms in magnetic traps. As we shall show, this effect leads to a reduction in the kinetic energy (cooling) of these atoms. This method of cooling is of special interest in connection with the recent discovery of the Bose condensation of alkali metal atoms.[7]

In this paper we examine the coherent repopulation of the levels of a three-level quantum system in the field of a resonant bichromatic radio frequency wave. It is assumed that the system has an impulse interaction with the rf wave, with the pulse duration shorter than any of the relaxation times. We consider the hyperfine level structure of gas atoms and a system of neutral-atom oscillator levels in magnetic traps as examples of quantum systems.

2. COHERENT REPOPULATION OF THE COMPONENTS OF A THREE-LEVEL SYSTEM IN THE FIELD OF A BICHROMATIC RADIO FREQUENCY WAVE

Let us consider a system consisting of three nonequidistant levels. One of the components of the bichromatic wave is in resonance with the transition between levels 1 and 3, while the other is in resonance with the transition between levels 2 and 3. Level 3 is common to both. It can be positioned arbitrarily relative to the other two levels.

The hamiltonian of the atom+rf field system has the form

$$H(t)=H_0+\hat{V}(t), \tag{1}$$

where H_0 is the hamiltonian of a three-level system with the characteristic wave functions F_i $(i=1,2,3)$ and $\hat{V}(t)$ is the interaction operator for this system with the bichromatic rf field.

We shall assume that the energy separation between the levels can be less than kT, so that, in general, their populations are not the same before the rf field is applied. Let us suppose that the time for this system to interact with the field is shorter than any of the relaxation times, in particular, than the longitudinal relaxation time T_1 and the transverse relaxation time T_2. This condition means that it is possible to examine the process assuming that the system is in a state described by a wave function which can be represented as a superposition of the functions F_i,

$$\Psi(t)=\sum_i a_i(t)F_i. \tag{2}$$

The amplitudes $a_i(t)$ are the population amplitudes of the ith levels, which satisfy the following initial conditions:

$$a_i(0)=\sqrt{A_i}\,\exp(i\alpha_i), \tag{3}$$

where A_i is the initial population of level i, with

$$A_i=|a_i(0)|^2, \tag{4}$$

and α_i is the initial phase of its population amplitude.

Since the interaction time is shorter than any of the relaxation times, there are essentially no stochastic perturbations of the system during the interaction process. Thus, the amplitude $a_i(t)$ at any time is proportional to the constant phase factor $\exp(i\alpha_i)$ and can be represented in the form

$$a_i(t)=\overline{a_i}(t)\exp(i\alpha_i), \tag{5}$$

where at any time the functions $\overline{a_i}(t)$ are independent of the α_i and at $t=0$ they are given by

$$\overline{a_i}(0)=\sqrt{A_i}. \tag{6}$$

As eigenfunctions F_i of the hamiltonian H_0 we can choose a set of functions of the form

$$F_i=\overline{F_i}\,\exp(-i\alpha_i), \tag{7}$$

where the $\overline{F_i}$ are independent of the phases α_i.

We obtain a system of equations for the amplitudes $a_i(t)$ from the Schrödinger equation for $\Psi(t)$,

$$i\hbar\,\frac{d}{dt}\,\Psi(t)=H(t)\Psi(t), \tag{8}$$

using the following assumptions: (1) both frequencies of the bichromatic wave coincide with the frequency of a transition between the corresponding levels (the resonance approximation); (2) throughout the entire interaction process, the phases of the components of the bichromatic wave φ_i $(i=1,2)$ and their relative phase $\Delta\varphi=\varphi_1-\varphi_2$ remain fixed. For simplicity, in the following we consider only the cases $\Delta\varphi=0$ and $\Delta\varphi=\pi$.

The system of equations for the amplitudes $a_i(t)$ has the form

$$\frac{da_1(t)}{dt}=-iV_{13}\exp(i\varphi_1)a_3(t),$$

$$\frac{da_2(t)}{dt}=-iV_{23}\exp(i\varphi_2)a_3(t),$$

$$\frac{da_3(t)}{dt}=-iV_{31}\exp(-i\varphi_1)a_1(t)-iV_{32}(-i\varphi_2)a_2(t), \tag{9}$$

where V_{13} and V_{23} are the matrix elements of the operator $\hat{V}(t)$, calculated using the functions F_i and corresponding to the resonant transitions, under the influence of the field, between levels 1 and 3 (V_{13}) and between levels 2 and 3 (V_{23}). In Eqs. (9) and below, we set $\hbar=1$. Let us transform to the amplitudes $\overline{a_i}(t)$ and functions $\overline{F_i}$ in Eqs. (9). Since the matrix elements are proportional to phase factors of the form $\exp[i(\alpha_i-\alpha_j)]$, it can be seen that the same phase factors will remain on the right and left hand sides of Eqs. (9) and these will cancel out. This means that the result of coherent repopulation of the levels of a three-level system is independent of the initial phases of the level populations.

The solution of the linear system (9) for the case in which all the levels are populated at $t=0$ has the form

$$\begin{aligned}
\overline{a_1}(t)&=[V_2A_-+V_1A_+\cos(\Omega t)]/\Omega^2\\
&\quad -i\sqrt{A_3}V_1\sin(\Omega t)/\Omega,\\
\overline{a_2}(t)&=(V_2A'_+\cos(\Omega t)-V_1A'_-)/\Omega^2\\
&\quad -i\sqrt{A_3}V_2\exp(i\Delta\varphi)\sin(\Omega t)/\Omega,\\
\overline{a_3}(t)&=-\sqrt{A_3}\cos(\Omega t)-iA_+\exp(-i\varphi_1)\sin(\Omega t)/\Omega,\\
A_+&=\sqrt{A_2}V_2\exp(-i\Delta\varphi)+\sqrt{A_1}V_1,\\
A_-&=\sqrt{A_1}V_2-\sqrt{A_2}V_1\exp(-i\Delta\varphi),\\
A'_+&=\sqrt{A_2}V_2+\sqrt{A_1}V_1\exp(i\Delta\varphi),\\
A'_-&=\sqrt{A_1}V_2\exp(i\Delta\varphi)-\sqrt{A_2}V_1.
\end{aligned} \tag{10}$$

Here Ω is the frequency of the Rabi oscillations, with $\Omega^2=V_1^2+V_2^2$, where V_1 and V_2 are the real parts of the matrix elements V_{13} and V_{23}, respectively, with $V_{i3}=|V_{i3}|\exp(i\varphi_i)$. It is clear from Eq. (10) that coherent population capture can be observed in a pulsed regime if the population of the common level satisfies $A_3=0$ at the initial time. Then $a_3(t)=0$ holds at any time if the parameters of the components of the bichromatic wave, the relative phase and strengths, obey the condition

$$\cos\Delta\varphi=-\frac{A_1V_1^2+A_2V_2^2}{2\sqrt{A_1A_2}V_1V_2}. \tag{11}$$

In general, when the populations of all the levels are nonzero at the initial time, there is no coherent population capture effect, but a radical readjustment of the system population is possible.

To estimate the magnitude of the repopulation, we proceed from the amplitudes $a_i(t)$ to the populations $\rho_{ii}=|a_i(t)|^2$. If the rf pulse is very much shorter than the Rabi period Ω^{-1}, then the actual populations can be found by averaging over the Rabi oscillations. For the populations after the end of the field pulse, we obtain

$$\overline{\rho_{11}}=V_2^2|A_-|^2/\Omega^4+V_1^2|A_+|^2/2\Omega^4+V_1^2A_3/2\Omega^2,$$

$$\overline{\rho}_{22}=V_1^2|A_-|^2/\Omega^4+V_2^2|A_+|^2/2\Omega^4+V_2^2A_3/2\Omega^2,$$

$$\overline{\rho}_{33}=|A_+|^2/2\Omega^2+A_3/2. \tag{12}$$

We shall examine two typical situations. Let the conditions $|A_+|^2=0$ and $A_3\neq 0$ be satisfied. This may occur, for example, for $\Delta\varphi=\pi$ and

$$V_1=V_2\sqrt{A_2/A_1}, \tag{13}$$

which is a special case of the condition (11). For the magnitudes of the average populations when $\Delta\varphi=\pi$ and Eq. (13) is satisfied, this yields

$$\overline{\rho}_{11}=A_1+A_2A_3/2(A_1+A_2),$$

$$\overline{\rho}_{22}=A_2+A_1A_3/2(A_1+A_2),$$

$$\overline{\rho}_{33}=A_3/2. \tag{14}$$

Let all the levels have the same populations initially, i.e., $A_1=A_2=A_3=1$ and $\Delta\varphi=\pi$. Then, after application of the bichromatic field pulse, the population of level 3 falls by a factor of two, while the populations of the other levels increase to 5/4. If the pulses are repeated, then, as can be seen from Eq. (14), after the N-th pulse the population of the third level decreases by a factor of 2^N. Thus, the common third level can be completely emptied.

We now consider the other situation: $|A_-|^2=0$ and $A_3\neq 0$. These equations can be satisfied when $\Delta\varphi=0$ and

$$V_1=V_2\sqrt{A_1/A_2}. \tag{15}$$

The expressions for the average populations will then have the form

$$\overline{\rho}_{11}=A_1(A_1+A_2+A_3)/2(A_1+A_2),$$

$$\overline{\rho}_{22}=A_2(A_1+A_2+A_3)/2(A_1+A_2),$$

$$\overline{\rho}_{33}=(A_1+A_2+A_3)/2. \tag{16}$$

According to Eq. (16), applying the bichromatic field makes the population of level 3 increase to 3/2, while the populations of levels 1 and 2 decrease to 3/4. As can be seen from Eq. (16), the picture is not changed by repeating the pulses.

There is a simple procedure by which a radical readjustment of the populations can be achieved, in this case as well. Suppose that after the end of the bichromatic field pulse, the component coupling levels 1 and 3 is shut off. Because of the effect on the atom of the remaining field component, the populations of levels 2 and 3 become balanced and equal to 9/8. We again turn on both components of the bichromatic field, varying their strength in accordance with Eq. (15) and the new values of the initial populations A_i. After the second bichromatic field pulse, the following level populations are established: $\overline{\rho}_{11}=6/10$, $\overline{\rho}_{22}=9/10$, and $\overline{\rho}_{33}=15/10$, i.e., there will be a further reduction in the populations of levels 1 and 2 relative to that of level 3.

Let us relabel the quantities A_i in Eq. (16) as $\overline{\rho}_{ii}^{(N-1)}$. For the population $\overline{\rho}_{11}$ of level 1 after completion of the N-th cycle of this procedure, we have the recurrence relation

$$\overline{\rho}_{11}^{(N)}=\overline{\rho}_{11}^{(N-1)}\Sigma/(\Sigma+\overline{\rho}_{11}^{(N-1)}), \tag{17}$$

where $\Sigma=\rho_{11}+\rho_{22}+\rho_{33}$ is the (constant) sum of the populations of all the levels. The solution of the recurrence relation (17) is

$$\overline{\rho}_{11}^{(N)}=\Sigma/(N+3). \tag{18}$$

With the normalization, this yields

$$\overline{\rho}_{22}^{(N)}=\overline{\rho}_{33}^{(N)}=(\Sigma/2)(N+2)/(N+3). \tag{19}$$

It is clear from these equations that after a sufficiently large number of cycles, level 1 is essentially empty, i.e., the atom will be cooled, if this level is the one with the highest energy.

This procedure can also be applied in the case of $\Delta\varphi=\pi$ when condition (13) is satisfied. Then the populations of level 3 after the Nth and $(N-1)$th cycles will be coupled by the recurrence relation

$$\overline{\rho}_{33}^{(N)}=\overline{\rho}_{33}^{(N-1)}-(\overline{\rho}_{33}^{(N-1)})^2/4(\Sigma-\overline{\rho}_{33}^{(N-1)}). \tag{20}$$

An approximate solution of Eq. (20) can be found, assuming that

$$\overline{\rho}_{33}^{(N)}\approx\overline{\rho}_{33}^{(N-1)}+d(\overline{\rho}_{33}^{(N)})/dN. \tag{21}$$

As a result, we obtain the approximate differential equation

$$\frac{d(\overline{\rho}_{33}^{(N)})}{dN}=-\frac{(\overline{\rho}_{33}^{(N)})^2}{4\Sigma}, \tag{22}$$

whose solution looks like

$$\overline{\rho}_{33}^{(N)}=4\Sigma/(4\Sigma\xi+N). \tag{23}$$

The quantity ξ in Eq. (23) is a constant determined by the initial conditions. Calculations show that Eq. (23) yields higher (by 5–10%) values than the exact value for $\overline{\rho}_{33}$. The error in the calculation will be smaller for larger N_0 employed as an initial value in determining the constant ξ. Here the values of $\overline{\rho}_{33}$ for N ranging from unity to N_0 must be calculated exactly using Eq. (20).

Up to now we have assumed that the field is strong enough that the Rabi frequency is rather high and the field pulses are much longer than the period of the Rabi oscillations. Let the parameters be chosen so that the condition $|A_+|^2=0$ is satisfied (see condition (11)). Then the population of the common third level will be $A_3\cos^2(\Omega t)$. Then it is clear that a pulse of length $\pi/2\Omega$ will empty the common level completely.

3. COHERENT REPOPULATION OF THE HYPERFINE LEVELS OF A GAS ATOM IN A BICHROMATIC RADIO FREQUENCY WAVE

In this section we examine the interaction of atoms in a gas with a pulsed bichromatic radio frequency wave. Compared to the case of an impurity atom in a magnetic matrix,[6] the atoms in a gas are significantly affected by both transverse relaxation (relaxation time $T_2=1/\Gamma$) and longitudinal relaxation (relaxation time $T_1=1/\gamma$), whose times are of the same order of magnitude in the gaseous phase.

It is appropriate to examine the interaction of a resonant field with the atoms in a gas including the effect of relaxation using the density matrix formalism. The interaction of the

three level system proposed in Sec. 2 with a resonant bichromatic field in the interaction representation in the resonance approximation is described by a ninth-order system of equations for the density matrix ρ_{ij}:

$$\dot\rho_{11}+\gamma(\rho_{11}-\rho_{11}^0)=-iV_1(\rho_{31}-\rho_{13}),$$
$$\dot\rho_{22}+\gamma(\rho_{22}-\rho_{22}^0)=-iV_2(\rho_{32}-\rho_{23}),$$
$$\dot\rho_{33}+\gamma(\rho_{33}-\rho_{33}^0)=-iV_1(\rho_{13}-\rho_{31})-iV_2(\rho_{23}-\rho_{32}),$$
$$\dot\rho_{13}+\Gamma\rho_{13}=-iV_1(\rho_{33}-\rho_{11})+iV_2\rho_{12},$$
$$\dot\rho_{31}+\Gamma\rho_{31}=-iV_1(\rho_{11}-\rho_{33})-iV_2\rho_{21},$$
$$\dot\rho_{23}+\Gamma\rho_{23}=-iV_2(\rho_{33}-\rho_{22})+iV_1\rho_{21},$$
$$\dot\rho_{32}+\Gamma\rho_{32}=-iV_2(\rho_{22}-\rho_{33})-iV_1\rho_{12},$$
$$\dot\rho_{12}+\Gamma\rho_{12}=-iV_1\rho_{32}+iV_2\rho_{13},$$
$$\dot\rho_{21}+\Gamma\rho_{21}=-iV_2\rho_{31}+iV_1\rho_{23}. \tag{24}$$

In Eq. (24) the ρ_{ii}^0 are the equilibrium values of the level populations ρ_{ii}. As in Sec. 2, we shall assume that there are no stochastic changes in the phases of the states during the interaction process. In this case, using a system of eigenfunctions of the hamiltonian H_0 of the form (7), it can be shown that the phase factors of the form $\exp(i\alpha_i)$ on the right and left hand sides of Eqs. (24) cancel out, just as in Eqs. (9) for the amplitudes. Furthermore, as before, we shall assume that the phases of the components of the rf field are constant during the interaction. For simplicity we only consider the cases $\Delta\varphi=0,\pi$. Then the matrix elements of the magnetic dipole interaction of the atom with the magnetic field of the i-th component of the bichromatic wave, $V_i=|V_i|\exp(i\varphi_i)$, can be regarded as real: $V_1=V_2$ for $\Delta\varphi=0$ and $V_1=-V_2$ for $\Delta\varphi=\pi$. We introduce the notation $|V_1|=|V_2|=V$.

To simplify the system of Eqs. (24) we use the normalization condition $\rho_{11}+\rho_{22}+\rho_{33}=\Sigma$ and introduce the quantities $\chi_{13}=\rho_{31}-\rho_{13}$, $\chi_{23}=\rho_{32}-\rho_{23}$, and $\chi_{12}=\rho_{12}+\rho_{21}$. As a result, we transform from a ninth-order system to a fifth-order system of equations:

$$\dot\rho_{11}+\gamma(\rho_{11}-\rho_{11}^0)=-iV_1\chi_{13},$$
$$\dot\rho_{22}+\gamma(\rho_{22}-\rho_{22}^0)=-iV_2\chi_{23},$$
$$\dot\chi_{13}+\Gamma\chi_{13}=-iV_1(4\rho_{11}+2\rho_{22}-2\Sigma)-iV_2\chi_{12},$$
$$\dot\chi_{23}+\Gamma\chi_{23}=-iV_2(4\rho_{22}+2\rho_{11}-2\Sigma)-iV_1\chi_{12},$$
$$\dot\chi_{12}+\Gamma\chi_{12}=-iV_2\chi_{13}-iV_1\chi_{23}. \tag{25}$$

The characteristic equation of the system of Eqs. (25) is a complete fifth-order algebraic equation whose exact solution is hard to obtain. We shall obtain an approximate solution of this equation, as well as the system (25), that is accurate to within first order terms in the small parameters Γ/V and γ/V, assuming that the field is low enough that $V\gg\Gamma,\gamma$. The solution of the characteristic equation then has the form

$$k_1=-3\Gamma/4-\gamma/4,\quad k_2=i\sqrt{2}V-\Gamma/2-\gamma/2,$$
$$k_3=-i\sqrt{2}V-\Gamma/2-\gamma/2,$$
$$k_4=2i\sqrt{2}V-5\Gamma/8-3\gamma/8,$$
$$k_5=-2i\sqrt{2}V-5\Gamma/8-3\gamma/8. \tag{26}$$

Let us assume that all the levels of a gas atom are populated at the initial time, but not equally, in general: the diagonal elements of the density matrix are $\rho_{ii}(0)=A_i$. For the gas it is also necessary to assume that they are equal to their equilibrium values: $A_i=\rho_{ii}^0$. Suppose the rf field is turned on instantaneously. Then the phases of the nondiagonal matrix elements $\rho_{ij}(0)=\sqrt{A_iA_j}\exp[i(\alpha_i-\alpha_j)]$ will be fixed at the initial time, but differ in absolute value from zero. These initial conditions correspond to the initial conditions (3) formulated for the amplitudes $a_i(t)$ and mean that the atom is in a state described by the superposition (2).

That the field is turned on instantaneously means that the switch-on time is much shorter than the time between the interactions which randomly change the phase of the amplitudes of the atomic state, i.e., it is less than all relaxation times. When the field is turned on adiabatically for a time much greater than T_1,T_2, the phases of the nondiagonal matrix elements can change many times over the switch-on time and the initial conditions must then have the form $\rho_{ij}=0$ $(i\neq j)$.[1] Thus, we shall assume that both the switch-on time and the duration of the field pulse are shorter than the relaxation times; this corresponds to the assumption that there are no stochastic changes in the phases of the wave functions during the interaction process. It is important to emphasize that, as noted above, the solution of the equations for the density matrix for this sort of impulsive interaction between the atom and the field does not depend on the initial phases of the amplitudes of the atomic states. On the other hand, in order to achieve a Rabi-oscillation regime, the quantum mechanical system must interact with the field coherently for a time no greater than T_1,T_2, i.e., the period of the Rabi oscillations must be less than T_1,T_2. Thus, the field has to be strong enough to make $V\gg\Gamma,\gamma$. As will be seen below, the solution of the system of equations for the matrix ρ in the impulse regime when $\rho_{ij}(0)\neq0$ for $i\neq j$ is the same as the corresponding solution obtained in the wave function formalism with the initial conditions (3).

A solution of the system (25) that is accurate to within first order terms in the parameters Γ/V and γ/V for instantaneous switch-on of a strong field when $\Delta\varphi=0,\pi$ has the form

$$\rho_{11}(t)=\frac{(A_1+A_2)\gamma}{2\mu}+\frac{3\Gamma A}{\mu}+\left[\frac{3(2A-A_1-A_2)(\gamma-\Gamma)}{8\mu}\mp\frac{\sqrt{A_1A_2}}{4}\right]\exp\left[-\frac{(3\Gamma+\gamma)t}{4}\right]+\frac{1}{256}\left\{128(A_1-A_2)\cos(\sqrt{2}Vt)+\left[\pm\frac{24(\gamma-\Gamma)\sqrt{2A_1A_2}}{V}+\frac{4\sqrt{2}A_1(5\gamma^2+26\Gamma\gamma+33\Gamma^2)}{\mu V}-\frac{4\sqrt{2}A_2(11\gamma^2+38\Gamma\gamma+15\Gamma^2)}{\mu V}\right]\sin(\sqrt{2}Vt)\right\}$$

$$\times\exp\left[-\frac{(\Gamma+\gamma)t}{2}\right]+\frac{1}{256}\left\{[96(A_1+A_2-2A)\pm 64\sqrt{A_1A_2}]\cos(2\sqrt{2}Vt)+\left[\frac{3(A_1+A_2)(3\gamma^2+14\Gamma\gamma+15\Gamma^2)}{\mu V}-\frac{6A(\gamma^2+10\Gamma\gamma+21\Gamma^2)}{\mu V}\mp 10\sqrt{A_1A_2}\,\frac{\gamma-\Gamma}{V}\right]\sqrt{2}\,\sin(2\sqrt{2}Vt)\right\}\times\exp\left[-\frac{(3\gamma+5\Gamma)t}{8}\right], \quad (27)$$

where we have set $\mu=\gamma+3\Gamma$ and $A=\Sigma/3$ is the average population of the levels. The upper sign in Eq. (27) corresponds to $\Delta\varphi=0$ and the lower to $\Delta\varphi=\pi$. Similar expressions exist for $\rho_{22}(t)$, but they differ in that the terms proportional to $\cos(\sqrt{2}Vt)$ and $\sin(\sqrt{2}Vt)$ have the opposite sign. The population $\rho_{33}(t)$ can be calculated using the normalization condition.

These formulas imply that the populations depend significantly on the phase difference $\Delta\varphi$ of the components of the bichromatic wave. In the case of a pulse whose duration satisfies the condition $V^{-1}\ll\tau\ll\Gamma^{-1},\gamma^{-1}$, the populations of levels 1 and 2 will be equal to

$$\overline{\rho}_{11}=\frac{(A_1+A_2)\gamma}{2\mu}+\frac{3A\Gamma}{\mu}+\left[\frac{3(2A-A_1-A_2)(\gamma-\Gamma)}{8\mu}\mp\frac{\sqrt{A_1A_2}}{4}\right]\left[1-\frac{(3\Gamma+\gamma)\tau}{4}\right]. \quad (28)$$

In Eq. (28) the negative sign in front of $\sqrt{A_1A_2}$ corresponds to $\Delta\varphi=0$ and the positive sign to $\Delta\varphi=\pi$. The overbar denotes averaging of the Rabi oscillations over time. It is clear that after a short pulse lasting less than the relaxation times (T_1,T_2), the population of the common level will differ substantially from the populations of the other levels, even when all the levels were populated equally before the field was turned on. It is important that this difference depends on the phase difference $\Delta\varphi$. This effect arises as a result of the coherent summation of the population amplitudes in the common third level, i.e., it is analogous to coherent population capture. If $\tau>T_1,T_2$ holds, then the populations of all three levels will become equal to one another with time under the influence of relaxation.

4. COHERENT REPOPULATION OF THE OSCILLATOR LEVELS OF A NEUTRAL ATOM IN A MAGNETIC TRAP

Recently there has been extensive development of magnetic traps for reducing the kinetic energy (cooling) of neutral atoms. In these traps atoms are localized within a small volume of space ($\sim 10^{-2}$ cm) as a result of the interaction of the magnetic moments of the atoms with a complicated magnetic field configuration and contained there for some time ($\sim$100 s); the atoms are cooled by various methods, such as interactions with lasers whose frequencies are in resonance with an optical transition of the atom. Very low temperatures of the atomic gas can be attained (much lower than 1 K). The Bose condensation of alkali metal atoms was recently discovered using these methods.[7]

Laser cooling cannot be used, however, if the atoms do not have excited states at optical frequencies. In this case, it is possible to utilize coherent repopulation of the oscillator levels which an atom has as a particle in a magnetic trap, in the field of a bichromatic radio frequency wave.

The magnetic traps which are used have various field configurations. The simplest is the so-called quadrupole trap. The field in this trap varies spatially as

$$B\propto(4z^2+\rho^2)^{1/2} \quad (29)$$

(cylindrical geometry). Clearly, it varies linearly in any direction and the potential of the interaction with the magnetic moment of an atom has a V shape. In the field of such a trap the oscillator levels of an atom as a particle will be equidistant:[8]

$$E_N=\frac{3}{2}\,(h^2Mv^4\rho^{-2})^{1/3}N^{2/3}=\frac{3}{2}\,h\nu N^{2/3}. \quad (30)$$

Here N is an integer quantum number, h is Planck's constant, the trajectory of a particle of mass M in the magnetic field is assumed to be roughly circular with radius ρ about the central axis of the trap, and v is the linear particle velocity. For large N the particle moves adiabatically, while the orientation of its magnetic moment relative to the direction of the field in the trap is conserved.

Of the levels (30) it is possible to choose three arbitrary nonequidistant levels which can be populated equally at the initial time. After a bichromatic rf electromagnetic field which is resonant with the system of chosen oscillator levels is turned on, coherent repopulation of these levels takes place. The repopulation process can be carried out in the impulse regime. For this case, the theoretical analysis of Section 2 based on the amplitude of states formalism is valid, since, because of the low density of states in magnetic traps, there is essentially no relaxation.

Different variants of the repopulation procedure are possible. For example, by applying a pulsed rf field with a pulse duration τ and strength such that the condition $\Omega\tau=\pi/2$ is satisfied, it is possible to empty the common level. As shown in Sec. 2, a single pulse is sufficient to completely empty the common third level. If this level does not have the highest energy, then a π-pulse of a resonant monochromatic field can be used to transfer the population of a level whose energy is higher than that of the emptied level to a free level. Repeating this process for the next, lower energy level of the three levels, one can greatly cool the atomic gas in the trap by transferring the atoms to ever lower energy levels.

Another means of coherent repopulation can be realized by applying bichromatic field pulses which last longer than the period of the Rabi oscillations and have a phase difference of π between the components. Then, as can be seen from Eq. (14), after a few pulses it is possible to greatly reduce the population of the common level. Using the procedure involving alternating switching-on of one or both of

the components of a bichromatic field with a phase difference $\Delta\varphi=0$, it is possible, according to Eq. (18), to empty one of the levels that is not common in the three level system. It is proposed that subsequent cooling be carried out as in the previous variant.

In order to carry out these cooling procedures, several conditions must be satisfied. Thus, the Rabi frequency $\Omega\sim V/h$ (V is the matrix element of the operator for the interaction of the atomic magnetic moment with the rf field) must be lower than the frequency corresponding to the separation between the oscillator levels (30). For large N, this will be of order $\nu N^{-4/3}/3$. Thus, the following inequality must hold:

$$\Omega<\nu N^{-4/3}/3. \tag{31}$$

On the other hand, the time for a single field pulse must be less than T_s/N, where T_s is the time a particle is confined in the trap ($T_s\sim 100$ s). This yields the condition

$$\Omega^{-1}<\tau<T_s/N \tag{32}$$

or, in other words,

$$N<T_s\Omega. \tag{33}$$

Equations (31) and (33) yield a condition for the maximum quantum number N at which the proposed repopulation procedure can be used:

$$N^{7/3}<\nu T_s/3. \tag{34}$$

For $\rho\approx 10^{-2}$ cm, $v\approx 10^2$ cm/s, and $M\approx 100$ amu, we obtain an estimate of $\nu\sim 100$ kHz. Assuming $T_s\sim 10^2$ s, from Eq. (34) we obtain a maximum estimate for N of $N\sim 10^3$. The resonant frequency of the rf field, equal to the difference in the energies of the levels (30), is, for large N, on the order of $\nu N^{-1/3}\sim 10$ kHz. The particle energy for these values of N will be on the order of 10^{-7} eV, which corresponds to an atomic gas temperature of $\sim 10^{-3}$ K. Thus, the atomic gas must somehow be cooled to these temperatures, then it can be cooled further by the method proposed here. Equation (31) implies an estimate for the field width $V\sim\Omega h$ of $3\cdot 10^{-14}$ eV. From this we obtain an estimate of the required strength H of the electromagnetic field. Assuming that the interaction of the magnetic moment μ of the atom with an electromagnetic field of wavelength λ is dipole in character, we can write $V\sim\mu H\rho/\lambda$ ($\rho/\lambda\sim 10^{-8}$). Finally, we obtain $H\sim 10^2$ G. We can obtain an estimate for the minimum period of the Rabi oscillations from Eq. (33): $\Omega^{-1}\sim 0.1$ s. This criterion corresponds to an rf field with a strength of ~ 75 G.

5. CONCLUSION

The preceding discussion has shown that the coherent repopulation of atomic levels using a pulsed bichromatic radio frequency field can have a fairly wide range of applications. We believe that the most promising approaches are the following:

1. Coherent repopulation of hyperfine structure components makes it possible to polarize atoms in the gaseous phase. Unlike optical pumping, the proposed method is not related to the existence of optical transitions in the atoms.

2. The proposed method opens up the possibility of polarizing atoms and nuclei of impurity centers in a matrix without using ultralow temperatures.[6]

3. Emptying one or several components of the hyperfine structure results in anomalous transparency of a medium for Mössbauer gamma rays. This effect can be used in experiments to observe stimulated emission for nuclear transitions.

4. A bichromatic rf wave can be used to induce cooling of neutral atoms in magnetic traps. In this case, deep cooling would become possible for atoms that have no excited levels in the optical range, such as atomic hydrogen.

For cooling atoms in magnetic traps it may turn out to be more convenient to use an oscillating component of the magnetic field of the trap itself as the repopulating variable field. This component should have two resonant frequencies corresponding to transitions between oscillator levels of the atom in the constant field of the trap, while the spatial variation in its amplitude and its polarization also should correspond to the parameters of the constant field.

The criteria for the strength of the repopulating variable field cited in Sec. 4 give an oscillating trap field component of this sort that is several orders of magnitude smaller than for an rf electromagnetic field, as the ratio ρ/λ will be of order unity in this case.

In conclusion, we thank Prof. R. Kuseman, Prof. G. Odors (Louvain, Belgium), and Prof. M. Levenstein (Saclay, France) for useful discussions.

This work was supported by the Russian Fund for Fundamental Research (Grant No. 96-02-17612a).

[1] B. D. Agap'ev, M. B. Gornyĭ, B. G. Matisov, and Yu. V. Rozhdestvenskiĭ, Usp. Fiz. Nauk **163**(9), 1 (1993) [Phys. Usp. **36**, 763 (1993)].
[2] G. Alzetta, L. Moi, and G. Orriols, Nuovo Cimento B **52**, 209 (1979).
[3] G. Orriols, Nuovo Cimento B **53**, 1 (1979).
[4] D. F. Zaretsky and S. B. Sazonov, Phys. Lett. A **198**, 55 (1995).
[5] D. F. Zaretskiĭ and S. B. Sazonov, JETP Lett. **60**, 699 (1994).
[6] D. F. Zaretskiĭ and S. B. Sazonov, Zh. Éksp. Teor. Fiz. **111**, 1236 (1997) [JETP **84**, 682 (1997)].
[7] K. D. Davies, M. O. Mewes, M. R. Andrews, N. J. van Druten, D. S. Durfee, D. M. Kurn, and W. Ketterle, Phys. Rev. Lett. **22**, 3969 (1992).
[8] H. Metcalf and P. van der Straten, Phys. Rep. **244**, 203 (1994).

Translated by D. H. McNeill

Effect of the state of a quantized electromagnetic field on the interaction with an atom with allowance for the continuum

A. D. Gazazyan

Institute of Physical Research, Armenian National Academy of Sciences, 378410 Ashtarak-2, Armenia
(Submitted 1 August 1997)
Zh. Éksp. Teor. Fiz. **113**, 1193–1205 (April 1998)

This paper studies the effect of a transition into the continuous spectrum on the ''collapse'' and ''revival'' of population oscillations in an atom. It is shown that at large values of the mean number of photons in a radiation field and in conditions of weak ionization the phenomena of collapse and revival can still be observed, but the amplitude of population oscillations decreases exponentially because of the damping of the level. The interaction of a quantized electromagnetic field with a Λ system of an atom when one state is continuous is examined. Expressions are derived for the probability of ''survival'' of the atom when the quantized field was initially in a state with a given number of photons and when it was in a coherent state. An approximate calculation of the sum in averaging over the photon number distribution in the case of a coherent field leads to expressions for the probabilities of survival of the atom that transform into expressions, as the mean number of photons tends to infinity, corresponding to the case of a field in the representation of a fixed number of photons. The possibility of a stable state existing in a coherent quantized field is examined. It is found that for a Λ system the condition for the existence of a stable state remains valid in the case of a coherent state of the field when the photon number is large. © *1998 American Institute of Physics.* [S1063-7761(98)00404-1]

1. INTRODUCTION

The study of various quantum states of electromagnetic radiation and the nature of the interaction are important in identifying optical fields and for their applications.

The first research in the interaction involving coherent quantized radiation was done in Refs. 1 and 2. New results followed. In Refs. 3–5, in particular, the phenomena of ''collapse'' and ''revival'' of atomic population oscillations in the interaction with coherent quantized radiation were predicted. Such phenomena cannot exist when a classical field is involved, which points to their quantum nature due to the discreteness of photons. Collapse and revival of population oscillations were observed first in the experiments of Rempe *et al.*[6,7] are were later studied by many researchers. Similar effects can be observed in squeezed quantized fields (see, e.g., Ref. 8).

Studying the interaction of a quantized electromagnetic field and atoms in the presence of transitions into the continuous spectrum is important because such transitions can dramatically change the pattern of the phenomena. In particular, in Refs. 9 and 10 it is shown that the decay of an isolated level initiated by quantized radiation definitely does not follow an exponential law and the probability of ''survival'' of the atom does not tend to zero for large times. Section 2 studies the population oscillations in the atom when there are transitions into the continuous spectrum.

Section 3 discusses the interaction of quantized electromagnetic radiation and an atom of the Λ type with one of the states being continuous. For a classical field this problem was discussed by Rzazewski and Eberly.[11]. When a certain condition is met[11–14] and a classical or quantum field interacts with the atom, the system always has a stable state in the photon number representation. If this condition is satisfied, the probability of survival of the atom does not tend to zero for large interaction times. Coleman *et al.*[12] studied this system in the case of a quantized radiation field. When initially the quantized field is in a coherent state, Coleman *et al.*[12] assume that the field is extremely strong and replace the number of photons by its mean value. Such replacement at the initial stage makes it impossible to study the effect of fluctuation of the number of photons in the coherent radiation on the process.

The present paper examines the interaction of a quantized electromagnetic field and an atom in the cases of a fixed number of photons and a coherent state. When the quantized field is coherent, the exact expressions for the probabilities are written in the form of a sum averaged over the photon number distribution. Expressions are derived for the probabilities of survival of the atom after this sum is calculated in an approximate manner. At large values of the mean number of photons, when $\bar{n}\to\infty$, the results correspond to the case of a quantized field with a fixed number of photons. In particular, it is found that if a certain condition is met, there is also a stable state when the atom interacts with coherent quantized radiation with a large mean number of photons in the beam. Fedorov and Movsesian[15] were the first to predict such stabilization of the atom in strong fields under certain conditions.

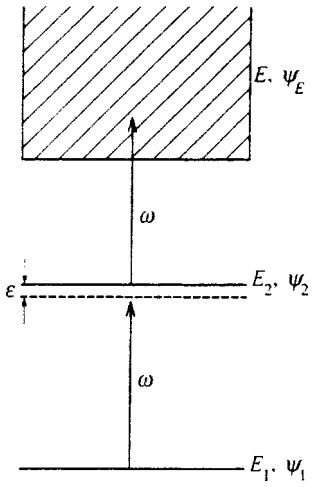


FIG. 1.

2. TWO-LEVEL SYSTEM WITH A CONTINUUM IN A QUANTIZED RADIATION FIELD

Let us examine a two-level system with a continuum (Fig. 1) in the interaction with a quantized radiation field of frequency ω close to the frequency of the transition between the discrete levels of the atoms with wave functions ψ_1 and ψ_2. A second photon in the radiation connects the second atomic state ψ_2 with the continuum. The system Hamiltonian has the following form:

$$H=H_{\rm at}+\omega c^{\dagger}c+\beta^{\dagger}c+c^{\dagger}\beta, \tag{2.1}$$

where ω is the frequency of the quantized radiation, c and $c^{\dagger}$ are the photon annihilation and creation operators, and β are the operators of the transitions in the atom.

Writing the wave function of the system consisting of the atom and the field as

$$\Phi(t)=\sum_{k=1}^{2} a_k(t)\psi_k+\int dE\ b_E(t)\psi_E, \tag{2.2}$$

we obtain, from the Schrödinger equation with the Hamiltonian (2.1), the following system of equations for the expansion coefficients in (2.2):

$$\begin{aligned}
i\frac{da_1(t)}{dt}&=(E_1+\omega c^{\dagger}c)a_1(t)+\beta_{12}c^{\dagger}a_2(t),\\
i\frac{da_2(t)}{dt}&=(E_2+\omega c^{\dagger}c)a_2(t)+\beta_{12}^{*}ca_1(t)\\
&\quad+\int dE\ \beta_{2E}c^{\dagger}b_E(t),\\
i\frac{db_E(t)}{dt}&=(E+\omega c^{\dagger}c)b_E(t)+\beta_{2E}^{*}ca_2(t),
\end{aligned} \tag{2.3}$$

where β_{12} and β_{2E} are, respectively, the matrix elements of the transitions between the discrete states of the atom and between state ψ_2 and the continuous spectrum.

Introducing the transformations

$$a_i(t)=\exp\{-i(E_i+\omega c^{\dagger}c)t\}\ f_i(t),\quad i=1,2,$$

$$b_E(t)=\exp\{-i(E+\omega c^{\dagger}c)t\}\ f_E(t) \tag{2.4}$$

and eliminating the third equation in (2.3), we arrive at the following system of equations:

$$\begin{aligned}
i\frac{df_1(t)}{dt}&=\beta_{12}\exp(-i\varepsilon t)c^{\dagger}f_2(t),\\
i\frac{df_2(t)}{dt}&=\beta_{12}^{*}\exp(i\varepsilon t)cf_1(t)-i\int dE\ |\beta_{2E}|^2c^{\dagger}c\int_0^t dt'\\
&\quad\times\exp[i(E-E_2-\omega)(t'-t)]\ f_2(t'),
\end{aligned} \tag{2.5}$$

where ε is the offset from resonance,

$$\varepsilon=E_2-E_1-\omega. \tag{2.6}$$

Expanding the amplitudes $f_1(t)$ and $f_2(t)$ in the photon number states,

$$f_i(t)=\sum_{n=0}^{\infty} f_i(t,n)|n\rangle, \tag{2.7}$$

we arrive at the following system of equations:

$$\begin{aligned}
i\frac{df_1(t,n)}{dt}&=\sqrt{n}\ \beta_{12}\exp(-i\varepsilon t)\ f_2(t,n-1),\\
i\frac{df_2(t,n-1)}{dt}&=\sqrt{n}\ \beta_{12}^{*}\exp(i\varepsilon t)\ f_1(t,n)-i\\
&\quad\times\int dE\ (n-1)|\beta_{2E}|^2\\
&\quad\times\int_0^t\exp[i(E-E_2-\omega)(t'-t)]\\
&\quad\times f_2(t',n-1)dt'.
\end{aligned} \tag{2.8}$$

If initially the atom was in the state ψ_1,

$$f_1(0,n)=\rho(n),\quad f_2(0,n-1)=0, \tag{2.9}$$

where $\rho(n)$ is the amplitude of the photon number distribution in the incident electromagnetic radiation, then the solution of the system (2.8) is

$$\begin{aligned}
f_1(t,n)&=\rho(n)\exp\left\{-\frac{i}{2}[\varepsilon+(n-1)\Delta]t\right\}\\
&\quad\times\exp\left\{-(n-1)\frac{\Gamma t}{4}\right\}\left\{\cos\frac{\Omega(n)}{2}t\right.\\
&\quad\left.+i\frac{\varepsilon+(n-1)\Delta-(n-1)i\Gamma/2}{\Omega(n)}\sin\frac{\Omega(n)}{2}t\right\},\\
f_2(t,n-1)&=-\rho(n)\frac{2i\sqrt{n}}{\Omega(n)}\beta_{12}^{*}\\
&\quad\times\exp\left\{\frac{i}{2}[\varepsilon-(n-1)\Delta]t\right\}\\
&\quad\times\exp\left\{-(n-1)\frac{\Gamma t}{4}\right\}\sin\frac{\Omega(n)}{2}t.
\end{aligned} \tag{2.10}$$

Here $\Omega(n)$ is the Rabi frequency, and Δ and Γ are, respectively, the shift and width of the upper discrete level:

$$\Omega(n)=\left\{\left[\varepsilon+(n-1)\left(\Delta-\frac{i\Gamma}{2}\right)\right]^2+4n|\beta_{12}|^2\right\}^{1/2},$$
$$\Delta=-P\int dE\,\frac{|\beta_{2E}|^2}{E-E_1-2\omega},\quad \Gamma=2\pi|\beta_{2,E_1+2\omega}|^2. \tag{2.11}$$

In deriving these expressions we assumed that Δ and Γ are weakly dependent on the argument.

Allowing for the expressions (2.10), we can write the following formulas for the probabilities of finding the atom in states ψ_1 and ψ_2, respectively:

$$W_1(t)=\sum_{n=0}^{\infty}|\rho(n)|^2\exp\left\{-(n-1)\frac{\Gamma t}{2}\right\}\left|\cos\frac{\Omega(n)}{2}t+i\,\frac{\varepsilon+(n-1)(\Delta-i\Gamma/2)}{\Omega(n)}\sin\frac{\Omega(n)}{2}t\right|^2,$$
$$W_2(t)=\sum_{n=0}^{\infty}|\rho(n)|^2\exp\left\{-(n-1)\frac{\Gamma t}{2}\right\}\left|\sin\frac{\Omega(n)}{2}t\right|^2. \tag{2.12}$$

When the quantized electromagnetic field is coherent, the distribution of the number of photons in the electromagnetic radiation is given by the Poisson law

$$|\rho(n)|^2=\exp(-\bar{n})\frac{\bar{n}^n}{n!}. \tag{2.13}$$

If initially the atom was in the state ψ_2,

$$f_1(0,n)=0,\qquad f_2(0,n-1)=\rho(n), \tag{2.14}$$

the probabilities of finding the atom in the states ψ_1 and ψ_2 are, respectively,

$$W_1(t)=\sum_{n=0}^{\infty}|\rho(n)|^2\,\frac{4(n+1)|\beta_{12}|^2}{|\Omega(n)|^2}\times\exp\left(-n\,\frac{\Gamma t}{2}\right)\left|\sin\frac{\Omega(n)}{2}t\right|^2,$$
$$W_2(t)=\sum_{n=0}^{\infty}|\rho(n)|^2\exp\left(-n\,\frac{\Gamma t}{2}\right)\left|\cos\frac{\Omega(n)}{2}t-i\,\frac{\varepsilon+n(\Delta-i\Gamma/2)}{\Omega(n)}\sin\frac{\Omega(n)}{2}t\right|^2, \tag{2.15}$$

where

$$\Omega(n)=\{[\varepsilon+n(\Delta-i\Gamma/2)]^2+4(n+1)|\beta_{12}|^2\}^{1/2}. \tag{2.16}$$

We see that because of damping the nature of collapse and revival of oscillations of the level populations changes considerably.

Let us examine Eq. (2.15) to find the probability of finding the atom on the upper level initially. For $\bar{n}\gg 1$ the Poisson distribution has a sharp peak at $n=\bar{n}$, so that we can use the expansion

$$\Omega(n)\approx\Omega(\bar{n})+\gamma(\bar{n})(n-\bar{n}), \tag{2.17}$$

where

$$\gamma(\bar{n})=\left.\frac{d\Omega(n)}{dn}\right|_{n=\bar{n}}=\left(\Delta-\frac{i\Gamma}{2}\right)\sqrt{1-\frac{4\bar{n}|\beta_{12}|^2}{\Omega^2(\bar{n})}}+\frac{2|\beta_{12}|^2}{\Omega(\bar{n})}. \tag{2.18}$$

Performing the summation in the expression for $W_2(t)$ in (2.15) via (2.17), we arrive at an approximate formula for the probability in the case of coherent radiation:

$$W_2(t)\approx\tfrac{1}{4}\{|1-\chi|^2\exp[\bar{n}(\exp[-(\Gamma/2+\mathrm{Im}\,\gamma(\bar{n}))t]-1)-t\ \mathrm{Im}\ p]+|1+\chi|^2\exp[\bar{n}(\exp[-(\Gamma/2-\mathrm{Im}\,\gamma(\bar{n}))t]-1)+t\ \mathrm{Im}\ p]+2\ \mathrm{Re}[(1+\chi)(1-\chi^*)\exp[\bar{n}\times(\exp[-(\Gamma/2+i\ \mathrm{Re}\ \gamma(\bar{n}))t]-1)-it\ \mathrm{Re}\ p]]\}, \tag{2.19}$$

where

$$\chi=\frac{\varepsilon+\bar{n}(\Delta-i\Gamma/2)}{\Omega(\bar{n})},\quad p=\Omega(\bar{n})-\bar{n}\gamma(\bar{n}). \tag{2.20}$$

In the limits $\beta_{12}\to 0$ or $\Gamma,\ \Delta\to 0$ the above expressions transform into the expressions obtained in Ref. 8 or Refs. 9 and 10.

To simplify (2.19) still further, we examine the special case of exact resonance, $\varepsilon=0$, and weak ionization, $\bar{n}\Delta,\bar{n}\Gamma\ll\sqrt{\bar{n}}\,|B_{12}|$. Then (2.19) yields

$$W_2(t)\approx\frac{1}{2}\left\{\exp[\bar{n}(\exp(-\Gamma t)-1)]+\exp\left[\bar{n}\left(\exp\left(-\frac{\Gamma t}{2}\right)\cos\frac{|\beta_{12}|t}{\sqrt{\bar{n}}}-1\right)\right]\times\left[\cos\left(\bar{n}\exp\left(-\frac{\Gamma t}{2}\right)\sin\frac{|\beta_{12}|t}{\sqrt{\bar{n}}}+\sqrt{\bar{n}}\,|\beta_{12}|t\right)-\frac{\bar{n}\Gamma}{\sqrt{\bar{n}}\,|\beta_{12}|}\sin\left(\bar{n}\exp\left(-\frac{\Gamma t}{2}\right)\sin\frac{|\beta_{12}|t}{\sqrt{\bar{n}}}+\sqrt{\bar{n}}\,|\beta_{12}|t\right)\right]\right\}. \tag{2.21}$$

We see from the expression for $W_2(t)$ that modulation of population oscillations leads to a collapse and revival of the population oscillations. The revival period is given by the following expression[8]:

$$T=\frac{2\pi\sqrt{\bar{n}}}{|\beta_{12}|}. \tag{2.22}$$

At the initial stage of the process, for small times, when $\Gamma t\ll|\beta_{12}|t/\sqrt{\bar{n}}\ll 1$, Eq. (2.21) becomes

$$W_2(t)\approx\frac{1}{2}\exp(-\bar{n}\Gamma t)\left\{1+\exp\left(-\frac{|\beta_{12}|^2t^2}{\sqrt{\bar{n}}}\right)\right.$$

$$\times\cos(2\sqrt{\bar{n}}\,|\beta_{12}|t)-\frac{\bar{n}\Gamma}{\sqrt{\bar{n}}\,|\beta_{12}|}$$

$$\times\sin(2\sqrt{\bar{n}}\,|\beta_{12}|t)\Bigg]\Bigg\}. \tag{2.23}$$

The damping of population oscillations initially follows a Gaussian law, as predicted by Cummings.[1] Subsequently, oscillations revive after each period T has elapsed, but the amplitude of these oscillations decreases exponentially due to atom ionization. For large times, as Eq. (2.21) implies, the probability of survival of an atom will not tend to zero, as it does in the case of an isolated level[9,10]; rather, it will be an exponentially small quantity, since $\bar{n}\gg 1$:

$$W_2(t)\approx\frac{e^{-\bar{n}}}{2}\Bigg[\cos(\sqrt{\bar{n}}\,|\beta_{12}|t)$$

$$-\frac{\bar{n}\Gamma}{\sqrt{\bar{n}}\,|\beta_{12}|}\sin(\sqrt{\bar{n}}\,|\beta_{12}|t)\Bigg]. \tag{2.24}$$

3. EFFECT OF THE STATES OF A QUANTIZED ELECTROMAGNETIC FIELD ON THE STABILIZATION OF ATOMIC LEVELS IN IONIZATION

When an atom is exposed to an external electromagnetic field, a new discrete level is induced in its continuous spectrum, and as a result of interference with the continuum states this level leads to a continuum with structure or to autoionization-like resonances. As shown in Refs. 11–14, under certain conditions, depending on the strength and frequency of the external field, near the Fano minimum the photoelectron spectrum undergoes a sharp change due to the destructive interference of various ionization channels. When a certain condition is met, one of the ''dressed'' states does not decay. Under certain conditions stabilization can be observed in strong fields, too. Stabilization in strong fields was predicted by Fedorov and Movsesian.[15] In all such studies the external field is assumed classical. In the present paper we examine how the state of a quantized electromagnetic field affects the formation of a stable state of an atom.

Let us consider a Λ system whose upper state has a continuous spectrum (Fig. 2) in an external quantized radiation field with frequencies ω_1 and ω_2. The field with frequency ω_1 connects the discrete state ψ_1 with the continuum, while the field with frequency ω_2 connects the discrete level ψ_2 with the continuum. The corresponding Hamiltonian of the system consisting of the atom and the quantized field is

$$H=H_{\rm at}+\sum_{k=1}^{2}\omega_k c_k^{\dagger}c_k+\sum_{k=1}^{2}(\beta_k^{\dagger}c_k+c_k^{\dagger}\beta_k), \tag{3.1}$$

where c_1, $c_1^{\dagger}$ and c_2, $c_2^{\dagger}$ are the photon annihilation and creation operators for the first and the second fields, respectively, and β_1 and β_2 are the operators of transitions from the states ψ_1 and ψ_2 into the continuum.

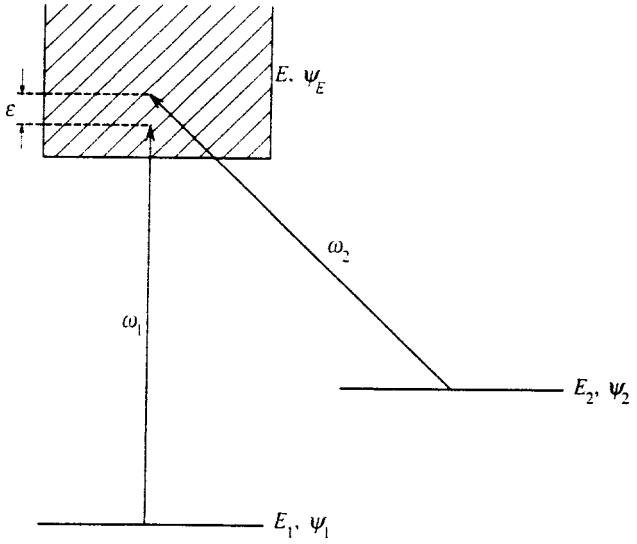


FIG. 2.

Representing the wave function of the atom–field system in the form (2.2) and proceeding from the Hamiltonian (3.1), we arrive at the following system of equations for the expansion amplitudes a_i $(i=1,2)$ and b_E:

$$i\frac{da_i(t)}{dt}=\Bigg(E_i+\sum_{k=1}^{2}\omega_k c_k^{\dagger}c_k\Bigg)a_i(t)+\int dE\;\beta_{iE}c_i^{\dagger}b_E(t),$$

$$i\frac{db_E(t)}{dt}=\Bigg(E+\sum_{k=1}^{2}\omega_k c_k^{\dagger}c_k\Bigg)b_E(t)+\sum_{k=1}^{2}\beta_{kE}^{*}c_k a_k(t). \tag{3.2}$$

Introducing the transformations

$$a_i(t)=\exp\Bigg\{-i\Bigg(E_i+\sum_{k=1}^{2}\omega_k c_k^{\dagger}c_k\Bigg)t\Bigg\}f_i(t),\quad i=1,2,$$

$$b_E(t)=\exp\Bigg\{-i\Bigg(E+\sum_{k=1}^{2}\omega_k c_k^{\dagger}c_k\Bigg)t\Bigg\}f_E(t) \tag{3.3}$$

into the system of equations (3.2), we arrive at the following system of equations:

$$i\frac{df_i(t)}{dt}=\int dE\;\beta_{iE}\exp[-i(E-E_i-\omega_i)t]$$

$$\times c_i^{\dagger}f_E(t),\quad i=1,2,$$

$$i\frac{df_E(t)}{dt}=\sum_{i=1}^{2}\beta_{iE}^{*}\exp[i(E-E_i-\omega_i)t]\;c_i f_i(t). \tag{3.4}$$

Next, expanding the amplitudes $f_i(t)$ and $f_E(t)$ in the photon number states,

$$f_i(t)=\sum_{n_1,n_2=0}^{\infty}f_i(n_1,n_2,t)|n_1,n_2\rangle,$$

$$f_E(t)=\sum_{n_1,n_2=0}^{\infty}f_E(n_1,n_2,t)|n_1,n_2\rangle, \tag{3.5}$$

we arrive at the following system of equations:

$$i\frac{df_1(n_1,n_2,t)}{dt}=\int dE\ \beta_{1E}\exp[-i(E-E_1-\omega_1)t]\ \sqrt{n_1}\ f_E(n_1-1,n_2,t),$$

$$i\frac{df_2(n_1-1,n_2+1,t)}{dt}=\int dE\ \beta_{2E}\exp[-i(E-E_2-\omega_2)t]\ \sqrt{n_2+1}\ f_E(n_1-1,n_2,t),\tag{3.6}$$

$$i\frac{df_E(n_1-1,n_2,t)}{dt}=\exp[i(E-E_1-\omega_1)t]\beta_{1E}^*\sqrt{n_1}\times f_1(n_1,n_2,t)+\exp[i(E-E_2-\omega_2)t]\ \beta_{2E}^*\sqrt{n_2+1}\times f_2(n_1-1,n_2+1,t).$$

If initially $(t=0)$ the atom was in the state ψ_1,

$$f_1(n_1,n_2,0)=\rho(n_1,n_2),\quad f_2(n_1-1,n_2+1,0)=0,\quad f_E(n_1-1,n_2,0)=0,\tag{3.7}$$

where $\rho(n_1,n_2)$ is the amplitude of the photon number distribution in the radiation field, then the solution of system (3.6) has the form

$$f_1(n_1,n_2,t)=\rho(n_1,n_2)\exp\left\{-\frac{i}{2}\left[\varepsilon+(n_2+1)\times\left(\Delta_2-\frac{i\Gamma_2}{2}\right)+n_1\left(\Delta_1-\frac{i\Gamma_1}{2}\right)\right]t\right\}\times\left\{\cos\frac{\Omega(n_1,n_2)}{2}t+i\times\frac{\varepsilon+(n_2+1)(\Delta_2-i\Gamma_2/2)-n_1(\Delta_1-i\Gamma_1/2)}{\Omega(n_1,n_2)}\times\sin\frac{\Omega(n_1,n_2)t}{2}\right\},$$

$$f_2(n_1-1,n_2+1,t)=-\frac{2i\sqrt{n_1(n_2+1)}\ F_{12}(1-i/q)}{\Omega(n_1,n_2)}\times\rho(n_1,n_2)\exp\left\{\frac{i}{2}\left[\varepsilon-n_1\left(\Delta_1-\frac{i\Gamma_1}{2}\right)+(n_2+1)\left(\Delta_2-\frac{i\Gamma_2}{2}\right)\right]t\right\}\times\sin\frac{\Omega(n_1,n_2)t}{2},\tag{3.8}$$

where

$$F_{12}=-P\int dE\ \frac{\beta_{1E}(\omega_1)\beta_{2E}^*(\omega_2)}{E-E_1-\omega_1},$$

$$\Delta_{1,2}=-P\int dE\ \frac{|\beta_{1,2E}|^2}{E-E_1-\omega_1},\tag{3.9}$$

$$\Gamma_{1,2}=2\pi|\beta_{1,2;E_1+\omega_1}|^2,\quad \varepsilon=E_2-E_1+\omega_2-\omega_1,$$

$\Omega(n_1,n_2)$ is the Rabi frequency,

$$\Omega(n_1,n_2)=\left\{\left[\varepsilon+(n_2+1)\left(\Delta_2-\frac{i\Gamma_2}{2}\right)-n_1\left(\Delta_1-\frac{i\Gamma_1}{2}\right)\right]^2+4n_1(n_2+1)|F_{12}|^2\left(1-\frac{i}{q}\right)^2\right\}^{1/2},\tag{3.10}$$

and q is the Fano parameter,

$$q=\frac{2F_{12}}{\sqrt{\Gamma_1\Gamma_2}}.\tag{3.11}$$

If we take into account the expressions (3.8), the probabilities of finding the atom in states ψ_1 and ψ_2 are, respectively,

$$W_1(t)=\sum_{n_1,n_2=0}^{\infty}|\rho(n_1,n_2)|^2\times\exp\left\{-\frac{1}{2}[n_1\Gamma_1+(n_2+1)\Gamma_2]t\right\}\left|\cos\frac{\Omega(n_1,n_2)}{2}t+i\frac{\varepsilon+(n_2+1)(\Delta_2-i\Gamma_2/2)-n_1(\Delta_1-i\Gamma_1/2)}{\Omega(n_1,n_2)}\times\sin\frac{\Omega(n_1,n_2)}{2}t\right|^2,\tag{3.12}$$

$$W_2(t)=\sum_{n_1,n_2=0}^{\infty}|\rho(n_1,n_2)|^2\frac{4n_1(n_2+1)|F_{12}|^2|1-i/q|^2}{|\Omega(n_1,n_2)|^2}\times\exp\left\{-\frac{1}{2}[n_1\Gamma_1+(n_2+1)\Gamma_2]t\right\}\times\left|\sin\frac{\Omega(n_1,n_2)}{2}t\right|^2.$$

In the case of coherent quantized fields, $|\rho(n_1,n_2)|^2$ is given by the Poisson distribution

$$|\rho(n_1,n_2)|^2=\exp[-(\bar{n}_1+\bar{n}_2)]\frac{\bar{n}_1^{n_1}\bar{n}_2^{n_2}}{n_1!\ n_2!}.\tag{3.13}$$

As with classical radiation, in quantized fields there is a stable state in the photon number representation, with

$$\varepsilon=\varepsilon_0=\frac{q}{2}[(n_2+1)\Gamma_2-n_1\Gamma_1]-(n_2+1)\Delta_2+n_1\Delta_1.\tag{3.14}$$

For further studies of the results of our investigation we limit ourselves to the case of a single field with frequency ω. Then the probabilities of finding the atom in states ψ_1 and ψ_2 become, respectively,

$$W_1(t)=\sum_{n=0}^{\infty}|\rho(n)|^2\exp\left(-n\frac{\Gamma t}{2}\right)$$

$$\times\left|\cos\frac{\Omega(n)}{2}t+\frac{\varepsilon+n(\Delta_2-\Delta_1-i(\Gamma_2-\Gamma_1)/2)}{\Omega(n)}\sin\frac{\Omega(n)}{2}t\right|^2, \tag{3.15}$$

$$W_2(t)=\sum_{n=0}^{\infty}|\rho(n)|^2\exp\left(-n\frac{\Gamma t}{2}\right)\times\frac{4n^2|F_{12}|^2|1-i/q|^2}{|\Omega(n)|^2}\left|\sin\frac{\Omega(n)}{2}t\right|^2,$$

where

$$\Omega(n)=\left\{\left[\varepsilon+n\left(\Delta_2-\Delta_1-\frac{i(\Gamma_2-\Gamma_1)}{2}\right)\right]^2+4n^2|F_{12}|^2(1-i/q)^2\right\}^{1/2}, \tag{3.16}$$

and

$$\varepsilon=E_2-E_1. \tag{3.17}$$

When the external quantized field is in a state with a fixed number of photons, i.e.,

$$\rho(n)=\delta_{n,n'}, \tag{3.18}$$

and the condition

$$\varepsilon=\varepsilon_0=\frac{q}{2}n(\Gamma_2-\Gamma_1)-n(\Delta_2-\Delta_1) \tag{3.19}$$

necessary for a stable state is met, the probabilities of finding the atom in the states ψ_1 and ψ_2 are, respectively,

$$W_1(t)=\frac{\Gamma_2^2}{(\Gamma_1+\Gamma_2)^2}\left\{1+\frac{\Gamma_1^2}{\Gamma_2^2}\exp[-n(\Gamma_1+\Gamma_2)t]+2\frac{\Gamma_1}{\Gamma_2}\exp\left[-\frac{n(\Gamma_1+\Gamma_2)t}{2}\right]\times\cos q\frac{n(\Gamma_1+\Gamma_2)t}{2}\right\}, \tag{3.20}$$

$$W_2(t)=\frac{\Gamma_1\Gamma_2}{(\Gamma_1+\Gamma_2)^2}\left\{1+\exp\left[-\frac{n(\Gamma_1+\Gamma_2)t}{2}\right]+2\exp[-n(\Gamma_1+\Gamma_2)t]\cos q\frac{n(\Gamma_1+\Gamma_2)t}{2}\right\}.$$

For large times, as $t\to\infty$, the above expressions yield

$$W_1(\infty)=\frac{\Gamma_2^2}{(\Gamma_1+\Gamma_2)^2},\quad W_2(\infty)=\frac{\Gamma_1\Gamma_2}{(\Gamma_1+\Gamma_2)^2}. \tag{3.21}$$

In conditions where $\Gamma_1\approx\Gamma_2$ we have $W_1(\infty)=W_2(\infty)=1/4$.

The total probability of atom survival is

$$W(t)=W_1(t)+W_2(t)=\frac{\Gamma_2}{\Gamma_1+\Gamma_2}\left\{1+\frac{\Gamma_1}{\Gamma_2}\exp[-n(\Gamma_1+\Gamma_2)t]\right\}, \tag{3.22}$$

and for large times, as $t\to\infty$, we have

$$W(\infty)=\frac{\Gamma_2}{\Gamma_1+\Gamma_2}. \tag{3.23}$$

When $\Gamma_1\approx\Gamma_2$ and the condition (3.19) is met, the probability of the atom being ionized is $W(\infty)\approx 1/2$.

Now let us examine the case of a coherent quantized field, where the photon number is described by the Poisson distribution (2.13). How will (3.15) look for $\bar{n}\gg 1$? As in Sec. 2, since the Poisson distribution has a sharp maximum for $\bar{n}\gg 1$, we can use the expansion (2.17) for the Rabi frequency (3.16), where

$$\gamma(\bar{n})=\sqrt{1-\frac{4\bar{n}^2|F_{12}|^2(1-i/q)^2}{\Omega^2(\bar{n})}}\left[\Delta_2-\Delta_1-\frac{i(\Gamma_2-\Gamma_1)}{2}\right]+\frac{4\bar{n}|F_{12}|^2(1-i/q)^2}{\Omega(\bar{n})}. \tag{3.24}$$

Performing the summation in (2.15) via (2.17), we arrive at the following approximate formulas for the probabilities:

$$W_1(t)\approx\frac{1}{4}\left\{|1+\chi|^2\exp\left[\bar{n}\left(\exp\left[-\left(\frac{\Gamma_1+\Gamma_2}{2}+\mathrm{Im}\ \gamma(\bar{n})\right)t\right]-1\right)-t\ \mathrm{Im}\ p\right]+|1-\chi|^2\times\exp\left[\bar{n}\left(\exp\left[-\left(\frac{\Gamma_1+\Gamma_2}{2}-\mathrm{Im}\ \gamma(\bar{n})\right)t\right]-1\right)+t\times\mathrm{Im}\ p\right]+2\ \mathrm{Re}\left[(1+\chi)(1-\chi^*)\times\exp\left[\bar{n}\left(\exp\left[-\left(\frac{\Gamma_1+\Gamma_2}{2}-i\times\mathrm{Re}\ \gamma(\bar{n})\right)t\right]-1\right)+it\ \mathrm{Re}\ p\right]\right]\right\}, \tag{3.25}$$

$$W_2(t)\approx\frac{1}{4}\left\{|1-\chi|^2\exp\left[\bar{n}\left(\exp\left[-\left(\frac{\Gamma_1+\Gamma_2}{2}+\mathrm{Im}\ \gamma(\bar{n})\right)t\right]-1\right)-t\ \mathrm{Im}\ p\right]+\exp\left[\bar{n}\left(\exp\left[-\left(\frac{\Gamma_1+\Gamma_2}{2}-\mathrm{Im}\ \gamma(\bar{n})\right)t\right]-1\right)+t\ \mathrm{Im}\ p\right]-2\ \mathrm{Re}\left[\exp\left[\bar{n}\left(\exp\left[-\left(\frac{\Gamma_1+\Gamma_2}{2}-i\ \mathrm{Re}\ \gamma(\bar{n})\right)t\right]-1\right)+it\ \mathrm{Re}\ p\right]\right]\right\},$$

where

$$\chi(\bar{n})=\frac{\varepsilon+\bar{n}[\Delta_2-\Delta_1-i(\Gamma_2-\Gamma_1)/2]}{\Omega(\bar{n})}, \tag{3.26}$$

and p is given by (2.20). In the limit $\Gamma_2=\Delta_2=0$, these expressions transform into the expressions for the decay of an isolated level.[9,10]

We examine the expressions (3.25) for the probabilities under the condition (3.19) that a stable state exists at $n=\bar{n}$, i.e.,

$$\varepsilon=\varepsilon_0=\frac{q}{2}\,\bar{n}(\Gamma_2-\Gamma_1)-\bar{n}(\Delta_2-\Delta_1). \tag{3.27}$$

Then Eqs. (3.25) yield

$$W_1(t)=\frac{\Gamma_2^2}{(\Gamma_1+\Gamma_2)^2}\Bigg\{1+\frac{\Gamma_1^2}{\Gamma_2^2}\exp[\bar{n}(\exp[-(\Gamma_1+\Gamma_2)t]-1)] +2\frac{\Gamma_1}{\Gamma_2}\,\mathrm{Re}\Bigg[\exp\Bigg[\bar{n}\Bigg(\exp\Bigg[-\Bigg(\frac{\Gamma_1+\Gamma_2}{2}-i\ \mathrm{Re}\ \gamma(\bar{n})\Bigg)t\Bigg] -1\Bigg)+it\ \mathrm{Re}\ p\Bigg]\Bigg]\Bigg\}, \tag{3.28}$$

$$W_2(t)=\frac{\Gamma_1\Gamma_2}{(\Gamma_1+\Gamma_2)^2}\Bigg\{1+\exp[\bar{n}(\exp[-(\Gamma_1+\Gamma_2)t]-1)] -2\,\mathrm{Re}\Bigg[\exp\Bigg[\bar{n}\Bigg(\exp\Bigg[-\Bigg(\frac{\Gamma_1+\Gamma_2}{2}-i\ \mathrm{Re}\ \gamma(\bar{n})\Bigg)t\Bigg] -1\Bigg)+it\ \mathrm{Re}\ p\Bigg]\Bigg]\Bigg\}.$$

The total probability of atom survival is

$$W(t)=W_1(t)+W_2(t) =\frac{\Gamma_2}{\Gamma_1+\Gamma_2}\Bigg\{1+\frac{\Gamma_1}{\Gamma_2}\exp[\bar{n}(\exp[-(\Gamma_1+\Gamma_2)t]-1)]\Bigg\}. \tag{3.29}$$

When $(\Gamma_1+\Gamma_2)t\ll 1$, we have the expression (3.22) with $n=\bar{n}$. For large times, as $t\to\infty$, from (3.28) we obtain

$$W_1(\infty)\approx\frac{\Gamma_2^2}{(\Gamma_1+\Gamma_2)^2}\Bigg\{1+\frac{\Gamma_1}{\Gamma_2}e^{-\bar{n}}\Bigg[\frac{\Gamma_1}{\Gamma_2}+2\cos(t\ \mathrm{Re}\ p)\Bigg]\Bigg\}, \tag{3.30}$$

$$W_2(\infty)\approx\frac{\Gamma_1\Gamma_2}{(\Gamma_1+\Gamma_2)^2}\{1+e^{-\bar{n}}[1-2\cos(t\ \mathrm{Re}\ p)]\}.$$

Since for $\bar{n}\gg 1$ the second terms in braces in (3.30) are small,

$$W_1(\infty)\approx\frac{\Gamma_2^2}{(\Gamma_1^2+\Gamma_2)^2},\quad W_2(\infty)\approx\frac{\Gamma_1\Gamma_2}{(\Gamma_1+\Gamma_2)^2}. \tag{3.31}$$

These expressions coincide with (3.21), and for $\Gamma_1\approx\Gamma_2$ we again find that $W_1(\infty)\approx W_2(\infty)=1/4$. From (3.29) we can derive an expression for the total probability of finding an electron in the atom:

$$W(\infty)=\frac{\Gamma_2}{\Gamma_1+\Gamma_2}\Bigg(1+\frac{\Gamma_1}{\Gamma_2}\,e^{-\bar{n}}\Bigg). \tag{3.32}$$

Since $\bar{n}\gg 1$, we arrive at (3.23), and at $\Gamma_1\approx\Gamma_2$ we find that $W(\infty)=1/2$.

In a similar way we can derive an expression for the case of a squeezed quantized field when the squeezing is weak.

I am grateful to M. L. Ter-Mikaelyan, M. V. Fedorov, and B. V. Kryzhanovskiĭ for the discussions of the results. This work was done under an agreement between the Theoretical Laboratory at the Institute of Physical Studies of the Armenian National Academy of Sciences and the Theoretical Section of the High-Power Lasers Division at the Institute of General Physics of the Russian Academy of Sciences, and was supported by the Ministry of Education of the Republic of Armenia (Project No. 96-772).

[1] F. W. Cummings, Phys. Rev. **140**, A1051 (1965).
[2] A. D. Gazazyan, Zh. Éksp. Teor. Fiz. **51**, 1863 (1966) [Sov. Phys. JETP **24**, 1254 (1967)].
[3] J. H. Eberly, N. B. Narozhny, and J. J. Sanchez-Mondragon, Phys. Rev. Lett. **44**, 1323 (1980).
[4] N. B. Narozhny, J. J. Sanchez-Mondragon, and J. H. Eberly, Phys. Rev. A **22**, 236 (1981).
[5] H. J. Yoo, J. H. Eberly, and J. J. Sanchez-Mondragon, J. Phys. A **14**, 1383 (1981).
[6] G. Rempe and H. Walther, Phys. Scr. **36**, 135 (1987).
[7] G. Rempe, H. Walther, and N. Klein, Phys. Rev. Lett. **58**, 353 (1987).
[8] A. D. Gazazyan, M. L. Ter-Mikaelyan, and V. G. Sherman, Izv. Nats. Akad. Nauk. Arm. **28**, 69 (1993).
[9] O. A. Alimov, V. P. Krainov, and A. A. Mikheev, Laser Phys. **4**, 551 (1994).
[10] A. D. Gazazyan, Laser Phys. **5**, 852 (1995).
[11] K. Rzążewski and J. Eberly, Phys. Rev. Lett. **47**, 408 (1981).
[12] P. Coleman, P. L. Night, and K. Burnett, Opt. Commun. **42**, 171 (1982).
[13] P. L. Knight, M. A. Lauder, and B. J. Dalton, Phys. Rep. **190**, No. 1 (1990).
[14] A. D. Gazazyan and R. G. Unanyan, Zh. Éksp. Teor. Fiz. **93**, 1590 (1987) [Sov. Phys. JETP **66**, 909 (1987)].
[15] M. V. Fedorov and A. M. Movsesian, J. Phys. B **21**, L155 (1988).

Translated by Eugene Yankovsky

Squeezed harmonic generation in cooperative resonance scattering of intense radiation by dipole molecules

V. A. Kovarskiĭ*) and O. B. Prepelitsa

Institute of Applied Physics, Moldavian Academy of Sciences, 277028 Kishinev, Moldova
(Submitted 1 August 1997)
Zh. Éksp. Teor. Fiz. **113**, 1206–1212 (April 1998)

We examine the processes of resonance Raman scattering of intense electromagnetic radiation by a lumped system of two-level atoms with a constant dipole moment. We calculate the intensity of the sth generated harmonic in a saturating electromagnetic field and the statistical characteristics of this harmonic. Finally, we show that the sth harmonic is squeezed at saturating field intensities. © *1998 American Institute of Physics.* [S1063-7761(98)00504-6]

1. INTRODUCTION

The generation of squeezed electromagnetic radiation is still drawing much attention. Generation of squeezed radiation in various spectral ranges, including vacuum UV, is especially important. On the other hand, observation of higher harmonics generated by short laser pulses applied to the atoms of inert gases, molecules, and crystals[1–6] proves that coherent radiation can be generated in a broad spectral range.

The aim of the present investigation is to study the possibility of squeezing in higher-harmonic generation using the example of resonance scattering of electromagnetic radiation by dipole molecules. One-photon squeezing in cooperative resonance fluorescence in two-level systems was studied by Q. V. and S. V. Lawande.[7] In our paper we focus on squeezing processes that take place during excitation of higher harmonics due to Raman scattering. We examine a lumped system of two-level molecules that have a dipole moment $\mathbf{d}_{22}$ in their excited state and are oriented by a given external field.

Multiphoton processes in the resonance scattering of light by an individual molecule have been discussed in Ref. 8. In contrast to Ref. 8, in the present paper we study the effects of cooperative scattering, which take into account the presence of a strong electromagnetic field and the possibility of cooperative spontaneous luminescence. We assume that in generating the sth harmonic, the system remains in the lumped state. We will show that the squeezing in cooperative luminescence in the hard UV range is a highly nonmonotonic function of the offset from resonance and the intensity of the scattered light and increases with the number of molecules in the system as long as the model remains lumped. In the vacuum UV range the advantage of the proposed multiphoton method of higher-harmonic generation of squeezed radiation over the one-photon method proposed in Ref. 7 is that one is able to use realistic Rabi frequencies. The nonmonotonic nature of the squeezing effect is due to the contribution of photon re-emission in the interaction of dipole moments $\mathbf{d}_{22}$ with the electromagnetic field, determined by a dimensionless parameter ρ greater than unity:

$$\rho=\frac{\mathbf{E}_0\cdot\mathbf{d}_{22}}{\hbar\omega_0},$$

where $\mathbf{E}_0$ and ω_0 are the amplitude and frequency of the incident radiation (the vector of the electric component of the linearly polarized wave is directed along the dipole moment $\mathbf{d}_{22}$ of the oriented molecules).

2. BASIC EQUATIONS FOR THE DENSITY MATRIX OF A SYSTEM OF ORIENTED DIPOLE MOLECULES IN AN EXTERNAL ELECTROMAGNETIC FIELD

The Hamiltonian of a system of N oriented dipole molecules in an external electromagnetic field for a lumped system ($\lambda \gg l$, with $\lambda=2\pi c/\omega$, and l the linear size of the system) has the form

$$H=H_0^{(M)}+H_0^{(\mathrm{ph})}+V(t)+W(t)+H_{\mathrm{int}}, \tag{1}$$

$$H_0^{(M)}=\hbar\omega_{21}R_z, \quad H_0^{(\mathrm{ph})}=\sum_{\mathbf{k}}\hbar\omega_{\mathbf{k}}b_{\mathbf{k}}^{\dagger}b_{\mathbf{k}},$$

$$V(t)=(\mathbf{E}_0\cdot\mathbf{d}_{22})R_z\cos\omega_0 t,$$

$$W(t)=(\mathbf{E}_0\cdot\mathbf{d}_{21})(R^{+}+R^{-})\cos\omega_0 t,$$

$$H_{\mathrm{int}}=i\sum_{\mathbf{k}}(\mathbf{g}_{\mathbf{k}}\cdot\mathbf{d}_{21})(R^{+}+R^{-})(b_{\mathbf{k}}^{\dagger}-b_{\mathbf{k}}).$$

Here ω_{21} is the frequency of the transition between two specified states of a molecule, which form a two-level system, $\mathbf{g}_{\mathbf{k}}=\sqrt{2\pi\hbar\omega_{\mathbf{k}}/V}\mathbf{e}_{\lambda}$, with V the quantization volume and $\mathbf{e}_{\lambda}$ the unit polarization vector, $R^{\pm}$ and R_z are the cooperative spin operators satisfying the commutation relations

$$[R^{+},\,R^{-}]=2R_z, \quad [R_z,\,R^{\pm}]=\pm R^{\pm},$$

and $b_{\mathbf{k}}^{\dagger}$ and $b_{\mathbf{k}}$ are the Bose operators of the electromagnetic field.

In what follows it will be convenient to pass to the Furry representation.[9] Then the term $V(t)$ is taken into account exactly in the zeroth-order Hamiltonian, and after performing simple transformations we can write the Hamiltonian (1) in the Furry representation as follows:

$$\tilde{H}=H_0^{(M)}+H_0^{(\mathrm{ph})}+\tilde{W}(t)+\tilde{H}_{\mathrm{int}},$$

$$\tilde{W}(t)=(\mathbf{E}_0\cdot\mathbf{d}_{21})(R^{+}\exp(i\rho\,\sin\,\omega_0 t)$$

$$+R^-\exp(-i\rho\ \sin\ \omega_0 t))\cos\ \omega_0 t,$$

$$\tilde{H}_{\rm int}=i\sum_{\mathbf{k}}\ (\mathbf{g_k}\cdot\mathbf{d}_{21})(R^+\exp(i\rho\ \sin\ \omega_0 t)$$

$$+R^-\exp(-i\rho\ \sin\ \omega_0 t))(b_{\mathbf{k}}^\dagger-b_{\mathbf{k}}).$$

Now let us introduce the density matrix of the atomic subsystem, $\sigma(t)=\mathrm{Tr}\ \chi(t)$, where $\chi(t)$ is the total density matrix, and the trace is taken over the variables of the photon subsystem.

We employ the projection-operator method (see, e.g., Ref. 10) and keep only terms of lowest order in the interaction with the photon-field vacuum. The result is the following equation of motion for the density matrix of the atomic subsystem:

$$\frac{d\sigma(t)}{dt}=i\delta[\sigma(t),\ R_z]+i\Omega_{n_0}[\sigma(t),\ R^++R^-]$$

$$+a([R^-,\ \sigma(t)R^+]+[R^-\sigma(t),\ R^+])$$

$$-b([R^-,\ R^+\sigma(t)]+[\sigma(t)R^-,\ R^+]), \qquad (2)$$

$$\delta=\omega_{21}-n_0\omega_0, \quad \Omega_{n_0}=\omega_{21}\frac{|\mathbf{d}_{21}|}{|\mathbf{d}_{22}|}J_{n_0}(\rho),$$

$$a=\frac{1}{2\tau_{\rm sp}}\sum_{m=1}^{\infty}\left(\frac{m}{n_0}\right)^3 J_{m-n_0}^2(\rho),$$

$$b=\frac{1}{2\tau_{\rm sp}}\sum_{m=1}^{\infty}\left(\frac{m}{n_0}\right)^3 J_{m+n_0}^2(\rho),$$

where n_0 is the resonance parameter, equal to the integer part of the ratio ω_{21}/ω_0, δ is the offset from resonance, and

$$\frac{1}{\tau_{\rm sp}}=\frac{4}{3}\frac{\omega_{21}^3}{\hbar c^3}|\mathbf{d}_{21}|^2$$

is the probability of one-photon spontaneous decay of the excited state.

Equation (2) was derived under the conditions that $\rho<2n_0$ and $(\mathbf{E}_0\cdot\mathbf{d}_{21})/\hbar\omega_{21}\ll 1$, which made it possible to consider the interaction of the dipole moment $\mathbf{d}_{21}$ and the external electromagnetic field in the resonance approximation. Clearly, Eq. (2) with $\mathbf{d}_{22}=0$ coincides with the well-known result of Ref. 11. The constants a and b are the rates of population depletion of the excited and ground states, respectively. The constant b is zero at $\mathbf{d}_{22}=0$ and is a specific feature of the generalized two-level systems with $\mathbf{d}_{22}\neq 0$ considered here. The emergence of a nonlinear dependence on the intensity of the external electromagnetic field of the $J_m^2(\rho)$ type ($J_m(\rho)$ is the Bessel function of a real argument) is due to photon re-emission processes and only in the limit $\rho\ll 1$ corresponds to perturbation-theory results with a minimum number of photons required by the given physical process.

3. STATISTICAL PROPERTIES OF THE GENERATED HARMONICS

Let us examine the power of the electromagnetic radiation generated by a quantum system:

$$P=\sum_{\mathbf{k}}\ \hbar\omega_{\mathbf{k}}\frac{d}{dt}\langle b_{\mathbf{k}}^\dagger b_{\mathbf{k}}\rangle. \qquad (3)$$

We write the equation of motion for the Bose variables obtained through a formal solution of the corresponding Heisenberg equation:

$$b_{\mathbf{k}}^\dagger(t)=b_{\mathbf{k}}^\dagger(0)\exp(i\omega_{\mathbf{k}}t)+\frac{(\mathbf{g_k}\cdot\mathbf{d}_{21})}{\hbar}\int_0^t d\tau$$

$$\times\exp(i\omega_{\mathbf{k}}\tau)(R^+(t-\tau)\exp(i\rho\ \sin\ \omega_0(t-\tau))$$

$$+R^-(t-\tau)\exp(-i\rho\ \sin\ \omega_0(t-\tau))). \qquad (4)$$

We insert (4) into (3) and use the Markov approximation, which amounts to ignoring retardation in the slow part of the spin operators:

$$R^\pm(t-\tau)\approx R^\pm(t)\exp(\mp i\omega_{21}\tau).$$

After simple transformations the expression for P becomes

$$P=\frac{\hbar\omega_{21}}{\tau_{\rm sp}}\sum_{s=1}^{\infty}\left(\frac{s}{n_0}\right)^4(J_{s-n_0}^2(\rho)\langle R^+R^-\rangle$$

$$+J_{s+n_0}^2(\rho)\langle R^-R^+\rangle),$$

where s stands for the number of the harmonic generated at the frequency $\omega=s\omega_0$. Finding the numerical value of P requires knowing the correlators $\langle R^+R^-\rangle$ and $\langle R^-R^+\rangle$, and the calculation of these correlators for a system consisting of a large number of molecules is extremely difficult mathematically due to the necessity of solving Eq. (2). Hence we consider only the steady-state case and the region of saturating fields $\Omega_{n_0}\gg N(a-b)$ (below we explain how these approximations can be used simultaneously). Since in this case $\langle R^+R^-\rangle$ and $\langle R^-R^+\rangle$ are of the same order, we conclude that the dependence of the intensity of the radiation on the harmonic number is determined primarily by the order of the Bessel function and that the intensity reaches its maximum value at $s\sim\rho+n_0$. When $s>\rho+n_0$, the Bessel function rapidly decreases, which corresponds to the cutoff effect in the theory of higher-order harmonic generation by atoms. To determine the statistical properties of the scattered radiation we examine the variance of the quadratures of the electromagnetic field:[12]

$$\langle{:}(\Delta M_i)^2{:}\rangle=\langle{:}M_i^2{:}\rangle-\langle M_i\rangle^2, \quad i=1,2,$$

$$M_1=\frac{1}{2}(b_{\mathbf{k}}^\dagger+b_{\mathbf{k}}), \quad M_2=\frac{i}{2}(b_{\mathbf{k}}^\dagger-b_{\mathbf{k}}),$$

where angle brackets indicate averaging with the total density matrix $\chi(t)$, and $:f:$ stands for normal ordering. The fact that the $\langle{:}(\Delta M_i)^2{:}\rangle$ ($i=1,2$) are nonzero is sufficient for the existence of squeezing in the spectrum of the scattered light.

After we eliminate the Bose variables via Eq. (4), and employ the Markov approximation, we arrive at the following expression:

$$\langle:(\Delta M_{1,2})^2:\rangle_s=\pm\frac{|\mu|^2}{4}(J^2_{s-n_0}(\rho)+J^2_{s+n_0}(\rho))[\langle R^+R^+\rangle$$
$$+\langle R^-R^-\rangle-\langle R^+\rangle^2-\langle R^-\rangle^2]$$
$$\mp\frac{|\mu|^2}{4}[J^2_{s-n_0}(\rho)(\langle R^+R^-\rangle-\langle R^+\rangle\langle R^-\rangle)$$
$$+J^2_{s+n_0}(\rho)(\langle R^-R^+\rangle-\langle R^+\rangle\langle R^-\rangle)],$$

where s is the number of a harmonic, and μ is a geometrical factor.[11]

Following Refs. 7 and 13, we introduce new quasispin operators $S^{\pm}$ and S_z such that

$$[S^+,\,S^-]=2S_z,\quad [S_z,\,S^{\pm}]=\pm S^{\pm};$$

they are related to the ''old'' operators via the following formulas:

$$R^+=\Delta_1S^+-\Delta_2S^-+2\Delta_3S_z,\quad R^-=(R^+)^{\dagger},$$
$$R_z=(\Delta_1-\Delta_2)S_z-\Delta_3(S^++S^-),$$

where $\Delta_1=\cos^2\eta$, $\Delta_2=\sin^2\eta$, $\Delta_3=\sin\eta\cos\eta$, and $\cot 2\eta=\delta/2\Omega_{n_0}$.

In terms of the new quasispin variables, the desired quantities, i.e., the variance of the quadratures of the electromagnetic field and the radiation power, become

$$\langle:(\Delta M_{1,2})^2:\rangle_s$$
$$=\pm\frac{|\mu|^2}{2}(J^2_{s-n_0}(\rho)+J^2_{s+n_0}(\rho))[4\Delta_3^2(\langle S_z^2\rangle-\langle S_z\rangle^2)$$
$$-\Delta_1\Delta_2(\langle S^+S^-\rangle+\langle S^-S^+\rangle-2\langle S^+\rangle\langle S^-\rangle)]$$
$$\mp\frac{|\mu|^2}{2}[J^2_{s-n_0}(\rho)(\Delta_1^2(\langle S^+S^-\rangle-\langle S^+\rangle\langle S^-\rangle)$$
$$+\Delta_2^2(\langle S^-S^+\rangle-\langle S^+\rangle\langle S^-\rangle)+4\Delta_3^2(\langle S_z^2\rangle$$
$$-\langle S_z\rangle^2))+J^2_{s+n_0}(\rho)(\Delta_1^2(\langle S^-S^+\rangle$$
$$-\langle S^+\rangle\langle S^-\rangle)+\Delta_2^2(\langle S^+S^-\rangle-\langle S^+\rangle$$
$$\times\langle S^-\rangle)+4\Delta_3^2(\langle S_z^2\rangle-\langle S_z\rangle^2))], \tag{5}$$

$$P=\frac{\hbar\omega_{21}}{\tau_{\rm sp}}\sum_{s=1}^{\infty}\left(\frac{s}{n_0}\right)^4[(\Delta_1^2J^2_{s-n_0}(\rho)+\Delta_2^2J^2_{s+n_0}(\rho))$$
$$\times\langle S^+S^-\rangle+(\Delta_2^2J^2_{s-n_0}(\rho)+\Delta_1^2J^2_{s+n_0}(\rho))\langle S^-S^+\rangle$$
$$+4\Delta_3^2(J^2_{s-n_0}(\rho)+J^2_{s+n_0}(\rho))\langle S_z^2\rangle]. \tag{6}$$

We assume that $\Omega_{n_0}\gg N(a-b)$. Then, if we ignore the small terms (of order $N(a-b)/\Omega_{n_0}$), Eq. (2) becomes

$$\frac{d\tilde\sigma(t)}{dt}=i\tilde\Omega_{n_0}[\tilde\sigma(t),S_z]+\tilde a([S^-,\tilde\sigma(t)S^+]$$
$$+[S^-\sigma(t),S^+])-\tilde b([S^-,S^+\tilde\sigma(t)]$$
$$+[\tilde\sigma(t)S^-,S^+])+\tilde\gamma([S_z\tilde\sigma(t),S_z]$$
$$+[S_z,\tilde\sigma(t)S_z]), \tag{7}$$
$$\tilde\Omega_{n_0}=\sqrt{\frac{1}{4}\,\delta^2+\Omega_{n_0}^2},\quad \tilde a=\Delta_1^2a+\Delta_2^2b,$$
$$\tilde b=\Delta_2^2a+\Delta_1^2b,\quad \tilde\gamma=\Delta_3^2(a+b).$$

Clearly, in the Dicke-state basis, the steady-state solution of Eq. (7) can be written as follows:[7]

$$\tilde\sigma_{\rm st}=Z^{-1}\exp(-\alpha S_z), \tag{8}$$

where Z is a normalization constant, and $\alpha=\ln(\tilde a^2/\tilde b^2)$.

A remark is in order. In Eqs. (2) and (7) we allowed only for the decay of the excited state, a process related to transitions to the ground state. Actually this means that we are employing a strictly two-level approximation, since we ignore decays into ''third'' states, whose transition probabilities are Γ. The two-level approximation works for times $t\ll\Gamma^{-1}$. Thus, the steady-state solution (8) is valid in the time interval $(a-b)^{-1}\ll t\ll\Gamma^{-1}$.

Using the solution (8), we can find the means in Eqs. (5)–(7):

$$\langle S^+S^-\rangle=-\langle S_{11}^2\rangle+(N-1)\langle S_{11}\rangle+N,\quad \langle S^{\pm}\rangle=0,$$
$$\langle S^-S^+\rangle=-\langle S_{11}^2\rangle+(N+1)\langle S_{11}\rangle,$$
$$\langle S_z^2\rangle=\langle S_{11}^2\rangle-N\langle S_{11}\rangle+\frac{N^2}{4},$$

where

$$\langle S_{11}^2\rangle=\frac{N^2L^{N+3}-(2N^2+2N+1)L^{N+2}(N+1)^2L^{N+1}-L^2-L}{Z(L-1)^3},$$
$$\langle S_{11}\rangle=\frac{NL^{N+2}-(N+1)L^{N+1}+L}{Z(L-1)^2},\quad L=\frac{\tilde a}{\tilde b}.$$

These expressions for the correlators make it possible to calculate the power of the generated harmonics and study the statistical properties of these harmonics in a saturating field. Note that the spectral distribution of the radiation intensity is a highly nonlinear (nonmonotonic) function of the frequencies $s\omega_0$ $(s=1,2,\dots)$ and strength of the applied electromagnetic field (Fig. 1). Here the intensity of some of the higher harmonics may exceed the radiation intensity at the fundamental frequency ω_0 (Fig. 1).

Numerical analysis shows (Fig. 2) that the degree of squeezing for the optimum harmonic $s\sim\rho+n_0$ reaches its maximum value at $(\delta/2\Omega_{n_0})^2\sim 0.36$. The size of this squeezing is of the same order as in one-photon resonance fluorescence at $\omega=s\omega_0$ (see Ref. 7). But in the event of such one-photon fluorescence the fields must be saturating, i.e., the Rabi frequency must be much larger than the product of the number of emitters and the probability of decay of the excited state of an individual emitter. This condition plays an important role in statistical properties of the scattered radiation, since in weak fields light scattering is coherent, with the statistical properties of the applied and scattered electromag-

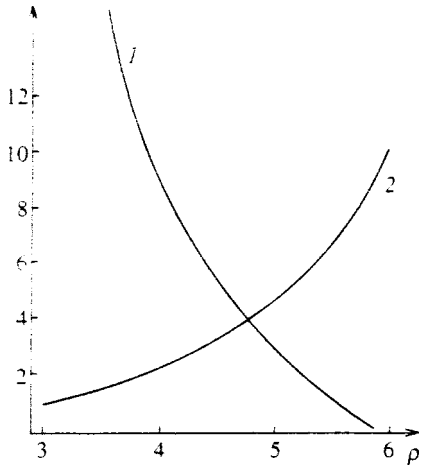


FIG. 1. The ratio of the radiation power at the frequencies $8\omega_0$ (curve *1*) and $10\omega_0$ (curve *2*) to the radiation power at ω_0 as a function of the parameter ρ at $n_0=5$.

netic fields being identical. Hence the intensities of the incident electromagnetic field at which one can expect a conversion of classical light into nonclassical (squeezed) light are sure to belong to the region of saturating fields. In the event of one-photon fluorescence at $\omega=s\omega_0$, saturation is possible only if the Rabi frequencies are extremely high (since the probability of decay of the excited state is proportional to $(s\omega_0)^3$), which at such high frequencies is essentially unattainable.

On the other hand, in our scheme saturation occurs for an energy gap $\omega_{21}\approx n_0\omega_0$, $n_0\ll s$. This difference makes it possible to squeeze high-frequency harmonics at moderate intensities of the incident radiation. The theory, however, does not address higher field strengths, at which the wavelength of the optimum harmonic becomes smaller than the size of the system being studied.

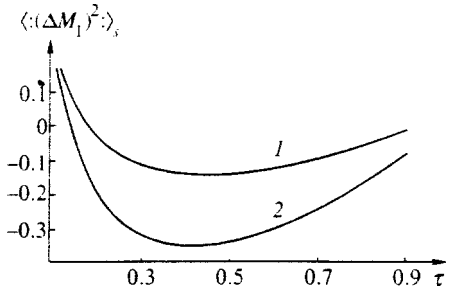


FIG. 2. $\langle:(\Delta M_1)^2:\rangle_s$ vs. $\tau=(\delta/2\Omega_{n_0})^2$ for the optimum harmonic $s=10$ ($\omega=10\omega_0$) at $n_0=5$, $\rho=6$, and $N=20$ (curve *1*) and $N=50$ (curve *2*).

*)E-mail: exciton@cc.acad.md

[1] A. L'Huillier, K. J. Schafer, and K. C. Kulander, Phys. Rev. Lett. **66**, 2200 (1991).
[2] A. L'Huillier and P. Balcou, Phys. Rev. Lett. **68**, 774 (1993).
[3] W. Becker, S. Long, and J. K. McIver, Phys. Rev. A **50**, 1540 (1994).
[4] C. T. Chen, Y. Wu, A. Jiang, B. Wu, G. You, R. Li, and S. Lin, J. Opt. Soc. Am. B **6**, 616 (1989).
[5] Y. Wu, T. Sasaki, S. Nakai, A. Yokotani, H. Tang, and C. Chen, Appl. Phys. Lett. **62**, 2614 (1993).
[6] L. J. Frasinski, M. Stankiewicz, P. A. Hatherly *et al.* Phys. Rev. A **46**, R6789 (1992).
[7] Q. V. Lawande and S. V. Lawande, Phys. Rev. A **38**, 800 (1988).
[8] V. A. Kovarskiĭ and N. F. Perel'man, Zh. Éksp. Teor. Fiz. **60**, 509 (1971) [Sov. Phys. JETP **33**, 274 (1971)].
[9] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 3rd ed., Pergamon Press, Oxford (1991).
[10] Ke-Hsueh Li, Phys. Rep. **134**, 1 (1986).
[11] G. S. Agarwal, Quantum Opt. **70** (1974).
[12] R. Loudon and P. L. Knight, J. Mod. Opt. **34**, 759 (1986).
[13] A. S. Shumovsky and Tran Quang, J. Phys. B **22**, 131 (1989).

Translated by Eugene Yankovsky

Diffusing-wave spectroscopy in randomly inhomogeneous media with spatially localized scatterer flows

S. E. Skipetrov*)

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

I. V. Meglinskiĭ

N. G. Chernyshevskiĭ Saratov State University, 410026 Saratov, Russia

(Submitted 6 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 1213–1222 (April 1998)

Multiple scattering of laser radiation in a randomly inhomogeneous turbid medium with a spatially localized flow of particles is studied. The time autocorrelation function of backscattered light is calculated for the case of a laminar flow of scatterers in a cylindrical capillary embedded in the medium. A new method is proposed and tested experimentally for determining the position of the dynamic region and the dominant form and characteristic velocity of the particle motion there. © *1998 American Institute of Physics.* [S1063-7761(98)00604-0]

1. INTRODUCTION

Problems involving the multiple-scattering of light in randomly inhomogeneous media have been attracting a great deal of attention in recent years.[1] This is due to both the diversity of beautiful physical effects observed under multiple-scattering conditions (coherent backscattering[2,3] and angular and temporal correlations of the scattered radiation[4]) and, in connection with the extensive use of optical diagnostics methods in modern medicine,[5–7] the extreme importance of correctly describing the processes occurring under these conditions.

So-called diffusing-wave spectroscopy has been developing rapidly during the last ten years. Its foundations were laid in Refs. 8 and 9. The method of diffusing-wave spectroscopy is based on measuring the time autocorrelation function $G_1(\tau)=\langle E(t)E^*(t-\tau)\rangle$ of light that is multiply scattered in a turbid medium. It is found that even under the conditions of strong multiple scattering $G_1(\tau)$ is sensitive to the character and intensity of scatterer motion in the medium. Moreover, by measuring $G_1(\tau)$ it is possible to detect very small displacements of the light-scattering particles (up to hundredths of a wavelength λ of the radiation employed), while the methods based on measuring the characteristics of single scattering of waves are sensitive only to scatterer displacements over a distance of the order of λ.[9] Another interesting possible application of diffusing-wave spectroscopy has been discovered in the last few years — the possibility of determining the location and performing diagnostics of dynamic inhomogeneities in turbid media on the basis of an analysis of the scattered radiation.[10–13]

In the present paper we propose a method of determining the location and measuring the characteristics of directed scatterer flows which are hidden deep in a turbid medium. The method makes it possible to obtain information about the motion of particles inside a medium on the basis of the dependence $G_1(\tau)$ measured at different points on the surface of the sample. The theoretical and experimental results presented are in good agreement with one another. Moreover, they agree with both the results of other experiments[13] and theoretical calculations.[12] The potential possibilities and limitations of the method of measuring the characteristics of scatterer flows in strongly light-scattering turbid media are evaluated. The method we propose could find interesting applications, for example, in hemodynamics, since it opens up the possibility of performing noninvasive measurements of the velocity of blood flow in blood vessels and detecting changes in blood volume in capillary ansae and other biological tissues.[14]

2. THEORETICAL ANALYSIS

Let us consider the scattering of light with wavelength λ in a sample of a turbid medium (particle size $\sim\lambda$), characterized by a photon transport mean-free path $\ell^*=(\mu_s'+\mu_a)^{-1}$, where μ_s' and μ_a are the scattering and absorption coefficients,[15] and in addition $\mu_a \ll \mu_s'$. Under strong multiple-scattering conditions ($\lambda \ll \ell^* \ll L$, where L is the characteristic size of the sample) the time correlation function $G_1(\mathbf{r},\tau)=\langle E(\mathbf{r},t)E^*(\mathbf{r},t-\tau)\rangle$ of depolarized multiply-scattered radiation measured at point $\mathbf{r}$ on the boundary S of the sample can be described in the diffusion approximation by solving the stationary diffusion equation[10,13]

$$[\nabla^2-\alpha^2(\tau)]G_1(\mathbf{r},\tau)=-\frac{F(\mathbf{r})}{D_p} \tag{1}$$

with the boundary condition[16]

$$G_1(\mathbf{r},\tau)-\frac{2}{3}\ell^*(\mathbf{n}\cdot\nabla G_1(\mathbf{r},\tau))=0, \quad \mathbf{r}\in S. \tag{2}$$

Here we have neglected light absorption in the medium, $F(\mathbf{r})$ describes the distribution of light sources, $D_p=c\ell^*/3$ is the light diffusion coefficient in the medium,[15] $\mathbf{n}$ is the inward unit normal to the surface S, and the specific form of the function $\alpha(\tau)$ depends on the character of the scatterer motion in the medium: $\alpha^2(\tau)=3\tau/2\tau_0\ell^{*2}$ in the case of

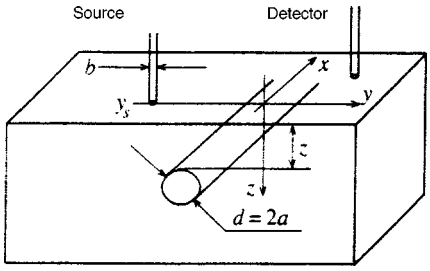


FIG. 1. Schematic diagram of the experiment. The sample ($8\times15\times15$ cm^3) consists of particles of rutile (TiO_2) suspended in resin (0.01 g of rutile per 100 ml of resin, particle diameter 0.25 μm, $\mu_s'=4$ cm^{-1}, $\mu_a=0.002$ cm^{-1}, $l^*=0.25$ cm, $D_B=0$). A transverse cylindrical opening with diameter $d=0.75$ cm, where a flow of a suspension of polystyrene beads in water (particle diameter 0.296 μm, concentration $\approx0.5\%$, $D_B=1.5\times10^{-8}$ cm^2/s, $\tau_0=6.32\times10^{-4}$ s) is produced, was made parallel to the surface inside the sample at a depth $z=0.925$ cm. Laser radiation is introduced into the medium by means of a narrow fiber-optic lightguide with diameter $b<\ell^*$. The time autocorrelation function of the intensity of the multiply scattered light in the medium is measured at the point $(x,y,0)$.

Brownian motion of particles in a medium with diffusion coefficient D_B,[9] where $\tau_0=(4k^2D_B)^{-1}$ and $k=2\pi/\lambda$, and $\alpha^2(\tau)=6(\tau/\tau_f\ell^*)^2$ in the case of laminar scatterer flow,[17] where the characteristic time τ_f depends on the flow geometry. In the case when the directed motion of the light-scattering particles is superimposed on the random walk of the particles, $\alpha^2(\tau)$ is given by a sum of terms corresponding to these two different types of motion.[18]

The boundary condition (2) can be approximately replaced by requiring $G_1(\mathbf{r},\tau)$ to vanish on the so-called extrapolated boundary[15] $z=-z_1=-\Delta\ell^*$, where Δ depends on the scattering conditions near the boundary. When the scattering is isotropic and the refractive index of the scattering medium equals that of the surrounding medium, the Milne theory gives $\Delta=0.7104$.[19]

Let the turbid medium fill the half-space $z>0$ and let the medium contain a hidden inclusion in the form of a cylindrical capillary with diameter $d=2a>\ell^*$. Let the capillary be directed along the x axis and located at a distance z from the boundary of the medium (Fig. 1). We denote by S_1 the surface of the capillary and by V_1 the volume enclosed by it. If a directed scatterer flow is produced in the capillary, while elsewhere in the medium the scatterers move as Brownian particles, then we can introduce in Eq. (1) a spatial dependence of $\alpha^2(\tau)$ of the form[13]

$$\alpha^2(\tau)=\begin{cases}\alpha_{\rm in}^2(\tau)=3\tau/2\tau_0\ell^{*2}+6(\tau/\tau_f\ell^*)^2, & \mathbf{r}\in V_1,\\ \alpha_{\rm out}^2(\tau)=3\tau/2\tau_0\ell^{*2}, & \mathbf{r}\notin V_1.\end{cases} \tag{3}$$

In our experimental situation (Fig. 1) a Poiseuille velocity profile can be assumed for the particles inside the capillary, which gives

$$\tau_f=\sqrt{30}/k\ell^*\Gamma_1, \quad \text{where } \Gamma_1=32Q/\sqrt{2}\pi d^3, \tag{4}$$

and Q is the volume flow rate of the liquid, equal to the volume of the liquid passing through the cross section of the capillary per unit time. It is easy to show that Q is related to the average velocity V of the directed motion of the particles in the capillary by the relation $Q=\pi a^2V$.

Finally, we write the boundary conditions on the surface S_1 of the capillary in the form[10]

$$G_1^{\rm in}(\mathbf{r},\tau)=G_1^{\rm out}(\mathbf{r},\tau), \quad \mathbf{r}\in S_1, \tag{5}$$

$$(\mathbf{n}\cdot\nabla G_1^{\rm in}(\mathbf{r},\tau))=(\mathbf{n}\cdot\nabla G_1^{\rm out}(\mathbf{r},\tau)), \quad \mathbf{r}\in S_1, \tag{6}$$

where $G_1^{\rm in,out}(\mathbf{r},\tau)$ are solutions of Eq. (1) inside and outside the volume V_1, respectively.

To complete the mathematical formulation of the problem we have only to specify the distribution $F(\mathbf{r})$ of the light sources in the medium. In an experiment the sample is ordinarily illuminated by a laser beam of finite width b. The cases $b\gg\ell^*$ and $b<\ell^*$ are easiest to describe theoretically, since in the first case one can assume approximately that a plane wave is incident on the surface of the medium, while in the second case one can study a point source of radiation positioned at a point determined by its coordinate y_s on the surface of the medium (see Fig. 1). Since the coherent laser radiation becomes diffuse at a depth $z_0\sim\ell^*$,[15] we shall write the source function in Eq. (1) in the form

$$F(\mathbf{r})\simeq\begin{cases}\delta(z-z_0), & b\gg\ell^*,\\ \delta(x)\,\delta(y-y_s)\,\delta(z-z_0), & b<\ell^*.\end{cases} \tag{7}$$

On this basis it is easy to obtain an expression for the correlation function G_1^0 of depolarized light backscattered from an infinite medium in the absence of a scatterer flow in the capillary ($\tau_f\to\infty$). In the limit $\tau\ll\tau_0$ we obtain

$$G_1^0(\tau)=\exp\{-\gamma\alpha_{\rm out}\ell^*\}, \tag{8}$$

for $b\gg\ell^*$ and

$$G_1^0(x,y,\tau)=\frac{1}{4\pi D_p}\left\{\frac{\exp(-\alpha_{\rm out}\,\rho)}{\rho}-\frac{\exp(-\alpha_{\rm out}\sqrt{\rho^2+4\gamma^2\ell^{*2}})}{\sqrt{\rho^2+4\gamma^2\ell^{*2}}}\right\} \tag{9}$$

for $b<\ell^*$. In these formulas $\gamma=1+\Delta$ is a numerical constant of the order of 2, $\rho=[x^2+(y-y_s)^2]^{1/2}$ (see Fig. 1), and the light source is assumed to be located at the point $(0,y_s)$ on the surface of the medium. The results (8) and (9) were obtained earlier by different methods, and they have also been confirmed experimentally.[8,9] We note that the correlation function (8) does not depend on the position $(x,\,y)$ of the detector on the surface of the medium or on the photon transport mean free path ℓ^*.

Now let a laminar flow of light-scattering particles be produced in the capillary. In this case we write the solution in the form $G_1(x,y,\tau)=G_1^0(x,y,\tau)+G_1^S(x,y,\tau)$, where the last term describes the effect of the flow on the correlation function. Neglecting the condition (2) on the boundary of the medium, we obtain for the case $b\gg\ell^*$

$$G_1^S(y,\tau)=-\frac{h-z_0}{2\pi\ell^*}\sum_{n=1}^{\infty}\int_{-\infty}^{\infty}\frac{dp}{\cos p}\cos(n(p-\theta))\times f_n\left(\alpha_{\rm out}\frac{h-z_0}{\cos p},\ \alpha_{\rm out}\sqrt{h^2+y^2},\ \alpha_{\rm out}a,\alpha_{\rm in}a\right), \tag{10}$$

where

$$f_n(\xi_1,\xi_2,\xi_3,\xi_4)=K_n(\xi_1)K_n(\xi_2)\times\left[\frac{\xi_3 I_n'(\xi_3)I_n(\xi_4)-\xi_4 I_n(\xi_3)I_n'(\xi_4)}{\xi_3 K_n'(\xi_3)I_n(\xi_4)-\xi_4 K_n(\xi_3)I_n'(\xi_4)}\right], \tag{11}$$

I_n and K_n are modified Bessel functions, primes denote differentiation of the corresponding function with respect to its argument, $h=z+a$, and $\theta=\tan^{-1}(y/h)$. A somewhat different result is obtained for $b<\ell^*$:

$$G_1^S(x,y,\tau)=-\frac{1}{2\pi^2}\sum_{n=1}^{\infty}\int_0^{\infty}dp\ \cos(n\theta)\cos(px)\times f_n(r\sqrt{p^2+\alpha_{\text{out}}^2},\ r_s\sqrt{p^2+\alpha_{\text{out}}^2},\ a\sqrt{p^2+\alpha_{\text{out}}^2},\ a\sqrt{p^2+\alpha_{\text{in}}^2}), \tag{12}$$

where $r=(h^2+y^2)^{1/2}$ and $r_s=(h^2+y_s^2)^{1/2}$.

To satisfy the zero boundary condition in the plane $z=-z_1$ and thereby obtain G_1^S for the experimental scheme shown in Fig. 1, we use the method of images.[19] We place the images of the capillary and radiation source on the other side of the plane $z=-z_1$ so that the geometry of the problem would become symmetric with respect to this plane. Then the desired solution can be written as a sum of the expressions (8) or (9) and terms of the form (10) or (12) corresponding to two different capillaries and light sources.

3. EXPERIMENTAL CONDITIONS

A schematic diagram of the experimental apparatus is shown in Fig. 1. The sample consists of rutile (TiO_2) particles suspended in resin. A cylindrical opening with diameter $d=0.75$ cm was made through the sample at distance $z=0.925$ cm from one of its faces. A laminar flow of a suspension of polystyrene beads in water is maintained in the cylindrical opening by means of communicating vessels placed at different heights. The optical properties (μ_s', μ_a) of the suspension are close to those of the sample material. Therefore the region inside the capillary differs from the surrounding medium only by the dynamics of the particles located in the capillary. For this reason, following the terminology of Ref. 13, we shall call this region dynamically heterogeneous.

Coherent laser radiation at wavelength $\lambda=514$ nm and power 1 W, generated in the TEM_{00} mode by an argon ion laser with a Fabry–Perot etalon placed inside the laser cavity, is injected by means of a system of mirrors and a lens into a multimode fiber-optic waveguide (core diameter 200 μm, numerical aperture 0.16). The Fabry–Perot etalon inside the laser cavity gives an adequate radiation coherence length (about 3 m), which is necessary in experiments on multiple scattering of light.[20] Passing along the waveguide the light is incident on the surface of the sample. The light scattered from the sample is collected by means of a single-mode fiber-optic waveguide (diameter 3.1 μm, numerical aperture 0.13), which allows the fluctuations of the light intensity within the coherence area of the scattered radiation to be recorded. Passing along the waveguide, the scattered radiation enters a photomultiplier,[1)] which operates in the photon counting mode and is connected with a digital multichannel autocorrelator.[2)] The use of fiber-optic waveguides to deliver the laser radiation to the sample and to record the scattered light and the use of a digital correlator make it possible to obtain a high signal/noise ratio in the measurement process.

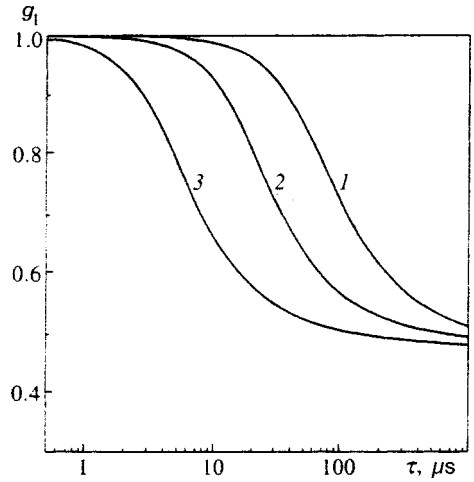


FIG. 2. Normalized time autocorrelation function of multiply scattered radiation, calculated in the diffusion approximation for the case when the radiation source and detector are arranged symmetrically relative to the capillary ($x=0$, $y=-y_s=2.5\ell^*$, $z=3.5\ell^*$, $d=3l^*$, $z_0=z_1=\ell^*$, $\tau_0=6\times10^{-4}$ s). Different curves correspond to different flow velocities: $V=0.1$ (*1*), 0.6 (*2*), 3 (*3*) cm/s.

The measured quantity in the present experiment is the normalized autocorrelation function $g_2(\tau)=\langle I(t)I(t-\tau)\rangle/\langle I\rangle^2$ of the intensity of the scattered light. The radiation scattered in the turbid medium has a Gaussian distribution, as a result of which $g_2(\tau)$ is related with $g_1(\tau)=G_1(\tau)/G_1(0)$ by the Siegert relation

$$g_2(\tau)=1+\beta|g_1(\tau)|^2,$$

where $0<\beta<1$ is the aperture function determined by the measurement system used.[5,21] Since the sample used in the experiment does not satisfy the condition of ergodicity, the product $I(t)I(t-\tau)$ is averaged over an ensemble of realizations by the method proposed in Ref. 21. The essence of this method is that the averaging is accomplished by moving the sample alternately in one direction and then in another relative to the stationary source and detector by means of an electric stepping motor. In our experiments the velocity of the sample motion was equal to about 50 μm/s. The direction of motion is changed automatically, as a result of which the sample is displaced by approximately 500 μm first in one and then in another direction parallel to the capillary axis.

4. BASIC RESULTS AND DISCUSSION

Figure 2 shows the normalized correlation function of the scattered radiation field, calculated for the case shown in Fig. 1 where the radiation is delivered and detected using thin fiber-optic waveguides ($b<\ell^*$). The calculation was performed using Eqs. (9) and (12) with $\alpha_{\text{out}}\equiv0$ and the val-

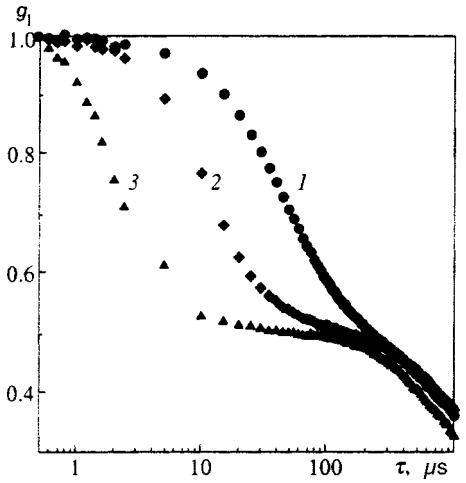


FIG. 3. Experimentally measured normalized time autocorrelation function of multiply scattered radiation for the experimental arrangement shown in Fig. 1. The radiation source and detector are arranged symmetrically with respect to the capillary ($x=0$, $y=-y_s=2.5\ell^*$). Different curves correspond to different flow velocities: $V=0.08$ (*1*), 0.62 (*2*), and 3.22 (*3*) cm/s. The measurement errors fall within the size of the symbols in the figure.

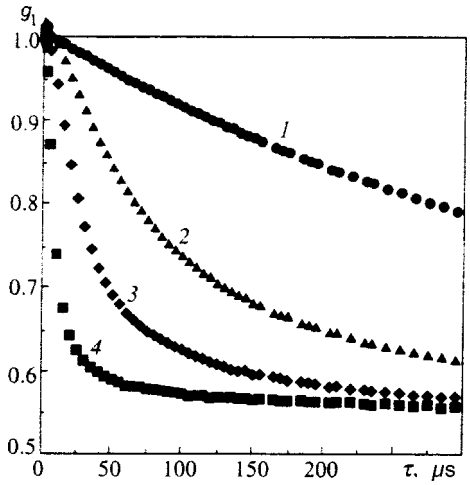


FIG. 4. Normalized time autocorrelation function of multiply scattered radiation, measured for different types of scatterer motion in the capillary: Brownian motion (*1*), laminar flow with velocity $V=0.08$ (*2*) and 0.24 (*3*) cm/s, turbulent flow with $V=0.88$ cm/s (*4*). The remaining parameters are the same as for the curves presented in Fig. 3. The measurement errors fall within the size of the symbols in the figure.

ues of the other parameters close to those used in the experiment. The corresponding experimental points are presented in Fig. 3. As one can see from these figures, the section of the correlation function in the bounded range of delay times τ ($1<\tau<400$ μs under the conditions of our experiment) is most sensitive to a change in the velocity of the fluid flow inside the dynamic region; this agrees with the results obtained in Ref. 13. For $\tau<1$ μs the behavior of the correlation function is determined mainly by the small but nonzero absorption of light in the medium, the absorption being the same both inside and outside the capillary. For $\tau\sim 50-200$ μs $g_1(\tau)$ tends to saturate at a constant level that is independent of the flow velocity. This fact was predicted theoretically in Ref. 12 and can be easily explained qualitatively on the basis of the correspondence between the short trajectories of photons in the medium and the long delay times τ:[8] For large τ the rate of decrease of $g_1(\tau)$ is determined mainly by photons with relatively short trajectories, since photons with long trajectories are now completely decorrelated. Photons with short trajectories consist mainly of photons which do not reach the capillary, and since the particles in the medium surrounding the capillary are immobile, the theoretically computed function $g_1(\tau)$ approaches a constant different from zero as $\tau\to\infty$. The value of this constant is determined solely by the depth at which the capillary is located. As shown in Ref. 12, for $d\gg\ell^*$ it can be estimated as $1-(z_0+z_1)/z\simeq 0.55$ for our sample ($z_0=\ell^*$, $z_1=0.7104\ell^*$, $z=3.7\ell^*$). Since in Ref. 12 a laminar flow of light-scattering particles in a plane-parallel layer and not in a cylindrical capillary is studied, the agreement of this estimate with reality is satisfactory. However, as one can see from Fig. 3, the experimentally measured correlation function, though it tends to saturate at $\tau\sim 50-200$ μs, still continues to decrease for $\tau>200$ μs in contrast to the function calculated theoretically (Fig. 2). This is due to effects which were neglected in our theoretical model (specifically, it is due to the method of averaging, vibration of the components of the optical system, photomultiplier noise, and so on).

We note that the theoretical and experimental results presented in Figs. 2 and 3 not only agree with one another qualitatively but they are also very close quantitatively. This attests to the fact that our theoretical approach to studying the multiple scattering of light in dynamically heterogeneous media is fully applicable to real experimental situations. Moreover, the quantitative agreement between the theoretical and experimental results makes it possible to suggest a method for determining the location and performing diagnostics of scatterer flows hidden inside strongly light-scattering media that is based on measurement and subsequent analysis of the temporal autocorrelation function of the multiply-scattered radiation. Indeed, the difference between the curves $g_1(\tau)$ corresponding to different average velocities V of the particle flow in the capillary is quite large and can be used to measure V directly. Quantitative calibration of this method of measuring the velocity can be performed on the basis of our theoretical model.

It is interesting that the method proposed for studying the dynamics of particles in turbid media can be used with equal success for different types of dynamics of scatterers in the medium. As an illustration, we present in Fig. 4 the results of measurements of the autocorrelation function of backscattered radiation for the cases of Brownian, directed, and turbulent motions of light-scattering particles in the capillary. The system of communicating vessels located at different heights was found to be inadequate for producing a turbulent fluid flow, since such a system does not make it possible to reach the critical Reynolds number on the experimental rectilinear section of the capillary. For this reason, turbulent scatterer motion in the cylindrical cavity inside the experimental sample was produced artificially with the aid of a nozzle placed at the entrance opening of the capillary.

As one can see from Fig. 4, different types of particle

motion lead to different time dependences $g_1(\tau)$. This conclusion also follows from the theoretical analysis: For Brownian motion the variance of the phase difference $\langle \Delta\phi^2(\tau)\rangle$ for two photons scattered successively by the same particles of the medium at times separated by an interval τ is proportional to τ.[8,9] For laminar and turbulent flows $\langle \Delta\phi^2(\tau)\rangle \propto \tau^2$,[18,22] but in the latter case the coefficient of proportionality is larger for the same characteristic particle velocity. Thus, there is a possibility of not only determining the location of the dynamically heterogeneous region (cylindrical capillary in our case) and estimating the average velocity of the particles inside this region but also drawing conclusions about the dominant character of the scatterer motion. We note that for the reasons explained above all curves in Fig. 4 approach the same asymptote for large τ.

5. CONCLUSIONS

In the present paper multiple scattering of laser radiation in a randomly inhomogeneous medium with a spatially localized flow of light-scattering particles was studied. The time autocorrelation function of the light backscattered from a semi-infinite medium with a directed particle flow localized in the region of a cylindrical capillary was calculated in the diffusion approximation. The results of the theoretical analysis are in good agreement with the experimental results at the flow velocity studied. This made it possible to suggest a new method for determining the position and size of the flow region as well as the velocity of the scatterers inside this region. We note that proposed method is the only optical method that permits measuring the velocity of relatively slow (up to 10^{-2} cm/s) flows under conditions of multiple scattering of light in the medium. However, this method can be used only for particle flows near the boundary of the medium ($z < (15-20)\ell^*$ in accordance with our theoretical calculations). This result agrees with the conclusions drawn in Refs. 12 and 13. For this reason, it can be asserted that this restriction is fundamental for turbid media and is not associated with the peculiarities of any specific measurement scheme.

Since the parameters of our sample are close to those of some biological media,[7,15] our method can be used for noninvasive measurement of the characteristics of blood flow.[14] Moreover, the velocity of scatterers can be measured similarly in experiments on laser acceleration of particles in dense media.[23] Thus, the range of potential applications of our method is very wide.

In closing, we note that this work was made possible only by our close collaboration with D. A. Boas, B. Chance, and A. G. Yodh at the University of Pennsylvania (USA), M. Heckmeier and G. Maret at the Center de Recherches sur les Macromolecules (Strasbourg, France), and Professor R. Maynard at the Université Joseph Fourier (Grenoble, France). We thank S. S. Chesnokov, V. V. Tuchin, and A. V. Priezzhev for assistance in preparing the manuscript for publication. We are grateful to the participants of the Joint Scientific Seminar of the Department of General Physics and Wave Processes in the School of Physics at Moscow State University and the International Laser Center at Moscow State University for their interest in this work and personally to N. I. Koroteev for providing the opportunity to speak at this seminar.

*)E-mail: skipetr@fort.phys.msu.su

1)An EMI, Gencom Inc., photomultiplier was used.

2)BI-9000AT, Brookhaven Instruments Corp.

[1] V. L. Kuz'min and V. P. Romanov, Usp. Fiz. Nauk **166**, 247 (1996).
[2] M. P. Van Albada and A. Lagendijk, Phys. Rev. Lett. **55**, 2692 (1985).
[3] P. E. Wolf and G. Maret, Phys. Rev. Lett. **55**, 2696 (1985).
[4] R. Berkovits and S. Feng, **238**, 135 (1994).
[5] A. V. Priezzhev, V. V. Tuchin, and L. P. Shubochkin, *Laser Diagnostics in Biology and Medicine* [in Russian], Moscow, Nauka, 1989.
[6] A. Yodh and B. Chance, Phys. Today **10**, 34 (1995).
[7] V. V. Tuchin, Usp. Fiz. Nauk **167**, 517 (1997).
[8] G. Maret and P. E. Wolf, Z. Phys. B **65**, 409 (1987).
[9] P. M. Chaikin, D. J. Pine, D. A. Weitz, and E. Herbolzheimer, Phys. Rev. Lett. **60**, 1134 (1988).
[10] D. A. Boas, L. E. Campbell, and A. G. Yodh, Phys. Rev. Lett. **75**, 1855 (1995).
[11] M. Heckmeier and G. Maret, Europhys. Lett. **34**, 257 (1996).
[12] S. E. Skipetrov and R. Maynard, Phys. Lett. A **217**, 181 (1996).
[13] M. Heckmeier, S. E. Skipetrov, G. Maret, and R. Maynard, J. Opt. Soc. Am. A **14**, 185 (1997).
[14] I. V. Meglinskiĭ, D. A. Boas, A. G. Yodh *et al.*, Izv. Vyssh. Uchebn. Zaved., Ser. PND **4**, 65 (1996).
[15] A. Ishimaru, *Wave Propagation and Scattering in Random Media*, Academic Press, N. Y., 1978 [Russian translation, Mir, Moscow, 1981].
[16] R. C. Haskell, L. V. Swaasand, T. Tsay *et al.*, J. Opt. Soc. Am. A **11**, 2727 (1994).
[17] D. Bicout and R. Maynard, Physica A **199**, 387 (1993).
[18] D. Bicout and G. Maret, Physica A **210**, 87 (1994).
[19] P. M. Morse and H. Feshbach, *Methods of Mathematical Physics*, McGraw-Hill, N. Y., 1953 [Russian translation, Izd. Inostr. Lit., Moscow, 1958, Vol. 1].
[20] T. Bellini, M. A. Glasser, N. A. Clark, and V. Degiorgio, Phys. Rev. A **44**, 5215 (1991).
[21] J.-Z. Xue, D. J. Pine, S. T. Milner *et al.* Phys. Rev. A **46**, 6550 (1992).
[22] D. J. Bicout and R. Maynard, Physica B **204**, 20 (1995).
[23] S. E. Skipetrov, M. A. Kazaryan, N. P. Korotkov, and S. D. Zakharov, Phys. Scr. **57**, 416 (1998).

Translated by M. E. Alferieff

Role of stationary photon statistics in a high-*Q* cavity mode while it is exciting the active laser medium

Yu. M. Golubev*) and G. R. Ershov

Physics Scientific-Research Institute, St. Petersburg State University, 198904 St. Petersburg, Russia
(Submitted 27 March 1997; resubmitted 11 September 1997)
Zh. Éksp. Teor. Fiz. **113**, 1223–1234 (April 1998)

It has been shown by Yu. M. Golubev, M. I. Kolobov, and I. V. Sokolov, Zh. Éksp. Teor. Fiz. **111**, 1579 (1997) [JETP **84**, 864 (1997)], that when an optical cavity is excited by external radiation from a sub-Poisson laser the cavity mode may be in either a sub-Poisson or a Poisson stationary state. This is not important for a resonant medium which is excited into the upper laser level while interacting with this mode inside the cavity. The degree of regularity of the excitation will be identical to that of the initial light flux incident on the cavity, and this ultimately ensures the same sub-Poisson lasing as for strictly regular pumping of the resonant medium. © *1998 American Institute of Physics.* [S1063-7761(98)00704-5]

1. INTRODUCTION

In the present paper we study the physical situation where in some natural processes, for example, in interatomic collisions, a laser medium is excited only into a nonlaser level. The population inversion required for lasing is provided by the additional action of radiation from an external sub-Poisson laser. The formulation of such a problem is entirely adequate, since we now have at our disposal not only theoretical suggestions for a source of nonclassical light[1] but also experimental implementations of this idea in Japan[2] and in Russia.[3] One can discuss two limiting situations. One is the situation where at the excitation frequency an optical cavity containing the excited medium does not manifest its resonance properties. This is the single-pass excitation variant. In the other limiting case external radiation excites an appropriate high-*Q* cavity mode and this mode in turn excites the laser medium. This is the multiple-pass excitation variant. The single-pass variant is of no interest to us, since the exciting sub-Poisson light will be ''spoiled'' as a result of absorption in the first few layers of the material so that now the main excitation in the further layers will actually be brought about by Poisson light. In such a situation regularization of the excitation of the active medium can only have a small effect.

The situation could be different in the multiple-pass variant, where the excitation in one pass of the radiation through the material is neglected and the required effect accumulates over a large number of passes, the process being uniform along the medium. In our opinion, the laser medium can be excited with the required regularity here. Indeed, according to Ref. 4, if interference conditions are secured such that the incident light flux is not reflected from the entrance cavity mirror but completely exits on the opposite side of the cavity (ideally transmitting optical cavity), then the outgoing photon flux is regularized to the same extent as at the cavity entrance. If the exit window is now closed and the flux is used to excite the laser medium, which is distributed uniformly inside the cavity, then the condition of total transmission of the optical cavity is preserved (the cavity does not reflect anything; the physical nature of the departure of the field from the cavity is, of course, not important). In this case there is hope that the excitation of the medium inside the cavity will be, first, uniform over the cavity volume and, second, regularized. In what follows we shall show this and we shall determine the measure of this regularization.

In addition to this, we must solve one other question of fundamental importance: Does stationary photon statistics in the exciting mode influence the excitation statistics of the medium? When we posed the problem in Ref. 5, we assumed that sub-Poisson photon statistics in the exciting mode must be established in order to accomplish regular excitation of the medium. We constructed our thought experiment so as to ensure that under given physical circumstances the mode photons would be sub-Poisson to the maximum degree possible. This corresponded to the generally held view that the predicted statistical pattern in one or another experiment and the stationary distribution in the cavity mode should correspond to one another. Now, however, in consequence of the results obtained in Ref. 4, we know that at least in some measurement procedures stationary photon statistics in a mode can be completely unimportant, and we intend to demonstrate this for the excitation of the active medium.

As we understand the situation,[4] depending on the ratios of the actual modal spectral widths, completely different statistical pictures can arise in a mode excited by the same sub-Poisson light. We can secure inside a cavity precisely the same photon number fluctuations as in the mode of the initial sub-Poisson laser. But the situation where these fluctuations are completely random is entirely realistic. At the same time it is obvious that what are more important for us than the stationary states of any group of particles are the statistical properties of the fluxes of particles which different groups of particles exchange with one another (for example, the photon flux from a stationary mode into a stationary medium) and which, strictly speaking, need not have the same statistical properties as the stationary states.

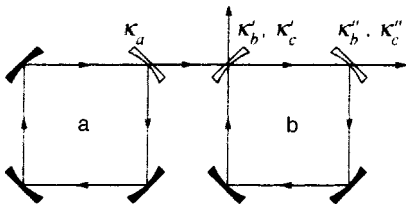


FIG. 1.

2. THOUGHT EXPERIMENT AND ITS QUALITATIVE DISCUSSION

Let us assume that the physical system consists of two high-Q optical cavities (see Fig. 1), standing next to one another. One cavity (the a cavity) lases at the frequency ω_a of a one-mode sub-Poisson laser (the a laser). Its radiation enters the b cavity through an entrance window and excites a c mode at frequency ω_c (we shall assume below that the frequencies ω_a and ω_c are the same). The c mode in turn acts on the medium placed in the cavity, giving rise to the required population inversion and hence the lasing of the b laser at the frequency ω_b. In order to accomplish all this the active media of both lasers must have the proper resonance properties (Fig. 2). We shall assume that the active medium of the a laser and its excitation system are the same as that used in Ref. 1 to describe a sub-Poisson laser: This is essentially a two-layer medium with regular and stationary excitation of the upper level. We shall assume the b-laser medium is a four-level medium. In the absence of external sub-Poisson radiation, the laser levels (1,2) will not be occupied at all. The atomic level (0) is populated entirely randomly, for example, as a result of collisions occurring in the laser plasma. We shall assume that the radiation frequency ω_a of the sub-Poisson laser equals the frequency of the 0–3 atomic transition. Then in the process of absorption of the external laser radiation and subsequent rapid spontaneous relaxation on the 3–2 transition the atoms are transferred back into the upper laser level, creating the required above-threshold inversion. For simplicity, we shall not introduce any frequency detuning: The mode frequencies ω_a, ω_b, and ω_c are equal to the frequencies of the corresponding atomic transitions in the medium of the a and b lasers.

It is necessary to introduce into the analysis the spectral widths of three actual cavity modes. We shall discuss the main results for the following physical conditions, optimal for our purposes, physical conditions: The spectral width $\kappa_c+\kappa$ of the c mode is formed by the passage of the field at

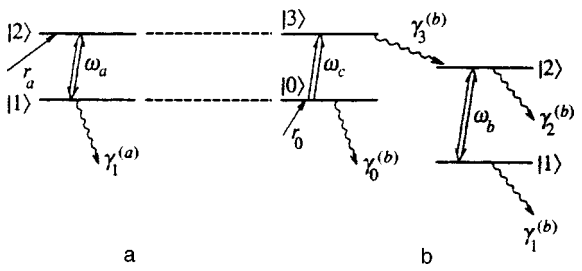


FIG. 2.

frequency ω_c through the entrance window (κ_c) and by absorption in the medium when it is excited (κ); the spectral width of the b mode is formed only by the passage of the field at frequency ω_b through the exit mirror. We shall study the physical situations with different ratios of all spectral widths κ_a (spectral width of the a mode), κ_b, κ_c, and κ.

We begin our discussion with the simplest situations, gradually making the system more complicated. We start from the situation when there is no b cavity at all and the radiation of the a laser, which we shall assume to be an ideal sub-Poisson laser, directly strikes the photodetector. In the present paper we shall assume that the electrical circuit for switching on the photodetector enables us to know the photocurrent spectrum at the exit. According to Ref. 1 this spectrum is given by the following explicit expression in the case of ideal photodetection:

$$i_\Omega^{(2)}=i_{\text{shot}}^{(2)}\left(1-\frac{\kappa_a^2}{\kappa_a^2+\Omega^2}\right). \tag{1}$$

As is well known, and as one can see here, at zero frequency shot noise is completely suppressed.

Let us now make the system more complicated by placing between the sub-Poisson laser and the photodetector a b cavity whose mode frequency ω_c, as we have already stipulated, equals the lasing frequency ω_a of the sub-Poisson laser. But now we shall assume that this cavity is empty, it contains no active medium, and both the entrance and exit mirrors transmit, i.e., $\kappa_c=\kappa_c'+\kappa_c''$. It is obvious that in the empty cavity a fixed number of photons will accumulate in the c mode and the entering flux of photons from the a laser will somehow separate into a flux reflected from the b cavity and a flux directed onto the photodetector. The question of what stationary photon statistics are formed inside the b cavity can now be easily answered.[4] For this, we make the following argument. The Mandel parameter ξ, characterizing the rms photon number fluctuations

$$\overline{n^2}-\overline{n}^2=\overline{n}(1+\xi) \tag{2}$$

is often used for the statistical description.

For the a laser, which in our analysis is an ideal sub-Poisson laser, the Mandel parameter is $\xi_a=-1/2$, which is established by the corresponding correlations between the Fourier components of the spectral contour of width κ_a. In the limit $\kappa_a\gg\kappa_c$, this correlation is transferred inside the b cavity to the Fourier components of the spectral contour of width κ_c. In this case it is natural to expect that the Mandel parameter ξ_c will also equal $-1/2$. In the opposite case $\kappa_a\ll\kappa_c$ the correctly correlated components occupy only a very small portion of the contour at its center, while a large number of Fourier components is excited by uncorrelated vacuum fluctuations. Then it can be expected that the Mandel parameters ξ_c will become close to zero. Therefore, depending on the ratios of the spectral widths, the stationary photon statistics inside the b cavity can be sub-Poisson, just as in the initial laser, or Poisson (and, of course, any intermediate statistics). Here we have the right to raise the questions: Will this stationary photon statistics influence the char-

acter of the excitation of the atoms? Is it necessary to strive for good photon statistics in order to secure regularity of excitation of the medium?

First, however, we shall finish our discussion for the case of an empty cavity. As we have already mentioned in the Introduction, for us it is most important to understand the statistical properties of the fluxes and not the stationary states of the particles. We can imagine in principle that when an ordered photon flux enters an empty cavity a completely random ensemble is created inside the cavity under stationary conditions, but the outgoing flux is nonetheless ordered to the same degree as the entering flux. This is confirmed in some sense in Ref. 4. In the case when the cavity is tuned for complete transmission (when the detuning of the frequency of the incident light from the c-mode frequency equals zero, this happens for $\kappa_c'=\kappa_c''$), the photon flux leaving the cavity is ordered to the same degree as the flux entering the cavity, irrespective of the statistics of the ensemble of intracavity photons. In this sense the cavity behaves just as a simple beam-splitting plate, ignoring the presence of any internal photon state.

Now let us place our four-level medium in the cavity (Fig. 2) and set $\kappa_c''=0$. We thereby redirect the outgoing photon flux: Before it was absorbed by the photodetector, whereas now it is absorbed by the active medium of the b laser. Obviously, the fate of the photons lost to the c mode cannot influence the c mode itself and thereby also our earlier qualitative conclusions. When $\kappa_c=\kappa$ holds, the cavity once again operates only in transmission, but the entire photon flux, which once again remains ordered, now goes to excitation of the active medium. In this case the same ordered excitation of the atomic medium can be expected, irrespective of the stationary photon statistics in the c mode. As a consequence of this, sub-Poisson lasing of the b laser can be expected.[5] This qualitative conclusion is confirmed by an analysis of the photocurrent spectrum. According to the calculations below, it can be represented in the form

$$i_\Omega^{(2)}=i_{\text{shot}}^{(2)}\left[1-\frac{4\kappa_c\kappa}{(\kappa_c+\kappa)^2}\,\frac{\kappa_a^2}{\kappa_a^2+\Omega^2}\,\frac{(\kappa_c+\kappa)^2}{(\kappa_c+\kappa)^2+4\Omega^2}\,\frac{\kappa_b^2}{\kappa_b^2+\Omega^2}\right]. \tag{3}$$

As one can see, near zero frequency a dip with a quite complicated shape and depth

$$\delta=\frac{|i_{\Omega=0}^{(2)}-i_{\text{shot}}^{(2)}|}{i_{\text{shot}}^{(2)}}=\frac{4\kappa_c\kappa}{(\kappa_c+\kappa)^2} \tag{4}$$

is present in the shot noise.

The depth reaches a maximum and the dip reaches zero in the case of an ideally transmitting cavity, where $\kappa_c=\kappa$.

3. BASIC KINETIC EQUATION

We are studying three field oscillators. One oscillator represents the intracavity single-mode lasing field of the a laser. The other two oscillators represent the two-mode field in the b cavity: One mode serves for excitation of the medium while the other is the lasing mode. The structure of the basic kinetic equation for the density matrix of a three-oscillator electromagnetic field can be determined as follows. The rate of change of the density matrix consists of the rate due to the interactions inside the a cavity, the rate due to the interactions inside the b cavity, and finally the rate associated with directed transfer of electromagnetic energy from the a to the b cavity. We shall express this formally as follows:

$$\dot\rho=r_a\left(\hat L_a-\frac{1}{2}\,\hat L_a^2\right)\rho-\frac{1}{2}\,\kappa_a\hat R_a\rho+\sqrt{\kappa_a\kappa_c}\hat L_{a\to c}\rho$$
$$+\kappa\underset{\leftarrow}{c}\underset{\rightarrow}{c}^{+}\hat L_b\rho-\frac{1}{2}\,(\kappa_c+\kappa)\hat R_c\rho-\frac{1}{2}\,\kappa_b\,\hat R_b\,\rho. \tag{5}$$

The first term in Eq. (5), containing the operator

$$\hat L_a=\left[2\underset{\rightarrow}{a}^{+}\underset{\rightarrow}{a}-\underset{\rightarrow}{a}\underset{\rightarrow}{a}^{+}-\underset{\leftarrow}{a}\underset{\rightarrow}{a}^{+}-\frac{1}{2}\,\beta_1^{(a)}(\underset{\rightarrow}{a}\underset{\rightarrow}{a}^{+}-\underset{\leftarrow}{a}\underset{\rightarrow}{a}^{+})^2\right]$$
$$\times\left[\underset{\rightarrow}{a}\underset{\rightarrow}{a}^{+}+\underset{\leftarrow}{a}\underset{\rightarrow}{a}^{+}+\frac{1}{2}\,\beta_1^{(a)}(\underset{\rightarrow}{a}\underset{\rightarrow}{a}^{+}-\underset{\leftarrow}{a}\underset{\rightarrow}{a}^{+})^2\right]^{-1}, \tag{6}$$

determines the development of the intracavity field of the a laser on account of the active medium (see Fig. 2) excited into the upper laser level with rate r_a with no fluctuations.[1,6] The expression is written under the assumption that the upper laser level does not decay spontaneously into extraneous levels. As is well known, the maximum possible squeezing of the light can be achieved in this case. The parameter

$$(\beta_1^{(a)})^{-1}=\frac{\gamma_1^{(a)}\gamma_{12}^{(a)}}{2|g_{12}^{(a)}|^2} \tag{7}$$

sets the number of photons saturating the laser transition. The active medium of the a laser is characterized by the following constants: $\gamma_1^{(a)}$, the longitudinal longitudinal relaxation constant of the lower laser level; $\gamma_{12}^{(a)}$, the transverse relaxation constant of the laser transition; and a and a^+, photon operators in the lasing mode of the a laser, $[a,a^+]=1$.

The second, last, and next-to-last terms in Eq. (5) predetermine the decay of the quantum oscillators. The operators $\hat R_a$ have the conventional form

$$\hat R_a=\underset{\leftarrow}{a}^{+}\underset{\leftarrow}{a}+\underset{\rightarrow}{a}^{+}\underset{\rightarrow}{a}-2\underset{\rightarrow}{a}\underset{\leftarrow}{a}^{+}, \tag{8}$$

$\hat R_b$ and $\hat R_c$ have the same form except that the photon operators a,a^+ are replaced by the photon operators b,b^+ or c,c^+, where $[b,b^+]=[c,c^+]=1$.

The b-laser medium excited into the upper laser level ensures the development of the b mode via the operator[6]

$$\hat L_b=\beta_2^{(b)}\left[2\underset{\rightarrow}{b}^{+}\underset{\rightarrow}{b}-\underset{\rightarrow}{b}\underset{\rightarrow}{b}^{+}-\underset{\leftarrow}{b}\underset{\rightarrow}{b}^{+}-\frac{1}{2}\,\beta_2^{(b)}(\underset{\rightarrow}{b}\underset{\rightarrow}{b}^{+}-\underset{\leftarrow}{b}\underset{\rightarrow}{b}^{+})^2\right]$$
$$\times\left[1+\frac{1}{2}\,\beta_+^{(b)}(\underset{\rightarrow}{b}\underset{\rightarrow}{b}^{+}+\underset{\leftarrow}{b}\underset{\rightarrow}{b}^{+})\right.$$
$$\left.+\frac{1}{4}\,\beta_1^{(b)}\beta_2^{(b)}(\underset{\rightarrow}{b}\underset{\rightarrow}{b}^{+}-\underset{\leftarrow}{b}\underset{\rightarrow}{b}^{+})^2\right]^{-1}. \tag{9}$$

Here the parameters determining the nonlinear properties of the b medium have the form

$$\beta_1^{(b)}=\frac{2|g_{12}^{(b)}|^2}{\gamma_1^{(b)}\gamma_{12}^{(b)}},\quad \beta_2^{(b)}=\frac{2|g_{12}^{(b)}|^2}{\gamma_2^{(b)}\gamma_{12}^{(b)}},\quad \beta_+^{(b)}=\beta_1^{(b)}+\beta_2^{(b)}. \tag{10}$$

The active medium of the b laser is characterized by the following constants: $\gamma_1^{(b)}$ and $\gamma_2^{(b)}$ are the longitudinal relaxation constants of the lower and upper laser levels, and $\gamma_{12}^{(b)}$ is the transverse relaxation constant of the laser transition. In formulating the basic kinetic equation (5) we adopted the requirement, optimal for our purposes, that the spontaneous decay rate of the upper laser level into extraneous levels equals zero. In what follows, we shall proceed in exactly the same way with respect to the b laser and we shall require that $\gamma_2^{(b)}=0$. Then the operator $\hat{L}_b$ will have exactly the same form as the operator $\hat{L}_a$, with all a indices replaced by b indices and all photon operators a,a^+ replaced by the photon operators b,b^+.

The physical meaning of the quantity $r_b=\kappa n_c$ is that of an average rate of excitation of the upper laser level of the b laser (n_c is the average number of photons in the c mode exciting the medium).

Finally, the term containing the operator

$$\hat{L}_{a\to c}=\underset{\sim}{a}\underset{\sim}{c}^+ +\underset{\sim}{a}^+\underset{\sim}{c}-\underset{\sim}{a}\underset{\sim}{c}^+ -\underset{\sim}{b}\underset{\sim}{a}^+, \tag{11}$$

in the basic kinetic equation gives directed transfer of electromagnetic radiation from the a cavity into the b cavity (from the a into the c mode).[7] All these expressions can be obtained by deriving the kinetic equation for the density matrix of the electromagnetic field using, for example, the procedure of Lamb and Scully.[8]

We shall make one more remark concerning the quantities κ_b and κ_c. In principle, we assume that the b cavity contains two working mirrors: One is the entrance mirror through which the external radiation enters the cavity and the second is the exit mirror through which the radiation exits the b cavity and reaches the photodetector. Correspondingly, we must write in the general form $\kappa_b=\kappa_b'+\kappa_b''$ and $\kappa_c=\kappa_c'+\kappa_c''$. However, as stated earlier, for us the optimal conditions are $\kappa_c''=0$, i.e., when the losses occur in the same channel as excitation, and $\kappa_b'=0$, i.e., when the losses occur only in the detection channel and as a result of detection. It is easy to see that these conditions are physically consistent, since the b- and c-mode frequencies can differ greatly and therefore the transmission of the mirrors for them can be completely different.

The absorption coefficient for the external radiation on the auxiliary transition 0–3 can be written in the form $\kappa=r_0\beta_0^{(b)}$, where r_0 is the average rate of completely random excitation of the auxiliary level 0. Equation (5) was written under the assumption that the nonlinear phenomena in absorption can be neglected. This is accomplished if

$$(\beta_0^{(b)}+\beta_3^{(b)})n_b\ll 1,$$

which contains the saturation parameters for the b atom

$$\beta_0^{(b)}=\frac{2|g_{03}^{(b)}|^2}{\gamma_0^{(b)}\gamma_{03}^{(b)}},\quad \beta_3^{(b)}=\frac{2|g_{03}^{(b)}|^2}{\gamma_3^{(b)}\gamma_{03}^{(b)}}.$$

4. BASIC KINETIC EQUATION IN THE DIAGONAL REPRESENTATION AND SEMICLASSICAL EQUATIONS

We now switch to the diagonal representation for the density matrix. This representation is introduced by the integral relation

$$\rho(t)=\int P(\alpha,\beta,\delta,t)|\alpha\rangle|\beta\rangle|\delta\rangle\langle\delta|\langle\beta|\langle\alpha|d^2\alpha d^2\beta d^2\delta, \tag{12}$$

where

$$a|\alpha\rangle=\alpha|\alpha\rangle,\quad b|\beta\rangle=\beta|\beta\rangle,\quad c|\delta\rangle=\delta|\delta\rangle.$$

This transformation leads to a partial differential equation for the weighting function $P(\alpha,\beta,\delta,t)$ that can be greatly simplified by assuming that the number of photons in each actual mode fluctuates very little around its stationary value:

$$\begin{aligned}&|\alpha|^2=n_a+\varepsilon_a,\quad \varepsilon_a\ll n_a,\\ &|\beta|^2=n_b+\varepsilon_b,\quad \varepsilon_b\ll n_b,\\ &|\delta|^2=n_c+\varepsilon_c,\quad \varepsilon_c\ll n_c.\end{aligned} \tag{13}$$

To a high degree of accuracy, the stationary solutions n_a, n_b, and n_c are identical to the semiclassical stationary solutions of the problem. The semiclassical equations can be found relatively easily, for example, from the initial equation (5) by neglecting all fluctuations of the fields. These equations are:

$$\begin{aligned}&\frac{d}{dt}n_a=r_a-\kappa_a n_a,\\ &\frac{d}{dt}n_b=-\kappa_b n_b+\kappa n_c,\\ &\frac{d}{dt}n_c=-(\kappa_c+\kappa)n_c+2\sqrt{\kappa_a\kappa_c}\sqrt{n_a n_c}.\end{aligned} \tag{14}$$

Under stationary conditions this system becomes a system of algebraic equations, the solutions of which can be written in the form

$$n_a=\frac{r_a}{\kappa_a},\quad \frac{n_c}{n_a}=\frac{4\kappa_a\kappa_c}{(\kappa_c+\kappa)^2},\quad \frac{n_b}{n_c}=\frac{\kappa}{\kappa_b}. \tag{15}$$

On this basis we can write for the photon density matrix in the Glauber representation

$$R(\varepsilon_a,\varepsilon_b,\varepsilon_c,t)=\int P(\alpha,\beta,\delta,t)d\varphi_a d\varphi_b d\varphi_c, \tag{16}$$

$$\alpha=\sqrt{n_a+\varepsilon_a}e^{i\varphi_a},\quad \beta=\sqrt{n_b+\varepsilon_b}e^{i\varphi_b},$$

$$\delta=\sqrt{n_c+\varepsilon_c}e^{i\varphi_c}$$

the following basic kinetic equation:

$$\frac{\partial R}{\partial t}=\kappa_a\frac{\partial}{\partial\varepsilon_a}\varepsilon_a R-\frac{1}{2}\kappa_a n_a\frac{\partial^2 R}{\partial\varepsilon_a^2}+\kappa_b\frac{\partial}{\partial\varepsilon_b}\times\left(\varepsilon_b-\frac{n_b}{n_c}\varepsilon_c\right)R+\frac{1}{2}(\kappa_c+\kappa)\frac{\partial}{\partial\varepsilon_c}$$

$$\times\left(\varepsilon_c-\frac{n_c}{n_a}\,\varepsilon_a\right)R+\{\cdots\}. \tag{17}$$

The expression $\{\cdots\}$ represents all higher-order ε derivatives which in principle must be taken into account for nonclassical fields. However, we shall show below that they do not contribute to the physical, measured quantities which we shall investigate here.

5. STATISTICAL PROPERTIES OF THE PHOTONS STORED IN A STATIONARY STATE IN THE *B* AND *C* MODES

Using the basic kinetic equation (17) and the conventional procedure, we write the equations for the effective averages as

$$\frac{d}{dt}\,\overline{\varepsilon_b^2}=-2\kappa_b\overline{\varepsilon_b^2}+2\kappa_b\,\frac{n_b}{n_c}\,\overline{\varepsilon_b\varepsilon_c}=0, \tag{18}$$

$$\frac{d}{dt}\,\overline{\varepsilon_a^2}=-2\kappa_a\overline{\varepsilon_a^2}-\kappa_a n_a=0, \tag{19}$$

$$\frac{d}{dt}\,\overline{\varepsilon_c^2}=-(\kappa_c+\kappa)\overline{\varepsilon_c^2}+(\kappa_c+\kappa)\,\frac{n_c}{n_a}\,\overline{\varepsilon_a\varepsilon_c}=0, \tag{20}$$

$$\frac{d}{dt}\,\overline{\varepsilon_b\varepsilon_c}=-\frac{1}{2}\,(\kappa_c+\kappa+2\kappa_b)\overline{\varepsilon_b\varepsilon_c}+\frac{1}{2}\,(\kappa_c+\kappa)\,\frac{n_c}{n_a}\,\overline{\varepsilon_a\varepsilon_b}+\kappa_b\,\frac{n_b}{n_c}\,\overline{\varepsilon_c^2}=0, \tag{21}$$

$$\frac{d}{dt}\,\overline{\varepsilon_a\varepsilon_b}=-(\kappa_a+\kappa_b)\overline{\varepsilon_a\varepsilon_b}+\kappa_b\,\frac{n_b}{n_c}\,\overline{\varepsilon_a\varepsilon_c}=0, \tag{22}$$

$$\frac{d}{dt}\,\overline{\varepsilon_a\varepsilon_c}=-\frac{1}{2}\,(\kappa_c+\kappa+2\kappa_a)\overline{\varepsilon_a\varepsilon_c}+\frac{1}{2}\,(\kappa_c+\kappa)\,\frac{n_c}{n_a}\,\overline{\varepsilon_a^2}=0. \tag{23}$$

As can be seen, we have obtained a closed system of differential equations which for stationary conditions becomes an easily solvable system of algebraic equations. We note that here the terms $\{\cdots\}$ from Eq. (17) in principle do not contribute here, and this is not due to any mathematical approximations. The solutions can be represented in the form

$$\overline{\varepsilon_a^2}=-\frac{1}{2}\,n_a, \tag{24}$$

$$\overline{\varepsilon_b^2}=-\frac{2\kappa_a\kappa_c\kappa(\kappa_c+\kappa+2\kappa_a+2\kappa_b)}{(\kappa_c+\kappa)(\kappa_c+\kappa+2\kappa_a)(\kappa_c+\kappa+2\kappa_b)(\kappa_a+\kappa_b)}\,n_b, \tag{25}$$

$$\overline{\varepsilon_c^2}=-\frac{2\kappa_a\kappa_c}{(\kappa_c+\kappa)(\kappa_c+\kappa+2\kappa_a)}\,n_c, \tag{26}$$

$$\overline{\varepsilon_b\varepsilon_c}=-\frac{2\kappa_a\kappa_c\kappa_b(\kappa_c+\kappa+2\kappa_a+2\kappa_b)}{(\kappa_c+\kappa)(\kappa_c+\kappa+2\kappa_a)(\kappa_c+\kappa+2\kappa_b)(\kappa_a+\kappa_b)}\,n_b, \tag{27}$$

$$\overline{\varepsilon_a\varepsilon_b}=-\frac{1}{2}\,\frac{\kappa_b(\kappa_c+\kappa)}{(\kappa_a+\kappa_b)(\kappa_c+2\kappa_a)}\,n_b, \tag{28}$$

$$\overline{\varepsilon_a\varepsilon_c}=-\frac{1}{2}\,\frac{\kappa_c+\kappa}{\kappa_c+\kappa+2\kappa_a}\,n_c. \tag{29}$$

We take the condition $\kappa_c=\kappa$, for which, as we know, the b cavity becomes completely transmitting in the sense that the photon flux entering the b cavity is completely expended in exciting the active medium of the b laser.

Then, since $\overline{\varepsilon_b^2}=\xi_b n_b$ and $\overline{\varepsilon_c^2}=\xi_c n_c$, we obtain the following expressions for the corresponding Mandel statistical parameters characterizing the stationary photon fluctuations in each active mode:

$$\xi_b=-\frac{1}{2}\,\frac{\kappa_a}{\kappa_a+\kappa_c}\,\frac{\kappa_c}{\kappa_b+\kappa_c}\,\frac{\kappa_a+\kappa_b+\kappa_c}{\kappa_a+\kappa_b}, \tag{30}$$

$$\xi_c=-\frac{1}{2}\,\frac{\kappa_a}{\kappa_a+\kappa_c}. \tag{31}$$

In the limit $\kappa_c\gg\kappa_a$,

$$\xi_c=0,\quad \xi_b=-\frac{1}{2}\,\frac{\kappa_a}{\kappa_a+\kappa_b}, \tag{32}$$

and in the limit $\kappa_c\ll\kappa_a$,

$$\xi_c=-\frac{1}{2},\quad \xi_b=-\frac{1}{2}\,\frac{\kappa_c}{\kappa_b+\kappa_c}. \tag{33}$$

The situation with the Mandel parameter ξ_c is, of course, in complete accord with the results of Ref. 4. But it turns out that even when $\xi_c=0$ holds with $\kappa_a\ll\kappa_c$, the Mandel parameter ξ_b for the lasing mode can still be a minimum for the present problem and equal to $-1/2$. For this it is sufficient to require that $\kappa_b\ll\kappa_a$.

6. SPECTRUM OF THE PHOTOCURRENT IN THE PROCESS OF DETECTING THE LASING OF THE *B* LASER

The spectrum of the photocurrent during observation of secondary lasing can be written in the form

$$i_\Omega^{(2)}=i_{\text{shot}}^{(2)}\left(1+2q\,\frac{\kappa_b}{n_b}\,\mathrm{Re}\int_0^\infty dt\;\overline{\varepsilon_b\varepsilon_b(t)}e^{i\Omega t}\right). \tag{34}$$

The solution of this problem gives a system of equations which can be obtained from the basic kinetic equation (17) by the standard procedure:

$$\frac{d}{dt}\,\overline{\varepsilon_b\varepsilon_b(t)}=-\kappa_b\,\overline{\varepsilon_b\varepsilon_b(t)}+\kappa_b\,\frac{n_b}{n_c}\,\overline{\varepsilon_b\varepsilon_c(t)}, \tag{35}$$

$$\frac{d}{dt}\,\overline{\varepsilon_b\varepsilon_c(t)}=-\frac{1}{2}\,(\kappa_c+\kappa)\overline{\varepsilon_b\varepsilon_c(t)}$$

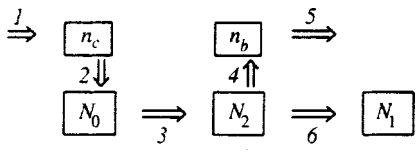


FIG. 3.

$$+\frac{1}{2}(\kappa_c+\kappa)\frac{n_c}{n_a}\overline{\varepsilon_b\varepsilon_a(t)}, \tag{36}$$

$$\frac{d}{dt}\overline{\varepsilon_b\varepsilon_a(t)}=-\kappa_a\overline{\varepsilon_b\varepsilon_a(t)}. \tag{37}$$

This system is easily solved directly, but in this case it is more efficient to switch to a system of algebraic equations for the Fourier components, especially since these equations are needed to write down the photocurrent spectrum:

$$x_\Omega=\int_0^\infty dt\ \overline{\varepsilon_b\varepsilon_b(t)}e^{i\Omega t}.$$

$$y_\Omega=\int_0^\infty dt\ \overline{\varepsilon_b\varepsilon_c(t)}e^{i\Omega t}. \tag{38}$$

$$z_\Omega=\int_0^\infty dt\ \overline{\varepsilon_b\varepsilon_a(t)}e^{i\Omega t}.$$

From the initial differential equations it is easy to obtain

$$\overline{\varepsilon_b^2}-i\Omega x_\Omega=-\kappa_b x_\Omega+\kappa_b\frac{n_b}{n_c}y_\Omega, \tag{39}$$

$$\overline{\varepsilon_b\varepsilon_c}-i\Omega y_\Omega=-\frac{1}{2}(\kappa_c+\kappa)y_\Omega+\frac{1}{2}(\kappa_c+\kappa)\frac{n_c}{n_a}z_\Omega, \tag{40}$$

$$\overline{\varepsilon_b\varepsilon_a}-i\Omega z_\Omega=-\kappa_a z_\Omega. \tag{41}$$

Solving this system of equations we can finally write down the explicit expression for the photocurrent spectrum in the form (3).

7. CONCLUSIONS

The entire b system as a whole can be schematically represented as a collection of boxes in which different types of particles are collected and fluxes between these boxes. Figure 3 shows only the basic elements, specifically, boxes with the average number of photons n_c (the number of photons stored in the stationary state in the c mode), the average number of photons n_a (the number of photons stored in the stationary state in the a mode), the average number of atoms N_0 (number of atoms stored under stationary conditions in the level 0), and the average number of atoms N_2 (the number of atoms stored under stationary conditions in level 2) and N_1 (number of atoms stored under stationary conditions in level 1). These boxes are all coupled through the particle fluxes, which to some extent form stationary states inside the boxes. The flux *1* is the photon flux from the external sub-Poisson laser to the entrance mirror of the b cavity, while the flux *5* is the photon flux from the b mode to the photodetector. The collection of particles in each box can be associated with a stationary state in which the number of particles fluctuates around its average value. But, on the whole, it is obvious that for us it is not so much the stationary states as the properties of the fluxes of the corresponding particles that are important. Of course, the properties of the fluxes and the properties of the local stationary states can be related in some way. At the same time, however, one can imagine a situation where this is not the case, and the main conclusion which we must draw from our arguments and calculations is that in a particular problem with $\kappa_c=\kappa$ all particle fluxes participating in the process are qualitatively identical and independent of the local stationary states, which disrupt these fluxes. If the initial flux 1 is somehow regulated, then to some extent each of the other five fluxes indicated in the diagram is also regulated to the same degree, though the character of the fluctuations in each box can be completely random.

This work was supported in part by INTAS (93-194 EXT) and INTAS-RFBR (95-0656), for which we are grateful.

*)E-mail:yuri_golubev@pop.convey.ru

[1] Yu. M. Golubev and I. V. Sokolov, Zh. Éksp. Teor. Fiz. **87**, 408 (1984) [Sov. Phys. JETP **60**, 234 (1984)].
[2] Y. Yamamoto, S. Mashida, and O. Nilson, Phys. Rev. A **34**, 4025 (1986); W. H. Richardson, Y. Yamamoto, and S. Mashida, Phys. Rev. Lett. **66**, 2867 (1991).
[3] A. S. Trifonov and P. A. Usachev, in *Proc. SPIE 2378* (1995), p. 122.
[4] Yu. M. Golubev, M. I. Kolobov, and I. V. Sokolov, Zh. Éksp. Teor. Fiz. **111**, 1579 (1997) [JETP **84**, 864 (1997)].
[5] Yu. M. Golubev, Zh. Éksp. Teor. Fiz. **103**, 832 (1993) [JETP **76**, 408 (1993)].
[6] Yu. M. Golubev, G. R. Ershov, A. N. Prokshin, and I. V. Sokolov, Opt. Spektrosk. **82**(6), 860 (1997) [Opt. Spectrosc. **82**, 936 (1997)].
[7] M. I. Kolobov and I. V. Sokolov, Opt. Spektrosk. **62**, 112 (1987) [Opt. Spectrosc. **62**, 69 (1987)].
[8] M. O. Scully and W. E. Lamb, Phys. Rev. A **159**, 208 (1967).

Translated by M. E. Alferieff

Quantum nondemolition measurements of the phase and polarization Stokes parameters of optical fields

A. P. Alodzhants and S. M. Arakelyan*)

Vladimir State University, 600026 Vladimir, Russia
(Submitted 19 July 1996; resubmitted 11 August 1997)
Zh. Éksp. Teor. Fiz. **13**, 1235–1252 (April 1998)

We propose a new method for performing continuous quantum nondemolition measurements (QNDM) of the polarization characteristics of light, the signal photon number, and the Stokes parameters of light. For devices that implement such measurements we take cubically nonlinear double-filament optical fibers with tunnel coupling of the filaments. Such a system employs a four-mode mixing of the waves (two probe modes and two signal modes). Linear optical elements needed for preliminary and subsequent conversion of the Stokes parameters are also used. We show that measurements of one Stokes parameter can be used for QNDM of the phase difference between the two initial modes. Here the accuracy of such measurements is determined by the initial fluctuations of the Stokes parameters of the probe modes and by the size of the nonlinear mode conversion coefficient in the optical fibers. © *1998 American Institute of Physics.* [S1063-7761(98)00804-X]

1. INTRODUCTION

Lately many papers have appeared concerning the phase problem in quantum optics.[1–8] The focus in such research is usually on the mathematics of the problem.[1–4] We believe that a big achievement here was the introduction of a new Hermitian operator by Pegg *et al.*,[2] which became known as the Pegg–Barnett operator. The use of this operator largely removed the drawbacks associated with the standard Susskind–Glogower representation. However, the analysis of the problem must not be separated from the specific procedure of phase measurements (cf. Refs. 5–7). The present paper discusses the possibilities of such quantum measurements of phase (the phase difference between two modes).

The phase problem proper is not specific to the quantum mode of description and is well-known in optics, where there are certain difficulties in measuring the phase of a light wave directly. This problem has a long history in classical optics (see, e.g., Ref. 8).

Of course, in various measurements involving light fields only the phase difference between two (or more) spatially distant points has physical meaning, and methods used to measure this difference abound.

Actually, all the methods are aimed at extracting phase information from intensity measurements in one or another optical scheme. In particular, the reconstruction of phase information is done in holographic images: the forming of a three-dimensional image is the result of such reconstruction. Any method of processing optical images is also related to this problem. For instance, one field of quantum optics being actively developed is quantum tomography.[7]

Since the phase parameters of optical radiation are directly related to light polarization, the study of quantum polarization states and the procedures used in measuring such states play an important in the given problem.[9]

One possible way to describe polarization problems (both classical[10] and quantum[11,12] polarization states of light) is to introduce real-valued (Hermitian) Stokes parameters of the light field (see Sec. 2). Here the specific features of the quantum polarization properties of light (and hence of the phase problem) are determined by uncertainty relations for the Stokes parameters. The existence of three noncommuting components of the operators of these parameters (similar to the operators of angular momentum) makes their simultaneous and exact measurement impossible (except in the trivial zero case).

For this reason the problem of quantum measurements of the Stokes parameters comes to the fore. Two basic methods of measuring the Stokes parameters are possible in quantum mechanics.

The first is related to the realization of the procedure of measuring all four Stokes parameters (see Ref. 13 for the method used to detect them), and as a result of these measurement the information about the phase characteristics of the two initial polarization modes is extracted. However, the precision with which all four Stokes parameters can be measured simultaneously is limited and is determined by the uncertainty relation between the fluctuation variances of the given quantities (see Sec. 2).

Another way to extract the information in the case of two modes is to conduct quantum nondemolition measurements (QNDM) for two phase-dependent Stokes parameters. We analyzed this problem for the first time in Ref. 14 for ''twisted'' optical fibers with tunnel coupling between the two propagating modes (see Refs. 15–17). The necessary condition for QNDM is the presence of polarization-squeezed light, which is generated under certain conditions in such fibers. For a given quantum polarization state of light, the fluctuation variance of one of the Stokes parameters is smaller than its value in a coherent state.[14,15] However, earlier schemes of QNDM of Stokes parameters based

on the interaction of two polarization modes are extremely difficult to implement in practice since, generally, repeated (or continuous) measurements of the observable must be carried out (see Refs. 18 and 19). In particular, to extract information about the measured Stokes parameter by this procedure, one must measure a large number of additional optical elements (linear and nonlinear), which reduces the effectiveness of QNDM in real experiments.

In this paper we propose new methods of QNDM of Stokes parameters, methods that can be employed together with continuous measurements fairly easily. Here we are interested in the basic aspects of the problem of such measurements, determined by general quantum mechanical laws, rather than in the technical limitations related to the sensitivity of the various devices or to the accuracy of detection methods, which can always be improved.

We also show that under certain conditions the given procedure of QNDM of Stokes parameters can be used in nondemolition measurements of the phase difference of the optical field. The plan of the paper is as follows.

In Sec. 2 we write the necessary introductory relations needed for the analysis of nonclassical polarization states of light in terms of Stokes parameters. The relationships between the phase-difference operators and these parameters are discussed in Sec. 3, and related problems and approaches to their solution (connected with the general phase problem in quantum optics) are analyzed in Sec. 5. Polarization QNDM for specific parameters of the vector optical field, the photon numbers and the Stokes parameters, are described in Secs. 3–5. The problem of using real media (cubically nonlinear tunnel-coupled optical fibers) for these purposes is discused in the Appendix.

2. NONCLASSICAL POLARIZATION STATES OF LIGHT IN TERMS OF STOKES PARAMETERS

A general quantum description of a two-mode (two orthogonal polarization components) light field can be achieved by using Hermitian Stokes parameter operators. The operators, which we denote by S_0, S_1, S_2, and S_3, can be expressed in terms of the photon creation and annihilation operators, $b_{1,2}^{\dagger}$ and $b_{1,2}$ in these two modes:[11–15]

$$\begin{aligned} &S_0=b_1^{\dagger}b_1+b_2^{\dagger}b_2, \quad S_1=b_1^{\dagger}b_1-b_2^{\dagger}b_2, \\ &S_2=b_2^{\dagger}b_1+b_1^{\dagger}b_2, \quad S_3=i(b_2^{\dagger}b_1-b_1^{\dagger}b_2), \end{aligned} \tag{1}$$

where the subscripts 1 and 2 in our polarization problem label modes with orthogonal linear (along the x and y axes) or circular polarizations (the details are given below). The operators b_1, $b_1^{\dagger}$, b_2, and $b_2^{\dagger}$ satisfy the following commutation relation, which holds for a bosonic system:

$$[b_i,b_j^{\dagger}]=\delta_{ij}, \quad [b_i,b_j]=0, \quad i,j=1,2, \tag{2}$$

where δ_{ij} is the Kronecker delta.

Thus, S_0 and S_1 are the operators of the sum and difference of photon numbers, $N_{\pm}\equiv N_1\pm N_2$, where $N_{1,2}\equiv b_{1,2}^{\dagger}b_{1,2}$ for each mode. The phase-dependent Stokes operators S_2 and S_3 are similar to the Hermitian quadrature components for the light-field amplitudes (see Ref. 15).

The formation and observation of nonclassical light states and QNDM of the quantities N_+ and N_- and Hermitian quadratures of two-mode fields (including optical fibers of a special type) are considered in Refs. 20 and 21.

Definition of the Stokes parameter operators in the form (1) with allowance for (2) leads directly to commutation relations characteristic of operators in the SU(2) algebra[12]:

$$\begin{aligned} &[S_1,S_2]=2iS_3, \quad [S_3,S_1]=2iS_2, \\ &[S_2,S_3]=2iS_1, \quad [S_0,S_i]=0. \end{aligned} \tag{3}$$

All this leads to uncertainty relations for the Stokes parameters:

$$\langle\Delta S_i^2\rangle\langle\Delta S_k^2\rangle\geqslant|\langle S_m\rangle|^2, \quad i,k,m=1,2,3, \quad i\neq k\neq m, \tag{4}$$

where $\langle\Delta S_i^2\rangle\equiv\langle S_i^2\rangle-\langle S_i\rangle^2$ $(i=1,2,3)$ are the fluctuation variances of the Stokes parameters. Thus, in quantum optics there is, in principle, no way in which all Stokes parameters can be measured exactly and simultaneously (with the exception of the trivial case in which all the parameters are zeros).

We define the vector polarization operator of a quantum state of the light field to be

$$\mathbf{p}=\mathbf{e}_1S_1+\mathbf{e}_2S_2+\mathbf{e}_3S_3, \tag{5}$$

where the $\mathbf{e}_j$ $(j=1,2,3)$ are orthogonal unit vectors. Clearly, the operator $\mathbf{p}^2=S_1^2+S_2^2+S_3^2$ and the Stokes parameters (1) commute, i.e.,

$$[\mathbf{p}^2,S_j]=0, \quad j=1,2,3. \tag{6}$$

The vector $\mathbf{p}$ in our description of the quantum state of light polarization plays the same role as the angular momentum vector $\mathbf{J}$ in quantum mechanics, so that by analogy we can call $\mathbf{p}^2$ the Casimir operator. Geometrically, the vector $\mathbf{p}$ determines the polarization state of light on the Poincaré sphere with coordinates S_1, S_2, and S_3 (see also Refs. 14 and 15).

If the operators $b_{1,2}$ correspond to the light field in a coherent state, the fluctuation variances of all Stokes parameters are the same, i.e.,

$$\langle\Delta S_0^2\rangle=\langle\Delta S_1^2\rangle=\langle\Delta S_2^2\rangle=\langle\Delta S_3^2\rangle=\langle N_1\rangle+\langle N_2\rangle, \tag{7}$$

where the averaging is done over the vector of a coherent quantum state of the two-mode field,

$$|\xi\rangle=|\beta_1\rangle|\beta_2\rangle. \tag{8}$$

Here $|\beta_{1,2}\rangle$ are eigenfunctions of the operators $b_{1,2}$ with eigenvalues $\beta_{1,2}$, respectively; $\langle N_{1,2}\rangle\equiv|\beta_{1,2}|^2$ are mean photon numbers in these modes.

For polarization-squeezed light, the fluctuation variance $\langle\Delta S_i^2\rangle$ of one of the Stokes parameters is always less than its value (7) corresponding to a coherent state, i.e., fluctuations are suppressed in comparison to their level for coherent light. Generation of such strictly quantum polarization states (see Refs. 14 and 15) is possible in special optical fibers (e.g., in cubically nonlinear twisted birefringent fibers) in the presence of both linear and nonlinear energy exchange between two orthogonally polarized modes (along the x and y axes, respectively).[16,17]

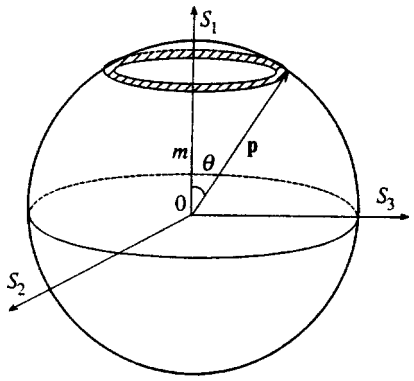


FIG. 1. Geometric representation of stationary polarization states of light on the Poincaré sphere; **p** is the polarization vector, and m is its projection on the S_1 axis. The angle between the S_1 axis and **p** is θ; the shaded area characterizes the quantum uncertainty of the Stokes parameters S_2 and S_3.

Let us determine the stationary polarization states of a light field, which are the eigenstates of the Casimir operator $\mathbf{p}^2$ and, simultaneously, of one of the three Stokes parameters S_j $(j=1,2,3)$:

$$\mathbf{p}^2|sm\rangle=\lambda_s|sm\rangle, \quad S_j|sm\rangle=m|sm\rangle, \quad j=1,2,3, \tag{9}$$

where λ_s and m are the eigenvectors of the operators $\mathbf{p}^2$ and S_j, respectively.

Equations (9) constitute an eigenvalue problem in the SU(2) algebra. The solutions of this problem are well known from quantum angular momentum theory.[22,23]

Let us examine one of the possible variants of its solution, when $j=1$. In this case the final expressions for the matrix elements of the Stokes operators in the representation in which the operators $\mathbf{p}^2$ and S_1 are diagonal are

$$\langle sm|\mathbf{p}^2|s'm'\rangle=s(s+2)\,\delta_{ss'}\delta_{mm'}, \tag{10a}$$

$$\langle sm|S_1|s'm'\rangle=m'\,\delta_{mm'}\delta_{ss'}, \tag{10b}$$

$$\langle sm|S_2|s'm'\rangle=0.5\sqrt{s(s+2)-m'(m'\pm 2)}\times\delta_{ss'}\delta_{mm'\pm 2}, \tag{10c}$$

$$\langle sm|S_3|s'm'\rangle=\mp 0.5i\sqrt{s(s+2)-m'(m'\pm 2)}\times\delta_{ss'}\delta_{mm'\pm 2}, \tag{10d}$$

where s is related to the eigenvalue λ_s by

$$\lambda_s=s(s+2) \tag{11}$$

and the quantum number m takes $2s+1$ discrete values: $-s, -s+1, \ldots, 0, \ldots, s-1, s$.

Using Eqs. (10), we can easily obtain relations for the fluctuation variances $\langle(\Delta\mathbf{p})^2\rangle=\langle\mathbf{p}^2\rangle-\langle\mathbf{p}\rangle^2$ and $\langle\Delta S_1^2\rangle$:

$$\langle\Delta S_1^2\rangle=0, \tag{12a}$$

$$\langle(\Delta\mathbf{p})^2\rangle=\langle\Delta S_2^2\rangle+\langle\Delta S_3^2\rangle=s(s+2)-m^2. \tag{12b}$$

Geometrically, the states (9)–(12) can be represented by the vector **p**, of length $p=\sqrt{s(s+1)}$, which rotates about the S_1 axis(see Fig. 1). Here the projection of the vector on the S_1 axis is equal to m.

The rotation angle θ of the polarization vector can be defined to be

$$\cos\,\theta=\frac{m}{\sqrt{s(s+2)}}. \tag{13}$$

For the maximum values $m=\pm s$, when the variance $(\Delta\mathbf{p})^2=2s$ is at its minimum (see (12b)), the angle θ is nevertheless finite, i.e.,

$$\cos\,\theta=\frac{1}{\sqrt{1+2/s^2}}. \tag{14}$$

Only in the semiclassical limit, where $s\gg 1$ (see below), Eq. (14) shows that **p** is directed along the S_1 axis, i.e., $\cos\,\theta=1$.

Fluctuations of the Stokes parameters (Eqs. (12)) lead to the appearance of an ''uncertainty layer'' on the Poincaré sphere (see Fig. 1). This uncertainty layer is different for different quantum polarization states of light. In our case of a fixed value of S_1, the fluctuation variance of this Stokes parameter is exactly zero (see Eq. (12a)).

Such a pictorial geometrical interpretation of these states is similar to the vector model of angular momentum in quantum mechanics.[22]

In the case of Fock states, where the polarization modes $b_{1,2}$ are in states with fixed photon numbers, $|n_1\rangle$ and $|n_2\rangle$ (respectively), and the general vector of state (8) is

$$|\xi\rangle=|n_1\rangle|n_2\rangle, \tag{15}$$

we obtain, by combining (1), (5), and (15),

$$\langle\mathbf{p}^2\rangle\equiv\langle S_0(S_0+2)\rangle=n_0(n_0+2), \tag{16a}$$

$$\langle S_1\rangle=\langle N_1\rangle-\langle N_2\rangle, \tag{16b}$$

where $n_0\equiv\langle N_1\rangle+\langle N_2\rangle$.

The states $|n_{1,2}\rangle$ and, consequently, $|\xi\rangle$ are eigenstates of the operators $N_{1,2}$ and $\mathbf{p}^2$, respectively. Hence, comparing (16a) and (16b) with (10a) and (10b), we have (cf. Ref. 23b)

$$s=n_0, \quad m=\langle N_1\rangle-\langle N_2\rangle. \tag{17}$$

Thus, the quantum number s gives the sum of the photon numbers of the two polarization modes, and the number m gives their difference. This suggests that the Fock state vector (15) is a special case of the solution of the eigenvalue problem (9) for $j=1$.

Below we analyze the ways in which the nonclassical polarization states of light considered here can be formed.

3. QUANTUM NONDEMOLITION MEASUREMENTS OF THE STOKES PARAMETER S_1

According to the main concept of QNDM, a parameter of the light field (the signal wave) observed in the process of such measurements must interact in a certain way in the QNDM device with a parameter of the other field, the measuring field (probe wave); it is the latter quantity that is detected.[20,24,25]

To employ such measurements, we assume that the QNDM device mixes four optical modes; the annihilation operators for two of these modes are denoted by $a_{1,2}$, and they belong (as before) to the measuring quantities; the other pair of operators, $b_{1,2}$, describes the measured quantities (see

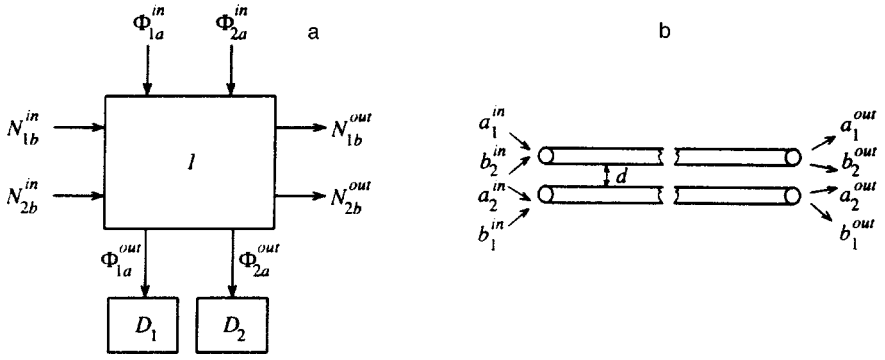


FIG. 2. a) Block diagram of parallel QNDM of the photon numbers $N_{1b}^{\rm in}$ and $N_{2b}^{\rm in}$: *1* is a QNDM device, which mixes the four modes $a_{1,2}$ (probe radiation) and $b_{1,2}$ (signal radiation), with the subscripts 1 and 2 indicating the polarization components; $D_{1,2}$ are the detectors of the phases $\Phi_{1,2a}^{\rm out}$ of the probe modes; and the superscripts in and out indicate the input and output modes, respectively. b) One of the possible realizations of a QNDM device based on cubically nonlinear tunnel-coupled optical fibers. We show two waveguides with tunnel coupling of the filaments (the distance d between the filaments is 1–10 μm).

Fig. 2). This sets the proposed QNDM scheme apart from those considered earlier for the two-mode problem (cf. Ref. 14). In the Appendix we discuss a specific implementation of such four-mode interaction in optical fibers of a special type, tunnel-coupled optical fibers.

We start by examining parallel nondemolition measurements of photon numbers in different optical-fiber filaments (cf. Refs. 26 and 27). Earlier (see Ref. 20) we discussed parallel nondemolition measurements, but for Hermitian quadratures. Figure 2 is a block diagram of measurements of the photon numbers in the signal modes, described by the operators $N_{1,2b}^{\rm in}=(b_{1,2}^{\rm in})^{+}b_{1,2}^{\rm in}$. Two detectors, D_1 and D_2, register the phases of the probe polarization modes, $\Phi_{1a}^{\rm out}$ and $\Phi_{2a}^{\rm out}$, described by the annihilation operators a_1 and a_2, respectively, at the output of the QNDM device (optical fiber).

The necessary relationships for the conversion of the measured photon number and the phase of the probe wave in the QNDM device can be obtained by employing the equations (A8) of the Appendix:

$$N_{1b}^{\rm out}=N_{1b}^{\rm in}, \tag{18a}$$

$$\Phi_{2a}^{\rm out}=\Phi_{2a}^{\rm in}-\gamma N_{1b}^{\rm in}. \tag{18b}$$

Similarly, for QNDM of the photon number $N_{2b}^{\rm in}$ in the other optical-fiber filament we have

$$N_{2b}^{\rm out}=N_{2b}^{\rm in}, \tag{19a}$$

$$\Phi_{1a}^{\rm out}=\Phi_{1a}^{\rm in}-\gamma N_{2b}^{\rm in}, \tag{19b}$$

where $N_{1,2n}^{\rm in}$ and $N_{1,2a}^{\rm in}$ are the photon number operators in the measured and probe (measuring) modes; $\Phi_{1,2a}^{\rm in}(\Phi_{1,2a}^{\rm out})$ are the phase operators of the probe modes at the input (output) of the filaments, respectively; and the parameter γ characterizes the QNDM device and determines the efficiency of the wave interaction in tunnel-coupled optical fibers (see (A8)).

The physical meaning of the quantities $\Phi_{1,2a}$ is easily grasped in a classical setting: they lead to the ordinary phases of classical (complex-valued) amplitudes of the corresponding components of the light field. In fact, the operators $\Phi_{1,2a}$ are introduced within what is known as the heuristic approach to the problem of phase in quantum theory (see Ref. 28). This approach is usually used in describing the procedure of QNDM of the photon number[26,27] or for optical fields of complex configuration, e.g., solitons. In the latter case, meaningful introduction of phase operators in the well-established form (i.e., in the form of Susskind–Glogower operators) is hardly possible, while the representation in the form of the operators $\Phi_{1,2a}$ is simple and instructive in the semiclassical approximation:

$$\langle N_{1,2j}\rangle \gg 1, \quad j=a,b. \tag{20}$$

The operators $\Phi_{1,2a}$ satisfy the commutation relations (see Refs. 26 and 27)

$$[N_{1,2j},\Phi_{1,2j}]=i, \quad [\Phi_{1j},\Phi_{2j}]=0, \quad j=a,b. \tag{21}$$

Clearly, the operators (18) and (19) fully satisfy the criteria for implementation of QNDM. Indeed, on the one hand, the measured quantities N_{1b} and N_{2b} are conserved in the measurement process (see Eqs. (18a) and (19a)), and on the other, the measuring parameters (the probe mode phases) are linearly related only to the measured photon number of the other mode. We can therefore limit our investigation to QNDM in each filament of the tunnel-coupled optical fiber independently, i.e., the problem is identical to the one considered in Refs. 26 and 27: parallel QNDM (Fig. 2) reduce to two independent schemes of measuring and detecting the probe-mode phases in each filament with their own heterodyne fields.

We now analyze QNDM based on the polarization aspects of the interacting modes. In this case the measured quantities are the Stokes parameters S_{ib} $(i=0,1,2,3)$ of the signal modes, which are described by the operators $b_{1,2}$ and $b_{1,2}^{\dagger}$. This means that by detecting the Stokes parameters S_{ia} (or the corresponding phase difference) belonging to the probe (measuring) modes we gather information about a linear combination of the signal polarization modes rather than information about each such mode individually. Here, due to the uncertainty relations (4), in each specific case only one Stokes parameter S_{ia} $(i=1,2,3)$ can be measured in a nondemolition manner.

Let us begin with the procedure for nondemolition measurements of the Stokes parameter S_{1b} (Fig. 3).

Using the definition (1) and Eqs. (18) and (19), we can express the necessary relationships between the measured Stokes parameter $S_{1b}\equiv N_{1b}-N_{2b}$ and the detected phase difference $\Psi_a\equiv\Phi_{2a}-\Phi_{1a}$ (or the Stokes parameter S_{3a}) in the form

$$S_{1b}^{\rm out}=S_{1b}^{\rm in}, \tag{22a}$$

$$\Psi_a^{\rm out}=\Psi_a^{\rm in}-\gamma S_{1b}^{\rm in}. \tag{22b}$$

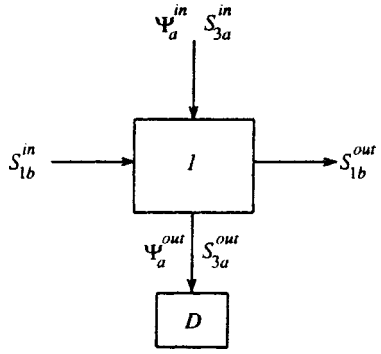


FIG. 3. Block diagram of QNDM of the Stokes parameter S_{1b}^{in}. The quantities S_{3a}^{in} (Ψ_a^{in}) and $S_{3a}^{\text{out}}(\Psi_a^{\text{out}})$ are the Stokes parameters (phase differences) of the probe modes at, respectively, the input and output of the tunnel-coupled optical fibers of the QNDM device *1*. Only the measured (signal) and probe Stokes parameters are shown; *D* is the detector.

Similar relationships exist for the conjugate quantities, i.e.,

$$S_{1a}^{\text{out}}=S_{1a}^{\text{in}}, \tag{23a}$$

$$\Psi_b^{\text{out}}=\Psi_b^{\text{in}}-\gamma S_{1a}^{\text{in}}, \tag{23b}$$

where $\Psi_{a,b}\equiv\Phi_{2a,b}-\Phi_{1a,b}$ is the phase difference between the probe and measured modes, respectively, at the input (or output) of the QNDM device.

Equations (22a) and (22b) satisfy (just as (18) and (19) did) the general criteria of realization of quantum nondemolition measurements for the Stokes parameters S_{1b} (see Refs. 14, 20, and 26). Indeed, first, this observable is conserved in the measuring process (see (22a)) and, second, the detected phase difference Ψ_a^{out} is a linear function of the measured Stokes parameter S_{1b}^{in} (see (22b)).

Note that the nonlinear correction to the initial phase difference (23b) associated with the Stokes parameter S_{1a}^{in} is precisely the quantity that determines the back action of the device on the corresponding quantity.

Let us examine the redistribution of quantum noise that takes place in the measuring process. To this end we use (22b) to find the observed value of $S_{1b}^{\text{obs}}\equiv\Psi_a^{\text{out}}/\gamma$.

For the variance of fluctuations of this quantity that emerges in the detection of the phase difference Ψ_a^{out} we have

$$\langle(\Delta S_{1b}^{\text{obs}})^2\rangle=\langle(\Delta S_{1b}^{\text{in}})^2\rangle+\frac{\langle(\Delta\Psi_a^{\text{in}})^2\rangle}{\gamma^2}. \tag{24}$$

As a result, the inaccuracy (error) of measurement, $\langle(\Delta S_{1b}^{\text{meas}})^2\rangle\equiv\langle(\Delta S_{1b}^{\text{obs}})^2\rangle-\langle(\Delta S_{1b}^{\text{in}})^2\rangle$, of the Stokes parameter S_{1b}^{in} (the difference between the photon numbers) has the form

$$\langle(\Delta S_{1b}^{\text{meas}})^2\rangle=\frac{\langle(\Delta\Psi_a^{\text{in}})^2\rangle}{\gamma^2}. \tag{25}$$

We see that the measurement inaccuracy $\langle(\Delta S_{1b}^{\text{meas}})^2\rangle$ in the QNDM considererd here is completely determined by the fluctuations in the phase difference between the polarization modes at the input of the QNDM device, and depends on the efficiency of the nonlinear phase buildup in tunnel-coupled optical fibers.

In the limit

$$2\bar{n}\langle(\Delta\Psi_a^{\text{in}})^2\rangle\ll 1, \tag{26a}$$

$$\bar{\gamma}\equiv 2\bar{n}\gamma\gg 1, \tag{26b}$$

with $\bar{n}$ the mean number of photons in the modes at the input of the QNDM device, we have $\langle(\Delta S_{1b}^{\text{meas}})^2\rangle\to 0$. Here

$$\langle(\Delta S_{1b}^{\text{obs}})^2\rangle\approx\langle(\Delta S_{1b}^{\text{out}})^2\rangle=\langle(\Delta S_{1b}^{\text{in}})^2\rangle, \tag{27}$$

i.e., we obtain the Stokes parameter S_{1b}^{in} at the input of the QNDM device, which has not been changed by the measuring process.

Thus, the condition (26a) means that the fluctuations of the phase difference Ψ_a^{in} between the probe modes must be suppressed, i.e., to employ QNDM we need polarization-squeezed light for the probe parameter S_{3a}^{in} at the output of the QNDM device (cf. Ref. 14). At the same time, an increase in the parameter $\bar{\gamma}$ which characterizes a QNDM device also reduces the inaccuracy of measurements (see (26b)). However, attaining large values of $\bar{\gamma}$ is extremely difficult from an experimental standpoint, since this means that the nonlinear interaction in the system must be highly efficient (see Appendix).

In the general case we must also examine the redistribution of noise in the other (conjugate) parameters (23) determining the given type of measurements. For the fluctuation variance of the phase difference Ψ_b (the conjugate of the measured Stokes parameter S_{1b}) Eq. (23b) yields

$$\langle(\Delta\Psi_b^{\text{out}})^2\rangle=\langle(\Delta\Psi_b^{\text{in}})^2\rangle+\gamma^2\langle(\Delta S_{1a}^{\text{in}})^2\rangle. \tag{28}$$

The corresponding uncertainty in the phase difference, $\langle(\Delta\Psi_b^{\text{imp}})^2\rangle=\langle(\Delta\Psi_b^{\text{out}})^2\rangle-\langle(\Delta\Psi_b^{\text{in}})^2\rangle$, introduced by the quantum nondemolition measurements of the Stokes parameter S_{1b}^{in} has the form

$$\langle(\Delta\Psi_b^{\text{imp}})^2\rangle=\gamma^2\langle(\Delta S_{1a}^{\text{in}})^2\rangle. \tag{29}$$

Hence the fluctuation variances (25) and (29) fully characterize the QNDM of the photon-number difference S_{1b}^{in} considered here.

If we allow for (25) and (29), we see that the product of these variances satisfies the inequality

$$\langle(\Delta S_{1b}^{\text{meas}})^2\rangle\langle(\Delta\Psi_b^{\text{imp}})^2\rangle=\langle(\Delta S_{1a}^{\text{in}})^2\rangle\langle(\Delta\Psi_a^{\text{in}})^2\rangle\geqslant 1. \tag{30}$$

Thus, the inaccuracy of measurements of the photon-number difference, $\langle(\Delta S_{1b}^{\text{meas}})^2\rangle$, and the phase-difference variance $\langle(\Delta\Psi_b^{\text{imp}})^2\rangle$ satisfy the corresponding uncertainty relation between the photon-number difference and the phase difference between two modes (cf. Ref. 9).

In the case of ''good'' QNDM, the lower bound in (30) is attained (see Refs. 26 and 27), i.e., the right-hand side of (30) becomes an equality. This occurs, for instance, if polarization-squeezed light for the probe parameter S_{3a}^{in}, which depends on the phase difference Ψ_a^{in}, is fed to the input of the QNDM device. Then the suppression of the fluctuations $\langle(\Delta\Psi_a^{\text{in}})^2\rangle$, on the one hand, leads to a reduction in the inaccuracy of measurements $\langle(\Delta S_{1b}^{\text{meas}})^2\rangle$ (see (25)) and, on the other, to an increase in $\langle\Delta S_{1a}^2\rangle$ due to the uncertainty relation (30). In turn, the latter factor leads in the measure-

ment process to an increase in the level of fluctuations for the phase difference, $\langle(\Delta\Psi_b^{\rm imp})^2\rangle$ (see (29)), and accordingly to a decrease in the inaccuracy of measurement, $\langle(\Delta S_{1b}^{\rm meas})^2\rangle$.

Thus, the QNDM of the Stokes parameter $S_{1b}^{\rm in}$ are done by redistributing the noise in the measuring process, i.e., the noise level in the conjugate quantity Ψ_b grows. However, to control this process in time, continuous nondemolition measurements must be employed.

Indeed, a characteristic feature of such QNDM is the possibility of periodically repeating the measurements.[18,24,26] Only as a result of a series of measurements can one judge the ''quality'' of the previous measurement.

In the schemes of QNDM of the Stokes parameters we considered earlier,[14] the implementation of recurrent measurements encounters experimental difficulties. On the other hand, in the problem of recurrent measurements considered here, such a procedure can easily be employed, the only requirement being that there must be negative feedback in the system (cf. Ref. 25).

In this case it also becomes possible to control the quantum fluctuations of the measured quantity in time, comparing each subsequent measurement with the previous one. Continuous QNDM of an observable constitute the limit of such recurrent measurements.[18] Here the quantity behaves classically with an ''uncertainty'' tube characterized by the value of $\langle(\Delta S_{1b}^{\rm meas})^2\rangle$.

However, such behavior within the given uncertainty region is strictly quantum, since fluctuations of the measured quantity at the output of the QNDM device are suppressed via the quantum interference of noise.

Note that negative feedback (without QNDM) employed, for example, in the electric circuit of the detection system makes it possible to alter the statistics of photons (within the negative feedback circuit) and of photoelectrons.[25] This result is due to the nontrivial role of negative feedback and of the amplitude (phase) modulator as a fundamental constituent of the system that produces the self-consistent field within the negative feedback loop. If to this system we add an element that performs QNDM in a certain optical parameter, the statistics can be ''shifted'' directly to the state of light for the given measured quantity, which state also becomes nonclassical.

Thus, by implementing negative feedback we can, after detecting the phase differences of the probe modes $\Psi_a^{\rm out}$ (or the corresponding Stokes parameter $S_{3a}^{\rm out}$), make continuous QND measurements of $S_{1b}^{\rm in}$. In particular, these measurement generate a new class of quantum states of light (see Sec. 2) that have a definite value of one of the Stokes parameters. The situation is similar to the case of Fock states in quantum optics, in which the number of photons is fixed, which is the limiting case of amplitude-squeezed light.[26]

4. QUANTUM NONDEMOLITION MEASUREMENTS OF THE STOKES PARAMETERS S_2 AND S_3

In this section we discuss the feasibility of QNDM of the phase-dependent Stokes parameters S_2 and S_3 (see Eqs. (1)).

Figure 4 presents a block diagram for the nondemolition measurement of the Stokes parameter S_{3c}. In this case, linear elements *2* and *3* (cf. Fig. 3) are added to the QNDM device. We first examine the role of the linear element *2* at the input of the QNDM device. This element performs preliminary linear conversion of the Stokes parameters for the measured modes (preparation of the quantum state). Standard linear optical elements, such as phase shifters and beam splitters, can be used for this purpose.

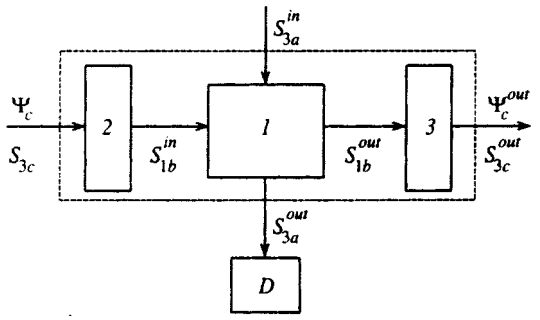


FIG. 4. Block diagram for QNDM of the Stokes parameter S_{3c} (phase difference Ψ_c). The quantities $S_{3a}^{\rm in}$ and $S_{3a}^{\rm out}$ denote the Stokes parameters of the probe modes at the input and output of the QNDM device, which consists of the linear elements *2* and *3* and the QNDM device *1*, respectively; $\Psi_c^{\rm out}$ is the phase difference after the measurements have been completed. Only the measured and probe Stokes parameters are shown; *D* is the detector.

As a result of such conversion of the measured modes, we have for the Stokes parameters at the input of the nonlinear system *1*

$$\begin{aligned} &S_{0b}^{\rm in}=S_{0c}, \quad S_{2b}^{\rm in}=S_{2c},\\ &S_{1b}^{\rm in}=S_{1c}\cos 2g-S_{3c}\sin 2g,\\ &S_{3b}^{\rm in}=S_{3c}\cos 2g+S_{1c}\sin 2g, \end{aligned} \tag{31}$$

where S_{jc} $(j=1,2,3)$ are the Stokes parameters at the input of the linear element *2*, and g is the linear mode conversion coefficient.

If for QNDM of the Stokes parameter S_{3c} we take $g=3\pi/4+\pi m$ $(m=0,1,2,\ldots)$, we have for the operators (31)

$$S_{0b}^{\rm in}=S_{0c}, \quad S_{1b}^{\rm in}=S_{3c}, \quad S_{2b}^{\rm in}=S_{2c}, \quad S_{3b}^{\rm in}=-S_{1c}. \tag{32}$$

Clearly, as a result of passage through the linear medium *2*, all information about the Stokes parameter S_{3c} is contained in $S_{1b}^{\rm in}$, the photon-number difference, which is then measured in a nondemolition manner.

Hence, after measurements in the nonlinear medium *1* have been completed, we must reinstate in a nondemolition manner the initial polarizations in the measured modes (which is the concept underlying QNDM). This is done in linear system *3* by a transformation of the Stokes parameters that is the inverse of (32).

As a result we have

$$\begin{aligned} &S_{0c}^{\rm out}=S_{0b}^{\rm out}=S_{0b}^{\rm in}=S_{0c}, \quad S_{3c}^{\rm out}=S_{1b}^{\rm out}=S_{1b}^{\rm in}=S_{3c},\\ &S_{2c}^{\rm out}=S_{2b}^{\rm out}, \quad S_{1c}^{\rm out}=-S_{3b}^{\rm out}, \end{aligned} \tag{33}$$

where S_{ic}^{out} are the Stokes parameters at the output of the QNDM device (see Fig. 4). Here the Stokes parameter S_{2c} does not change in the conversion process, and the measured quantity S_{3c} is conserved.

Thus, QND measurements of the Stokes parameter S_{3c} can be implemented by a proper selection of the conversion coefficient in the linear systems *2* and *3*. This constitutes the essence of the process of controlling the quantum system in the given case.

The parameter S_{2c} can be measured in a nondemolition manner in a similar way. However, in this case we must select the proper optical elements for the device *2* in Fig. 4 to ensure preliminary conversion of the Stokes parameters S_{1b} and S_{2b} similar to (31) and (32). Only then will the measured Stokes parameter S_{1b}^{in} contain complete information on S_{2c}.

5. QUANTUM NONDEMOLITION MEASUREMENTS OF THE PHASE DIFFERENCE

We consider QNDM of the phase difference (which for the sake of definiteness is denoted by Ψ_c) based on nondemolition measurements of the phase-dependent Stokes parameter S_{3c} (see Fig. 4). In the semiclassical approximation (20) we represent this Stokes parameter (see also Ref. 14) as:

$$S_{3c}=2\sqrt{N_{1c}}\sqrt{N_{2c}}\,\sin\Psi_c, \tag{34}$$

where $N_{1,2c}$ are photon-number operators for the polarization modes $c_{1,2}$ at the input of the QNDM device, and Ψ_c is the measured phase difference between these modes.

According to the above procedure of measuring the Stokes parameter S_{3c}, we have for this quantity at the input of the nonlinear medium (see also (32))

$$S_{1b}^{\text{in}}=S_{3c}=2\sqrt{N_{1c}}\sqrt{N_{2c}}\,\sin\Psi_c. \tag{35}$$

In this case the relationship between the detected phase difference Ψ_a^{out} (or the Stokes parameter S_{3a}^{out}) and the corresponding measured quantity (35) is (cf. (22b), (34))

$$\Psi_a^{\text{out}}=\Psi_a^{\text{in}}-2\gamma\sqrt{N_{1c}}\sqrt{N_{2c}}\,\sin\Psi_c, \tag{36a}$$

$$S_{3a}^{\text{out}}=2\sqrt{N_{1a}^{\text{in}}}\sqrt{N_{2a}^{\text{in}}}\,\sin\Psi_a^{\text{out}}. \tag{36b}$$

We see from (36) that the information about the measured phase difference Ψ_c is contained in the detected phase difference Ψ_a^{out} and hence in the Stokes parameter S_{3a}^{out}. Subsequent reasoning is similar to that in our analysis of QNDM of the Stokes parameter S_{1b}^{in} (see Sec. 4). Therefore, here we touch only on the specific features of QNDM of the phase difference Ψ_c.

Concerning these features, we note that the detected quantity Ψ_a^{out} also depends on the amplitudes (photon numbers) of the measured modes (see (36a)). This in turn means that there is additional amplitude noise, which destroys the process of measuring the phase difference Ψ_c. In the ideal case, the QNDM device must be tuned only to the measured phase difference (cf. Ref. 6).

Let us examine the conditions under which we can ignore the amplitude fluctuations of the measured modes. To this end we write the photon-number and phase-difference operators for the measured modes as (cf. Refs. 9 and 27)

$$N_{1,2c}=\langle N_{1,2c}+\Delta N_{1,2c}\rangle,\quad \Psi_c\equiv\langle\Psi_c\rangle+\Delta\Psi_c, \tag{37}$$

where angle brackets indicate averaging over the initial states of the operators of photon numbers and phase difference; $\Delta N_{1,2c}$ and $\Delta\Psi_c$ are the corresponding operator (small fluctuation) parts. When the number of photons in the modes is large (see (20)), we have for the fluctuations of the measured Stokes parameter S_{3c} (Eq. (34))

$$\Delta S_{3c}=\sum_{j=1,2}\left.\frac{\partial S_{3c}(N_{jc},\Psi_c)}{\partial N_{jc}}\right|_{\substack{N_{jc}=\langle N_{jc}\rangle\\ \Psi_c=\langle\Psi_c\rangle}}\Delta N_{jc}+\left.\frac{\partial S_{3c}(N_{jc},\Psi_c)}{\partial\Psi_c}\right|_{\substack{N_{jc}=\langle N_{jc}\rangle\\ \Psi_c=\langle\Psi_c\rangle}}\Delta\Psi_c. \tag{38}$$

As a result, if we allow for (34), the fluctuation variance of the detected phase difference (36a) is

$$\langle(\Delta\Psi_a^{\text{out}})^2\rangle=\langle(\Delta\Psi_a^{\text{in}})^2\rangle+4\gamma^2\langle N_{1c}\rangle\langle N_{2c}\rangle\cos^2(\langle\Psi_c\rangle)\times\langle(\Delta\Psi_c^2)^2\rangle+\gamma^2\left(\frac{\langle N_{2c}\rangle}{\langle N_{1c}\rangle}\langle\Delta N_{1c}^2\rangle+\frac{\langle N_{1c}\rangle}{\langle N_{2c}\rangle}\langle\Delta N_{2c}^2\rangle\right)\sin^2\langle\Psi_c\rangle. \tag{39}$$

Defining the observed value of the phase as $\Psi_c^{\text{obs}}=\Psi_a^{\text{out}}/\overline{\gamma}$, where $\overline{\gamma}\equiv 2\gamma(\langle N_{1c}\rangle\langle N_{2c}\rangle)^{1/2}\cos\langle\Psi_c\rangle$ is the nonlinear parameter characterizing the measurement efficiency, we find from Eq. (39) that the inaccuracy of phase-difference measurements, $\langle(\Delta\Psi_c^{\text{meas}})^2\rangle=\langle(\Delta\Psi_c^{\text{obs}})^2\rangle-\langle(\Delta\Psi_c^2)\rangle$ (at $\langle\Psi_c\rangle\neq\pi/2+\pi m$, $m=0,1,2,\ldots$), is

$$\langle(\Delta\Psi_c^{\text{meas}})^2\rangle=0.25\left[\frac{\langle\Delta N_{1c}^2\rangle}{\langle N_{1c}\rangle^2}+\frac{\langle\Delta N_{2c}^2\rangle}{\langle N_{2c}\rangle^2}\right]\tan^2\langle\Psi_c\rangle+\frac{\langle(\Delta\Psi_a^{\text{in}})^2\rangle}{\overline{\gamma}^2}. \tag{40}$$

This implies that in the general case, where $\langle\Psi_c\rangle\neq\pi m$, $m=0,1,2,\ldots$, the amplitude noise of the modes (the two terms in square brackets) has an additional destructive effect on the measuring process.

To be able to ignore this destructive effect, we assume that

$$\langle\Delta N_{1,2c}^2\rangle\ll\langle N_{1,2c}\rangle. \tag{41}$$

Physically, this condition means that there is amplitude-squeezed (in photon numbers) light in the measured modes at the input of the linear element *2* in Fig. 4.

Thus, QND measurements of the phase difference Ψ_c are identical to those of Fig. 4 used in measuring S_{3c} under conditions in which the fluctuations of the number of photons in the corresponding modes are suppressed (cf. Ref. 14).

Note that the measured phase difference Ψ_c (see (36a)) is shifted by $\pi/2$ relative to its fluctuations (see the second term on the right-hand side of Eq. (39)). This fact has been noted in Ref. 5. Of course, the shift must be taken into account in the measurements, because otherwise the information about the measured phase difference might be lost.

6. CONCLUSION

We have shown that continuous QND measurements of the Stokes parameter of a light field and the phase difference between the initial polarization modes are possible. We believe that the nondemolition measurements considered here are promising both from a fundamental standpoint and for practical applications. We now list the most important aspects of this approach.

First, precision measurements of vector light fields may prove useful in areas where light polarization is important. In particular, we would like to mention the possibility of observing quantum polarization instabilities and chaos.[9,17] At the same time, such measurements could be used in problems of quantum cryptography, where light polarization plays an important role in exchanging information between the corresponding components (cf. Ref. 29).

Second, the nonclassical polarization states of light predicted in the present paper and the polarization-squeezed light we considered earlier in Ref. 15 may play an important role in various phase-sensitive interference measurements of the highest possible sensitivity.[13,30,31]

Third, there exists a deep analogy between the Stokes parameters (1) and angular momenta in quantum mechanics, which are described by the formalism of the SU(2) algebra on the basis of the commutation relations (3). This analogy may bring to light entirely new experiments in atomic physics based on the general principles of QNDM of Stokes parameters examined in this paper. Indeed, as shown in Ref. 14b, QND measurements of the components of the angular momentum of an atomic system are possible in a modified Stern–Gerlach experiment, where a beam of atoms propagating in the space between two magnets interacts with parametric photons, which are used to form squeezed states for the probe component of the angular momentum (see also Ref. 32). This measuring scheme is identical to the QNDM procedure for Stokes parameters in the absence of negative feedback.[14] Combined methods of QNDM of angular (spin) momenta of atoms interacting with a probe optical field are also possible (see, e.g., Ref. 33).

Fourth, the approach discussed in this paper can be used directly to describe the polarization characteristics of boson-like atoms in quantum theory.[34] In this case the Stokes parameters determine the polarization properties of the atomic system proper.

Let us briefly discuss the possibility of experimentally implementing the quantum nondemolition measurements proposed in this paper. Most suitable for this purpose are the high-stability single-mode chip lasers with semiconductor pumps.[35] The latter fact is important for introducing negative feedback in the measuring channel (cf. Refs. 25 and 26). Tunnel-coupled optical fibers can be used to mix the four modes (see the Appendix and Ref. 16). In this respect InSb-based waveguides, which allow using moderate pumping powers due to the large nonlinearity of the waveguide material, have great potential. A similar system consists of $GaAs/Ga_{0.3}Al_{0.7}As$-based optical fibers with a nonlinearity mechanism (a cubic nonlinearity of order 10^{-4} esu) based on quantum wells (an MQW structure). In this case, for an optical fiber roughly 1.5-mm long and the pump laser emitting 0.9-μm light, the necessary effective phase buildup $\overline{\gamma \approx 5}$ can be attained at intensities I of roughly 7.1 kW cm^{-2}.

In conclusion we note that in the four-mode problem with optical fibers considered here, polarization-squeezed states of light for the Stokes parameters S_{jb} and S_{ja} can be formed only if we allow for linear energy exchange between the modes (see Eqs. (A1) and (A4)). Indeed, in this case redistribution of energy and hence redistribution of fluctuations between the Stokes parameters become possible, as is the case in the two-mode problem.[15]

This work was supported by the Russian Fund for Fundamental Research.

APPENDIX. FOUR-MODE INTERACTIONS IN NONLINEAR TUNNEL-COUPLED OPTICAL FIBERS

Here we discuss the problem of QNDM in a tunnel-coupled optical fiber with a nonlinearity of the Kerr type.

The propagation of classical fields of distinct frequencies ω and ν in cubically nonlinear tunnel-coupled optical fibers can be described by the following equations for two (different polarization) complex-valued amplitudes $A_{1,2}$ and $B_{1,2}$ in each fiber (subscripts 1 and 2), respectively (see Ref. 16):

$$
\begin{aligned}
2i\beta\frac{c}{\omega}\frac{dA_1}{dz} &= -K_{12}^{\omega}\exp(i\varphi_a)A_2-(\kappa_{\omega 1}|A_1|^2 \\
&\quad +2\kappa_{\omega\nu 1}|B_2|^2)A_1, \\
2i\beta\frac{c}{\omega}\frac{dA_2}{dz} &= -K_{21}^{\omega}\exp(-i\varphi_a)A_1-(\kappa_{\omega 2}|A_2|^2 \\
&\quad +2\kappa_{\omega\nu 2}|B_1|^2)A_2, \\
2i\beta\frac{c}{\nu}\frac{dB_1}{dz} &= -K_{21}^{\nu}\exp(-i\varphi_b)B_2-(\kappa_{\nu 2}|B_1|^2 \\
&\quad +2\kappa_{\nu\omega 2}|A_2|^2)B_1, \\
2i\beta\frac{c}{\nu}\frac{dB_2}{dz} &= -K_{12}^{\nu}\exp(i\varphi_b)B_1-(\kappa_{\nu 1}|B_2|^2 \\
&\quad +2\kappa_{\nu\omega 1}|A_1|^2)B_2,
\end{aligned}
\tag{A1}
$$

where $K_{12}^{\omega,\nu}$ and $K_{21}^{\omega,\nu}$ are the linear mode-coupling coefficients at the corresponding frequencies; $\varphi_a=\alpha_\omega\omega z/c$ and $\varphi_b=\alpha_\nu\nu z/c$ are the mode phases determined by the difference of the effective refractive indices $\beta_{\omega,\nu;j}$ $(j=1,2)$, i.e., $\alpha_\omega-\beta_{\omega 2}-\beta_{\omega 1}$ and $\alpha_\nu=\beta_{\nu 2}-\beta_{\nu 1}$; $\beta=(\beta_{\omega 1}+\beta_{\omega 2}+\beta_{\nu 1}+\beta_{\nu 2})/4$; c is the speed of light in vacuum; and $\kappa_{\omega j}$, $\kappa_{\nu j}$, $\kappa_{\omega\nu;j}$, $\kappa_{\nu\omega;j}\propto\chi^{(3)}$ $(j=1,2)$ are the nonlinear coefficients of the material of the optical fibers. In contrast to the case that we considered earlier,[14] here we ignore linear energy exchange between the modes, i.e., we assume that the corresponding mode-coupling coefficients are zero, i.e., $K_{12}^{\omega}\approx K_{12}^{\nu}\approx K_{21}^{\omega}\approx K_{21}^{\nu}\approx K_{21}^{\nu}=0$. We also assume that $\omega\approx\nu$. From an experimental standpoint, the latter condition means that the modes A_j and B_j have almost equal frequencies, but each pair (A and B) differ in polarization (e.g., the modes can be circularly polarized[16]).

Thus, the system of equations (A1) in this approximation has the form

$$i\frac{dA_{1,2}}{dz}=-(\kappa|A_{1,2}|^2+2\kappa_{ab}|B_{2,1}|^2)A_{1,2},$$

$$i\frac{dB_{1,2}}{dz}=-(\kappa|B_{1,2}|^2+2\kappa_{ab}|A_{2,1}|^2)B_{1,2}, \qquad \text{(A2)}$$

where $\kappa\equiv\kappa_{\omega 1}\omega/2\beta c\approx\kappa_{\omega 2}\omega/2\beta c\approx\kappa_{\nu 1}\nu/2\beta c\approx\kappa_{\nu 2}\nu/2\beta c$ and $\kappa_{ab}\equiv\kappa_{\omega\nu 1}\omega/2\beta c\approx\kappa_{\omega\nu 2}\omega/2\beta c\approx\kappa_{\nu\omega 1}\nu/2\beta c\approx\kappa_{\nu\omega 2}\nu/2\beta c$.

When we go over to a quantum description, we must replace the complex-valued classical amplitudes A_j and B_j in Eqs. (A2) with operators according to the standard procedure (cf. Ref. 14), i.e.,

$$A_j\to i\tilde{C}a_j,\quad A_j^*\to -i\tilde{C}a_j^\dagger,$$

$$B_j\to i\tilde{C}b_j,\quad B_j^*\to -i\tilde{C}b_j^\dagger,\quad j=1,2, \qquad \text{(A3)}$$

where $\tilde{C}=(2\pi\hbar\omega/\varepsilon_0 V)^{1/2}$, with V the quantization volume. As a result we obtain at the quantum equations of motion for the annihilation operators of the two orthogonally polarized modes a_j and b_j:

$$i\frac{da_{1,2}}{dz}=-(\bar{\kappa}a_{1,2}^\dagger a_{1,2}+2\bar{\kappa}_{ab}b_{2,1}^\dagger b_{2,1})a_{1,2},$$

$$i\frac{db_{1,2}}{dz}=-(\bar{\kappa}b_{1,2}^\dagger b_{1,2}+2\bar{\kappa}_{ab}a_{2,1}^\dagger a_{2,1})b_{1,2}, \qquad \text{(A4)}$$

where $\bar{\kappa}\tilde{C}^2\kappa$, and $\bar{\kappa}_{ab}\equiv\tilde{C}^2\kappa_{ab}$.

The operators a_j and b_j satisfy the usual commutation relations for a bosonic system:

$$[a_j,a_k^\dagger]=[b_j,b_k^\dagger]=\delta_{jk},\quad [a_j,b_k^\dagger]=[a_j,b_k]=0, \qquad \text{(A5)}$$

where $j,k=1,2$.

Equations (A4) can also be obtained from the equation of evolution of the operators a_j and b_j in the Heisenberg picture:

$$i\hbar\frac{da_j}{dt}=[a_j,H_{\rm int}],\quad i\hbar\frac{db_j}{dt}=[b_j,H_{\rm int}], \qquad \text{(A6)}$$

where the interaction Hamiltonian

$$H_{\rm int}=\frac{\hbar c}{2n_1}[\bar{\kappa}(a_1^{\dagger 2}a_1^2+a_2^{\dagger 2}a_2^2+b_1^{\dagger 2}b_1^2+b_2^{\dagger 2}b_2^2)+4\bar{\kappa}_{ab}(a_1^\dagger a_1 b_2^\dagger b_2+b_1^\dagger b_1 a_2^\dagger a_2)] \qquad \text{(A7)}$$

with $d/dt\to-(c/n_1)d/dz$ (n_1 is the linear refractive index).

Thus, the system of equations (A4) describes the self-action and cross-interaction between the four modes in a cubically nonlinear tunnel-coupled optical fiber. In the general case, self-action has a destructive effect on the type of QNDM considered here (cf. Refs. 26 and 27). In particular, the terms with self-action are present in (18), (19), (22), and (23), which leads to additional phase fluctuations for the measuring (probe) and measured modes.

To reduce the effect of self-action, specially prepared optical fibers with $\bar{\kappa}\ll\bar{\kappa}_{ab}$ can be used. However, in standard tunnel-coupled optical fibers the anisotropy of cubic susceptibility is, for all practical purposes, low.[16] It is also possible to place in front of the QNDM device an additional cubically nonlinear medium, with the sign of the nonlinear susceptibility, however, being opposite that of the susceptibility of the given fiber (cf. Refs. 20 and 26).

In the ideal case, with $\bar{\kappa}=0$, the solution of the system of equations (A4) is

$$a_{1,2}^{\rm out}=(\exp\{-i\gamma(b_{2,1}^{\rm in})^\dagger b_{2,1}^{\rm in}\})a_{1,2}^{\rm in},$$

$$b_{1,2}^{\rm out}=(\exp\{-i\gamma(a_{2,1}^{\rm in})^\dagger a_{2,1}^{\rm in}\})b_{1,2}^{\rm in}, \qquad \text{(A8)}$$

where $\gamma\equiv-2\bar{\kappa}_{ab}l$ (here l is the length of the tunnel-coupled optical fiber), and $a_{1,2}^{\rm in}(a_{1,2}^{\rm out})$ and $b_{1,2}^{\rm in}(b_{1,2}^{\rm out})$ are the operators of the probe and measured modes, respectively, at the input (output) of the fiber, operators that satisfy the commutation relations (A5) (see also Fig. 2).

Thus, using (A8), we can easily obtain the transformations (18) and (19) needed to implement quantum nondemolition measurements.

*)E-mail: wlad%rtf@vpti.vladimir.su

[1] Physica Scripta, Special issue T **48** (1993).
[2] D. T. Pegg, S. M. Barnett, and J. A. Voccaro, in *Quantum Optics V*, J. D. Harvey and D. F. Walls (eds.), Springer, Berlin (1989), p. 122.
[3] R. Tanaś, A. Miranowicz, and Ts. Gantsog, in *Progress in Optics*, E. Wolf (ed.), Vol. 35, North-Holland, Amsterdam (1996), p. 355.
[4] V. P. Belavkin and C. Bendjaballah, Quantum Opt. **6**, 169 (1994).
[5] J. W. Noh, A. Fougères, and L. Mandel, Phys. Rev. Lett. **67**, 1426 (1991); Phys. Rev. A **45**, 424 (1992); **46**, 2840 (1992).
[6] V. B. Braginsky, F. Ya. Khalili, and A. A. Kulaga, Phys. Lett. A **202**, 1 (1995).
[7] D. T. Smithey, M. Beck, J. Cooper, and M. G. Raymer, Phys. Rev. A **48**, 3159 (1993).
[8] T. I. Kuznetsova, Usp. Fiz. Nauk **154**, 677 (1988) [Sov. Phys. Usp. **31**, 364 (1988)].
[9] (a) A. P. Alodzhants and S. M. Arakelyan, Zh. Éksp. Teor. Fiz. **107**, 1792 (1995) [JETP **80**, 995 (1995)]; (b) A. P. Alodzhants and S. M. Arakelian, Laser Phys. **5**, 812 (1995).
[10] M. Born and E. Wolf, *Principles of Optics*, 6th ed., Pergamon Press, Oxford (1980).
[11] F. A. Kaempffer, *Concepts in Quantum Mechanics*, Academic Press, New York (1965).
[12] V. P. Karassiov and V. I. Puzyrevsky, J. Sov. Res. **10**, 229 (1989); V. P. Karassiov, E-print:quant-ph/9503011 (1995).
[13] A. P. Alodzhants and S. M. Arakelian, Proc. SPIE (1997), Vol. 3076, p. 97.
[14] (a) A. P. Alodzhants and S. M. Arakelyan, Izv. Ross. Akad. Nauk, Ser. Fiz. **59** No. 6, 62 (1995) [Izv. Ross. Akad. Nauk, Ser. Fiz. **59**, 977 (1995)]; A. P. Alodjants, S. M. Arakelian, and A. S. Chirkin, Quantum Semiclassic. Opt. **9**, 311 (1997); (b) A. P. Alodjants, S. M. Arakelian, and A. S. Chirkin, Appl. Phys. B: Photophys. Laser Chem. **66**, 53 (1988).
[15] A. P. Alodzhants, S. M. Arakelyan, and A. S. Chirkin, Zh. Éksp. Teor. Fiz. **108**, 63 (1995) [JETP **81**, 34 (1995)]; A. S. Chirkin, A. P. Alodzhants, and S. M. Arakelyan, Opt. Spektrosk. **82**, 1001 (1997) [Opt. Spectrosc. **82**, 919 (1997)].
[16] A. A. Maĭer, Usp. Fiz. Nauk **165**, 1037 (1995) [Phys. Usp. **38**, 991 (1995)].
[17] S. F. Feldman, D. A. Weinberger, and H. G. Winful, J. Opt. Soc. Am. B **10**, 1191 (1993).
[18] M. B. Mensky, *Continuous Quantum Measurements and Path Integrals*, IOPP, Bristol (1993).
[19] M. Ban, Phys. Rev. A **49**, 5078 (1994).
[20] A. P. Alodjants and S. M. Arakelian, Laser Phys. **4**, 765 (1994).
[21] C. Fabre, E. Giacobino, A. Heidmann *et al.*, Quantum Opt. **2**, 159 (1990); P. J. Horowics, Europhys. Lett. **10**, 537 (1989).

[22] R. N. Zare, *Angular Momentum. Understanding Spatial Aspects in Chemistry and Physics*, Wiley, New York (1988).
[23] (a) C. Brif, E-print: quant-ph/9701003 (1997); (b) D. N. Klyshko, Zh. Eksp. Teor. Fiz. **111**, 1955 (1997) [JETP **84**, 1065 (1997)].
[24] V. B. Braginsky and F. Ya. Khalili, *Quantum Measurements*, Cambridge Univ. Press, London (1992); Yu. I. Vorontsov, Usp. Fiz. Nauk **164**, 89 (1994) [Phys. Usp. **37**, 81 (1994)].
[25] D. N. Klyshko and A. V. Masalov, Usp. Fiz. Nauk **165**, 1249 (1995) [Phys. Usp. **38**, 1203 (1995)].
[26] N. Imoto, H. A. Haus, and Y. Yamamoto, Phys. Rev. A **32**, 2287 (1985); Y. Yamamoto, N. Imoto, and Y. Machida, Phys. Rev. A **33**, 3243 (1986); S. Watkins and Y. Sasaki, Opt. Commun. **61**, 159 (1987).
[27] H. A. Haus, R. Watanabe, and Y. Yamamoto, J. Opt. Soc. Am. B **6**, 1138 (1989).
[28] H. A. Haken, *Laser Theory*, Vol. 25/2c of *Encyclopedia of Physics*, S. Flügge (ed.), Springer, Berlin (1970), pp. 85–87.
[29] M. J. Werner and G. J. Milburn, Phys. Rev. A **47**, 639 (1993); C. H. Bennett, F. Bessette, G. Brassard *et al.*, J. Cryptology **5**, 3 (1992).
[30] B. Yurke, S. L. McCall, and J. R. Klauder, Phys. Rev. A **33**, 4033 (1986); C. Brif, and A. Mann, Phys. Rev. A **54**, 4505 (1996).
[31] B. Yurke, Phys. Rev. Lett. **56**, 1515 (1986).
[32] G. S. Agarwal and R. R. Puri, Phys. Rev. A **49**, 4968 (1994); J. P. Dowling, G. S. Agarwal, and W. P. Schleich, Phys. Rev. A **49**, 4101 (1994).
[33] J. L. Sorensen, J. Jorgensen, and J. Erland, and E. S. Polzik, Quantum Semiclassic. Opt. **9**, 239 (1997); A. Kuzmich, K. Mølmer, and E. S. Polzik, Phys. Rev. Lett. **79**, 4782 (1997).
[34] G. Lenz, P. Meystre, and E. M. Wright, Phys. Rev. A **50**, 1681 (1994).
[35] N. V. Kravtsov and O. E. Naniĭ, Kvant. Elektron. (Moscow) **20**, 322 (1993) [Quantum Electron. **23**, 272 (1993)].

Translated by Eugene Yankovsky

Stability of strongly localized excitations in discrete media with cubic nonlinearity

S. Darmanyan and A. Kobyakov

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia

F. Lederer

Institute of Solid State Theory and Theoretical Optics, Friedrich-Schiller-Universität Jena, 07743 Jena, Germany

(Submitted 6 August 1997)
Zh. Éksp. Teor. Fiz. **113**, 1253–1260 (April 1998)

By using a linear analysis it is analytically shown that the stability of strongly localized modes depends on their symmetry, the sign of nonlinearity, and the degree of localization. The existence of a stable, bright, even mode of the discrete nonlinear Schrödinger equation is demonstrated and confirmed by direct numerical simulations. Possible applications to all-optical switching are discussed. © *1998 American Institute of Physics.* [S1063-7761(98)00904-4]

In the past decade many investigations have been devoted to intrinsic localized modes in discrete nonlinear systems due to their relevance to different branches of science, e.g., solid state physics, nonlinear optics, and biology. The fundamental properties of localized structures were used to explain some thermodynamic effects in solids (e.g., nonexponential energy relaxation), polaron and defect dynamics in anharmonic lattices and quantum crystals (see Refs. 1–6 and the bibliography cited there). Many physical phenomena such as modulational instability of plane waves,[7,8] formation and stability of temporal solitons,[9,10] and the recurrence effect[11] occur in discrete systems in a quite different way compared to those in extensively studied continuum systems. The discreteness of the medium is responsible for new physical effects that could not be forecast in studying the continuum model. Some of the theoretically predicted properties of discrete systems, in particular, modulational instability of plane waves, existence and dynamics of bright and dark localized states, have already been verified experimentally.[12,13]

In many cases the evolution of the initial excitation may be described by the discrete nonlinear Schrödinger equation (DNLSE), which is one of the fundamental equations in nonlinear physics. For instance, it governs electron-phonon interaction in a one-dimensional ionic crystal or mediates nonlinear processes in biology, where it is called a discrete self-trapping equation.[1] Another spectacular example is the evolution of the electromagnetic field in an array of linearly coupled waveguides, which have a great potential in applications for performing all-optical switching, steering, and demultiplexing. The use of such waveguide arrays for power- and phase-controlled, all-optical information processing was discussed in many papers (see Refs. 14 and 15 and the bibliography cited there). However, from the point of view of obtaining a practical device the number of excited channels in the array should be minimized. Fortunately, discrete systems are able to support the so-called strongly localized modes (SLMs), which contain only a few excited components and hence exactly suit the above-mentioned criterion. In contrast to an inhomogeneous discrete system, this intrinsic localization is a pure nonlinear effect which appears to be very promising in optical information processing. However, to optimize the switching process, the boundaries between stable and unstable propagation of the SLM must be identified.

As far as the structure of the SLM is concerned, two basic types of SLMs can be distinguished, i.e., odd (centered on-site) and even modes (centered between sites). In each case the adjacent components may oscillate either in-phase (unstaggered modes) or out-of-phase (staggered modes),[16,17] depending on the sign of the nonlinearity. As was already mentioned, the stability of SLMs against perturbations affects substantially the dynamics of the mode and is therefore an important issue to be addressed. The problem can be tackled by using various approaches, e.g., direct numerical calculations or a method based on the so-called Peierls-Nabarro (PN) potential.[16,17] It is evident that the former method cannot cover the entire problem; i.e., study of the effect of variation of all parameters involved on the stability. The latter method relies on the PN potential (PN barrier) of both types of solutions, providing no information about the instability gain. Moreover, as was demonstrated in Ref. 17, to consistently interpret the results obtained, one must introduce concepts such as the negative mass for staggered modes. Another technique, which is based on a variational approach, was applied to investigate the existence and stability of relatively weak localized modes of the generalized DNLSE.[18] Finally, the onset of chaos, including the so-called microchaos for three coupled oscillators, has been studied by calculating the Lyapunov exponent.[19]

As a result of these previous studies, all even SLMs of the DNLSE with the Kerr-like nonlinearity have been assumed to be unstable. In this paper we prove for the first time the existence of a stable even mode in the system described by the DNLSE and give an analytical criterion for its stability. We show that a direct linear analysis can be exploited to straightforwardly investigate the stability of the entire family of SLMs. This technique provides a clear physical picture of the onset of SLM dynamics. The analytical results concern-

ing the regions of instability as well as the respective gain permit us to draw conclusions for all-optical switching in waveguide arrays.

The DNLSE under consideration is

$$i\,\frac{dE_n}{dt}+c(E_{n+1}+E_{n-1})+\lambda|E_n|^2E_n=0, \tag{1}$$

where t and n represent the evolution parameter and the site index, respectively; E_n represents the excitation at the nth site, c is the linear coupling coefficient, and λ is the effective nonlinear coefficient. All quantities are dimensionless. This can be achieved by a convenient normalization using characteristic scales for the evolution variable and the amplitude of the excitation. In case of waveguide arrays t denotes the propagation distance along the waveguide.

In order to identify SLMs we take advantage of a method reported in Refs. 1 and 4. Inserting $E_n=e_n\exp(i\omega t)$ into (1), where e_n represent the respective amplitudes of a bright localized mode, we obtain a system of a few algebraic equations. Thus, for the even mode $e_n=A(\ldots,0,\alpha_3,\alpha_2,1,s,s\alpha_2,s\alpha_3,0,\ldots)$, $|n|=1m.2,3,\ldots$, $s=\pm1$ we obtain the following equation with the requirement for strong localization $|\alpha_3|\ll|\alpha_2|\ll1$, $a_n\approx0$ for $n>3$:

$$\omega\equiv\omega_e=\lambda A^2+sc+\frac{c^2}{\lambda A^2},$$

$$\alpha_2\equiv\alpha=\frac{c}{\lambda A^2}-s\left(\frac{c}{\lambda A^2}\right)^2,\quad \alpha_3=\left(\frac{c}{\lambda A^2}\right)^2, \tag{2}$$

where for symmetry reasons the subscript $n=0$ has been dropped.

Analogously, for the odd mode the ansatz $e_n=B(\ldots,0,\beta_2,\beta_1,\beta_0,s\beta_1,s\beta_2,0,\ldots)$, $|\beta_2|\ll|\beta_1|\ll1$, gives

$$\omega\equiv\omega_{os}=\lambda B^2+\frac{2c^2}{\lambda B^2},\quad \beta_0=1,\quad \beta_1\equiv\beta=\frac{c}{\lambda B^2},$$

$$\beta_2=\beta_1^2,\quad s=1 \tag{3a}$$

for the symmetric mode and

$$\omega\equiv\omega_{oa}=\lambda B^2+\frac{2c^2}{\lambda B^2},\quad \beta_0=0,\quad \beta_1=1,$$

$$\beta_2\equiv\beta=\frac{c}{\lambda B^2},\quad s=-1 \tag{3b}$$

for the antisymmetric mode. Here the subscripts e and o represent the even and the odd mode, respectively, and the parameter $s=\pm1$ defines the symmetry of the mode. In deriving (2) and (3) we restricted the analysis to the second-order terms concerning the small parameters α and β for no more than six excitations. In concentrating on the physical aspect of the problem we restrict the discussion to the first-order approximation. For sufficiently strong localization they provide a reasonable accuracy, which was confirmed by a direct numerical solution of (1). The difference between numerical and approximate analytical solutions merely amounts to a few percents. A detailed study which takes higher-order terms into account represents a separate subject and is beyond the scope of this paper. Hence, in what follows we mainly deal with *strongly* localized modes and assume that for α, $\beta<\nu_{sl}\approx0.2$ the second-order terms can be ignored.

To study the stability of SLMs we impose complex perturbations $\delta_n(t)$ on each nonzero excitation amplitude.[20] We begin with the even mode and insert the perturbed profile $e_n=A(\ldots,0,\alpha+\delta_{-2},1+\delta_{-1},s+\delta_{+1},\alpha+\delta_{+2},0,\ldots)$ into (1). A subsequent linearization yields an eighth-order system of equations for the real-valued variables, which is only numerically solvable. However, a considerable simplification can be achieved by a proper decomposition of the perturbations into symmetric and antisymmetric components as $\delta_j^{\pm}=\delta_{+j}\pm\delta_{-j}$ $(j=1,2)$,[19] which leads to a decoupling of the system. Separating real and imaginary parts of the perturbations $\delta_j^{\pm}=\delta_{jr}^{\pm}+i\,\delta_{ji}^{\pm}$ and introducing the scaled time $\tau_e=\omega_e t$, we obtain two independent systems for the column vector $\bar{\delta}^{\pm}=(\delta_{1r}^{\pm},\delta_{1i}^{\pm},\delta_{2r}^{\pm},\delta_{2i}^{\pm})$

$$\frac{d\bar{\delta}^{\pm}}{d\tau_e}=\begin{pmatrix}0 & (s-p)\alpha & 0 & -\alpha\\ 2-(3s-p)\alpha & 0 & \alpha & 0\\ 0 & -\alpha & 0 & 1\\ \alpha & 0 & -1 & 0\end{pmatrix}\bar{\delta}^{\pm}, \tag{4}$$

where $p=\pm1$ stands for the symmetric (δ_j^{+}) and antisymmetric (δ_j^{-}) perturbation, respectively. If we introduce $\bar{\delta}^{\pm}\propto\exp(g\tau_e)$, then the eigenvalues g of (4) are given by the biquadratic equation

$$g^4+[1+2(p-s)\alpha+2(3-2ps)\alpha^2]g^2+2(p-s)\alpha+2(3-2ps)\alpha^2+2(p-2s)\alpha^3+\alpha^4=0. \tag{5}$$

If the symmetry of the perturbation coincides with that of the SLM ($s=p$), Eq. (5) does not exhibit real-valued solutions provided that α is small, as required ($\alpha\leqslant\nu_{sl}$). Thus, the SLM is always stable against those perturbations, which was numerically verified. In contrast, if the perturbation has the opposite symmetry of the SLM ($p=-s$), the SLM can become unstable [$\mathrm{Re}(g)\neq0$].

We observed two basically different kinds of SLMs dynamics. Both staggered and unstaggered modes are always unstable with respect to symmetric and antisymmetric perturbations, respectively, whereas SLMs with $\alpha s<0$ are unstable only if the modulus of the amplitude α exceeds a critical value, i.e., $|\alpha|>\alpha_{\mathrm{cr}}$. These particular even SLMs are neither staggered nor unstaggered and can be obtained from those by changing the phase of excitations on sites $n\geqslant1$ by π. Hence, we call these modes *twisted staggered* (TS) ($s=1,\alpha<0$) and *twisted unstaggered* (TU) ($s=-1,\alpha>0$) SLM. It is worth noting that the continuous NLSE limit does not exhibit a solution of that topology.

In analyzing the solutions of (5) we can ignore higher than quadratic terms in α and thus obtain a compact expression for the instability gain. If the linear coupling (c) and the nonlinearity (λ) have the same sign (i.e., $\alpha>0$), the unstaggered SLM ($s=1$) is always unstable against antisymmetric perturbations, where the gain of instability

$$g\approx2\sqrt{s\alpha}(1-5s\alpha/4),\quad s\alpha>0 \tag{6}$$

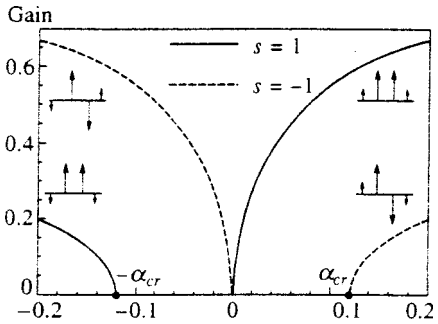


FIG. 1. Instability gain [Re(g)] plotted as a function of the amplitude α for even SLMs. The insets show the shape of the respective SLMs, where the twisted modes are sketched at the bottom.

increases with α (see Fig. 1). The instability of unstaggered modes is confirmed by a direct numerical solution of (1). The decay of the antisymmetrically perturbed, unstaggered SLM and its subsequent transformation into an odd mode can be clearly recognized in Fig. 2. As can be anticipated from (6) (see also Fig. 1), the transition time decreases with α due to the increase in the instability gain. A change of α from 0.13 (Fig. 2a) to 0.15 (Fig. 2b) causes a significant reduction of that transition time to the stable odd mode. We mention that

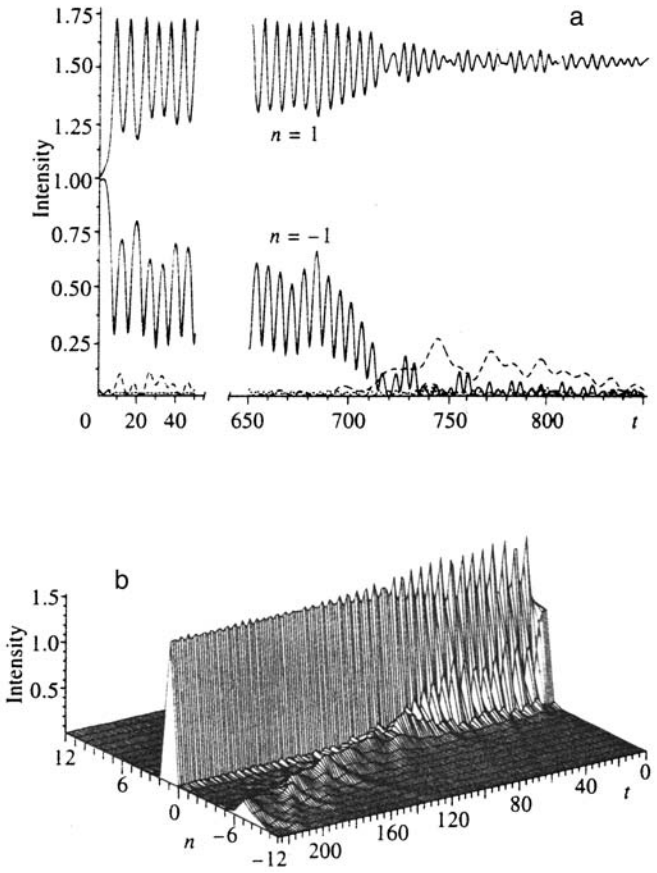


FIG. 2. Evolution of an unstaggered SLM ($s=1$) antisymmetrically perturbed; $\lambda=1$, $A=1$, $\bar{\delta}^{-}=(0.01,\ 0,\ 0,\ 0)$; (a) $\alpha=0.13$, evolution of the four initial excitations (solid lines $n=\pm1$, dashed line $n=-2$, dotted line $n=2$); (b) $\alpha=0.15$, evolution of the mode.

for relatively small amplitudes ($\alpha\approx0.1$) the intermediate, asymmetric, oscillating state is fairly persistent and can be thus considered a quasi-stationary state.

In contrast to the behavior of the unstaggered SLM, which is in agreement with the results previously reported,[16–18] the TU mode ($s=-1$) becomes unstable against symmetric perturbations only beyond the critical amplitude and the corresponding gain is

$$g\approx\sqrt{-s\alpha-\alpha_{\rm cr}},\quad s\alpha+\alpha_{\rm cr}<0,\quad \alpha_{\rm cr}\approx0.12. \tag{7}$$

This has the consequence that TU SLMs are stable against *any* perturbation if $\alpha<\alpha_{\rm cr}$. For the case where c and λ have opposite signs (i.e., $\alpha<0$) the situation is reversed (see the left side of Fig. 1) and the TS mode exhibits stability for that particular region of $|\alpha|$. These predictions were double-checked by numerically solving (1), imposing an *asymmetric* perturbation on the TU SLM. If α does not exceed the critical value $\alpha_{\rm cr}$, the TU mode is stable, exhibiting only slight oscillations produced by the perturbation (see Fig. 3a). If α grows larger and exceeds the critical value, the TU SLM becomes unstable and decays eventually (see Fig. 3b). Thus, the existence of a stable even SLM of the DNLSE has been proven. The stability of the twisted modes might be explained by the fact that neither the TU nor the TS variant has a topological counterpart among odd SLMs. Hence, such a twisted SLM cannot transform to an odd SLM and stability arguments based on the PN barrier do not apply here. Beyond the critical value $\alpha_{\rm cr}$ instability manifests itself in a spreading of the mode and sets in if the localization becomes weaker due to an increasing secondary amplitude $|\alpha|$ (see Fig. 3b). We note that allowance for second-order terms in conjunction with the excitations at the sites $|n|=3$ does not significantly change the instability regions and the gain. The transition from stability to instability, which is caused by a slight change of α at the input, can be potentially exploited for all-optical switching (e.g., see the drastic change of the output intensity in the waveguide labeled $n=-1$ in Figs. 3a and 3b, respectively).

Following the same procedure, one can also study the stability of odd SLMs. For example, if we ignore the second-order corrections for the odd symmetric SLM in Eq. (3a) and impose complex perturbations $\varepsilon_n(t)$, we obtain from (1) and the subsequent linearization a six-order system of ordinary differential equations. Decomposing the perturbation into the symmetric and antisymmetric components $\varepsilon_1^{\pm}=\varepsilon_{+1}\pm\varepsilon_{-1}$, we can easily infer that the equation for ε_1^{-} can be separated, and that it yields the solution $\varepsilon_1^{-}(t)=\varepsilon_1^{-}\exp(-i\omega_o t)$. Obviously, this type of perturbation does not provoke any instability of the system. Thus, one needs only to study the stability with respect to symmetric perturbations. Separating real and imaginary parts of the perturbations $\varepsilon_0=\varepsilon_{0r}+i\varepsilon_{0i}$, $\varepsilon_1^{+}=\varepsilon_{1r}^{+}+i\varepsilon_{1i}^{+}$, we obtain a system of four linear equations

$$\frac{d\bar{\varepsilon}}{d\tau_{os}}=\begin{pmatrix}0 & 0 & 0 & -\beta\\ 2 & 0 & \beta & 0\\ 0 & -2\beta & 0 & 1\\ 2\beta & 0 & -1 & 0\end{pmatrix}\bar{\varepsilon}, \tag{8}$$

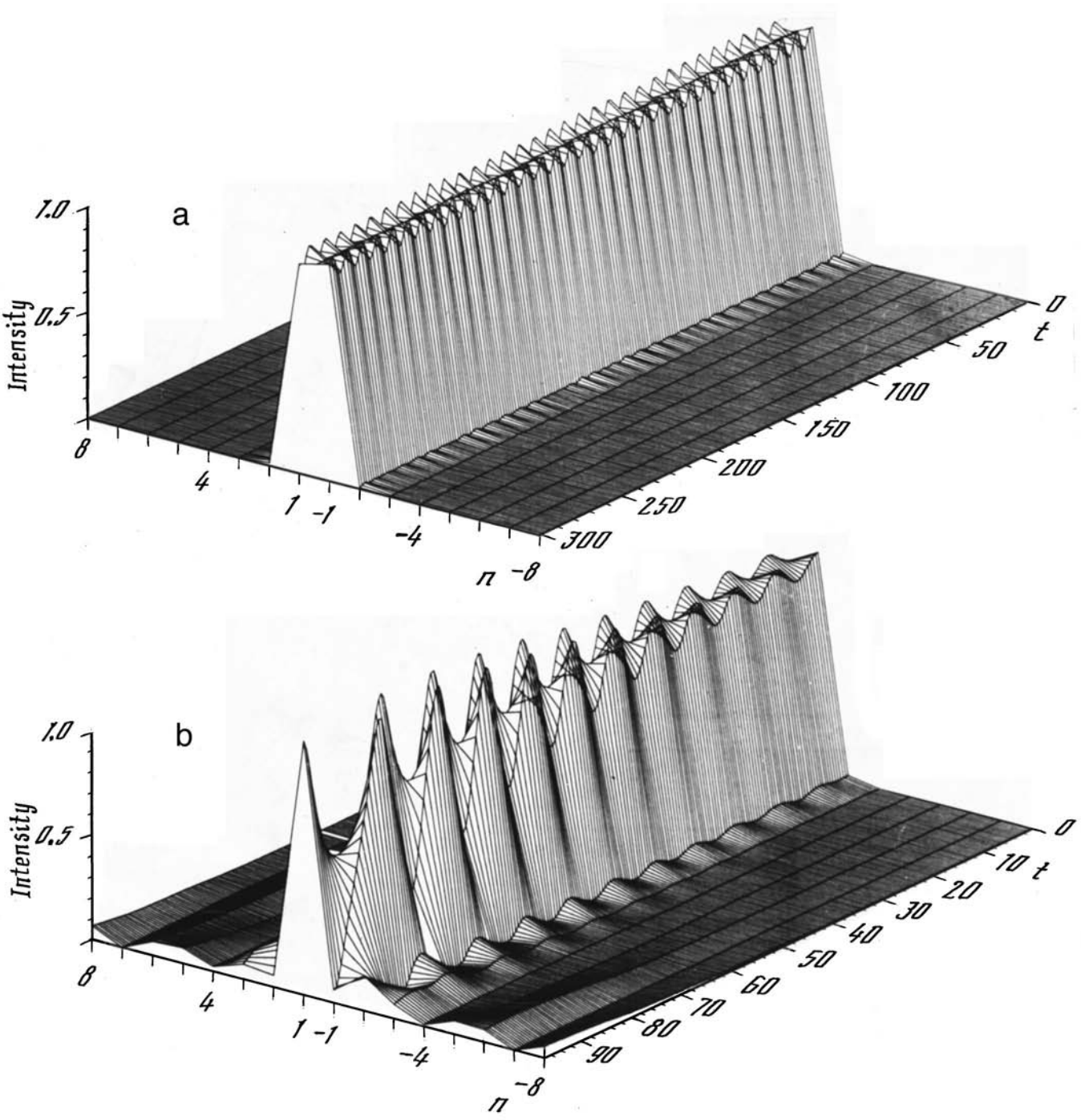


FIG. 3. Propagation of the perturbed twisted unstaggered SLM ($s=-1$), $\lambda=1$, $A=1$; (a) $\alpha=0.11<\alpha_{cr}$, $\bar{\delta}=(0.04,\ -0.02,\ 0.02,\ -0.02)$, $\bar{\delta}^{-}=(0,\ -0.04,\ -0.04,\ 0.04)$; (b) $\alpha=0.16>\alpha_{cr}$, $\bar{\delta}^{+}=(0.04,\ 0,0,\ 0)$.

where $\tau_{os}=\omega_{os}t$ is the scaled time, and $\bar{\varepsilon}=(\varepsilon_{0r},\varepsilon_{0i},\varepsilon_{1r}^{+},\varepsilon_{1i}^{+})$ is the perturbation vector. Again, the corresponding eigenvalue problem represents a simple biquadratic equation, which now reads as

$$g^4+(1+4\beta^2)g^2+4\beta^2(1+\beta^2)=0, \tag{9}$$

where $\mathrm{Re}(g)$ also represents the instability gain. We straightforwardly obtain a nonzero gain $\mathrm{Re}(g)$ only provided that the secondary excitation $|\beta|>1/\sqrt{8}\approx 0.35$. Such an instability causes the spreading of the mode in both directions in n. The larger the instability gain, the faster the unstable SLM decays and the excitation is spread over the entire array. However, the above values for the secondary excitation are beyond the required small-parameter limit for β. Thus, we may draw the conclusion that the odd, *strongly* localized mode (3a) is stable against small perturbations. This result was confirmed numerically. Figure 4 shows the evolution of a perturbed odd SLM, where a complex perturbation was superimposed on a solution of (3a). Obviously, the perturbation results only in

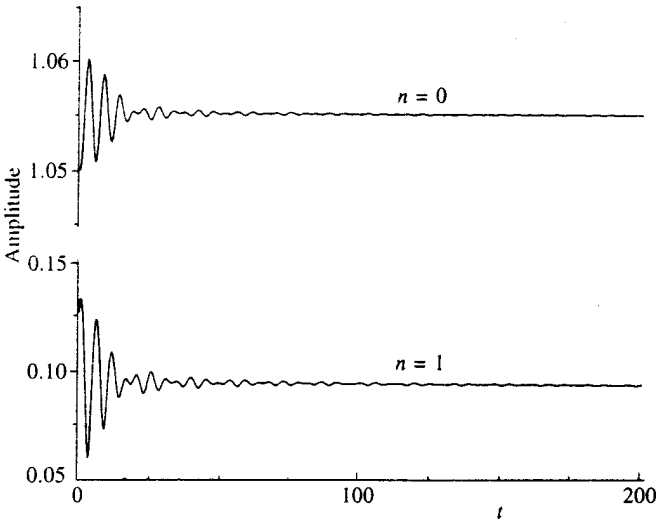


FIG. 4. Evolution of the perturbed odd SLM. The amplitudes of the central (upper curve) and the secondary (lower curve) excitations are shown for $\lambda = 1$, $B=1$, $\beta=0.1$, $\overline{\varepsilon}=(0.05,\ 0,\ 0.05,\ 0)$; $\varepsilon^{-}=i\cdot 0.05$.

the quickly damping oscillations near the SLM. Because (1) has a continuum set of SLM solutions, which depend on the amplitude B, the perturbed solution eventually transforms into a stable SLM with a new amplitude which is determined by the strength of perturbation. Analogously, one can show that the odd antisymmetric SLM is also stable.

In conclusion, we have demonstrated that by using a direct linear analysis the stability behavior of intrinsic, strongly localized modes of the discrete nonlinear Schrödinger equation can be analytically predicted. The regions of instability and the respective gain have been explicitly calculated. The familiar stability of odd modes was confirmed. It was shown for the first time that twisted *even* modes can also be stable provided that the secondary amplitudes are below a certain critical value.

S.D. and F.L. gratefully acknowledge a grant from the Deutsche Forschungsgemeinschaft. Bonn. S.D. is also thankful for the partial support through Grant 96-0334049 of the Russian Fund for Fundamental Research. The authors are grateful to E. Schmidt for invaluable help with the numerics.

[1] A. C. Scott and L. Macneil, Phys. Lett. A **98**, 87 (1983); J. C. Eilbeck, P. S. Lomdahl, and A. C. Scott, Physica D **16**, 318 (1985).
[2] A. S. Dolgov, Fiz. Tverd. Tela **28**, 1641 (1986), [Sov. Phys. Solid State **28**, 902 (1986)].
[3] A. J. Sievers and S. Takeno, Phys. Rev. Lett. **61**, 970 (1988).
[4] J. B. Page, Phys. Rev. B **41**, 7835 (1990).
[5] V. M. Burlakov and S. A. Kiselev, Zh. Eksp. Teor. Fiz. **99**, 1526 (1991) [Sov. Phys. JETP **72**, 854 (1991)].
[6] G. P. Tsironis and S. Aubry, Phys. Rev. Lett. **77**, 5225 (1996).
[7] Yu. S. Kivshar and M. Peyrard, Phys. Rev. A **46**, 3198 (1992).
[8] V. M. Burlakov, S. A. Darmanyan, and V. N. Pyrkov, Zh. Éksp. Teor. Fiz. **108**, 904 (1995) [JETP **81**, 496 (1995)].
[9] A. A. Aceves, G. G. Luther, C. De Angelis, A. M. Rubenchik, and S. K. Turitsyn, Phys. Rev. Lett. **75**, 73 (1995).
[10] S. A. Darmanyan, I. Relke, and F. Lederer, Phys. Rev. E **55**, 7662 (1996).
[11] V. M. Burlakov, S. A. Darmanyan, and V. N. Pyrkov, Phys. Rev. B **54**, 3257 (1996).
[12] B. Denardo, B. Galvin, A Greenfield, A. Larraza, S. Putterman, and W. Wright, Phys. Rev. Lett. **68**, 1730 (1992).
[13] P. Marquie, J. M. Bilbaut, and M. Remoissenet, Phys. Rev. E **51**, 6127 (1995).
[14] A. Aceves, C. De Angelis, T. Peschel, R. Muschall, F. Lederer, S. Trillo, and S. Wabnitz, Phys. Rev. E **53**, 1172 (1996).
[15] W. Krolikowski and Yu. S. Kivshar, J. Opt. Soc. Am. B **13**, 876 (1996).
[16] Yu. S. Kivshar and D. K. Campbell, Phys. Rev. E **48**, 3077 (1993).
[17] D. Cai, A. R. Bishop, and N. Gronbech-Jensen, Phys. Rev. Lett. **72**, 591 (1994).
[18] E. W. Laedke, O. Kluth, and K. H. Spatschek, Phys. Rev. E **54**, 4299 (1996).
[19] N. Finlayson, K. J. Blow, L. J. Bernstein, and K. W. DeLong, Phys. Rev. A **48**, 3863 (1993).
[20] K. W. Sandusky, J. B. Page, and K. E. Schmidt, Phys. Rev. B **46**, 6161 (1992).

Published in English in the original Russian journal. Reproduced here with stylistic changes by the Translation Editor.

Second-harmonic generation in the interior of an isotropic medium with quadratic nonlinearity by a focused inhomogeneously polarized pump beam

S. N. Volkov, N. I. Koroteev,*) and V. A. Makarov

Physics Department and International Laser Center, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

(Submitted 16 September 1997)
Zh. Éksp. Teor. Fiz. **113**, 1261–1276 (April 1998)

This paper presents a theoretical study of second-harmonic generation (SHG) by a focused pump beam in the interior of an isotropic medium, experimentally observed earlier. It shows that the spatial dispersion of the quadratic optical response of the substance can be responsible for this nonlinear process even when a macroscopic inversion center is present in the medium. It is established that this effect, which is impossible in the plane-wave approximation, also does not occur when a Gaussian pump beam with uniform distribution of the polarization state of the wave field over the cross section is used, but that the presence in the pump beam of higher transverse modes with polarization different from the fundamental mode can cause an SHG signal to appear. The conditions for a wave to appear at the doubled frequency are found, analytical dependences for its electric field and total power on the propagation coordinate, the degree of focusing, and the other parameters of the problem are obtained, and the requirements on the optimum experimental geometry are formulated. The dependence of the signal-beam power on the wavevector detuning is studied, and it is shown that interference effects can cause the wave at the second harmonic to disappear when a normal dispersion law is obeyed in the region between the fundamental and doubled frequencies. © *1998 American Institute of Physics.* [S1063-7761(98)01004-X]

1. INTRODUCTION

The conversion of laser pump radiation into the second harmonic in the interior of noncentrosymmetric crystals has been known since the first days of nonlinear optics.[1] Second-harmonic generation (SHG) is also widely used in the spectroscopy of the surfaces and interfaces of media,[1–3] since this three-wave-mixing process can be efficient only when there is no macroscopic inversion center in the test object. Volume SHG is also possible in an isotropic medium, for example a liquid or a gas, when a static electric field is imposed on the medium,[1] removing the inversion center in it. At the same time, in the absence of external fields, the appearance of a signal at the doubled frequency due to local electric-dipole optical susceptibility $\hat{\chi}^{(2)}$ is forbidden not only in centrosymmetric but also in noncentrosymmetric liquids and gases, since the corresponding tensor $\hat{\chi}^{(2)}$ equals zero because of symmetry under permutation of its last two subscripts.[4,5] Nevertheless, SHG was experimentally produced several years ago in the interior of a noncentrosymmetric solution of arabinose using noncollinear interaction of two laser beams with identical frequency.[6,7] This result is because the signal obtained at the second harmonic was not quadratic but fourth-order in the field of the pump wave, which is explained by the fact that it displayed nonlinear optical susceptibility $\hat{\chi}^{(4)}(2\omega;\omega,\omega,\omega,-\omega)$.

A number of experimental papers[8–14] have recently been published in which a signal was observed at the doubled frequency, generated by one pump beam, in a suspension of arbitrarily oriented fragments of the purple membranes of *Halobacterium halobium*, containing chiral molecules of bacteriorhodopsin. In a number of cases, when coarse membrane fragments (about 1 μm in diameter) are studied, the appearance of the second harmonic can be ascribed to partially coherent hyper-Rayleigh scattering.[8–11] Actually, each membrane fragment is an ordered structure of significant size, which can generate a coherent SHG signal. When it is averaged over the volume of the solution, the total signal at the doubled frequency does not go to zero, because the membrane fragments illuminating it are macroscopic in size (of the order of the wavelength of the laser radiation being used). A theory based on the present model of partially coherent hyper-Rayleigh scattering was constructed in Ref. 15; its conclusions give a fairly good description of the experiments on the study of a suspension of coarse fragments of purple membranes.[8–11] However, by no means all the experimental results can be explained in this way. Membrane fragments reduced to about 50 nm were used in Refs. 12–14; in this case, the directionality diagram of the observed SHG signal and its polarization dependences did not match the prediction of its appearance due to hyper-Rayleigh scattering (both completely incoherent and partially coherent). Moreover, the presence of two components that interfere with each other was distinctly observed in the signal at the doubled frequency in these experiments. These were quadratic and fourth-order in the field of the pump wave.

There has been no consensus among investigators regarding the nature of the SHG signal component quadratic in field, and moreover there is no satisfactory theory that can explain all the experimental results so far obtained. We in-

tend to show in this article that the quadratic part of the signal wave observed in Refs. 12–14 can be caused by spatial dispersion of the nonlinear-optical response of the interior of the substance, since, even when the size of the investigated membrane fragments is much less than the wavelengths of the interacting fields (about 50 nm), the nonlocal behavior of the nonlinear optical response can be significant. Theoretical treatment of this phenomenon is not a simple problem and requires that the spatial limitation of the pump beam and the inhomogeneity of its polarization state with respect to the cross section be treated, since, as will be seen from the formulas given below, such an effect is impossible in the plane-wave approximation.

2. MAIN CONSIDERATIONS

In this paper, we consider the signal component at the second harmonic generated in the interior of a nonabsorbing homogeneous isotropic medium that is quadratic in the field of a monochromatic pump beam. The material equation for the generalized polarization of a medium at the doubled frequency, allowing for spatial dispersion of the quadratic nonlinearity, can be written as

$$P_i^{(2)}(\mathbf{r})=\gamma_{ijkl}^{(2)}E_l(\mathbf{r})\,\frac{\partial}{\partial r_j}\,E_k(\mathbf{r}), \tag{1}$$

where $\mathbf{E}(\mathbf{r})\propto\exp(-i\omega t)$ is the electric field of the pump beam. As pointed out above, the local electric-dipole optical susceptibility $\hat{\chi}^{(2)}$ equals zero in the present case even when there is no macroscopic inversion center in the medium.

In this paper, we use the approach to the electrodynamics of complex media in which the magnetic induction is identically equal to the magnetic field. The generalized polarization of the medium unambiguously describes the optical response of the latter in this case, since it is the only cause for the appearance of the electromagnetic signal wave. All the magnetic effects are completely taken into account in the framework of the assumption that spatial dispersion is present in the $\mathbf{P}(\mathbf{E})$ dependence. We shall restrict ourselves in what follows to terms that are linear in the small spatial-dispersion parameter d'/λ (d' is the characteristic scale of the nonlocal optical response, and λ is the wavelength of the light). When we take this into account to describe the quadratic optical response of the material, it is sufficient in our problem to know only the tensor $\hat{\gamma}^{(2)}$ (we recall once again that $\hat{\chi}^{(2)}=0$).

The literature exhibits frequent attempts to separate the generalized polarization into electric-dipole, magnetic-dipole, electric-quadrupole, etc., contributions. In such an approach, $\hat{\gamma}^{(2)}$ would be represented in terms of a number of tensors describing the nonlinear-optical response of a ''different nature.'' It should be pointed out that this is not possible for arbitrary media and by no means for all cases (see Ref. 16). Moreover, even when such a separation can be made, the tensor $\hat{\gamma}^{(2)}$ can include not only terms of magnetic-dipole and electric-quadrupole origin, but also terms formed by its manifestation of the nonlocal nature of the electric-dipole contribution, caused by the interaction of the dipoles that compose the substance (see, for example, Ref. 17). Moreover, this separation is extremely conventional and arbitrary, since it is the generalized polarization of the medium, including the electric-dipole and all the other possible contributions, that enters into the right-hand part of the wave equation as the source of the electromagnetic wave. In this connection, in the case considered here, it is impossible in principle to obtain more spectroscopic information than is contained in the independent components of the tensor $\hat{\gamma}^{(2)}$. This is why it seems to us to be more convenient and correct to use the tensor $\hat{\gamma}^{(2)}$ introduced in accordance with Eq. (1) to describe the nonlocal behavior of the quadratic optical response of a substance in the problem considered here.

The material tensor $\hat{\gamma}^{(2)}$ possesses no additional internal symmetry. Therefore, for the media considered here, with $\infty\infty$ or $\infty\infty m$ symmetry, it has three independent Cartesian components, which we denote by

$$\gamma_1^{(2)}=\gamma_{xxyy}^{(2)},\quad \gamma_2^{(2)}=\gamma_{xyyx}^{(2)},\quad \gamma_3^{(2)}=\gamma_{xyxy}^{(2)}.$$

Taking this into account, the material Eq. (1) can be written in vector form as

$$\mathbf{P}^{(2)}(\mathbf{r})=\frac{1}{2}\,\gamma_1^{(2)}\boldsymbol{\nabla}(\mathbf{E}\cdot\mathbf{E})+\gamma_2^{(2)}\mathbf{E}(\boldsymbol{\nabla}\cdot\mathbf{E})+\gamma_3^{(2)}(\mathbf{E}\cdot\boldsymbol{\nabla})\mathbf{E}, \tag{2}$$

where we have written $\boldsymbol{\nabla}=\partial/\partial\mathbf{r}$ and omitted the argument $\mathbf{r}$ in the field $\mathbf{E}$ everywhere for brevity. We should point out that the prescribed-pump approximation should be used when $\mathbf{E}(\mathbf{r})$ is substituted in Eq. (2), since we hope to restrict ourselves to a study of the optical response of the medium that is quadratic in the field. Taking into account the action of the signal beam on the pump wave and also the self-action of the latter does not affect the form of the terms in the expression for the polarization of the medium that are quadratic in field, but only adds terms proportional to higher powers of the field of the pump beam.

Let us examine what each of the three terms on the right-hand side of Eq. (2) contributes to the field of the signal wave. The first of these is the gradient of a scalar function, and therefore is a purely irrotational[1)] vector field of the polarization of the medium. The electric field created by it at the doubled frequency is found from the Maxwell equation $\mathrm{div}\,\mathbf{D}=0$:

$$\mathbf{E}_{\mathrm{pot}}^{(2)}(\mathbf{r})=-\frac{4\pi}{\varepsilon_{SH}}\,\mathbf{P}_{\mathrm{pot}}^{(2)}(\mathbf{r}). \tag{3}$$

[This is the form that results from the reduction of the wave equation for longitudinal plane electromagnetic waves (see Ref. 1, Sec. 3.3), and consequently also for beams of irrotational type.] Here ε_{SH} is the permittivity of the medium at the doubled frequency, and the subscript ''pot'' recalls the irrotational character of the field (a potential field). The last equation shows that the irrotational field of the vector $\mathbf{P}_{\mathrm{pot}}^{(2)}$ generates an electromagnetic wave of irrotational type, which, as is well known, cannot propagate freely but exists only inside a medium, being coupled to the wave $\mathbf{P}_{\mathrm{pot}}^{(2)}(\mathbf{r})$ by Eq. (3). When it encounters the surface of the medium, $\mathbf{E}_{\mathrm{pot}}^{(2)}$

contributes to a free wave because the boundary conditions for an electromagnetic field must be satisfied, but this contribution must be taken into account[1] (and is taken into account—see, for example, Refs. 18 and 19) as part of the SHG signal from the surface. In this article, we shall consider only electromagnetic waves of the solenoidal type at the doubled frequency, since they are the only purely bulk effect (i.e., the signal is formed in the interior of the medium).

The second term in Eq. (2) is proportional to div $\mathbf{E}$, and therefore equals zero, which follows from Maxwell's equation div $\mathbf{D}=0$ taking into account the constant-pump approximation. Thus, the SHG signal of interest to us can be generated only by the solenoidal component of the third term in Eq. (2). Below, we consider it in more detail.

Note first that the case in which the medium lacks a macroscopic inversion center (symmetry $\infty\infty$) differs from the case of a centrosymmetric substance by the presence of linear gyration in the medium. This has the result, in particular, that normal waves with right- and left-circular polarization have different wave vectors in such a medium. However, this difference is a first-order correction in the spatial dispersion parameter d'/λ, and therefore the only result of including it is that additional terms quadratic in d'/λ appear in Eq. (2), which already includes the tensor $\hat{\gamma}^{(2)}$, which is linear in the small spatial dispersion parameter; these quadratic terms are outside the scope of our treatment. Thus, including the influence of linear gyration does not change the general form of the expression for $\mathbf{P}^{(2)}$ and consequently cannot affect the conclusion that SHG is possible or impossible in any specific situation. Of course, if the rotation of the polarization ellipse that arises in this way at distances of the order of the length of the medium is sufficiently large, it should be taken into account when writing out the expression for the pump-beam field and the wave equation in order to correctly describe the characteristics of the signal wave (especially its polarization). However, in this paper, we shall consider the linear rotation of the polarization ellipse to be small at all lengths of interest to us, which agrees quite well with the real situation that occurs in the experiments considered above for studying a suspension of fragments of purple membranes. Taking into account what has been said, we shall neglect the influence of linear gyration for the particular case of an isotropic noncentrosymmetric medium. This makes our further calculations for isotropic media with $\infty\infty$ symmetry (with no inversion center) and $\infty\infty m$ (a centrosymmetric medium) completely identical.

If we consider a plane pump wave with wave vector $\mathbf{k}$, the third term in Eq. (2) takes the form $i\gamma_3^{(2)}(\mathbf{E}\cdot\mathbf{k})\mathbf{E}$, which equals zero because the field of a free plane wave is transverse. As will be seen from the subsequent formulas, a Gaussian beam with a homogeneous distribution of the polarization state of the wave over its cross section also cannot generate an SHG signal, and therefore we shall assume that, besides the zeroth-order (Gaussian) mode $\mathbf{E}_G(\mathbf{r})$, the pump wave contains two additional higher-order transverse modes $\mathbf{E}^{X,Y}(\mathbf{r})$, where all three of these modes can have different polarizations. We direct the z axis of the xyz coordinate system along the beam axis (see Fig. 1), and we shall write all subsequent formulas up to and including terms linear in $1/kw$ ($k=n\omega/c$ is the wave number of the pump wave in the medium, $n=\sqrt{\varepsilon}$ is the refractive index, ε is the permittivity of the medium at frequency ω in the absence of linear absorption, and w is the half-width of the Gaussian mode of the pump beam). The field of the beam at the fundamental frequency thus has the form

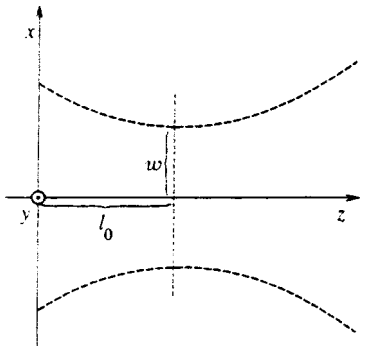


FIG. 1. Propagation of a pump beam with SHG through a medium. The zero initial conditions for the field of the signal wave are set in the $z=0$ plane; w is the half-width of the beam at the fundamental frequency in the plane of the beam waist, $z=l_0$.

$$\mathbf{E}(\mathbf{r})=\mathbf{E}_G(\mathbf{r})+\mathbf{E}^X(\mathbf{r})+\mathbf{E}^Y(\mathbf{r}), \tag{4}$$

where

$$\mathbf{E}_G(\mathbf{r})=E_0\left[\mathbf{e}_0-\frac{\mathbf{e}_z}{ik}(\mathbf{e}_0\cdot\boldsymbol{\nabla})\right]e_G(\mathbf{r}),$$

$$\mathbf{E}^X(\mathbf{r})=E_0^X\left[\mathbf{e}_0^X-\frac{\mathbf{e}_z}{ik}(\mathbf{e}_0^X\cdot\boldsymbol{\nabla})\right]\frac{x}{w\beta(z)}e_G(\mathbf{r}),$$

$$\mathbf{E}^Y(\mathbf{r})=E_0^Y\left[\mathbf{e}_0^Y-\frac{\mathbf{e}_z}{ik}(\mathbf{e}_0^Y\cdot\boldsymbol{\nabla})\right]\frac{y}{w\beta(z)}e_G(\mathbf{r}). \tag{5}$$

Here $e_G(\mathbf{r})$ describes a scalar normalized Gaussian beam

$$e_G(\mathbf{r})=\frac{1}{\beta(z)}\exp\left[-i\omega t+ik(z-l_0)-\frac{r_\perp^2}{w^2\beta(z)}\right], \tag{6}$$

where $\beta(z)=1+i(z-l_0)/l_d$, $l_d=kw^2/2$ is the diffraction length of the Gaussian beam, l_0 is the distance from the $z=0$ plane to its waist, and $\mathbf{r}_\perp$ is the component of radius vector $\mathbf{r}$ that lies in the xy plane. In Eqs. (5), E_0 and $E_0^{X,Y}$ are scalar amplitudes, while $\mathbf{e}_0$ and $\mathbf{e}_0^{X,Y}$ are the polarization vectors of the three transverse modes of the pump beam under consideration, and the absolute magnitudes of these vectors equal unity. All six of the indicated quantities are in general complex, while $\mathbf{e}_0,\mathbf{e}_0^{X,Y}\perp\mathbf{e}_z$, where $\mathbf{e}_z$ is the unit vector along the z axis. In the framework of the assumptions made here, all three transverse modes of the beam with frequency ω satisfy not only the wave equation but also the condition div $\mathbf{E}=0$, which follows from the Maxwell equation div $\mathbf{D}=0$ when the constant-pumping approximation is taken into account. This is why, as already pointed out in endnote 1, the fields $\mathbf{E}_G(\mathbf{r})$ and $\mathbf{E}^{X,Y}(\mathbf{r})$ of these three modes, which are solenoidal free waves, have components along the z axis,

which are additions to the main (x and y) components of the field of the corresponding transverse mode that are linear in the small parameter $1/kw$.

Substituting the field $\mathbf{E}(\mathbf{r})$ given by Eq. (4) into the third term of the material equation (2) and using Eqs. (5) and (6), we obtain that, taking into account the approximations made here, this term will have a purely solenoidal character in our case. It was shown above that neither the first nor the second term of Eq. (2) contributes to the solenoidal component of the vector polarization field of the medium, and therefore

$$\mathbf{P}^{(2)}_{\rm vor}(\mathbf{r})=\gamma_3^{(2)}(\mathbf{E}\cdot\nabla)\mathbf{E}=\gamma_3^{(2)}\frac{e_G^2(\mathbf{r})}{w\beta(z)}\left\{E_0E_0^X[\mathbf{e}_x\times(\mathbf{e}_0^X\times\mathbf{e}_0)]+E_0E_0^Y[\mathbf{e}_y\times(\mathbf{e}_0^Y\times\mathbf{e}_0)]-\frac{\mathbf{r}_\perp}{w\beta(z)}E_0^XE_0^Y(\mathbf{e}_z\cdot(\mathbf{e}_0^X\times\mathbf{e}_0^Y))\right\}, \tag{7}$$

where $\mathbf{e}_{x,y}$ are the unit vectors along the x,y axes, respectively. We have used the subscript ''vor'' to designate the solenoidal part of the vector field $\mathbf{P}^{(2)}(\mathbf{r})$ (a vortex field). Only $\mathbf{P}^{(2)}_{\rm vor}$ can generate a free electromagnetic wave of the solenoidal type in the volume of a medium, since the irrotational component $\mathbf{P}^{(2)}_{\rm pot}$ does not contribute to it (this follows from the form of the wave equation, which can be found in the book cited above; see Ref. 1, Sec. 3.3).

3. SOLUTION OF THE WAVE EQUATION AND ANALYSIS OF THE TRANSVERSE FIELD DISTRIBUTION OF THE SIGNAL WAVE OVER THE BEAM CROSS SECTION

Let us write the wave equation for the electromagnetic field $\mathbf{E}^{(2)}_{\rm vor}$ at the doubled frequency, generated by the polarization $\mathbf{P}^{(2)}_{\rm vor}$ of the medium, in abbreviated form, for which we introduce a slowly varying amplitude $\mathbf{A}^{(2)}_{\rm vor}$ according to

$$\mathbf{E}^{(2)}_{\rm vor}(\mathbf{r})=\mathbf{A}^{(2)}_{\rm vor}(\mathbf{r})\exp[-2i\omega t+ik_{SH}(z-l_0)]. \tag{8}$$

where $k_{SH}=2n_{SH}\omega/c$, and $n_{SH}=\sqrt{\varepsilon_{SH}}$ is the refractive index of the medium at the doubled frequency. The reduced wave equation for $\mathbf{A}^{(2)}_{\rm vor}$ has the following form:

$$\left(\frac{\partial}{\partial z}-\frac{i}{2k_{SH}}\Delta_\perp\right)\mathbf{A}^{(2)}_{\rm vor}(\mathbf{r})=\frac{2\pi ik_{SH}}{\varepsilon_{SH}}\exp[2i\omega t-ik_{SH}(z-l_0)]\mathbf{P}^{(2)}_{\rm vor}(\mathbf{r}), \tag{9}$$

where $\Delta_\perp=\partial^2/\partial x^2+\partial^2/\partial y^2$, while $\mathbf{P}^{(2)}_{\rm vor}$ is given by Eq. (7), using Eq. (6).

Since the SHG signal from the surface is not under discussion in this paper, we set zero initial conditions for Eq. (9) at $z=0$ (the boundary of the medium). Its exact solution can be written in quadratures as

$$\mathbf{A}^{(2)}_{\rm vor}(\mathbf{r})=\frac{2\pi ik_{SH}}{\varepsilon_{SH}}\frac{\gamma_3^{(2)}}{w}\int_0^z dz'\left\{E_0E_0^X[\mathbf{e}_x\times(\mathbf{e}_0^X\times\mathbf{e}_0)]+E_0E_0^Y[\mathbf{e}_y\times(\mathbf{e}_0^Y\times\mathbf{e}_0)]-\frac{\mathbf{r}_\perp}{w}\left[\beta(z')-4i\frac{z'-z}{k_{SH}w^2}\right]^{-1}E_0^XE_0^Y(\mathbf{e}_z\cdot(\mathbf{e}_0^X\times\mathbf{e}_0^Y))\right\}\left[\beta(z')-4i\frac{z'-z}{k_{SH}w^2}\right]^{-1}\frac{1}{\beta^2(z')}\exp\left\{i\Delta k(z'-l_0)-\frac{2r_\perp^2}{w^2}\left[\beta(z')-4i\frac{z'-z}{k_{SH}w^2}\right]^{-1}\right\}, \tag{10}$$

where $\Delta k=2k-k_{SH}$ is the detuning of the wave vectors. Fortunately, by imposing not very burdensome limitations on the parameters of the problem, Eq. (10) can be greatly simplified and the result can be written in analytical form in terms of special functions. We require that

$$|\Delta k||z-l_0|\ll k_{SH}l_d \tag{11}$$

be satisfied for all z inside the medium. Since usually $|\Delta k|/k_{SH}\ll 1$ holds (small detuning of the wave vectors), this approximation can obviously be used for a thickness L of the medium of at least several diffraction lengths (provided that the pump-beam waist lies inside the medium). However, in the limit $|z-l_0|/l_d\gg 1$, the expression under the integral in Eq. (10) is small by comparison with its value in the plane of the beam waist; i.e., as should be expected, regions of the medium that far from the beam waist contribute little to the SHG process. Therefore, there is every reason to assume that, in all cases of practical importance, the simplified formula given below for $\mathbf{A}^{(2)}_{\rm vor}(\mathbf{r})$ is valid even in the limit $L\gg l_d$:

$$\mathbf{A}^{(2)}_{\rm vor}(\mathbf{r})=i\frac{\pi kk_{SH}w}{\varepsilon_{SH}}\gamma_3^{(2)}\left\{E_0E_0^X[\mathbf{e}_x\times(\mathbf{e}_0^X\times\mathbf{e}_0)]+E_0E_0^Y[\mathbf{e}_y\times(\mathbf{e}_0^Y\times\mathbf{e}_0)]-\frac{\mathbf{r}_\perp}{w\beta_{SH}(z)}E_0^XE_0^Y\times\mathbf{e}_z\cdot(\mathbf{e}_0^X\times\mathbf{e}_0^Y)\right\}\frac{1}{\beta_{SH}(z)}\exp\left[-\frac{2r_\perp^2}{w^2\beta_{SH}(z)}\right]\times\mathscr{T}^{(2)}(-l_0,z-l_0), \tag{12}$$

where

$$\beta_{SH}(z)=1+i\frac{z-l_0}{l_d}\frac{2k}{k_{SH}}$$

is in fact an analog of the coefficient $\beta(z)$ for a Gaussian beam at the doubled frequency with half-width $w/\sqrt{2}$, and $\mathscr{T}^{(2)}$ is used to denote the following dimensionless integral:

$$\mathscr{T}^{(2)}(z_1,z_2)=\int_{z_1/l_d}^{z_2/l_d}\frac{\exp(i\nu\zeta)}{(1+i\zeta)^2}d\zeta=i\left\{\frac{1}{1+i\zeta}\exp(i\nu\zeta)-\nu\exp(-\nu)\mathrm{Ei}(i\nu\zeta+\nu)\right\}\Bigg|_{\zeta=z_1/l_d}^{\zeta=z_2/l_d}, \tag{13}$$

where $\nu=\Delta kl_d$ is a dimensionless parameter that characterizes the detuning of the wave vectors, and

$$\mathrm{Ei}(x)=\int_{-\infty}^{x}(e^t/t)dt$$

is the exponential integral function.

It should be pointed out that an integral similar to Eq. (13) was encountered in Refs. 20–22 when four-wave mixing processes (in particular, third-harmonic generation) in focused Gaussian beams were considered. At the same time, the corresponding integral for the SHG case in a weakly anisotropic medium[20] differs appreciably from the formula that we have obtained for $\mathscr{T}^{(2)}$. The cause of this difference is that, in the problem considered here, SHG does not occur under the action of a Gaussian beam, as in Ref. 20, but because of the interaction of various transverse modes of a multimode pump beam. The analogy with third-harmonic generation in isotropic media is therefore rather unexpected, and we shall return to this point in Sec. 4.

As follows from the basic Eq. (12), SHG in the interior of an isotropic medium in the case considered in this paper is possible only with noncoincident polarization vectors in three transverse modes of the pump beam. Generation of the second harmonic is impossible in the case of a single-mode beam, for instance a purely Gaussian beam with uniform polarization over the cross section. It also follows from Eq. (12) that the field distribution over the cross section of the signal beam is the same as if it consisted of a Gaussian beam and higher transverse modes, but the z dependence in $\mathbf{A}^{(2)}_{\mathrm{vor}}(\mathbf{r})$ because of the factor $\mathscr{T}^{(2)}(-l_0, z-l_0)$ is different than in a free electromagnetic wave. If it is assumed that the structure of the pump beam is such that $\mathbf{E}^{X,Y}(\mathbf{r})$ constitute only a small correction to $\mathbf{E}_G(\mathbf{r})$, it follows from Eq. (12) that the Gaussian component will dominate in the signal beam. It is also interesting to point out that the two higher transverse modes in the SHG signal form an axisymmetric ''speckle'' distribution of the wave polarization state, in which the electric field vector at any point of the cross section of the beam is directed along its radius. The total polarization of the signal beam will in general be elliptical, with the parameters of the polarization ellipse being different at different points of the cross section.

Figure 2 illustrates the transverse spatial distribution of the electromagnetic field of a three-mode pump beam in the far field of diffraction. In drawing this figure, we assumed that the field of the Gaussian component of the beam is directed along the x axis, while the small addition from the two upper transverse modes, with speckle polarization, has a phase shift of $\pi/2$ relative to the field of the Gaussian mode, in order that the polarization of the wave would be strictly linear and not elliptical at all points of the figure (this is done only for greater clarity of the image). In other words, Fig. 2 is drawn with the following values of the parameters that characterize the pump beams:

$$E_0^X = E_0^Y = 0.2iE_0, \quad \mathbf{e}_0 = \mathbf{e}_0^X = \mathbf{e}_x, \quad \mathbf{e}_0^Y = \mathbf{e}_y$$

($\mathbf{e}_{x,y}$ are the unit vectors directed along the x,y coordinate axes). The quantities $x/r_0(z)$ and $y/r_0(z)$ are plotted along the axes in the figure, where $r_0(z) = 2|z-l_0|/kw$ is the half-width of the Gaussian mode of the pump beam in the far field of diffraction. The transverse field distribution in the pump beam shown in Fig. 2 is remarkable in that the transverse field distribution of the signal wave created by it in the far zone of diffraction will have almost the same form (to within some phase factor), except that the half-width of the beam at the doubled frequency will be about a factor of $\sqrt{2}$ less.

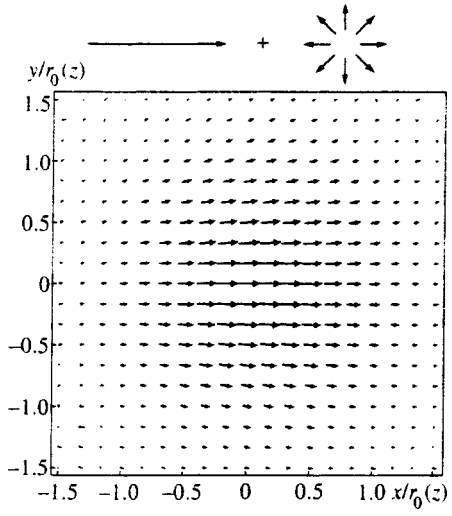


FIG. 2. Transverse distribution of the electric field of the pump beam over its cross section in the far field of diffraction. The arrows show the relative amplitude and direction of the linear polarization of the field at the point with the corresponding x and y coordinates; the latter are normalized to the half-width $r_0(z)$ of the Gaussian mode of the pump beam. Above the figure is schematically illustrated the transverse modal structure of the pump beam, in which the zeroth (Gaussian) mode, which is homogeneous and linearly polarized along the x axis, predominates, and which also contains a small admixture of the first order mode, the direction of whose linear polarization at each point is perpendicular to the axis of the beam.

4. FORMULA FOR THE SHG SIGNAL POWER AND DISCUSSION OF THE RESULTS

We start by obtaining expressions that connect the pump beam power W with the amplitudes of its three transverse modes. As a result of integrating the intensity of a wave with frequency ω, the expression for which can easily be found from Eqs. (4)–(6), it is found that over the beam cross section W equals the sum of the powers of these three modes W_G, $W^{X,Y}$ (i.e., these three modes are in some sense orthogonal):

$$W = \frac{cn}{8\pi} \int\int |\mathbf{E}(\mathbf{r})|^2 d\mathbf{r}_\perp = W_G + W^X + W^Y, \tag{14}$$

where

$$W_G = \frac{cn}{16} w^2 |E_0|^2, \quad W^{X,Y} = \frac{cn}{64} w^2 |E_0^{X,Y}|^2. \tag{15}$$

All the powers in Eqs. (14) and (15) are independent of z, which is associated with the use of the constant-pump approximation.

The total SHG signal power is obtained similarly to integrating the square of the slow field amplitude $\mathbf{A}^{(2)}_{\mathrm{vor}}(\mathbf{r})$ at the doubled frequency, given by Eq. (12), over the beam cross section:

$$W^{(2)}(z) = \frac{128\pi^2 k^4}{cn_{SH}\varepsilon^2} |\gamma_3^{(2)}|^2 \{W_G W^X |\mathbf{e}_0^X \times \mathbf{e}_0|^2 + W_G W^Y |\mathbf{e}_0^Y$$

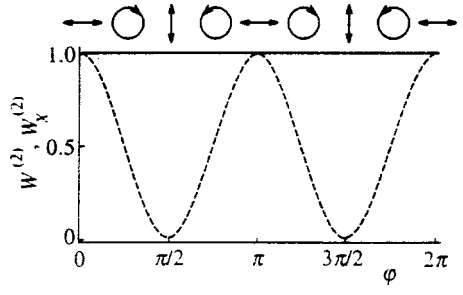


FIG. 3. Normalized dependence of the total power $W^{(2)}$ of the signal wave (solid curve) and the power $W_X^{(2)}$ of its x-polarized component (dashed curve) vs parameter φ, which determines the polarization state of the Gaussian transverse mode of the pump beam according to the formula $\mathbf{e}_0 = \mathbf{e}_x \cos\varphi + i\mathbf{e}_y \sin\varphi$, where $\mathbf{e}_0$ is the polarization vector of this Gaussian mode, and $\mathbf{e}_{x,y}$ are the unit vectors along the x,y axes. The polarization states of the Gaussian component of the pump beam for different φ are schematically shown in the upper part of the figure. It was assumed that the first-order transverse mode of the beam at the fundamental frequency is significantly weaker than the Gaussian component, while the direction of linear polarization of its electric field at any point of the cross section and for any φ is perpendicular to the beam axis.

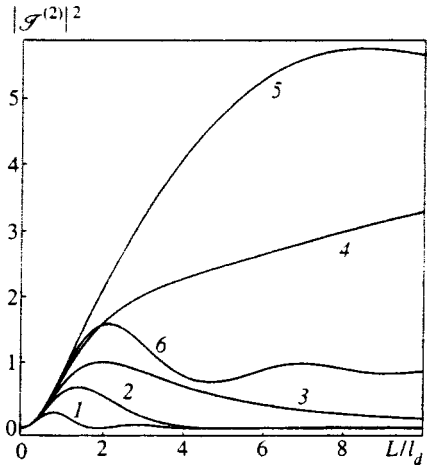


FIG. 4. Dependence of $|\mathscr{T}^{(2)}(-L/2,L/2)|^2$, the dimensionless power of the SHG signal when the pump beam is focused at the center of the medium, on the length L of the medium, normalized to the diffraction length l_d of the pump beam. The values of the parameter $\nu = \Delta k l_d$, which characterizes the detuning Δk of the wave vectors for the SHG process, were taken as follows: -2 (curve *1*), -0.5 (*2*), 0 (*3*), 0.5 (*4*), 1 (*5*), and 3 (*6*).

$$\times \mathbf{e}_0|^2 + W^X W^Y |\mathbf{e}_0^X \times \mathbf{e}_0^Y|^2\} |\mathscr{T}^{(2)}(-l_0, z-l_0)|^2. \tag{16}$$

If an analyzer that transmits only the x-polarized component of the electric field of the signal wave is put at the output from the medium (as was done in the experiment described in Ref. 14), the power of the transmitted beam at the doubled frequency will be determined by

$$W_X^{(2)}(z) = \frac{128\pi^2 k^4}{c n_{SH} \varepsilon^2} |\gamma_3^{(2)}|^2 \left\{ W_G W^Y |\mathbf{e}_0^Y \times \mathbf{e}_0|^2 + \frac{1}{2} W^X W^Y |\mathbf{e}_0^X \times \mathbf{e}_0^Y|^2 \right\} |\mathscr{T}^{(2)}(-l_0, z-l_0)|^2. \tag{17}$$

In Eqs. (16) and (17), the factors in braces describe the dependence of the SHG signal power on the polarization states of the three transverse modes of the pump beam, while the square of the absolute value of the dimensionless integral $\mathscr{T}^{(2)}$ determines how $W^{(2)}$ and $W_X^{(2)}$ depend on the z coordinate and the position l_0 of the beam waist.

Figure 3 shows the normalized dependence of the powers $W^{(2)}$ and $W_X^{(2)}$ on parameter φ, which gives the polarization state of the Gaussian component of the pump beam according to

$$\mathbf{e}_0 = \mathbf{e}_x \cos\varphi + i\mathbf{e}_y \sin\varphi.$$

For clarity, polarization states of this Gaussian mode corresponding to certain values of φ are shown symbolically plotted against φ. Horizontal arrows show the electric field linearly polarized along the x axis, vertical arrows show it linearly polarized along the y axis, and circles with an arrow correspond to right- or left-circular polarizations. In constructing Fig. 3, we assumed that the two transverse modes $\mathbf{E}^{X,Y}(\mathbf{r})$ have constant (independent of φ) speckle polarization with $W^{X,Y} = 0.01 W_G$, which corresponds to the ratio of the amplitudes of these modes used in constructing Fig. 2.

The theoretical curves shown in Fig. 3 are in qualitative agreement with the recently obtained experimental dependence of the signal power of bulk SHG on the polarization state of the pump beam,[14] although there is no detailed quantitative agreement between them. In fact, in Ref. 14, besides signal components that are quadratic in the field, a sizable fourth-order component relative to the pump-wave field was also observed. The latter results from the presence of nonlinear optical susceptibility $\hat{\chi}^{(4)}$ in an isotropic noncentrosymmetric medium, but is also possible because of certain cascade processes. The indicated effects have remained virtually uninvestigated theoretically, the more so when spatially bounded wave packets are involved. This is why it is not currently possible for us to give a definitive interpretation of the experimental results obtained in Ref. 14. Nevertheless, it is noteworthy that an SHG signal is present in the dependences shown in Fig. 3 even for circular polarization of the pump wave (in our case, of course, we are dealing with its Gaussian fundamental mode). This effect was observed experimentally,[11,14] but it has not yet been theoretically explained in relation to Ref. 14. The theoretical calculations carried out in this article make it possible to explain the appearance of such a ''doubly forbidden'' signal (according to the expression of Allcock *et al.*[11] and Balakin *et al.*[14]) at the doubled frequency, created in the interior of an isotropic medium by an (almost) circularly polarized pump beam.

The four remaining figures are constructed assuming $z = L$; i.e., the SHG signal at the output from the medium is considered. These figures show how the dimensionless coefficient $|\mathscr{T}^{(2)}|^2$ entering into Eqs. (16) and (17) depends on the length L of the medium, on the coordinate l_0 of the plane of the pump-beam waist, and on the parameter $\nu = \Delta k l_d$ that characterizes the detuning of the wave vectors. We recall that $|\mathscr{T}^{(2)}|^2$ completely determines the dependence of the signal-wave power on L, l_0, and ν.

Figure 4 shows $|\mathscr{T}^{(2)}(-L/2,L/2)|^2$ vs. L for the following values of ν: -2 (curve *1*), -0.5 (*2*), 0 (*3*), 0.5 (*4*), 1 (*5*),

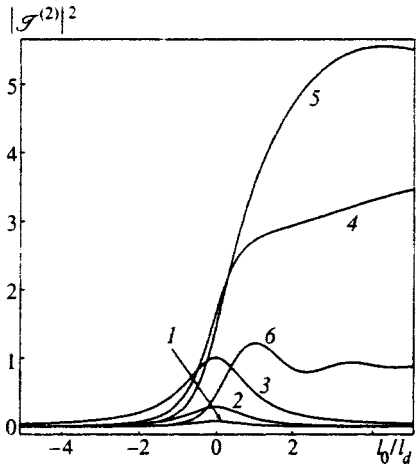


FIG. 5. Dependences of $|\mathscr{T}^{(2)}(-l_0,+\infty)|^2$, the dimensionless power of the wave at the doubled frequency, vs the distance l_0 from the plane of the pump-beam waist to the surface of a long medium, normalized to the diffraction length l_d of the pump beam. The values of parameter ν are the same as in constructing the corresponding curves in Fig. 4.

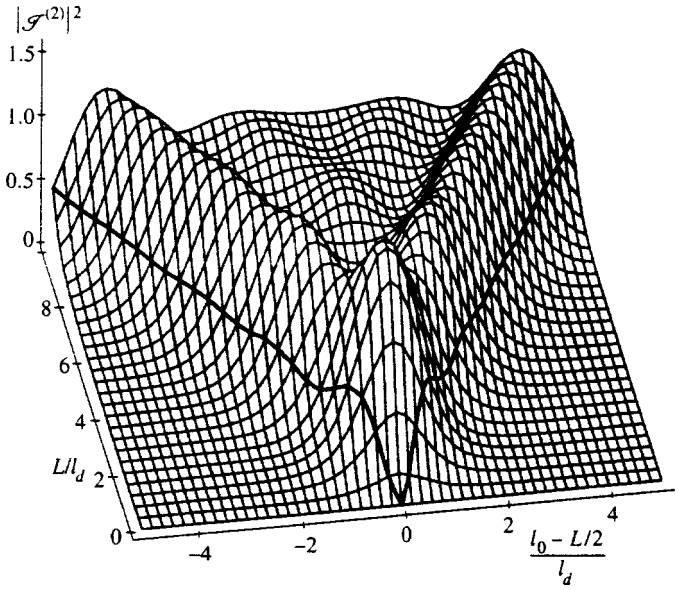


FIG. 6. The dimensionless power $|\mathscr{T}^{(2)}(-l_0,L-l_0)|^2$ of the SHG signal vs the length L of the medium and vs the distance $l_0-L/2$ from the pump beam waist to the center of the medium, expressed in diffraction lengths l_d of the pump beam. The two heavy curves correspond to those L and $l_0-L/2$ values for which the plane of the pump beam waist coincides with one of the surfaces of the medium. The figure is drawn for a dimensionless detuning of the wave vectors of $\nu=3$.

and 3 (*6*). The given case corresponds to focusing of the pump beam at the center of the medium ($l_0=L/2$) for all values of L. Note that, for $\nu\approx 1$, a sort of quasisynchronism occurs, characterizing the increase of the relative SHG signal amplitude. We shall discuss this fact in more detail below when studying the dependence of $\mathscr{T}^{(2)}$ on the parameter ν. Figure 5 shows the dependence of $|\mathscr{T}^{(2)}(-l_0,+\infty)|^2$ on l_0 for the same values of ν. Of course, under real conditions, the length of the medium is finite; here it only important to us that it be so large that $|\mathscr{T}^{(2)}|^2$ differs little from its asymptotic value as $L\to+\infty$. As can be understood from Fig. 4, to satisfy this requirement, it is sufficient to take L equal to several tens of diffraction lengths l_d.

Figures 4 and 5 show that the larger part of the SHG signal, as expected, is formed in the region of the pump beam waist. Actually, increasing the length L of the medium beyond several diffraction lengths no longer increases the power of the wave at the doubled frequency, and, if we bring the region of the beam waist beyond the boundary of the medium (negative l_0 in Fig. 5), the relative value of $W^{(2)}$ sharply decreases. Because of interference effects, it turns out that the maximum SHG signal power from the volume can be attained by placing the pump beam waist close to the surface of the medium (the $z=0$ plane) zero to several l_d away, depending on the value of ν). Although such a requirement on the geometry of the problem is not critical for the appearance of the second harmonic when ν is positive, for $\nu\leqslant 0$ the power of the wave at the doubled frequency formed in the volume of a (long) medium falls to zero when the plane of the beam waist is significantly remote from the surface of the medium. All these circumstances should be taken into account in setting up the experiment. To choose the optimum geometry of the latter, it can be useful to construct the three-dimensional dependence of the SHG signal power on L and l_0 in order to find the values of the parameters for which the power is a maximum. Figure 6 shows the dependence of the dimensionless coefficient $|\mathscr{T}^{(2)}(-l_0,L-l_0)|^2$ on the length L of the medium and on the distance $l_0-L/2$ from the pump beam waist to the center of the medium, plotted for $\nu=3$. The two heavy curves in Fig. 6 correspond to the values L and $l_0-L/2$ for which this plane of the beam waist coincides with one of the surfaces of the medium. Figure 6 clearly confirms the conclusions that we reached above on the basis of an analysis of Figs. 4 and 5.

Figure 7 illustrates the dependence of the SHG signal power on the parameter $\nu=\Delta k l_d$, which characterizes the wave vectors. The solid curve corresponds to a position of the pump beam waist well within a fairly long medium (i.e., the distance from the plane of the waist to both surfaces is at least several tens of diffraction lengths—see the explanation to Fig. 5). The dashed curve is plotted for the case in which

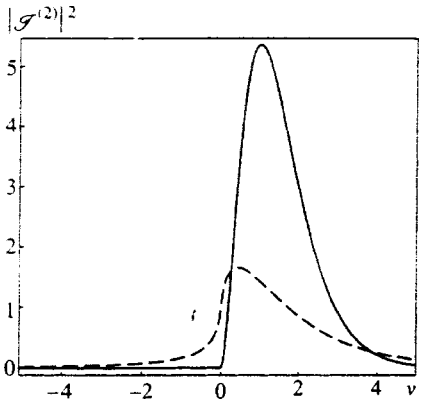


FIG. 7. Dimensionless power of the signal wave when the pump beam is focused at the center of a long medium ($|\mathscr{T}^{(2)}(-\infty,+\infty)|^2$, solid curve) and when it is focused on the input surface of a long medium ($|\mathscr{T}^{(2)}(0,+\infty)|^2$, dashed curve) vs the parameter $\nu=\Delta k l_d$, which characterizes the wave-vector mismatch Δk (l_d is the diffraction length of the pump beam).

the plane of the waist coincides with the input surface (the $z=0$ plane) of a fairly long medium. An interesting feature of the phenomenon we have investigated that is indicated by the form of the solid curve in Fig. 7 is that when the pump beam waist lies in the depth of the medium the SHG signal completely disappears for $\nu \leqslant 0$, i.e., $n_{SH} \geqslant n$, which corresponds to normal dispersion in the frequency region from ω to 2ω. The analytical formula for the indicated dependence is given by the square of the following expression:

$$\mathscr{T}^{(2)}(-\infty,+\infty)=\begin{cases} 2\pi\nu e^{-\nu}, & \nu \geqslant 0. \\ 0, & \nu < 0. \end{cases} \tag{18}$$

The disappearance of the SHG signal for negative ν is caused by the destructive interference of waves at the doubled frequency generated at different points of the medium illuminated by the pump beam. For $\nu > 0$ this interference becomes constructive and we can see the signal wave even when the pump beam waist lies deep within the medium. When the plane of the beam waist approaches the surface of the medium, part of the sources of the wave at the second harmonic as it were disappear, which makes it possible to partially escape the destructive influence of the interference. That is why, when the pump beam waist lies at the surface of the medium, the SHG signal appears even when ν is negative, as is demonstrated by the dashed curve in Fig. 7.

Of course, the signal in the $\nu < 0$ case can be said to completely disappear when the pump beam is focused in the interior of the medium only to within the approximations used in this treatment. In practice, only a significant weakening of the wave at the doubled frequency should be expected. Moreover, at very large values of $|\nu|$, when the inequality $|\Delta k|/k \ll 1$ breaks down, our formulas can become invalid because the condition given in inequality (11) is not obeyed and the method of slowly varying amplitudes is not applicable. However, the latter limitations do not seem crucial to us, since, when $|\nu|$ increases significantly, it is impossible to obtain any perceptible SHG signal because of the sharp decrease of the coherent-interaction length for this nonlinear optical process.

The solid curve in Fig. 7 reaches its maximum at $\nu = 1$. For this value of ν, a sort of quasisynchronism is observed, for which the pump radiation is transformed into the second harmonic with the greatest efficiency. This quasisynchronism, which appears for nonzero detuning of the wave vector mismatch, was noted in the references cited above[20–22] devoted to the study of four-wave processes in focused Gaussian beams. Moreover, Eq. (18) is completely analogous to the corresponding expression obtained in the indicated articles for the case of third-harmonic generation. This fact is a consequence of the mathematical similarity noted in Sec. 3 between certain formulas that describe two extremely different problems. We shall present a qualitative explanation of the phenomenon of quasisynchronism for SHG in a focused beam, based on the discussions of Ward and New.[20] The wave vectors of the spatial Fourier harmonics of the pump beam are tilted both with respect to the axis of the beam and relative to each other. Each pair of these plane-wave harmonics generates its own contribution to the signal wave at the doubled frequency. Phase synchronism is attained in this case at some nonnegative mismatch of the moduli of the wave vectors ($2k \geqslant k_{SH}$), since the interaction of the two spatial Fourier harmonics of the pump beam is noncollinear in general. The mean angle between the wave vectors of the different harmonics of a beam with half-width w is on the order of $1/kw$, and therefore phase quasisynchronism for the SHG process is attained when the wave-vector mismatch is

$$\Delta k \sim 2k[1-\cos(1/kw)] \sim 1/kw^2 \sim 1/l_d,$$

which agrees well in order of magnitude with the exact value $\nu = \Delta k l_d = 1$ obtained above.

5. CONCLUSION

This paper has presented the first study of SHG caused in the interior of an isotropic medium by spatial dispersion of the quadratic optical nonlinearity. A coherent signal at the doubled frequency in this case results from the interaction of different transverse modes of a pump beam possessing polarization that varies over the cross section. It can be shown that the described SHG mechanism is unrealistic because the signal amplitude is directly proportional to three small parameters: the spatial-dispersion parameter, the divergence angle of the pump beam, and the quantity $\sqrt{W^{X,Y}/W}$, which characterizes the fraction of higher-order transverse modes in the total pump power. As a matter of fact, the detection of such a weak signal is by no means a hopeless task, since, as shown in this article, more effective mechanisms for the appearance of a quadratic-in-field SHG signal in the volume of an isotropic medium are forbidden for one reason or another. Indeed, the local electric-dipole optical susceptibility $\hat{\chi}^{(2)}$ of an isotropic medium equals zero even in the absence of a inversion center in the medium. Spatial dispersion of the quadratic optical response of the substance also cannot cause a volume SHG signal if the pumping is represented by a plane light wave or by a Gaussian beam homogeneously polarized over the cross section. This is because the polarization wave formed in the medium at the doubled frequency has a pure irrotational character and cannot generate a free electromagnetic wave in the interior of the medium. A study of the dependence of the signal power at the second harmonic on the wave-vector mismatch Δk showed that quasisynchronism is observed for some nonzero value $\Delta k > 0$ inversely proportional to the diffraction length l_d of the beam. The SHG process occurs most efficiently close to this value, and therefore choosing the optimum Δk and the optimum degree of focusing of the pump beam can make it significantly easier to experimentally observe the effect. Moreover, we have investigated in great detail the question of optimizing the other geometrical parameters of the problem.

We hope that our work will promote further advances in experimental studies of SHG in the interior of isotropic liquids. It would be desirable in subsequent experiments to separate the signal into two components quadratic and fourth order in the field. In this case, the characteristics of the part of the beam at the doubled frequency that is quadratic in field (i.e., the dependence of the SHG signal power on the polarization states of the pump beam and the signal wave, the

transverse distributions of the intensity and polarization state of the latter, and the dependence of the SHG signal power on the parameters that specify the experimental geometry) could be correlated with the formulas obtained in this paper. This would make it possible to check the validity of our hypothesis that the second harmonic in the volume of the medium results from the presence of higher-order transverse modes in the pump beam in the case that we have considered. No less important is further theoretical study of SHG in the interior of an isotropic noncentrosymmetric medium. In particular, the authors regard as the next problem the study of the signal at the doubled frequency, whose occurrence is due to the fourth-order optical susceptibility $\hat{\chi}^{(4)}$. Interesting results can be obtained by investigating the contribution of certain cascade nonlinear optical processes to the experimentally observed SHG signal. The solution of these problems would make it possible, in our opinion, to completely interpret the experimental data—both those already waiting to be explained, and those that will certainly be obtained in the near future.

The authors are grateful to A. P. Shkurinov and A. V. Pakulev for useful discussions, as well as to the Russian Fund for Fundamental Research for partial financial support of our work (Project Code 96-02-16596).

*)E-mail: koroteev@nik.phys.msu.su

1)The concepts of longitudinal or transverse polarization of a vector field are suitable only for the case of plane waves. The beam can be represented as a superposition of plane-wave spatial Fourier harmonics whose wave vectors are slightly noncollinear, and therefore, even if the field in the beam is polarized across its axis, it still has a small component along almost all of these wave vectors. A fairly obvious generalization of the concept of longitudinal or transverse plane electromagnetic waves is provided by irrotational or solenoidal beams, given, respectively, by the conditions curl $\mathbf{E}=0$ and div $\mathbf{E}=0$. It is easy to show that with this definition each irrotational or solenoidal spatial Fourier harmonic of the beam is, respectively, a longitudinal or a transverse plane wave, and therefore all the specific properties of longitudinal or transverse electromagnetic waves will be inherited by irrotational or solenoidal beams, respectively. Note that an irrotational beam has a small field component directed across its axis, while a solenoidal beam has a small longitudinal component. These components are first-order in the beam-divergence angle. The concepts of irrotational and solenoidal can be applied to the wave beams of any vector quantity, including the wave of the polarization of a medium.

[1] Y. R. Shen, *Principles of Nonlinear Optics* (Wiley, New York, 1984; Nauka, Moscow, 1989).
[2] S. A. Akhmanov, N. I. Koroteev, and I. L. Shumay, *Nonlinear Optical Diagnostics of Laser-Excited Semiconductor Surfaces* (Harwood Academic Press, Chur, 1989).
[3] Y. R. Shen, Nature **337**, 519 (1989).
[4] J. A. Giordmaine, Phys. Rev. A **138**, 1599 (1965).
[5] N. I. Koroteev, Zh. Éksp. Teor. Fiz. **106**, 1260 (1994) [JETP **79**, 681 (1994)].
[6] A. V. Dubrovskiĭ, N. I. Koroteev, and A. P. Shkurinov, JETP Lett. **56**, 551 (1992).
[7] A. P. Shkurinov, A. V. Dubrovskiĭ, and N. I. Koroteev, Phys. Rev. Lett. **70**, 1085 (1993).
[8] P. K. Schmidt and G. W. Rayfield, Appl. Opt. **33**, 4286 (1994).
[9] Q. Song, C. Wan, and C. K. Johnson, J. Phys. Chem. **98**, 1999 (1994).
[10] E. Hendrickx, K. Clays, A. Persoons *et al.*, J. Am. Chem. Soc. **117**, 3547 (1995).
[11] P. Allcock, D. L. Andrews, S. R. Meech, and A. J. Wigman, Phys. Rev. A **53**, 2788 (1996).
[12] A. V. Balakin, D. Boucher, N. I. Koroteev *et al.*, JETP Lett. **64**, 718 (1996).
[13] A. V. Balakin, D. Boucher, E. Fertein *et al.*, Opt. Commun. **141**, 343 (1997).
[14] A. V. Balakin, D. Boucher, N. I. Koroteev *et al.*, Zh. Éksp. Teor. Fiz. **112**, 97 (1997) [JETP **85**, 52 (1997)].
[15] D. L. Andrews, P. Allcock, and A. A. Demidov, Chem. Phys. **190**, 1 (1995).
[16] S. R. de Groot and L. G. Suttorp, *Foundations of Electrodynamics* (Elsevier, New York, 1973; Nauka, Moscow, 1982), chap. 2.
[17] S. N. Volkov, N. I. Konovalov, N. I. Koroteev, and V. A. Makarov, Kvant. Elektron. (Moscow) **22**, 71 (1995).
[18] S. N. Volkov, N. I. Koroteev, and V. A. Makarov, Kvant. Elektron. (Moscow) **22**, 1220 (1995).
[19] N. I. Koroteev, V. A. Makarov, and S. N. Volkov, Nonlinear Opt. Princ. Mater. Phenom. Devices **17**, 247 (1997).
[20] J. F. Ward and G. H. C. New, Phys. Rev. **185**, 57 (1969).
[21] R. B. Miles and S. E. Harris, IEEE J. Quantum Electron. **QE-9**, 470 (1972).
[22] G. Bjorklund, IEEE J. Quantum Electron. **QE-11**, 287 (1975).

Translated by W. J. Manthey

Trapping of an electromagnetic wave by the boundary of created plasma

M. I. Bakunov*) and S. N. Zhukov

N. I. Lobachevskiĭ Nizhniĭ Novgorod State University, 603600 Nizhniĭ Novgorod, Russia
(Submitted 16 June 1997)
Zh. Éksp. Teor. Fiz. **113**, 1277–1288 (April 1998)

We discuss a new phenomenon of the electrodynamics of transient media, the trapping of electromagnetic radiation by the boundary of a transient plasma due to the conversion of the radiation into surface waves localized at the boundary. Calculations are done for an initial plane wave and for a beam of finite width in conditions where the boundary of the suddenly created (because of ionization) plasma half-space is perpendicular to the initial wavefront. Two frequency down-shifted surface waves traveling along the boundary in opposite directions are shown to be excited, as well as frequency up-shifted outgoing radiation and a time-independent mode in the form of a spatially inhomogeneous structure of dc currents and a magnetic field within the plasma half-space. We study the associated kinematic, amplitude, and energy relations. Finally, we establish that the most efficient trapping (up to 40% in energy) can be achieved with the forward (with respect to the direction of the initial wave propagation) surface mode and that the trapping is accompanied by concentration of electromagnetic energy at the plasma boundary. © *1998 American Institute of Physics.* [S1063-7761(98)01104-4]

1. INTRODUCTION

Recently there has been an upsurge of interest in studies of transformations of electromagnetic waves in nonstationary media (see, e.g., Refs. 1–8). The reason lies largely in the successful experiments in conversion of the frequency of microwave radiation in devices with transient plasma[9] (at an ionization front moving in the plasma), and in experiments on simultaneous ionization of the entire gaseous medium in the active volume (flash ionization).[10,11] In a recent review by Faĭnberg[12] this avenue of research in quantum electronics was mentioned as one of the most rapidly developing.

As for the theoretical work in the field of the electrodynamics in transient media, the bulk of it, starting with the pioneering work of Einstein[13] and Morgenthaler,[14] has developed in two directions: the study of the interaction of electromagnetic waves and of moving inhomogeneities (the medium's boundary or a parameter wave in an immobile medium),[2,3,5,13,15–18] and the study of conversion of waves in a homogeneous infinite medium with purely temporal variations of the parameters.[7,8,14,19–21] Much less developed is the theory of the interaction of electromagnetic radiation with bounded transient media. The first to study some features of the reflection of electromagnetic waves from the planar boundary of a transient medium (nondispersive insulators and plasmas) was Fante[22] in 1971. Borisov[23,24] and Kalluri[25] thoroughly studied steady-state and transient processes in the interaction of an electromagnetic wave and a suddenly created (because of ionization) plasma half-space (for a transient insulator half-space the necessary calculations were done by Nerukh).[26] A more realistic case of sudden ionization in a plasma slab was examined by Kalluri and Goteti.[27]

All this work, however, touched only on problems with the simplest geometry, where the boundary of the created plasma was assumed to be parallel to the wavefront. A richer case from a physical standpoint is the absence of the condition that the wavefront be parallel to the plasma's boundary. Indeed, in this case, with a TM polarized initial wave, we expect a conversion of this wave into surface waves driven by the created plasma boundary. At a steady-state boundary there can be no direct conversion of the incident wave into surface waves, because it is impossible to satisfy two conservation laws: of the wave's frequency (the photon energy), and of the component of the wave vector (momentum) tangential to the boundary.[28] Existing methods of feeding electromagnetic radiation to planar waveguide structures are based upon the creation of a longitudinal spatial inhomogeneity in the system (via a diaphragm, a smoothly profiled or rough section, etc.), or upon spatial synchronization of the volume and surface waves via coupling prisms (the method of frustrated total internal reflection).[28,29] For a transient boundary there exists an entirely different way of inputting radiation, since here the requirement that the wave frequency be conserved does not hold.

In the present paper we demonstrate the effect of conversion of electromagnetic radiation into surface waves at a transient plasma boundary for the case where the boundary of the suddenly created (because of ionization) plasma half-space is orthogonal to the initial wavefront (this geometry considerably simplifies the analysis of the energy relations in the conversion). We show that two frequency down-shifted surface waves traveling along the boundary in opposite directions are excited. We establish that the most efficient conversion (up to 40% in energy) can be achieved for the forward (with respect to the direction of the initial wave propagation) surface mode. Trapping of the electromagnetic wave by the boundary of the created plasma is accompanied by dissipation of a fraction of the energy into volume radia-

tion with a continuous frequency spectrum, moving away from the boundary, and by excitation of a time-independent mode in the form of a spatially inhomogeneous structure of dc currents and a magnetic field within the plasma half-space. We also find that the energy density at the boundary in the forward surface wave is much higher than the energy density in the initial wave, over a wide range of densities of the created plasma, i.e, the trapping is accompanied by a concentration of electromagnetic energy at the plasma boundary.

Some preliminary results of the present study were reported in Ref. 30.

2. STATEMENT OF THE PROBLEM. CALCULATION OF LAPLACE TRANSFORMS

We assume that initially, at $t<0$, in an nonionized medium (gas) with a dielectric constant $\varepsilon\approx 1$ there is a linearly polarized plane electromagnetic wave of frequency ω_0 with fields

$$E_y(x,t)=B_z(x,t),$$
$$B_z(x,t)=B_0\ \exp(i\omega_0 t-ih_0 x),\quad h_0=\frac{\omega_0}{c}, \tag{1}$$

propagating along the x axis.

At time $t=0$ external factors suddenly ionize the half-space $y<0$, with the result that cold collisionless plasma of concentration N is formed in that half-space. We begin our calculation of the conversion of the wave (1) at the given space–time inhomogeneity by using Maxwell's equations, in which at time $t=0$ the current of the produced electrons is switched on, along with the equations of electron motion, with the electrons having zero initial velocity.[15,16] Applying the Laplace transformation to the given system of equations, we arrive at an equation for the transform of the magnetic field, $b(x,y,p)$ (here p is the complex-valued ''Laplace'' variable):

$$\varepsilon\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial b}{\partial y}\right)-\frac{1}{c^2}(p^2\varepsilon+\omega_0^2)b$$
$$=-\frac{1}{c^2}(p\varepsilon+i\omega_0)B_0\ \exp(-ih_0 x), \tag{2}$$

where $\varepsilon(y,p)=1+\eta(-y)\omega_p^2/p^2$, $\eta(-y)$ is the unit step function, and $\omega_p=\sqrt{4\pi e^2 N/m}$ is the plasma frequency. Finding the solutions of Eq. (2) in the homogeneous regions $y>0$ and $y<0$ satisfying the radiation conditions, and matching them at the boundary $y=0$ by requiring that b and $\varepsilon^{-1}\partial b/\partial y$ be continuous at the boundary, we arrive at the following expressions for $b(x,y,p)$:

$$b(x,y,p)=\frac{B_0\ \exp(-ih_0 x)}{p-i\omega_0}+A_{\rm I}(p)$$
$$\times\exp\left(-ih_0 x-\frac{y}{c}\sqrt{p^2+\omega_0^2}\right) \tag{3}$$

for $y>0$, and

$$b(x,y,p)=\frac{p(p+i\omega_0)+\omega_p^2}{p(p^2+\omega_0^2+\omega_p^2)}B_0\ \exp(-ih_0 x)$$
$$+A_{\rm II}(p)\exp\left(-ih_0 x+\frac{y}{c}\sqrt{p^2+\omega_0^2\omega_p^2}\right) \tag{4}$$

for $y<0$, with

$$A_{\rm I}(p)=-B_0\frac{ip\omega_0\omega_p^2}{(p-i\omega_0)\sqrt{p^2+\omega_0^2+\omega_p^2 D(p)}}, \tag{5}$$

$$A_{\rm II}(p)=B_0\frac{i\omega_0\omega_p^2(p^2+\omega_p^2)\sqrt{p^2+\omega_0^2}}{p(p-i\omega_0)(p^2+\omega_0^2+\omega_p^2)D(p)}, \tag{6}$$

$$D(p)=(p^2+\omega_p^2)\sqrt{p^2+\omega_0^2}+p^2\sqrt{p^2+\omega_0^2+\omega_p^2}; \tag{7}$$

here we select the roots whose real parts are positive.

Under an inverse Laplace transformation, the first terms in (3) and (4), which are forced solutions of Eq. (2), contribute at any point in space to the field $B_z(x,y,t)$, starting at $t=0$. In the region $y>0$ this contribution has the form (1), i.e., coincides with the initial wave, while in the region $y<0$ it is determined by the formula

$$B_z(x,t)=B_0\ \exp(-ih_0 x)\left[\frac{\omega_p^2}{\omega_0^2+\omega_p^2}\right.$$
$$\left.+\sum_{\pm}\frac{\omega_0}{2}\,\frac{\omega_0\pm\sqrt{\omega_0^2+\omega_p^2}}{\omega_0^2+\omega_p^2}\exp(\pm it\sqrt{\omega_0^2+\omega_p^2})\right], \tag{8}$$

which agrees with the results of conversion of an electromagnetic wave under instantaneous ionization of the medium in the entire space.[20,21]

Under an inverse Laplace transformation, the second terms in (3) and (4) (free solutions) contribute to the field at a given point only starting at $t=|y|/c$. This means that starting at the moment of ionization $t=0$, the region of conversion of the wave fields (1) and (8) propagates in both directions from the emerging boundary $y=0$.

To calculate the steady-state solutions (as $t\to\infty$) and the angular distribution of the outgoing radiation ($y\to\pm\infty$), in applying the inverse Laplace transformation to the second terms in (3) and (4) it is convenient to select the integration contour as depicted in Fig. 1. The poles of the functions $A_{\rm I}(p)$ and $A_{\rm II}(p)$ contribute, together with (1) and (8), to the steady-state solutions, while the integrals along the edges of the cuts determine transient processes, including the radiation field. In Fig. 1 the poles $i\omega_0$ and $\pm i\sqrt{\omega_0^2+\omega_p^2}$ are depicted to the right of the cuts, as follows from the case where the initial wave is a wave beam of finite width rather than a plane wave. The contribution of the pole $i\omega_0$ of the function $A_{\rm I}(p)$ cancels (''turns off'') the initial wave (1) in the region $y>0$, and the contributions of the poles $\pm i\sqrt{\omega_0^2+\omega_p^2}$ of the function $A_{\rm II}(p)$ cancel the waves with frequencies $\pm\sqrt{\omega_0^2+\omega_p^2}$, the waves determined by (8), in the region $y<0$.

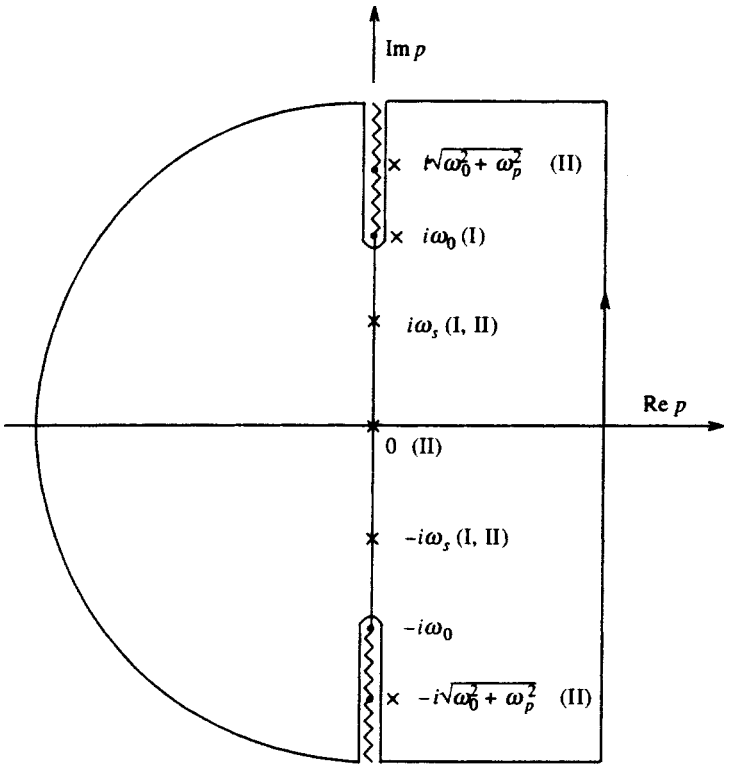


FIG. 1. The integration contour for $t>|y|/c$. The wavy lines represent the cuts ($\pm i\omega_0$ and $\pm i\sqrt{\omega_0^2+\omega_p^2}$ are the branch points of the roots), and $\times$ indicate the poles of the functions $A_{\rm I}(p)$ and $A_{\rm II}(p)$ (the corresponding label is in parentheses).

3. SURFACE WAVES

The effect of formation of steady-state solutions as surface waves is described by the contribution of the poles of $A_1(p)$ and $A_{\rm II}(p)$, which are determined by the equation $D(p)=0$. This equation has two solutions, $p=\pm i\omega_s$ (Fig. 1), where

$$\omega_s=\sqrt{\omega_0^2+\tfrac{1}{2}\omega_p^2-\sqrt{\omega_0^4+\tfrac{1}{4}\omega_p^4}}. \tag{9}$$

Thus, the electromagnetic wave trapped by the plasma boundary consists of two surface waves: one (with frequency $\omega_s>0$) propagating in the same direction as the initial plane wave, the other (with frequency $-\omega_s<0$) propagating in the opposite direction. This result can be illustrated by a kinematic diagram (Fig. 2). In this diagram the coordinates of the points of intersection of the two branches of the dispersion curve of the surface waves (see Ref. 31) and the straight line $h=h_0$ representing the invariance (at a temporal discontinuity) of the spatial structure of the waves in the boundary's plane correspond to the frequencies of the surface waves. Figure 2 shows that the conversion of the initial wave into surface waves entails a down-shift in frequency ($\omega_s<\omega_0$). The frequency conversion coefficient ω_s/ω_0 increases with the density of the created plasma, approaching unity when $N\gg N_c$, with $N_c=m\omega_0^2/4\pi e^2$ the critical density (Fig. 3). When $N\ll N_c$, the frequency ω_s is close to the upper limit on the frequency of the surface wave, $\omega_p/\sqrt{2}$ (see Fig. 2).[31]

The amplitudes $B_\pm$ of the surface waves are given by the residues of the functions $A_{\rm I}(p)$ and $A_{\rm II}(p)$ at the poles $p=\pm i\omega_s$ and have the following form:

$$B_\pm=B_0\frac{(\omega_p^2-\omega_s^2)(\omega_0\pm\omega_s)}{2\omega_0\sqrt{\omega_0^4+(1/4)\omega_p^2}}. \tag{10}$$

The amplitude of the backward wave is always smaller than that of the forward wave ($B_-<B_+$).

The amplitude conversion coefficients $B_\pm/B_0$ are depicted in Fig. 4 as functions of the density of the created plasma. The ratio B_-/B_0 reaches its maximum value $(B_-/B_0)_{\max}\approx 0.12$ at $N/N_c\approx 1.7$ and approaches zero as $N/N_c\to\infty$. The ratio B_+/B_0 increases monotonically with plasma density, approaching a limit of 2 as $N/N_c\to\infty$.

At first glance, the asymptotic behavior of the conversion coefficient B_+/B_0 seems to violate energy conservation. Indeed, in the limit $N/N_c\to\infty$, the ''plus'' surface wave becomes a plane wave with the same frequency ω_0 as the initial wave and propagates in vacuum ($y>0$) along the perfectly conducting half-space $y<0$. Since its amplitude is

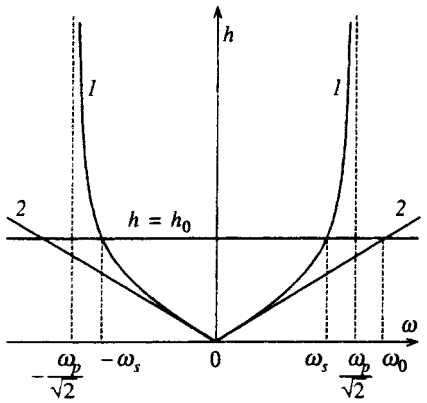


FIG. 2. The kinematic diagram for the surface wave: curves *1* represent the branches of the dispersion curve of the surface waves at the boundary of the plasma half-space, and curves *2* represent the light lines ($h=\pm\omega/c$). The points where the curves *1* intersect the straight line $h=h_0$ correspond to the frequencies of the trapped surface waves, $\pm\omega_s$.

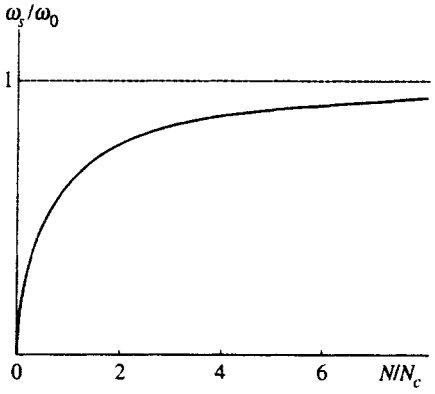


FIG. 3. Frequency conversion coefficient ω_s/ω_0 as a function of the density of the created plasma.

twice the amplitude of the initial wave, the energy density at each point of the region $y>0$ is quadrupled. Here the energy of the ''plus'' wave is twice the energy of the initial wave (it is assumed that the latter is a large but still spatially bounded wave packet). To resolve this contradiction, we must dismiss the ideal picture of a plane wave acting as an initial wave and must do the necessary calculations with a (wide but limited in y) wave beam. For the initial beam it is convenient to take

$$B_z(x,y,0)=B_0\frac{\sin(y/a)}{y/a}\exp(-ih_0x), \tag{11}$$

a beam of width $a\gg 1/h_0$ with a uniform spatial frequency spectrum $\Phi(\kappa)$:

$$B_z(x,y,0)=B_0\ \exp(-ih_0x)\int_{-\infty}^{\infty}\Phi(\kappa)e^{-i\kappa y}d\kappa, \tag{12}$$

$$\Phi(\kappa)=\begin{cases} a/2, & |\kappa|<1/a,\\ 0, & |\kappa|>1/a. \end{cases} \tag{13}$$

Substituting $B_z(x,y,0)$ from (12) for $B_0\exp(-ih_0x)$ on the right-hand side of Eq. (2) and proceeding in the same way as for an initial plane wave, we arrive at the following expression for the amplitude B_+:

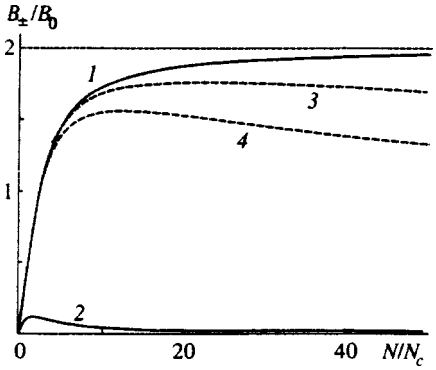


FIG. 4. Amplitude conversion coefficients of a plane wave into surface waves, B_+/B_0 (curve *1*) and B_-/B_0 (curve *2*), as functions of the density of the created plasma. The dashed curves represent the results of calculating B_+/B_0 for a wave beam with $h_0a=10$ (curve *3*) and $h_0a=5$ (curve *4*).

$$B_+=B_0\frac{(\omega_p^2-\omega_s^2)(\omega_0+\omega_s)}{2\omega_0\sqrt{\omega_0^4+(1/4)\omega_p^4}}\ \gamma\ \arctan\frac{1}{\gamma}, \tag{14}$$

where $\gamma=h_0a\sqrt{2(1-\omega_s/\omega_0)}$, $\omega_0=ch_0$. The results of calculations of the conversion coefficient B_+/B_0 via (14) are depicted by dashed curves in Fig. 4. Both Eq. (14) and Fig. 4 imply that $B_+/B_0\to 2$ as N/N_c increases as long as $N/N_c \ll (h_0a)^2$, i.e., as long as the region of localization of the surface wave in vacuum is small compared to the width a of the initial beam. When $N/N_c\geqslant(h_0a)^2$, i.e., when the transverse scale of the surface wave becomes comparable to the width of the initial beam and exceeds it, the conversion coefficient B_+/B_0 decreases, approaching zero as $N/N_c\to\infty$. As a result, no violation of energy conservation occurs (see Sec. 5).

4. STATIC MODE. ANGULAR DISTRIBUTION OF RADIATION

The static mode excited in the plasma half-space and determined by the contribution of the time-independent term in (8) and the residue at the pole $p=0$ (see Fig. 1) of the second term in (4) is also a steady-state solution (in addition to the surface waves discussed earlier):

$$B_z^{\rm st}(x,y)=\frac{\omega_p^2}{\omega_0^2+\omega_p^2}\left[B_z(x,y,0)-B_0\ \exp\left(\frac{y}{c}\sqrt{\omega_0^2+\omega_p^2}-ih_0x\right)\right],\quad y<0. \tag{15}$$

The magnetic field $B_z^{\rm st}(x,y)$ is associated with a spatial distribution of dc currents $\mathbf{j}^{\rm st}(x,y)=(c/4\pi)\nabla\times\mathbf{B}^{\rm st}$ in the plasma.

The radiation field is determined by the integrals along the right edges of the cuts (see Fig. 1) in the intervals $\omega_0<|\omega|<\infty$ for $y>0$ and $\sqrt{\omega_0^2+\omega_p^2}<|\omega|<\infty$ for $y<0$ ($\omega=\mathrm{Im}\ p$). Following the method developed in Refs. 32 and 33, we calculate the energy of the radiation field (per unit surface area of the boundary) in vacuum ($W_{\rm I}$) and plasma ($W_{\rm II}$) as $t\to\infty$ by integrating the electromagnetic energy density expressed in terms of the above integrals with respect to y. As a result we arrive at a representation

$$W_{\rm I,II}=\int_{-\pi/2}^{\pi/2}w_{\rm I,II}(\theta)\ d\theta \tag{16}$$

with angular density of the radiant energy

$$w_{\rm I,II}(\theta)=\frac{c\omega_0}{16\pi^2}\cot^2\theta|A_{\rm I,II}(p=i\omega)|^2, \tag{17}$$

where the radiation angle θ in vacuum and in the plasma is measured from the normal to the boundary ($-\pi/2<\theta<\pi/2$; $\theta=\pi/2$ corresponds to the direction of the x axis), and

$$\omega=\begin{cases}\omega_0/\sin\ \theta, & y>0,\\ \dfrac{1}{\sin\ \theta}\sqrt{\omega_p^2\ \sin^2\theta+\omega_0^2}, & y<0.\end{cases} \tag{18}$$

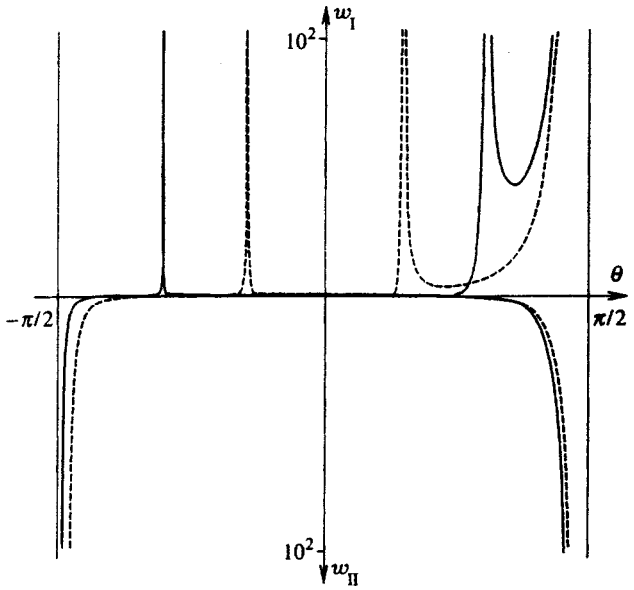


FIG. 5. Angular densities of the radiant energy in vacuum, $w_{\mathrm{I}}(\theta)$, and in the plasma, $w_{\mathrm{II}}(\theta)$, at $N/N_c=0.5$ (solid curves) and $N/N_c=4$ (dashed curves). All curves are normalized to the same quantity $cB_0^2/16\pi^2\omega_0$.

Negative frequencies correspond to backward radiation with respect to the direction of propagation of the initial wave (at angles $\theta<0$).

The function $w_{\mathrm{I,II}}(\theta)$ is plotted in Fig. 5 for two values of the density of the created plasma. The divergences of the functions w_{II} as $\theta\to\pm\pi/2$ ($w_{\mathrm{II}}(\theta)\propto(\pi/2\mp\theta)^{-2}$) and $w_{\mathrm{I}}(\theta)$ at the total reflection angles $\theta\to\pm\theta_*$, $\sin\theta_*=\omega_0/\sqrt{\omega_0^2+\omega_p^2}$ ($w_{\mathrm{I}}(\theta)\propto|\theta\mp\theta_*|^{-1}$) correspond to the emission of parts of the energies of the ''halves'' of the plane waves not trapped by the plasma boundary; these waves have frequencies $\pm\sqrt{\omega_0^2+\omega_p^2}$ and are generated in the plasma ($y<0$) immediately after creation of the plasma (second term in (8)). The divergence of the function $w_{\mathrm{I}}(\theta)$ as $\theta\to\pi/2$ ($w_{\mathrm{I}}(\theta)\propto(\pi/2-\theta)^{-2}$) corresponds to the emission of a part of the energy of the ''half'' (for $y>0$) of the plane wave of type (1) not trapped by the plasma boundary (the forced solution in the vacuum). When the initial beam (11) is involved, the divergences become logarithmic, which means that they are integrable.

5. ENERGY RELATIONS. THE EFFECT OF ENERGY CONCENTRATION

The plane-wave idealization is inadequate in studies of the energy efficiency of trapping of a wave by the created plasma boundary. We conduct our analysis with a beam of type (11). The energy of the initial beam (per unit surface area of the boundary), $W_0=B_0^2a/8$, is converted into the energy of surface waves,[31]

$$W_{\pm}=\frac{cB_{\pm}^2}{16\pi\omega_s}\,\frac{(1-\varepsilon_s)(1+\varepsilon_s^2)}{\varepsilon_s^2\sqrt{-1-\varepsilon_s}} \tag{19}$$

($\varepsilon_s=1-\omega_p^2/\omega_s^2$, and the amplitudes B_+ and B_- are defined in Eqs. (14) and (10), respectively), the radiant energy $W_r=W_{\mathrm{I}}+W_{\mathrm{II}}$ (see Eqs. (16) and (17)), and the energy of the static mode,

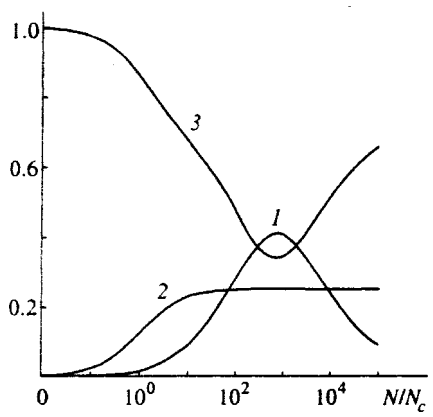


FIG. 6. Energy conversion coefficients W_+/W_0 (*1*), W_{st}/W_0 (*2*), and W_r/W_0 (*3*) as functions of the density of created plasma for an initial beam with $h_0a=20$.

$$W_{\mathrm{st}}=\frac{B_0^2a}{32}\,\frac{\omega_p^2}{\omega_p^2+\omega_0^2}, \tag{20}$$

which consists of the magnetic field energy (15) and the kinetic energy of the electrons calculated on the basis of the expression for the current $\mathbf{j}^{\mathrm{st}}(x,y)$. The following asymptotic formulas, for $N/N_c\to\infty$, have also proved useful:

$$\frac{W_{\pm}}{W_0}\to 0,\quad \frac{W_{\mathrm{st}}}{W_0}\to 0.25,\quad \frac{W_r}{W_0}\to 0.75. \tag{21}$$

Figure 6 depicts the dependence of the distribution of energy between the secondary solutions on the density of the created plasma for an initial beam with $h_0a=20$. Conversion to the backward surface wave has low efficiency: $(W_-/W_0)_{\max}\approx 7\times10^{-4}$ at $N/N_c\approx 1$, while the efficiency of conversion to the forward surface wave may exceed 40%:

$$\left(\frac{W_+}{W_0}\right)_{\max}\approx 0.41 \quad \text{at } \frac{N}{N_c}\approx 800. \tag{22}$$

The highest energy efficiency of trapping is achieved at $N/N_c\sim(h_0a)^2$, when the scale of localization of the forward surface wave in the vacuum, $(1/h_0)\sqrt{N/N_c}$, coincides in order of magnitude to the width a of the initial beam (cf. the results of Sec. 3). Here the depth of penetration of the plasma by the surface wave is small ($\sim(h_0^2a)^{-1}$) and the energy of the wave is localized mostly in vacuum. As h_0a decreases, maximum conversion efficiency is reached at lower densities of the created plasma and hence at greater localization of the surface wave in the vacuum.

It is important here that the trapping of electromagnetic radiation be accompanied by energy concentration near the created plasma boundary: the time averages of the energy density in the forward surface wave in the vacuum, $w_v(y=+0)$, and in the plasma, $w_p(y=-0)$, exceed the energy density in the initial wave, $w_0=B_0^2/8\pi$, over a wide range of concentrations of the created plasma (Fig. 7). In other words, when surface waves are produced by transient processes, energy flows toward the created plasma boundary.

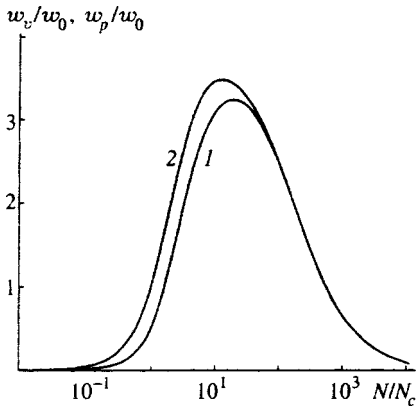


FIG. 7. Relative energy densities in the forward surface wave: curve *1*, in vacuum, w_v/w_0; and curve *2*, in the plasma, w_p/w_0.

6. CONCLUSION

We have used a plasma system with simple geometry to demonstrate the effect of conversion of electromagnetic radiation into surface waves at the boundary of a highly transient medium (with a characteristic ''switching'' time much shorter than the period of the radiation). We have found that the efficiency of such conversion (the trapping of radiation by the boundary) may be high (up to 40% in energy), with conversion accompanied by a down-shift in frequency of the waves. We have also studied the accompanying processes: scattering into frequency up-shifted volume radiation with a continuous spectrum and excitation of a static mode.

The trapping effect can be observed in experiments involving a device of the type described in Ref. 6, where it takes less than 1 ns to create transient plasma by sending an electric discharge between the plates of a long plane-parallel capacitor. If the created plasma layer is illuminated from the side by a beam of microwave radiation, the surface wave generated as a result of trapping can be detected at a considerable distance along the plasma layer from the spot where the beam ''illuminates'' the plasma.

The pattern of conversion of the electromagnetic wave in a bounded transient plasma, constructed within the scope of a parametric (linear) problem, may prove useful in studying nonlinear (self-consistent) regimes of radiation trapping in gas discharges in regions where high-power microwave beams overlap (as in the experiments discussed in Refs. 6 and 11) and in processes of formation of self-sustaining plasma structures that drive the surface waves created as a result of trapping (such structures were observed by Sola *et al.*[34]).

The most promising application of the trapping effect lies in developing entirely new methods of feeding radiation to planar waveguide structures with semiconductor layers. The transient nature of the semiconductor medium can be maintained through various mechanisms, such as carrier injection,[35], photoionization by a laser pulse,[36,37] and ''switching'' of the effective mass.[35] Since the attained characteristic transient times are shorter than 1 ps, these mechanisms can be used to achieve transient input of the radiation of submillimeter radiation, which plays an important role in microwave electronics. The simplest device of this kind incorporates a semiconductor waveguide layer in which the carrier concentration rapidly increases due to the action of a master laser pulse, which results in the trapping by that layer of a signal wave incident on the layer. Another possible device incorporates a waveguide layer that is steady-state (for example, an insulator), while the transient condition is created in a plasma film covering the part of the waveguide ''illuminated'' by the signal wave.

This work was supported by the Ministry of Education of the Russian Federation (Grant No. KTsFE 95-0-8.2-33).

*)E-mail: bakunov@rf.unn.runnet.ru

[1] C. H. Lai, R. Liou, T. C. Katsouleas, P. Muggli, R. Brogle, C. Joshi, and W. B. Mori, Phys. Rev. Lett. **77**, 4764 (1996).
[2] W. B. Mori, T. C. Katsouleas, J. M. Dawson, and C. H. Lai, Phys. Rev. Lett. **74**, 542 (1995).
[3] R. L. Savage Jr., R. P. Brogle, W. B. Mori, and C. Joshi, IEEE Trans. Plasma Sci. **21**, 5 (1993).
[4] V. B. Gildenburg, A. V. Kim, V. A. Krupnov *et al.*, IEEE Trans. Plasma Sci. **21**, 34 (1993).
[5] C. H. Lai, T. C. Katsouleas, W. B. Mori, and D. Whittum, IEEE Trans. Plasma Sci. **21**, 45 (1993).
[6] S. P. Kuo and A. Ren, IEEE Trans. Plasma Sci. **21**, 53 (1993).
[7] A. Banos Jr., W. B. Mori, and J. M. Dawson, IEEE Trans. Plasma Sci. **21**, 57 (1993).
[8] D. K. Kalluri, V. R. Goteti, and A. M. Sessler, IEEE Trans. Plasma Sci. **21**, 70 (1993).
[9] R. L. Savage Jr., C. Joshi, and W. B. Mori, Phys. Rev. Lett. **68**, 946 (1992).
[10] C. Joshi, C. E. Clayton, K. March *et al.*, IEEE Trans. Plasma Sci. **18**, 814 (1990).
[11] S. P. Kuo, Phys. Rev. Lett. **65**, 1000 (1990).
[12] Ya. B. Faĭnberg, Fiz. Plazmy **20**, 613 (1994) [Plasma Phys. Rep. **20**, 549 (1994)].
[13] A. Einstein, Ann. Phys. **17**, 891 (1905) [English translation in *Special Theory of Relativity*, C. W. Kilmister (ed.), Pergamon Press, Oxford (1970), pp. 891–921].
[14] F. R. Morgenthaler, IEEE Trans. Microwave Theory Tech. **MTT-6**, 167 (1958).
[15] G. I. Freĭdman, Zh. Éksp. Teor. Fiz. **41**, 226 (1961) [Sov. Phys. JETP **14**, 165 (1962)].
[16] L. A. Ostrovskiĭ and N. S. Stepanov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **14**, 489 (1971).
[17] S. N. Stolyarov, in *Einstein: Collected Papers 1975–1976* [in Russian], Nauka, Moscow (1978), p. 152.
[18] W. B. Mori, Phys. Rev. A **44**, 5118 (1961).
[19] L. B. Felsen and G. M. Whitman, IEEE Trans. Antennas Propag. **AP-18**, 242 (1970).
[20] C. L. Jiang, IEEE Trans. Antennas Propag. **AP-23**, 83 (1975).
[21] S. C. Wilks, J. M. Dawson, and W. B. Mori, Phys. Rev. Lett. **61**, 337 (1988).
[22] R. L. Fante, IEEE Trans. Antennas Propag. **AP-19**, 417 (1971).
[23] V. V. Borisov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **25**, 1053 (1982).
[24] V. V. Borisov, *Transient Electromagnetic Waves* [in Russian], Leningrad State Univ. Press, Leningrad (1987).
[25] D. K. Kalluri, IEEE Trans. Plasma Sci. **16**, 11 (1988).
[26] A. G. Nerukh, Radio Sci. **3**, 481 (1995).
[27] D. K. Kalluri and V. R. Goteti, J. Appl. Phys. **72**, 4575 (1992).
[28] V. M. Agranovich and V. L. Ginzburg, *Crystal Optics with Spatial Dispersion and Excitons*, 2nd ed., Springer, Berlin (1984), Sec. 11.
[29] V. M. Agranovich and D. L. Mills (eds), *Surface Polaritons*, North-Holland, Amsterdam (1982).
[30] M. I. Bakunov and S. I. Zaitsev, in *Abstract of 24th General Assembly of URSI*, Kyoto (1993), p. 448.
[31] A. N. Kondratenko, *Surface and Volume Waves in a Bounded Plasma* [in Russian], Énergoatomizdat, Moscow (1985), Sec. 1.2.
[32] V. L. Ginzburg and V. N. Tsytovich, *Transition Radiation and Transition Scattering (Theoretical Aspects)* [in Russian], Nauka, Moscow (1984), Sec. 2.2.

[33] V. L. Ginzburg, *Applications of Electrodynamics in Theoretical Physics and Astrophysics*, Gordon & Breach, New York (1989), Chap. 8.
[34] A. Sola, J. Cotrino, A. Gamero, and V. Colomer, J. Phys. D **20**, 1250 (1987).
[35] M. Kuijk and R. Vounckx, Electron. Lett. **25**, 231 (1989).
[36] K. H. Schoenbach, V. K. Lakdawala, R. Germer, and S. T. Ko, J. Appl. Phys. **63**, 2460 (1988).
[37] A. Kost, L. West, T. C. Hasenberg, J. O. White, M. Matloubian, and G. C. Valley, Appl. Phys. Lett. **63**, 3494 (1993).

Translated by Eugene Yankovsky

Ionization self-channeling of modulated plasma-wave beams in a magnetic field

G. A. Markov*)

Nizhegod State University, 603600 Nizhniĭ Novgorod, Russia
(Submitted 11 August 1997)
Zh. Éksp. Teor. Fiz. **113**, 1289–1298 (April 1998)

The dynamics of ionization self-channeling of modulated beams of plasma waves forming a solitary plasma-wave channel in an external magnetic field is investigated. It is shown that electromagnetic wave processes at the modulation frequencies of the ionizing radiation can be excited in the background plasma and in the channel. © *1998 American Institute of Physics.* [S1063-7761(98)01204-9]

1. INTRODUCTION

Ionization self-action of the fields of electromagnetic sources in the whistler frequency range ($\Omega_H < \omega < \omega_H$, where ω_H and Ω_H are the electron and ion gyrofrequencies) admits simultaneous concentration of the ionizing radiation and the resulting plasma.[1–4] As a result of such self-action, narrow plasma-wave channels, which localize the radiation energy, and the release of energy in a narrow magnetic field tube are observed.[2,3] The plasma inhomogeneity formed with such a discharge automatically[5] matches the short exciting antenna with the pump generator and the surrounding background plasma. This increases the efficiency of energy transfer from the source into the plasma in the field tube.[6] In the nighttime ionosphere, the plasma-wave channel[7–9] excited at altitudes $h \geqslant 150$ km by an on-board radio source is a source of powerful fluxes of hot particles along the flux lines of the geomagnetic field. It is evident from theoretical estimates that within several minutes hot particles from the discharge can fill a large portion of a magnetic field tube, resting on the discharge, and form an artificial magnetospheric duct.[10] As a result of this, the geophysical situation near a perturbed tube can change substantially.[8,9] For example, the conditions for the excitation, propagation, and reflection of whistler and Alfvén electromagnetic waves change.[8,11] The fluxes of precipitating charged particles from a perturbed tube change.[9] Modulation of the fluxes of precipitating high-energy ($\mathscr{E}_e \geqslant 40$ keV) electrons signifies excitation of an ac current along the entire length of the magnetospheric tube perturbed by the modulated discharge.[10] Under laboratory conditions such discharges, called helicons,[12–14] are of great interest for the development of ultraclean plasma chemical reactors and for investigating the nonlinear interaction between the fields of powerful HF sources and the resulting plasma.

The present paper reports the results of an experimental investigation of the dynamic processes of ionization self-channeling of plasma-wave beams produced in a magnetic field by a HF dipole source whose output is amplitude-modulated. It is shown that when a plasma channel forms, beams of fast electrons from the discharge are injected along the external magnetic field. The ionization of the neutral gas by the beam electrons allows the channel to grow in length. The rate of growth of the plasma-wave channel is determined by the dispersion of the wave fields producing and supporting the channel. The currents generated by the plasma-wave beams excite an azimuthal magnetic field and radial and longitudinal polarization electric fields of the plasma in the channel. When the HF source is amplitude-modulated, the plasma parameters and the dimensions of the plasma-wave channel are observed to vary with a frequency equal to the modulation frequency and LF wave fields are generated. The relaxation time of the LF oscillations was found to be adequate for observing echo signals after the modulation of the HF source is switched off.

2. EXPERIMENTAL CONDITIONS AND RESULTS

The experiments were performed in a glass discharge chamber 1800 mm long and 200 mm in diameter. The exciting rings of the dipole source (60 mm in diameter) were positioned at the end of the chamber coaxially with respect to the axis of the chamber and the longitudinal magnetic field and separated from one another by a distance $l = 60$ mm. A GST-2 generator supplied to the rings a HF voltage ($f_0 = 240$ MHz, $U_0 = 50$ V) which was amplitude-modulated at two frequencies $f_1 = 6$ kHz, $f_1 < f_2 \leqslant 70$ kHz. The modulation frequency f_2 could be varied over the range 6–70 kHz. The air pressure in the chamber was maintained at the level $p \leqslant 3 \times 10^{-4}$ torr. The magnitude B_0 of the external magnetic field could be varied from 60–750 G. The plasma parameters were measured with mobile Langmuir probes and a four-wall electrostatic analyzer positioned on the axis of the discharge chamber at the opposite end from the HF source. The HF radiation from the source loaded on the plasma channel was detected with a P6–22 measuring antenna and a S4-27 analyzer. The LF fields excited by the discharge were picked out with the S4-73 analyzer using electric (probe) antennas and a ferrite antenna in the form of a Rogowski loop. The dynamics of the emission of the discharge plasma was recorded with FD-21 KP photodiodes.

The structure of the HF discharge formed during ionization self-channeling of plasma waves contains a resonance cone (Fig. 1), which rests on the exciting rings, and a narrow filament (plasma-wave channel) extending from the focus of the resonance cone in the direction of the magnetic field $\mathbf{B}_0$.[2] At low pressures, when the electron mean free path is greater

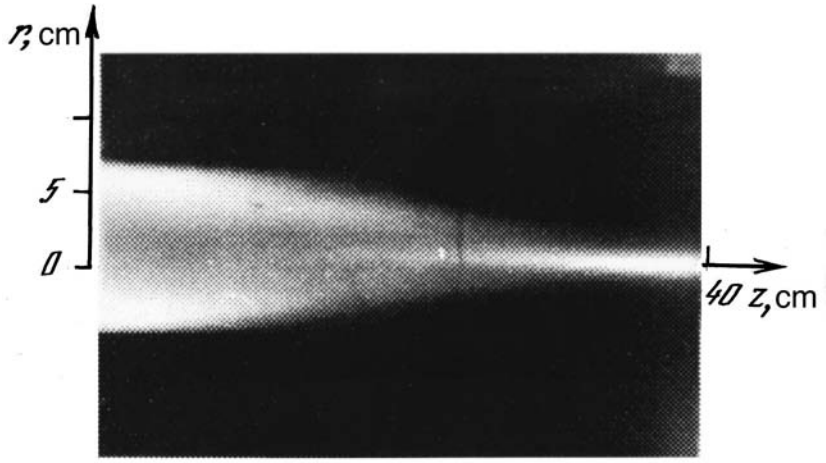


FIG. 1. Photograph of a HF discharge formed under the conditions of ionization self-channeling by the field of plasma waves excited by a dipole source for $p=2\times10^{-2}$ torr, $B_0=500$ G, and $U_0=50$ V (z — distance from source).

than the dimensions of the cone, the cone becomes diffuse and joins more smoothly with the plasma-wave channel.[3] Under stationary conditions the length L of the channel is determined by the decay length of the waves that maintain the channel. In the case of collisional damping

$$L\sim\frac{\omega}{\nu_e}\lambda_z(\omega_p/\omega)^2,$$

where λ_z is the longitudinal wavelength, ν_e is the effective electron collision frequency, and ω_p is the plasma frequency corresponding to the plasma density in the column. The transverse length scale $r_\perp$ ($r_\perp\ll L$) of the plasma-wave channel is determined by the amplitude E_0 of the wave field and the losses of particles from the discharge. For the case of diffusive losses

$$r_\perp\sim\sqrt{D_\perp/\nu_i},$$

where $D_\perp$ is the ambipolar diffusion coefficient and $\nu_i\sim E_0^{2\beta}$ is the ionization frequency.[15] The plasma density distribution in the channel under stationary conditions, which was measured with a mobile electric probe, is displayed in Fig. 2a for the case $p=2\times10^{-2}$ torr, $B_0=500$ G. The distribution of the relative magnitude $E_z/E_{z\,\max}$ of the longitudinal component of the HF electric field in the channel is shown in Fig. 2b for the same conditions. Figure 3 shows the retardation curve of the electron current as a function of the retarding potential. This curve was obtained using a multigrid analyzer at pressure $p\approx3\times10^{-4}$ torr, $B_0=250$ G. It should be noted that the number of high-energy electrons decreases quite rapidly with increasing distance r from the channel axis to the chamber wall and with increasing pressure p.

Modulation of the HF fields that produce the discharge gives rise to an entire spectrum of dynamical processes in the plasma of the plasma-wave channel and the surrounding background. Figure 4 shows traces of the time dependence of the power $W(t)$ of the HF radiation producing the discharge, the plasma density $N(t)$ at the center of the discharge, the plasma density $N_s(t)$ near the wall, the LF magnetic field $H_\varphi(t)$ excited by the discharge, and the flux $J_e(t)$ of fast

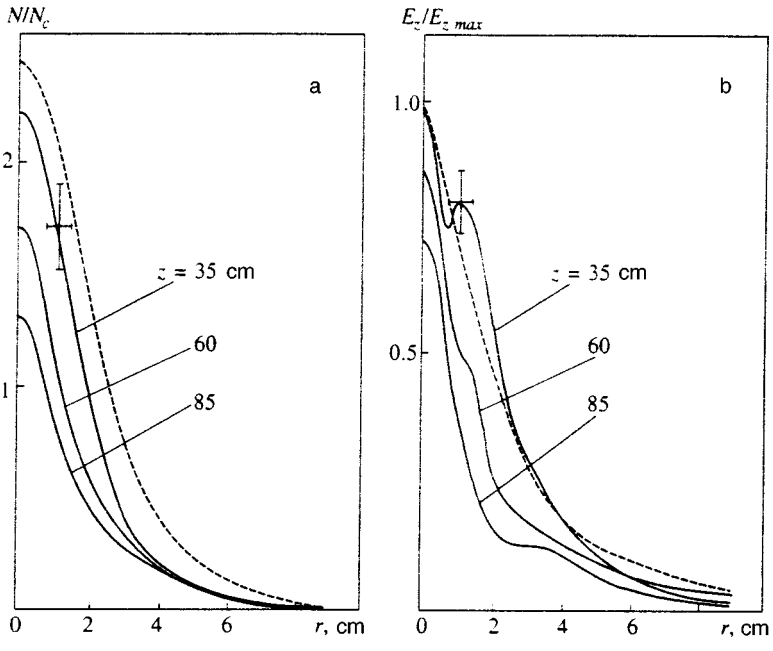


FIG. 2. Transverse distributions of the plasma density $N(r)$ (a) and relative amplitude $E_z/E_{z\,\max}$ (b) of the longitudinal component of the HF electric field in the plasma-wave channel. The distributions were measured at different distances from the dipole source under stationary conditions corresponding to Fig. 1. The dashed curves were obtained by solving the system of equations (1) and (2) numerically for the conditions of the present experiment.

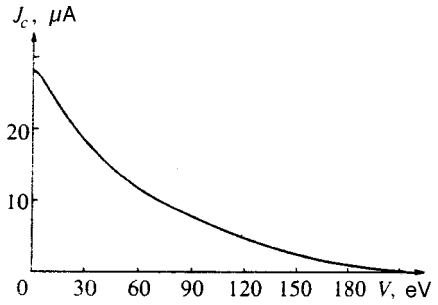


FIG. 3. Curve of retardation of the electron flux ejected from the plasma-wave channel in a direction along the external magnetic field with $p=3\times10^{-4}$ torr, $B_0=250$ G, and $U_0=50$ V.

($\mathscr{E}_e>30$ eV) electrons from the discharge which are emitted along the axis $\mathbf{z}_0\|\mathbf{B}_0$.

The large difference in the behavior of the $N(t)$ and $N_s(t)$ curves reflects the difference in the dynamics of the plasma at the center of the channel (the curve $N(t)$ was obtained with a negative probe according to the ion saturation current) and the background plasma near the wall (the curve $N_s(t)$ was obtained with a photodiode with a wide-aperture window, oriented toward the region near the wall $r>0.5\,a$). It was found that for the present modulation frequencies the density of the comparatively cold ($\mathscr{E}_e\sim3$ eV) discharge plasma near the wall continuously increased during the HF pulse, virtually did not respond to fast modulation (with frequency f_2) of the input HF power $W(t)$, and did not have enough time to relax to zero during pauses. The constant component of the photocurrent of the FD-21KP diode amounted to >80% of the variable component presented in Fig. 4. The plasma density at the center of the plasma-wave channel followed closely the fast (f_2) modulation of the input HF power. The shift of the main maximum in $N_s(t)$ relative to the main maximum in $N(t)$ characterizes the formation time of the transverse structure of the plasma-wave channel ($\tau\sim6$ μs).

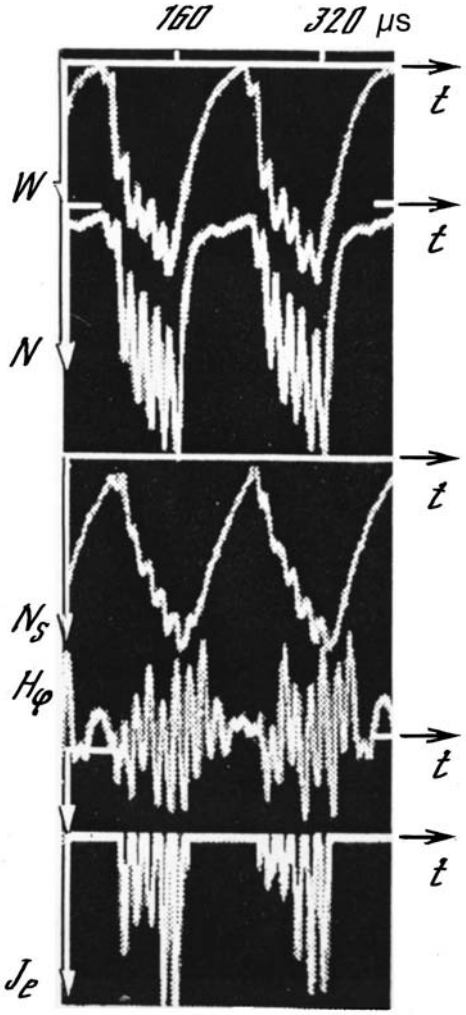


FIG. 4. Dynamics of the oscillations of the plasma density $N(t)$ near the axis and $N_s(t)$ in the peripheral (near-wall) regions of the discharge, generation of the azimuthal magnetic field $H_\varphi(t)$ and fast-electron fluxes $J_e(t)$ associated with the modulation of the HF power $W(t)$ supplied to the discharge with $p=3\times10^{-4}$ torr, $B_0=250$ G, and $U_0=50$ V.

The formation of plasma-wave channels is accompanied by injection of fast-electron beams (the curve $J_e(t)$ in Fig. 4) and excitation of LF electromagnetic fields. The oscillogram of $H_\varphi(t)$ in Fig. 4 demonstrates excitation of an azimuthal component of the alternating (with frequency f_2) magnetic field. Excitation of H_φ is possible[16] because of both the diamagnetism of the nonuniform plasma in the channel and the plasma polarization currents in the channel. It is significant that the oscillations of $H_\varphi(t)$ were also observed after the fast modulation (f_2) was switched off. Echo signals are seen especially clearly in the top panel of Fig. 5, which displays oscillograms of $W(t)$ and $H_\varphi(t)$ for the case $f_2=f_1$ (fast modulation of input power is absent). Two frequencies of the excited oscillations $f'\sim70$ kHz during the HF pulse and $f''\sim24$ kHz during the pauses between the pulses are clearly seen in the curve $H_\varphi(t)$ in this panel.

The polarization of the channel plasma by charge-particle beams emanating from the discharge region during the formation and relaxation of the plasma-wave channel gives rise to oscillations of the electric field near the channel. Oscillograms of the oscillations of $H_\varphi(t)$ and the radial electric field $E_r\simeq\Delta U/\Delta r$, where ΔU is the potential difference between the electric probes located on the axis of the plasma-wave channel and in the background plasma at a distance $\Delta r=6$ cm from the axis, are displayed at the center and bottom of Fig. 5. The amplitude of the oscillations of $E_r(t)$ is maximum when $\omega_{He}/\omega_0=n$, $n=2,3$, and reaches values $\geqslant50$ V/m. The amplitude of the oscillations of $H_\varphi(t)$ $\sim10^{-2}$ A/m corresponds to a longitudinal polarization current in the plasma-wave channel $J_e\leqslant10$ mA ($W<10$ W).

The beams of fast electrons from the discharge which emerge in a direction along the external magnetic field and polarize the plasma-wave channel are generated on the leading edge of the HF pulse forming the plasma-wave channel. Figure 6 displays oscillograms of $J_e(t)$ and $W(t)$ that demonstrate the characteristic fine structure of the generated electron beams. It was found that for HF pulse rise times $\Delta\tau>20$ μs the peak of the electron flux $J_e(t)$ splits in two.

It would appear that the first peak in $J_e(t)$ is due to motion of the region of the focus of the resonance cone, where the plasma density gradient $\nabla N(\mathbf{r})$, the external mag-

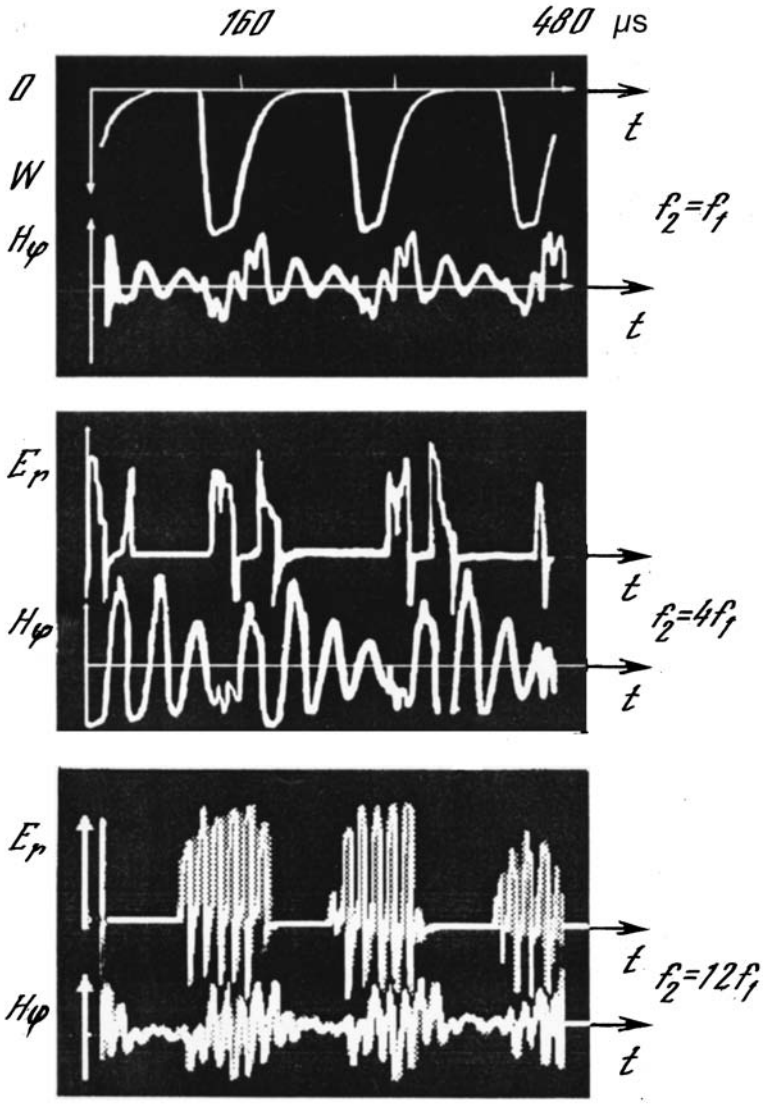


FIG. 5. Oscillograms of the radial electric $E_r(t)$ and azimuthal magnetic $H_\varphi(t)$ components of the LF fields generated by a modulated plasma-wave discharge with $p=3\times10^{-4}$ torr, $B_0=250$ G, and $U_0=50$ V and different values of the fast-modulation frequency f_2.

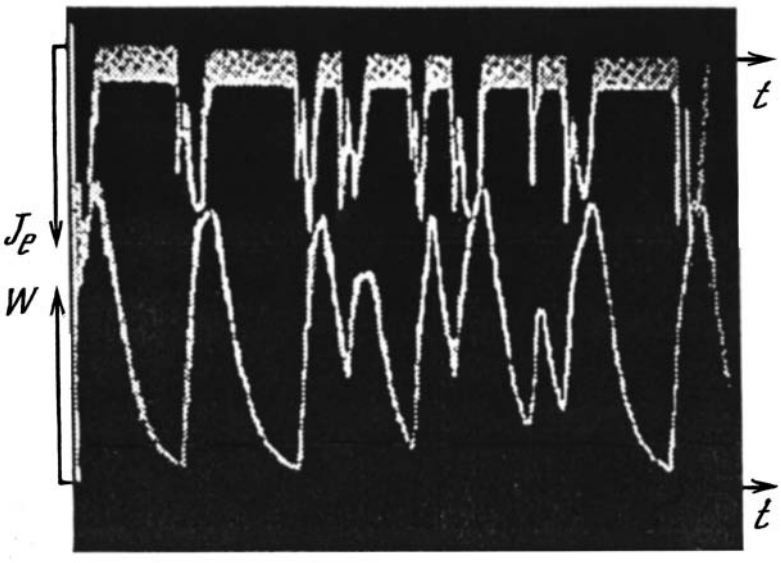


FIG. 6. Oscillograms of $J_e(t)$ and $W(t)$ measured with $p=3\times10^{-4}$ torr, $B_0=250$ G, $U_0=50$ V, and $f_2\leqslant12$ kHz and demonstrating the structure and distribution of the electron beams generated on the leading edge of the HF pulse forming the discharge.

netic field $\mathbf{B}_0$, and the source electric field $\mathbf{E}(\mathbf{r})$ are all oriented in the same direction, away from the source. The field $\mathbf{E}(\mathbf{r})$ in this region has a maximum associated with a characteristic singularity of the refractive index and with the excitation of plasma waves which accelerate electrons.[17] The rise time of the first peak of $J_e(t)$ characterizes the growth time of the plasma-wave channel from the source to the measuring probe and makes it possible to estimate the rate of growth of the channel along the magnetic field ($v_\parallel \sim v_e \geqslant 10^8$ cm/s).

The second peak in $J_e(t)$ is formed by the flux of electrons ejected from the positive column of the plasma-wave channel by the longitudinal electric field $E_\parallel \sim \nabla_\parallel N$. The flux exists so long as the positive potential continues to increase on the leading edge of the HF pulse and the longitudinal nonuniformity of the plasma in the plasma-wave channel remains. As the amplitude of the channel-forming HF pulse decreases, the longitudinal nonuniformity in the channel decreases as a result of rapid expansion of the plasma and can change sign, as observations of the changes in the longitudinal component of the electric field showed.

3. DISCUSSION OF THE EXPERIMENTAL RESULTS

Taken together, the experimental data presented above show the following physical picture of the dynamical phenomena that are observed. The discharge starts to form in the near-field of the HF source. The structure of this quasistatic field changes substantially when the plasma density increases above the critical density N_c and the plasma frequency ω_p exceeds the frequency ω of the exciting field. For $\omega_p > \omega$ conical resonance surfaces, resting on the rings of the exciting antenna, appear (Fig. 1). The field of the plasma waves, which are efficiently excited along the surface of the resonance cone, forms a plasma column (a plasma-wave channel) at the focus, extending along the external magnetic field. A beam of plasma waves (small-scale whistlers), which transport sufficient energy along the filament to produce and maintain the filament, is trapped in the channel. We note that a short ($l \ll \lambda_0$, where λ_0 is the wavelength in vacuum) HF source efficiently excites plasma waves over a wide range of wave numbers.[18] For this reason, the wave fields with wavelengths λ_p ranging in scale from the diameter of the wire of the exciting rings to twice the distance between the the rings ($\lambda_p \leqslant 2l$) form a channel. The good spatial resolution of the resonance cone (see Fig. 1) confirms that small-scale plasma waves with large amplitude are efficiently excited by the short dipole antenna. Beats of the total field of the plasma waves with different longitudinal scales can explain[18] the excitation of low-frequency oscillations with frequency f' (see Fig. 5) during a long HF pulse. The plasma nonuniformity in the resonance cone becomes more smeared out with decreasing gas pressure, i.e., it is determined by the electron mean free path. At high plasma densities whistlers whose transverse structure contains both small and large scales can be excited in the channel.[5,6,19]

The stationary self-consistent plasma density distribution $N_{(r_\perp)}$ in the transverse section of an axisymmetric channel and the HF potential $\varphi = \Psi(r_\perp) e^{i\omega t - ik_z z}$ ($\mathbf{E} = -\nabla\varphi$) at short distances from the source ($z < \lambda_0$) are determined by the system of equations[3]

$$\nabla_\perp \Psi - k_z^2 (1 - N/N_c)\Psi = 0, \tag{1}$$

$$D_\perp \Delta_\perp N + (\nu_i - v_a) N = 0, \tag{2}$$

where ν_a is the attachment frequency. Here we took into account the fact that in the parameter range considered the components of the permittivity tensor are $\varepsilon_{11} \approx \varepsilon_{22} \approx 1$ and $\varepsilon_{33} = 1 - N/N_c$, while charged-particle losses are due mainly to electron attachment to air molecules. The weak longitudinal nonuniformity of the stationary channel can be taken into account by adding to Eq. (2) an additional loss factor $\sim D_\parallel / L^2$ corresponding to particle diffusion to the end wall. To construct a simplified theoretical model of a stationary plasma-wave channel, the frequency ν_i of electron-impact ionization of air molecules can be represented by a model function $\nu_i = \alpha (k_z \Psi)^{2\beta}$ of the amplitude of the longitudinal field, determining the electron temperature in the channel.[15] The system of equations (1) and (2) admits self-localized axisymmetric[2] and multihump solitary field and plasma distributions.[3] The multihump (multisoliton) solutions of Eqs. (1) and (2) correspond to multibeam self-channeling of plasma waves, which is observed when asymmetric sources of ionizing radiation are used.[3] An axisymmetric single-soliton solution describes a solitary plasma-wave channel formed by a symmetric, medium-power source.[2] Figure 2 shows the computational results (dashed curves) for the system (1) and (2) with $\beta = 2$, $D_\perp / \nu_a = 63$ mm^2, and $\lambda_z = 120$ mm.

The solution of a self-consistent system of equations of the type (1) and (2) for the time-dependent case, taking account of losses and the longitudinal nonuniformity of the plasma-wave channel, presents substantial difficulties. For this reason, to explain the experimental results on the dynamics of ionization self-channeling of plasma-wave beams, we confine our attention to well-known theoretical considerations and estimates. We note first that because the field of the wave beam does not heat the plasma uniformly over the cross section, and because an external magnetic field is present ($r_{He} \leqslant 0.1$ mm $\ll r_{Hi} \leqslant 10$ mm $\ll a = 100$ mm) two essentially different regions of the discharge are observed in the dynamical regime of ionization self-channeling. A quasistationary region of comparatively cold plasma, whose density varies approximately by a factor of 2 over the time of the HF pulse ($\tau_1 \sim 1/f_1$), always remaining somewhat less than the critical density ($N_s \leqslant N_c$), is present at the periphery ($r \geqslant 50$ mm). The average electron energy in this region varies in the range $\mathscr{E}_e < 3$ eV. In the central, dynamical part of the discharge (in the region of the plasma-wave channel $r \leqslant 30$ mm) N and $\mathscr{E}_e$ vary more strongly (see Fig. 4) and closely follow the modulation of the supplied power with both frequencies (f_1 and f_2), and under appropriate conditions they also follow the beat frequency f' of the total field of the plasma waves with different longitudinal scales.

The low-frequency oscillations observed in the channel occur as follows. As the input HF power increases, the plasma density in the channel increases and a positive col-

umn of increasing length forms. The electrons are heated by the field of the plasma waves trapped in the column and are ejected in a direction along the external magnetic field. In the process, the positive potential of the column increases and ion diffusion through the lateral surface of the column accelerates. The column continues to increase in length, even as the input power decreases, on account of the energy of the trapped radiation. In the process, the potential and plasma density in the expanding column decrease. The electron beam ejected from the column excites an azimuthal magnetic field, while the growing positively charged plasma column forms radial and longitudinal electric fields which move together with it. As a result, when the discharge-forming radiation is subjected to fast amplitude modulation, a wave of perturbations of the parameters of the plasma channel and electromagnetic fields and currents associated with these perturbations can form. The propagation velocity of such a wave equals the propagation velocity of the channel in the background plasma ($v_c \sim v_e$). The characteristic longitudinal scale of the wave of perturbations is $\lambda_c \sim v_e / f_{\text{mod}}$, while the frequency of the wave equals the modulation frequency f_{mod} and satisfies the condition

$$f_{\text{mod}} > \omega (\nu_e / \omega_p)^2 (l_e / \lambda_z),$$

where $\lambda_c < L$. Here l_e is the electron mean free path, while $\lambda_z \sim 2l$. Under the conditions of our experiments, $v_e \geqslant 10^8$ cm/s, $f_{\text{mod}} \leqslant 10^5$ Hz, the dimensions of the apparatus are $\mathscr{L} \sim 2$ m $\leqslant \lambda_c/4$, which made it impossible to investigate the structure of the fields and currents of the wave of perturbations. However, observations of the sign changes of the longitudinal component of the LF electric field in phase with the sign changes of the longitudinal component of the gradient of the plasma density in the channel and the sign changes of the E_r and H_φ components of the excited LF fields make it possible to talk about observation of an induced LF wave process formed by the modulated wave beam. We recall that in the presence of ionization nonlinearity in the frequency range $\Omega_H < \omega < \omega_H$ the field of the plasma waves is localized as a result of total internal reflection at the crest of the dense plasma.[2] The observed sign changes of the longitudinal component of the gradient of the plasma density in the channel are due to the motion of the crests of the dense plasma away from the HF source along the longitudinal axis of the apparatus. These properties of the observed effect differentiate it substantially from the outwardly similar effects consisting of the formation of moving plasma nonuniformities by the field of Langmuir solitons and space-charge solitons.[21]

Oscillations of the parameters of the plasma-wave channel excite characteristic oscillatory and wave processes with frequency close to f_{mod} in the background plasma and the positive column. For example, the echo signals $H_\varphi(t)$ observed under the experimental conditions during the pauses between the HF pump pulses ($W \sim 0$) are explained by excitation of magnetohydrodynamic H waves whose dispersion curves at high frequencies ($\omega_H \ll \omega < \omega_H$) pass into the dispersion curves of helical waves (whistlers).[20] The increase in the amplitude of the echo signal at the modulation frequency $f_2 \sim 24$ kHz is due to a resonance of a MHD wave in the discharge chamber, whose longitudinal scale equals the wavelength of the excited TE wave. These results make it possible to explain the mechanism of the operation of a plasma antenna, observed during the ''Active filament'' ionospheric experiment,''[10,11] by the excitation of VLF radiation by the wave of perturbations in the plasma-wave channel formed by a discharge in the field of the modulated beam of plasma waves.

4. CONCLUSIONS

The experimental results presented in this paper show the complicated nature of the dynamical phenomena arising during ionization self-channeling of beams of plasma waves in a magnetic field. Amplitude modulation of the radiation forming the plasma-wave channel gives rise to modulated electron fluxes, whose current is closed by the background-plasma currents and ion fluxes from the surface of the channel, directed along the magnetic field and LF oscillations of the potential and density of the plasma in the channel that move in a direction away from the HF source and along the external magnetic field, i.e., it leads to the formation of waves of perturbations. Under the nighttime ionosphere conditions the plasma-wave channel excited by the on-board HF source is a source of electron fluxes whose velocity can be higher than the phase velocity of Alfvén and VLF whistler waves.[7,8] Amplitude modulation of the HF source makes it possible to produce current structures, which move together with the electron fluxes and increase in size, together with the low-frequency electromagnetic waves associated with them, i.e., to form plasma-dynamical antennas in the ionosphere.[11] The scales of such dynamical antennas and the wavelength of the perturbations in them can be comparable to the wavelengths of the VLF waves excited by the antenna. For this reason they could be interesting for purposes of LF radio communication and wave diagnostics of the magnetospheric plasma. Investigations of the dispersion characteristics and the structure of the fields and currents of the wave of perturbations that is excited by the modulated wave beam in a magnetized plasma are of interest in themselves.

This work was supported by the Russian Fund for Fundamental Research (grant No. 95-02-05816a) and the ''Universities of Russia'' program.

*)E-mail: markov@rf.unn.rannet.ru

[1] J. Musil and F. Zacek, Czech. J. Phys. B **22**, 1108 (1972).
[2] G. A. Markov, V. A. Mironov, and A. M. Sergeev, JETP Lett. **29**, 617 (1979).
[3] G. A. Markov, V. A. Mironov, A. M. Sergeev, and I. A. Sokolov, Zh. Éksp. Teor. Fiz. **80**, 2264 (1981) [Sov. Phys. JETP **53**, 1183 (1981)].
[4] I. A. Vdovichenko, G. A. Markov, V. A. Mironov, and A. M. Sergeev, JETP Lett. **44**, 275 (1986).
[5] G. A. Markov, Fiz. Plazmy **14**, 1094 (1988) [Sov. J. Plasma Phys. **14**, 641 (1988)].
[6] A. V. Kudrin and G. A. Markov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **34**, 163 (1991).
[7] Yu. N. Agafonov, A. P. Babaev, V. S. Bazhanov *et al.*, Pis'ma Zh. Tekh. Fiz. **15**, 1 (1989) [Sov. Tech. Phys. Lett. **15**, 661 (1989)].
[8] Yu. N. Agafonov, V. S. Bazhanov, Yu. I. Gal'perin *et al.*, Pis'ma Zh. Tekh. Fiz. **16**, 65 (1990) [Sov. Tech. Phys. Lett. **16**, 627 (1990)].
[9] Yu. N. Agafonov, V. S. Bazhanov, V. Ya. Isyakaev *et al.*, JETP Lett. **52**, 530 (1990).

[10] G. A. Markov and Yu. V. Chugunov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **37**, 791 (1994).
[11] Yu. N. Agafonov, G. V. Bashilov, G. A. Markov, and Yu. V. Chugunov, Geomagn. Aéron. **36**, 206 (1996).
[12] R. W. Boswell, Plasma Phys. Controlled Fusion **26**, 1147 (1984).
[13] F. F. Chen, J. Vac. Sci. Technol. A **10**, 1389 (1991).
[14] K. P. Shamrai and V. B. Taranov, Plasma Sources Sci. Technol. **5**, 474 (1996).
[15] J. T. Mayhan, R. L. Fante, R. O'Keefe *et al.*, J. Appl. Phys. **42**, 536 (1971).
[16] A. V. Kudrin, G. A. Markov, V. Yu. Trakhtengerts *et al.*, Geomagn. Aéron. **31**, 334 (1991).
[17] V. I. Arkhipenko, V. N. Budnikov, and A. A. Obukhov, Zh. Tekh. Fiz. **43**, 113 (1973) [Sov. Phys. Tech. Phys. **18**, 72 (1973)]; *ibid.* **43**, 313 (1973) [*ibid.* **18**, 204 (1973)].
[18] E. A. Mareev and Yu. V. Chugunov, *Antennas in Plasma* [in Russian], Institute of Applied Physics of the Soviet Academy of Sciences, N. Novgorod, 1991.
[19] T. M. Zaboronkova, A. V. Kudrin, and G. A. Markov, Fiz. Plazmy **19**, 769 (1993) [Plasma Phys. Rep. **19**, 397 (1993)].
[20] A. N. Kondratenko, *Plasma Waveguides* [in Russian], Atomizdat, Moscow, 1976.
[21] M. V. Nezlin, *Physics of Intense Beams in Plasmas* (IOP Press, Philadelphia, 1993; Énergoizdat, Moscow, 1982.

Translated by M. E. Alferieff

Nonlinear properties of a weakly collisional plasma at low radiation intensities

A. V. Maksimov, K. N. Ovchinnikov, V. P. Silin,*) and S. A. Uryupin

P. N. Lebedev Institute of Physics, Russian Academy of Sciences, 117924 Moscow, Russia
(Submitted 24 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 1299–1312 (April 1998)

In this paper we evaluate theoretically the modification of the distribution function of slow subthermal electrons heated by radiation in a completely ionized plasma. The new solution we have found to the kinetic equation under conditions typical of weak-collisional plasmas can be used to predict new nonlinear behavior arising from perturbations of the nonuniform electron density and the coefficient of nonlocal heat transfer that depend on the intensity of the radiation heating the plasma, which is absorbed via inverse bremsstrahlung. It is predicted that this new nonlinear behavior manifests itself at unexpectedly small radiation intensities.
 [S1063-7761(98)01304-3]

1. INTRODUCTION

The task of this paper is to discuss nonlinear effects that appear when a completely ionized plasma is subjected to a comparatively weak electromagnetic field. Let us write the electric field of the radiation in the form

$$(1/2)\mathbf{E}\exp(-i\omega_0 t)+\text{c.c.},$$

where $\mathbf{E}$ is an amplitude that is slowly varying over the period $2\pi/\omega_0$. We will refer to a field as ''high-frequency'' when ω_0 significantly exceeds the effective electron–ion collision frequency

$$\nu_{ei}=4\sqrt{2\pi}Ze^4 n\Lambda/3m^2 v_T^3. \tag{1.1}$$

Here e and m are the charge and mass of an electron, $v_T=\sqrt{k_B T/m}$ is the thermal velocity of an electron (where k_B is Boltzmann's constant), Λ is the Coulomb logarithm, n is the electron number density, and Z is the effective ionization state of the ions defined by the relation

$$Z=\sum_i \frac{e_i^2 n_i}{e^2 n}. \tag{1.2}$$

The sum runs over all ion species; e_i and m_i are the charge and mass of the corresponding species.

As a measure of the effect of electromagnetic radiation on the electrons, it is customary[1] to use the velocity with which the electrons oscillate in the pump field:

$$v_E=|e\mathbf{E}|/m\omega_0. \tag{1.3}$$

When speaking of a weak high-frequency electromagnetic field, we will always assume that the oscillatory velocity v_E is small compared to electron velocities characteristic of the processes under study. It is in this approximation that we obtain the kinetic equation (2.3) used below (see Ref. 2). Therefore, in every case we will assume that the following inequality holds:

$$v_E^2 \ll v_T^2, \tag{1.4}$$

which is often chosen as the condition for weakness of the pump field. However, even when the inequality (1.4) holds, at high degrees of ionization $Z\gg 1$ (which will be used below) it is well known[3,4] that the electron distribution function can be strongly modified over the entire velocity phase space when the Langdon parameter

$$\alpha=\frac{Zv_E^2}{v_T^2} \tag{1.5}$$

is not small compared to unity. Our discussion will apply to conditions where the field is so weak that

$$\alpha\ll 1. \tag{1.6}$$

This is the case of interest to us, i.e., a weak high-frequency radiation field, for which no one has predicted nonlinear effects due to readjustment of the electron distribution until now.

The Langdon parameter (1.5) reflects a competition between two processes, and arises as a ratio of two times: the electron–electron collision time for thermal electrons

$$t_{eT}=Z/\nu_{ei},$$

which determines when a Maxwellian distribution is established in the range of velocities $v\sim v_T$, and the time for heating thermal electrons

$$t_{HT}=(v_T^2/v_E^2)/\nu_{ei}.$$

Our interest will be in the distribution of electrons with velocities less than the thermal velocity. For these cold subthermal electrons the electron–electron relaxation time is determined by collisions with thermal electrons, and according to Eq. (3.21) is proportional to the square of the velocity:

$$t_{eC}(v)\propto t_{eT}(v^2/v_T^2).$$

In contrast to this, the characteristic time for heating cold electrons by inverse bremsstrahlung absorption is proportional, according to (3.5), to the fifth power of the velocity:

$$t_{HC}\propto t_{HT}(v^5/v_T^5).$$

Therefore, for velocities such that

$$v\leqslant v_T(t_{eT}/t_{HT})^{1/3}\sim v_T\alpha^{1/3}, \tag{1.7}$$

conversion of the distribution law for cold electrons to Maxwellian form via electron-electron collision turns out to be a rather slow process compared to heating by a high-frequency field via inverse bremsstrahlung absorption. It is this property that distinguishes the nonlinear processes we are discussing for the plasma.

Cold subthermal electrons play an important role in the transport theory of weakly collisional plasmas,[5–7] when the collisionless condition is satisfied for thermal electrons:

$$k\ell_{ei} \gg 1. \tag{1.8}$$

Here k is the wave vector of the spatially nonuniform plasma perturbations, and $\ell_{ei}=v_T/\nu_{ei}$ is the mean-free path of a thermal electron. In this case, since the free range of an electron with velocity v is proportional to the fourth power of the velocity:

$$\ell(v) \propto (v/v_T)^4 \ell_{ei},$$

cold subthermal electrons with velocities

$$v < v_T (k\ell_{ei})^{-1/4} \equiv v_* \ll v_T \tag{1.9}$$

are found to be strongly collisional.

In this report, the theoretical results we have used apply to a situation where collisional electrons satisfy the following inequality over the entire velocity space:

$$v < v_* \leqslant v_T \left(\sqrt{\frac{\pi}{8}}\, Z\, \frac{v_E^2}{v_T^2} \right)^{1/3} \equiv v_L \sim v_T \alpha^{1/3}. \tag{1.10}$$

This inequality defines a new range of physical parameters that specify the interaction of a plasma with radiation, in which we predict new behavior for the electron transport. In this case, using the terminology of flux limiting of electron heat transport in a plasma, the results we will obtain below can be briefly characterized as a relaxation of this limit. Now, however, we extend results established previously for fairly strong fields with Langdon parameters that are not small[8,9] to the new conditions where inequality (1.6) is satisfied.

In the second section of the article we derive a starting kinetic equation for the slowly varying electron distribution function in a high-frequency electromagnetic field. In the third section, we find the quasistationary distribution of electrons in a spatially uniform electromagnetic field, which follows a law that is quite different from Maxwellian in the low-velocity range. The electron distribution function in the presence of nonuniform electromagnetic fields is found in the fourth section. This distribution allows us to obtain an expression for the nonlinear perturbation of the electron density. Conditions are derived in which this perturbation dominates. In the fifth section an expression is obtained for the effective nonlocal thermal conductivity, which is characterized by a nonlinear dependence on the pump field. The sixth section is devoted to discussing the results of the article.

2. STARTING KINETIC EQUATION

In discussing processes in a fully ionized plasma in a high-frequency electromagnetic field with fairly low intensity, an approximate description is used which is based on an expansion in the small parameter

$$\frac{v_E^2}{v_T^2} \ll 1. \tag{2.1}$$

In this case the electron distribution function is written in the form

$$f = f_0 + [(1/2) f_1 \exp(-i\omega_0 t) + \text{c.c.}], \tag{2.2}$$

where the functions f_0 and f_1 change very little within one period of the high-frequency oscillation $2\pi/\omega_0$. The corresponding kinetic equation for the slowly time-varying function f_0 has the following form[2] to first-order accuracy in the parameter (2.1):

$$\begin{aligned}
&\frac{\partial f_0}{\partial t} + \mathbf{v}\frac{\partial f_0}{\partial \mathbf{r}} + \frac{e\mathbf{E}_0}{m}\frac{\partial f_0}{\partial \mathbf{v}} - J_{ei}[f_0] - J_{ee}[f_0, f_0] \\
&= \frac{e^2}{4\omega_0^2 m^2}\left\{ \frac{\partial |\mathbf{E}|^2}{\partial \mathbf{r}}\frac{\partial f_0}{\partial \mathbf{v}} + \frac{1}{2}\frac{\partial^2 f_0}{\partial v_i \partial v_j}\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)\right. \\
&\quad \times (E_i E_j^* + E_i^* E_j) + (E_i E_j^* + E_i^* E_j)\left(\frac{\partial^2 f_0}{\partial r_i \partial v_j} + \left(\frac{\partial}{\partial t}\right.\right. \\
&\quad \left.\left. \left. + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)\frac{\partial^2 f_0}{\partial v_i \partial v_j} - \frac{\partial}{\partial v_i} J_{ei}\left[\frac{\partial f_0}{\partial v_j}\right] - J_{ee}\left[\frac{\partial f_0}{\partial v_i}, \frac{\partial f_0}{\partial v_j}\right]\right)\right\},
\end{aligned} \tag{2.3}$$

where $\mathbf{E}_0$ is the intensity of the quasistationary electric field. For the electron–ion collision integral we use the approximate expression

$$J_{ei}[f_0] = \nu(v)\frac{\partial}{\partial v_r}\left\{ [v^2 \delta_{rs} - v_r v_s]\frac{\partial f_0}{\partial v_s}\right\}, \tag{2.4}$$

where

$$\nu(v) = 3\sqrt{\frac{\pi}{8}}\frac{\nu_{ei} v_T^3}{v^3} \equiv \frac{A}{v^3}. \tag{2.5}$$

Expression (2.4) does not take into account the exchange of energy between electrons and ions, which is described by terms we have neglected of the same order as the electron-ion mass ratio. Finally, we will write the electron-electron collision integral in the Landau–Fokker–Planck form

$$J_{ee}[f_0, f_0] = \frac{\partial}{\partial v_r}\left(D_{rs}\frac{\partial f_0}{\partial v_s}\right) - \frac{\partial}{\partial v_r}(A_r f_0), \tag{2.6}$$

from which the following expressions are obtained for the diffusion coefficient and frictional force in the velocity phase space:

$$\begin{aligned}
D_{rs} &= \frac{2\pi e^4 \Lambda}{m^2}\int d\mathbf{v}'\, \frac{(\mathbf{v}-\mathbf{v}')^2 \delta_{rs} - (\mathbf{v}-\mathbf{v}')_r (\mathbf{v}-\mathbf{v}')_s}{|\mathbf{v}-\mathbf{v}'|^3} \\
&\quad \times f_0(\mathbf{v}'),
\end{aligned} \tag{2.7}$$

$$A_r = \frac{2\pi e^4 \Lambda}{m^2}\int d\mathbf{v}'\, \frac{(\mathbf{v}-\mathbf{v}')^2 \delta_{rs} - (\mathbf{v}-\mathbf{v}')_r (\mathbf{v}-\mathbf{v}')_s}{|\mathbf{v}-\mathbf{v}'|^3}$$

$$\times\frac{\partial f_0(\mathbf{v}')}{\partial v_s'}. \tag{2.8}$$

It should be emphasized that the term appearing on the right side of Eq. (2.3) that contains the electron–ion collision integrals describes, among other things, inverse bremsstrahlung absorption of electromagnetic radiation by electrons as they collide with ions. Equation (2.3) will be used below to analyze the kinetic properties of electrons in the radiation field.

3. ELECTRON GROUND-STATE DISTRIBUTION FUNCTION

In this section we obtain the distribution function for plasma electrons heated by inverse bremsstrahlung absorption, under conditions where nonuniformity of the heating radiation is insignificant. That is, we will assume that

$$\frac{\partial}{\partial\mathbf{r}}E_iE_j^*=0. \tag{3.1}$$

Using Eq. (2.3) next, we will assume that the degree of ionization is large:

$$Z\gg 1. \tag{3.2}$$

Then we obtain from Eq. (2.3)

$$\frac{\partial f_0}{\partial t}-J_{ei}[f_0]-J_{ee}[f_0,f_0]=-\frac{e^2}{4\omega_0^2m^2}\,(E_iE_j^*+E_i^*E_j)\,\frac{\partial}{\partial v_i}\,J_{ei}\left[\frac{\partial f_0}{\partial v_j}\right]. \tag{3.3}$$

In studying the solutions to this equation we write the electron distribution function in the form of two terms:

$$f_0=f_s+f_a, \tag{3.4}$$

where f_s is the symmetric part of the electron distribution function, obtained by averaging the function f_0 over the angles of the velocity vector $f_s=\langle f_0\rangle$, and $f_a=f_0-f_s$ is its asymmetric part.

Assuming that the characteristic time for the electron distribution function to change is large compared to the time it takes electrons with velocities v to become isotropic ($t_i(v)=\nu^{-1}(v)$), it is not difficult to see that as a result of averaging Eq. (3.3) we obtain

$$\frac{\partial f_s}{\partial t}-J_{ee}[f_s,f_s]=\frac{1}{3}\,v_E^2\,\frac{1}{v^2}\,\frac{\partial}{\partial v}\left(v^2\nu(v)\,\frac{\partial f_s}{\partial v}\right). \tag{3.5}$$

In this case, for the distribution function f_s that depends on the absolute magnitude of the velocity we have

$$J_{ee}[f_s,f_s]=\frac{1}{v^2}\,\frac{\partial}{\partial v}\left\{v^3\left[\frac{d(v)}{v}\,\frac{\partial f_s}{\partial v}-a(v)f_s\right]\right\}, \tag{3.6}$$

where

$$d(v)=\frac{8\pi}{3}\,B\left[\int_v^\infty dv'v'f_s(v')+\frac{1}{v^3}\int_0^v dv'(v')^4f_s(v')\right], \tag{3.7}$$

$$a(v)=\frac{8\pi}{3}\,B\left[\int_v^\infty dv'\,\frac{\partial f_s(v')}{\partial v'}+\frac{1}{v^3}\int_0^v dv'(v')^3\,\frac{\partial f_s(v')}{\partial v'}\right]=-\frac{8\pi B}{v^3}\int_0^v dv'(v')^2f_s(v'), \tag{3.8}$$

$$B=\frac{2\pi e^4\Lambda}{m^2}=\frac{A}{Zn}. \tag{3.9}$$

When the Langdon parameter (1.5) is small it is usually assumed that the solution to Eq. (3.5) is a Maxwell distribution:[3]

$$f_m(v,t)=\frac{n}{(2\pi)^{3/2}v_T^3(t)}\,\exp\left(-\frac{v^2}{2v_T^2(t)}\right), \tag{3.10}$$

where the time dependence of the thermal velocity is described by the equation

$$v_T\,\frac{dv_T}{dt}=\frac{1}{6}\,\nu_{ei}v_E^2, \tag{3.11}$$

which corresponds to the characteristic heating time for thermal electrons

$$t_{HT}=(v_T^2/v_E^2)/\nu_{ei}. \tag{3.12}$$

The argument for this widely-held assumption is that for $v\sim v_T$ smallness of the Langdon parameter implies that the time (3.12) for heating thermal electrons is much larger than the time $t_{eT}=Z/\nu_{ei}$ for their distribution to become Maxwellian. From this same expression, however, it follows that the distribution (3.10) will be incorrect for cold subthermal electrons, where the heating time is determined by the expression

$$t_{HC}(v)=(v^2/v_E^2)/\nu(v), \tag{3.13}$$

while the time for conversion of the distribution law to Maxwellian form, mediated by collisions of cold electrons with thermal electrons, is

$$t_{eC}(v)=Z(v^2/v_T^2)/\nu_{ei}. \tag{3.14}$$

Therefore, this time turns out to be much longer than the heating time for velocities that satisfy inequality (1.7).

In identifying the form of the distribution function for electrons with velocities much smaller than the thermal velocity

$$v\ll v_T \tag{3.15}$$

we are interested in a small portion of the electrons in the velocity phase space when the distribution of electrons in the remainder of that space is the Maxwellian (3.10). This latter assumption allows us to use the following approximation:

$$d(v)\approx -v_T^2a(v), \tag{3.16}$$

which obviously follows from Eq. (3.7) and the first form of expression (3.8) after substituting (3.10) into the latter. The second way of writing expression (3.8), taking into account the approximation

$$f_s(v)=f_s(0), \tag{3.17}$$

allows us to obtain

$$a=-\frac{8\pi}{3}Bf_s(0). \tag{3.18}$$

In our discussion, we can assume with fair accuracy that

$$f_s(0)=f_m(v=0). \tag{3.19}$$

Therefore, for cold subthermal electrons we have

$$a=-\frac{\nu_{ei}}{Z}\equiv-\nu_{ee}. \tag{3.20}$$

Equations (3.6)–(3.9) and (3.16)–(3.20) allow us to represent Eq. (3.5) for cold subthermal electrons with velocities that satisfy the inequality (3.15) in the form of the following differential equation:

$$\frac{\partial f_s}{\partial t}-\frac{\nu_{ee}}{v^2}\frac{\partial}{\partial v}\left\{v^3\left[\frac{v_T^2}{v}\frac{\partial f_s}{\partial v}+f_s\right]\right\}=\sqrt{\frac{\pi}{8}}\,v_E^2\nu_{ei}\frac{v_T^3}{v^2}\frac{\partial}{\partial v}\left(\frac{1}{v}\frac{\partial f_s}{\partial v}\right). \tag{3.21}$$

In discussing the consequences of this equation we will take into account the fact that the electron–electron relaxation time of cold electrons Eq. (3.14), which characterizes the second term on the left side of Eq. (3.21), is much smaller than the characteristic time for heating of thermal electrons given by Eq. (3.12), which characterizes the change of the thermal velocity $v_T(t)$ with time. This allows us to neglect the time derivatives in Eq. (3.21) and write the following ordinary differential equation for the quasistationary case when the thermal velocity varies slowly with time:

$$\frac{1}{v^2}\frac{d}{dv}\left\{v^3\left[\frac{1}{v}\frac{df_s}{dv}+\frac{f_s}{v_T^2}\right]+\frac{v_L^3}{v}\frac{df_s}{dv}\right\}=0, \tag{3.22}$$

where the Langdon velocity has the form

$$v_L=\left(\sqrt{\frac{\pi}{8}}\,Zv_E^2v_T\right)^{1/3}. \tag{3.23}$$

We must find a solution to Eq. (3.22) that satisfies the boundary condition

$$\frac{1}{v}\frac{df_s}{dv}=0 \tag{3.24}$$

at $v=0$, which comes from the last term on the left side of Eq. (3.22) and corresponds to the absence of sources of electrons for $v=0$. The corresponding solution to Eq. (3.22) has the form

$$f_s(v)=f_s(0)\exp\left[-\frac{1}{v_T^2}\int_0^v\frac{u^4du}{u^3+v_L^3}\right]. \tag{3.25}$$

From this, we find that when condition (3.15) holds,

$$f_s(v)=\frac{n}{(2\pi)^{3/2}v_T^3}\left\{1-\frac{1}{v_T^2}\int_0^v\frac{u^4du}{u^3+v_L^3}\right\}. \tag{3.26}$$

In this case, the approximation (3.19) was used. The accuracy of this latter approximation can be estimated if we note that for $v_T\gg v\gg v_L$ it follows from (3.26) that

$$f_s(v)=\frac{n}{(2\pi)^{3/2}v_T^3}\left\{1-\frac{v^2}{2v_T^2}+\frac{2\pi}{3\sqrt{3}}\frac{v_L^2}{v_T^2}\left[1+O\left(\frac{v_L}{v}\right)\right]\right\}. \tag{3.27}$$

In view of the smallness of the Langdon parameter expressed by Eq. (1.6) and our neglect of corrections of order $\alpha^{2/3}$, we may assume approximation (3.19) is justified. Expression (3.26), which describes the distribution of subthermal electrons of a plasma heated by inverse bremsstrahlung absorption is a new result of the theory, which will be the starting point for deriving the new results described in the following sections of this paper.

4. PERTURBATION OF THE ELECTRON DENSITY BY A NONUNIFORM FIELD

In the theory of parametric instability the nonlinear interaction between modes is determined by the perturbation of the electron density by a spatially nonuniform electromagnetic field. For this reason, we choose to write the electric field in a new form, different from the form used in the previous section:

$$E_iE_j^*\to E_iE_j^*+\delta(E_iE_j^*)\exp(i\mathbf{kr}). \tag{4.1}$$

The first term on the right side of this expression is determined by the spatially nonuniform pump field. The second term is determined by a superposition of the electromagnetic fields of the pump and the perturbed modes, which are interacting at parametric resonance. Then picking

$$f_0=f_s+\delta f\,\exp(i\mathbf{kr}), \tag{4.2}$$

where δf is a small perturbation of the electron distribution function determined by the spatially nonuniform perturbation (4.1), we can write the following linearized equation that follows from Eq. (2.3):

$$\begin{aligned}i(\mathbf{kv})\delta f-J_{ei}[\delta f]-J_{ee}[\delta f]&=\frac{e^2}{4\omega_0^2m^2}\left(i\mathbf{k}\frac{\partial f_s}{\partial\mathbf{v}}\,\delta|\mathbf{E}|^2+\frac{1}{2}\,i(\mathbf{kv})\delta(E_iE_j^*\right.\\&+E_i^*E_j)\frac{\partial^2f_s}{\partial v_i\partial v_j}-\delta(E_iE_j^*+E_i^*E_j)\frac{\partial}{\partial v_i}J_{ei}\left[\frac{\partial f_s}{\partial v_j}\right]\\&\left.-(E_iE_j^*+E_i^*E_j)\frac{\partial}{\partial v_i}J_{ei}\left[\frac{\partial\delta f}{\partial v_j}\right]\right).\end{aligned} \tag{4.3}$$

In writing this equation we have used the assumption that the degree of ionization (3.2) is large, allowing us to neglect the contribution of the electron–electron collision integral on the right side of Eq. (4.3). In addition, we have used condition (1.4) everywhere. Finally, we have neglected time derivatives of the function δf everywhere, following the usual argument that its influence on the electron density perturba-

tions caused by the nonuniform intensity of the pump field is small. For simplicity, we will assume that $\delta|\mathbf{E}|^2$ is purely real in what follows.

In analyzing the consequences of Eq. (4.3), the most important thing to note is that, since inequality (1.8) is satisfied for the case of interest to us, the electrons are collisionless over a large region of the velocity phase space, when

$$v \gg v_T (k\ell_{ei})^{-1/4} \ll v_T . \tag{4.4}$$

This implies that in this range of velocities, in looking for δf we can neglect terms that contain collision integrals which have δf as an argument. Then it follows from Eq. (4.3) that for the thermal electrons (4.4) we have

$$\delta f_T = \frac{e^2}{4\omega_0^2 m^2}\left(\frac{1}{v}\frac{df_s}{dv}\,\delta|\mathbf{E}|^2 + \frac{1}{2}\times\delta(E_iE_j^*+E_i^*E_j)\,\frac{\partial^2 f_s}{\partial v_i \partial v_j} - \frac{1}{(i(\mathbf{kv})}\,\delta(E_iE_j^* + E_i^*E_j)\,\frac{\partial}{\partial v_i}\,J_{ei}\left[\frac{\partial f_s}{\partial v_j}\right]\right). \tag{4.5}$$

This solution allows us to write the following expression for the contribution of thermal collisionless electrons to the perturbation of the electron density:

$$\delta n_T = \frac{e^2\delta|\mathbf{E}|^2}{4\omega_0^2 m^2}\int d\mathbf{v}\,\frac{1}{v}\frac{df_s}{dv} \approx -\frac{e^2\delta|\mathbf{E}|^2}{4\omega_0^2 m^2 v_T^2}\,n \equiv -\frac{\delta v_E^2}{4v_T^2}\,n. \tag{4.6}$$

In computing the integral in Eq. (4.6) we use the fact implied by Eq. (3.26) that if we neglect small quantities of order $\alpha^{2/3}$ the electron distribution function $f_s(v)$ coincides with a Maxwellian distribution over the entire velocity phase space except a small region $v \leqslant v_L$.

For cold subthermal electrons with velocities that satisfy inequality (1.9), particle collisions are decisive. In this case the largest term in the kinetic equation is the electron–ion collision integral. However, the corresponding operator acts only on the asymmetric part of the distribution function which depends on the angle between velocities. Therefore, the asymmetric part of the distribution function will be relatively small. Consequently, expressing the perturbation of the distribution function for cold electrons in the form

$$\delta f_c = \delta f_0 + \delta f_a , \tag{4.7}$$

where $\delta f_0 = \langle \delta f_c \rangle$ is a distribution function averaged over the angles of the velocities, we can assume in the spirit of the usual approach due to B. I. Davydov that δf_a is relatively small, and write the following two equations from Eq. (4.3):

$$\langle i\mathbf{kv}\delta f_a\rangle - v_T^2\nu_{ee}\,\frac{1}{v^2}\frac{d}{dv}\left(v^3\left[\frac{1}{v}\frac{d\delta f_0}{dv} + \frac{\delta f_0}{v_T^2}\right] + \frac{v_L^3}{v}\frac{d\delta f_0}{dv}\right) = \sqrt{2\pi}v_T^3\nu_{ei}\,\frac{1}{v^2}\frac{d}{dv}\left(\frac{1}{v}\frac{df_s}{dv}\right)\frac{e^2\delta|\mathbf{E}|^2}{4m^2\omega_0^2}, \tag{4.8}$$

$$i\mathbf{kv}\delta f_0 - J_{ei}[\delta f_a] = \frac{e^2}{4m^2\omega_0^2}\left(i\mathbf{kv}\,\frac{1}{v}\frac{df_s}{dv}\,\delta|\mathbf{E}|^2 + \frac{1}{2}\,i\mathbf{kv}\,\frac{\partial^2 f_s}{\partial v_i \partial v_j}\,\delta(E_iE_j^*+E_i^*E_j) + \delta(E_iE_j^*+E_i^*E_j)\left(v_iv_j - \frac{1}{3}\,\delta_{ij}v^2\right)\frac{1}{v}\frac{d}{dv}\,\frac{2\nu(v)}{v}\,\frac{df_s}{dv}\right). \tag{4.9}$$

Keeping in mind inequality (1.9), we write the solution to Eq. (4.9) in the form

$$\delta f_a = -\frac{i\mathbf{kv}}{2\nu(v)}\left(\delta f_0 - \frac{1}{3v^5}\frac{d}{dv}\left(v^5\frac{df_s}{dv}\right)\frac{e^2\delta|\mathbf{E}|^2}{4m^2\omega_0^2}\right) + \left(v_iv_j - \frac{1}{3}\,\delta_{ij}v^2\right)\frac{e^2}{12m^2\omega_0^2\nu(v)}\,\delta(E_iE_j^* + E_i^*E_j)\,\frac{1}{v}\frac{d}{dv}\left(\frac{\nu(v)}{v}\frac{df_s}{dv}\right). \tag{4.10}$$

The second term of this expression is irrelevant if we only want to find an equation for the symmetric part of the perturbation of the distribution function of cold collisional electrons. Substituting Eq. (4.10) into Eq. (4.8) gives

$$\frac{k^2v^2}{6\nu(v)}\,\delta f_0 - v_T^2\nu_{ee}\,\frac{1}{v^2}\frac{d}{dv}\left(v^3\left[\frac{1}{v}\frac{d\delta f_0}{dv} + \frac{\delta f_0}{v_T^2}\right] + \frac{v_L^3}{v}\frac{d\delta f_0}{dv}\right) = \sqrt{2\pi}v_T^3\nu_{ei}\,\frac{1}{v^2}\frac{d}{dv}\left(\frac{1}{v}\frac{df_s}{dv}\right)\frac{e^2\delta|\mathbf{E}|^2}{4m^2\omega_0^2} + \frac{k^2v^2}{6\nu(v)}\,\frac{1}{3v^5}\frac{d}{dv}\left(v^5\frac{df_s}{dv}\right)\frac{e^2\delta|E|^2}{4m^2\omega_0^2}. \tag{4.11}$$

Because the region of interest to us is where the distribution (3.26) differs from a Maxwellian, we will assume that the characteristic velocities of cold electrons, which determine the perturbation of the electron density, are small compared to v_L. Furthermore, we will assume that velocities are smaller than v_L throughout the entire region of cold collisional electrons. This implies that condition (1.10) is satisfied, from which we have

$$\alpha = \frac{Zv_E^2}{v_T^2} > \sqrt{\frac{8}{\pi}}\,\frac{1}{(k\ell_{ei})^{3/4}}. \tag{4.12}$$

The latter condition determines the magnitude of the pump field intensity at which radiation heating of the plasma gives rise to the new nonlinear effects we have identified in this paper. For cold collisional electrons, we have according to Eq. (3.26)

$$f_s(v) = \frac{n}{(2\pi)^{3/2}v_T^3}\left(1 - \frac{v^5}{5v_T^2v_L^3}\right). \tag{4.13}$$

This expression makes it clear that the last term ($\propto k^2$) in Eq. (4.11) gives a contribution to δn of the same order as the small discrepancy from which the integral (4.6) is calculated. Therefore, in what follows we will omit this term. Moreover, for velocities smaller than v_L electron–electron collisions are

less important in redistributing the electrons than bremsstrahlung absorption. Therefore, neglecting electron–electron collision integrals we obtain the following differential equation:

$$\frac{4}{9\pi}\frac{k^2v^5}{v_E^2v_T^6\nu_{ei}^2}\,\delta f_0-\frac{1}{v^2}\frac{d}{dv}\left(\frac{1}{v}\frac{d}{dv}\left[\delta f_0+\frac{\delta v_E^2}{v_E^2}f_s\right]\right)=0. \tag{4.14}$$

This equation is very simple and can be treated analytically. In fact, making the variable substitution $x=(v/v_k)^5$, where

$$v_k=v_T[(15\sqrt{\pi}/2)(v_E/v_T)k\ell_{ei}]^{1/5}, \tag{4.15}$$

and writing δf_0 in the form

$$\delta f_0=x^{1/5}\Psi(x)\,\frac{9n}{10\pi^{3/2}v_T^3}\,\frac{\nu_{ee}}{kv_E}\,\frac{\delta v_E^2}{v_E^2}, \tag{4.16}$$

we obtain for the function $\Psi(x)$ the following equation:

$$x^2\frac{d^2\Psi}{dx^2}+x\frac{d\Psi}{dx}-\left(\frac{1}{25}+x^2\right)\Psi=x^{4/5}. \tag{4.17}$$

We emphasize that Eq. (4.15) determines the characteristic velocity of electrons, which is an important parameter of our theory.

The solution to Eq. (4.17) that is regular at infinity has the form

$$\Psi(x)=C_1K_{1/5}(x)-I_{1/5}(x)\int_x^\infty\frac{dz}{z^{1/5}}K_{1/5}(z)$$
$$-K_{1/5}(x)\int_0^x\frac{dz}{z^{1/5}}I_{1/5}(z), \tag{4.18}$$

where $K_{1/5}$ and $I_{1/5}$ are Bessel function of imaginary argument. In order to determine the constant of integration C_1 we use the boundary condition $v^{-1}d\delta f_0/dv=0$ for $v=0$, which, like Eq. (3.24), corresponds to the absence of a source of particles with zero velocity. We emphasize that in this case in particular there are qualitative differences between the approach given here and previous theories, in which the primary process that establishes the symmetric part of the distribution function is electron–electron collisions.[5] As a result, we obtain

$$C_1=-2^{-1/5}\pi^{-1/2}\Gamma\left(\frac{3}{10}\right)\sin\frac{\pi}{5}. \tag{4.19}$$

Equations (4.16)–(4.19) allow us to find the following expression for the contribution of cold subthermal electrons to the density perturbation:

$$\frac{\delta n_c}{n}=-\frac{\delta v_E^2}{4v_T^2}\,\frac{1}{k\ell_{ee}(k\ell_{ei})^{3/5}}\left(\frac{v_T}{v_E}\right)^{12/5}C_0, \tag{4.20}$$

where $\ell_{ee}=Z\ell_{ei}$,

$$C_0=-\left(\frac{2^{12}3^{13}}{\pi5^7}\right)^{1/5}\int_0^\infty\frac{dx}{x^{1/5}}\,\Psi(x)=44. \tag{4.21}$$

Equations (4.6) and (4.20) give the following final result for nonlinear perturbations of the density of electrons by a nonuniform electromagnetic field:

$$\frac{\delta n}{n}=-\frac{\delta v_E^2}{4v_T^2}\left[1+\frac{44}{k\ell_{ee}(k\ell_{ei})^{3/5}}\left(\frac{v_T}{v_E}\right)^{12/5}\right]. \tag{4.22}$$

The weakly collisional contribution (4.20) to this expression exceeds the ponderomotive contribution (4.6) when

$$\frac{Zv_E^2}{v_T^2}\equiv\alpha<\frac{23Z^{1/6}}{(k\ell_{ei})^{4/3}}. \tag{4.23}$$

The latter inequality combined with (Eq. 4.12) is realized when

$$k\ell_{ei}<100Z^{2/7}. \tag{4.24}$$

Even when the usual condition (1.8) for the plasma to be collisionless holds, inequality (4.24) still defines a very sizable wavelength region where weak-collisional nonlinearities can occur. The new nonlinearity introduced by Eq. (4.22) corresponds to decreasing the weak-collisional contribution as the pump field grows.

5. COEFFICIENT OF NONLOCAL EFFECTIVE ELECTRON HEAT CONDUCTIVITY

The results of the previous section allow us to derive the nonlinear change in the effective electron thermal conductivity induced by a weak field heating the plasma. In doing so, we focus on perturbations of the kinetic energy density of the electrons by the nonuniform radiation field

$$\frac{3}{2}\,\delta(nk_BT)\equiv\int d\mathbf{v}\,\frac{mv^2}{2}\,\delta f. \tag{5.1}$$

Because of the low velocities of the cold subthermal electrons, their kinetic energy density is small compared to $(3/2)k_BT\delta n_c$. Therefore, the perturbation of the temperature of the cold electrons δT_c will be

$$\frac{\delta T_c}{T}=-\frac{\delta n_c}{n}=\frac{\delta v_E^2}{4v_T^2}\,\frac{1}{k\ell_{ee}(k\ell_{ei})^{3/5}}\left(\frac{v_T}{v_E}\right)^{12/5}C_0. \tag{5.2}$$

Comparing this expression with the corresponding result of the linear theory,[5] we note that in addition to the new dependence on the wave vector a new nonlinear dependence on the spatially uniform pump feeding the field appears as well: as the pump intensity increases, the perturbation of the temperature decreases.

For kinetic energy perturbations of the thermal electrons, according to Eq. (4.5) we have

$$\frac{3}{2}\,\delta(nk_BT)_T=\frac{3}{2}\,k_B(n\,\delta T_T+T\delta n_T)=-\frac{\delta v_E^2}{4}\,\frac{3}{2}\,mn. \tag{5.3}$$

In light of Eq. (4.6), it is not difficult to see that (4.5) implies $\delta T_T=0$ to this order of approximation. The perturbation of the effective temperature of thermal electrons acquires a nonzero value when we take their collisions into account, and turns out to be of order

$$\delta T_T\sim T(\delta v_E^2/v_T^2)(k\ell_{ei})^{-1}.$$

The latter expression is small compared to Eq. (5.2) by virtue of the inequality

$$\frac{Zv_E^2}{v_T^2} \equiv \alpha < \frac{23Z^{1/6}}{(k\lambda_{ei})^{1/2}}, \tag{5.4}$$

which in view of inequality (1.8) is always satisfied when condition (4.23) holds. Therefore, under the conditions of interest to us the primary contribution to heating of the electrons by a nonuniform electromagnetic field comes from the cold electrons, and is determined by Eq. (5.2).

The divergence of the electron heat flux density **q** arising from heating of the plasma by spatially nonuniform radiation, as follows from Eq. (4.3), is determined in the following way:

$$\text{div } \mathbf{q} = Q_0 + \delta Q, \tag{5.5}$$

where

$$Q_0 = \frac{e^2}{4\omega_0^2 m^2} \int d\mathbf{v} v_i J_{ei} \left[\frac{\partial f_s}{\partial v_j} \right] \delta (E_i E_j^* + E_i^* E_j), \tag{5.6}$$

$$\delta Q = \frac{e^2}{4\omega_0^2 m^2} \int d\mathbf{v} v_i J_{ei} \left[\frac{\partial \delta f_s}{\partial v_j} \right] (E_i E_j^* + E_i^* E_j). \tag{5.7}$$

Since, the function f_s does not differ from Maxwellian over most of the velocity phase space, the heat Q_0 generated per unit time by the nonuniform radiation field acting on the unperturbed electron distribution can be written in the form

$$Q_0 = \frac{1}{2} mn \delta v_E^2 \nu_{ei}. \tag{5.8}$$

This expression arises from the contribution of the thermal electrons and coincides with the result obtained by direct calculation of div **q** using the distribution of collisionless thermal electrons Eq. (4.5). The contribution of thermal electrons to Eq. (5.7) is smaller than Eq. (5.8) by a factor of v_E^2/v_T^2. The contribution of subthermal electrons to Eq. (5.7) is given by the following expression:

$$\delta Q = -\frac{1}{2} mn \delta v_E^2 \nu_{ei} \frac{\nu_{ee}}{k v_E} C_2, \tag{5.9}$$

where

$$C_2 = \frac{9}{5\sqrt{2\pi}} \Gamma\left(\frac{1}{5}\right) \Gamma\left(\frac{3}{10}\right) \sin\frac{\pi}{5} \approx 5.8. \tag{5.10}$$

Expression (5.9) indicates that the redistribution of cold electrons caused by the nonuniform electromagnetic field releases heat from the plasma, which corresponds to bremsstrahlung of the nonequilibrium subthermal electrons. Condition (4.12) implies that Eq. (5.9) is small compared to Eq. (5.8) by the small factor $[Z^{1/2}(k\lambda_{ei})^{5/8}]^{-1}$. Therefore, we may neglect the contribution (5.9) to the heat balance equation. Then we have

$$i\mathbf{kq} = \frac{1}{2} mn \delta v_E^2 \nu_{ei}$$

$$= k^2 n k_B \delta T v_T \lambda_{ei} \frac{Z}{(k\lambda_{ei})^{2/5}} \left(\frac{v_E}{v_T} \right)^{12/5} \frac{2}{C_0}. \tag{5.11}$$

Here the relation defined by Eq. (5.2) between δv_E^2 and the temperature increment $\delta T = \delta T_c$ has been used. Expression (5.11) is written in such a way that it allows us to determine the effective nonlocal thermal conductivity for the irrotational part of the heat current $\mathbf{q} = -\chi \text{ grad}(\delta T)$

$$\chi_{\text{eff}}(k) = \frac{\chi_{SH}}{1 + 300(Zv_E^2/v_T^2)^{-6/5}(Zk^2\lambda_{ei}^2)^{1/5}}. \tag{5.12}$$

The 1 in the denominator of Eq. (5.12) has been added in order to make this expression interpolate between Eq. (5.11) and the well-known expression for the Spitzer–Härm thermal conductivity, $\chi_{SH} = 13.6 n v_T k_B \lambda_{ei}$ as $k \to 0$.

6. CONCLUSION

Expression (5.12) modifies the results of the linear theory of nonlocal thermal conductivity[5] by introducing a new fractional-power dependence on the wave vector k and a nonlinear dependence on the intensity of the heating field.

However, the principal result of our article is Eq. (4.22), which describes the new nonlinear behavior of perturbations of the plasma density modified by radiation-induced heating. Expression (4.22) requires reexamination of the theory of a number of parametric instabilities in plasmas under conditions that are typical of laser-generated plasma experiments directed toward controlling laser-induced thermonuclear fusion.

Thus, our theory reveals a new effect: a nonlinear contribution to the heat transport by the field that heats the plasma through reverse bremsstrahlung absorption.

This work was financially supported by the Russian Fund for Fundamental Research (Grant No. 2.46) and the State Program (Grant 96-02-17002) on ''Optics and Laser Physics.''

*)E-mail: silin@sci.lpi.ac.ru

[1] V. P. Silin, *Parametric Excitation of High-Power Radiation by a Plasma* (Nauka, Moscow, 1973) [in Russan].
[2] A. V. Maksimov, V. P. Silin, and M. V. Chegotov, Fiz. Plazmy **16**, 575 (1990) [Sov. J. Plasma Phys. **16**, 331 (1990)].
[3] A. B. Langdon, Phys. Rev. Lett. **44**, 575 (1980).
[4] R. Balescu, J. Plasma Phys. **23**, 553 (1982).
[5] A. V. Maksimov and V. P. Silin, Zh. Éksp. Teor. Fiz. **103**, 73 (1993) [JETP **76**, 39 (1993)].
[6] A. V. Maksimov and V. P. Silin, Zh. Éksp. Teor. Fiz. **105**, 1242 (1994) [JETP **78**, 669 (1994)].
[7] V. P. Silin, Zh. Éksp. Teor. Fiz. **106**, 1398 (1994) [JETP **79**, 756 (1994)].
[8] E. M. Epperlein and R. W. Short, Phys. Rev. E **50**, 1697 (1994).
[9] V. P. Silin, Zh. Éksp. Teor. Fiz. **108**, 193 (1995) [JETP **81**, 103 (1995)].

Translated by Frank J. Crowne

"Shallow-water" and "deep-water" approximations in the theory of the disruptive instability of thin current-carrying layers

S. K. Zhdanov*)

Moscow State Institute of Engineering Physics, 115409 Moscow, Russia

V. P. Vlasov

"Kurchatov Institute" Russian Science Center, 123182 Moscow, Russia

(Submitted 31 September 1997)
Zh. Éksp. Teor. Fiz. **113**, 1313–1318 (April 1998)

Within the framework of the two-fluid hydrodynamics of plasmas it is shown that the problem of instability of a thin current-carrying layer admits two limiting cases which allow analytic solutions and complement one another. These limits are analogous to the well-known shallow-water and deep-water approximations in the fluid mechanical "wave-breaking" instability. In this case, the long-wave limit coincides with the "quasi-Chaplygin" dynamic system of Bulanov and Sasorov, Fiz Plazmy **4**, 746 (1978) [Sov. J. Plasma Phys. **4**, 418 (1978)], while the short-wavelength limit corresponds to the phenomenological model of Trubnikov, Usp. Fiz. Nauk **160**, 167 (1990) [Sov. Phys. Usp. **33**, 87 (1990)], for the clumping of "elementary" currents. In the latter case, strong collapse is unavoidable with the appearance of current filaments that trap a finite current. © *1998 American Institute of Physics.* [S1063-7761(98)01404-8]

1. The tearing instability and the associated phenomena of reconnection of magnetic field lines and formation of magnetic islands are fundamental aspects of the physics of neutral current layers. Many original papers and reviews have been devoted to them (see, for example, Refs. 2–5 and the papers cited in them) in which the instability is analyzed in all of its manifestations. At present, the physics of this instability is clear in broad outline. It is well known that the tearing mode can be interpreted from a physical point of view as the result of a pinch in the distributed current layer—a kind of instability that leads to "clumping" due to the mutual attraction of individual elementary current filaments that remain when the layer inevitably breaks up.[2–4] Thus, the disruption and clumping, which would appear to be mutually exclusive phenomena, are organically combined as this instability evolves. However, this fundamental feature of the theory of unstable thin current-carrying layers manifests itself in a most peculiar way, in that two approaches that have seemed at first glance to contradict one another are now combined.[1,2] In fact, as we will show in this paper, the nonlinear dynamic models proposed in Refs. 1 and 2 are two sides of the same coin, not mutually exclusive but rather complementing one another. We show here that they correspond to two limiting cases with respect to a certain parameter that depends on the product of the line density and the magnitude of the characteristic scale of the perturbation. This fact—which, as far as we know, has not been noted before—is an informative contribution to the theory of the tearing instability of thin current layers.

2. The model first proposed by Bulanov and Sasorov in Ref. 1 can be written in the following way, incorporating the correction pointed out in Ref. 6:

$$\rho_t' + (\rho v)_x' = 0, \quad v_t' + v v_x' = -(1/2) c_0^2 ((\rho_0/\rho)^2)_x', \quad (1)$$

where $c_0 = 2(c_A \delta/a)_0$, $\rho = na$ is the line density of the plasma in the current layer, a is the layer thickness, c_A is the Alfvén velocity, $\delta_0 = c/\omega_{pe}$ is the vacuum skin depth, and the label "0" denotes unperturbed values of quantities. The perturbations are assumed to be long-wavelength, so that their characteristic scale λ greatly exceeds the thickness of the current layer, $\lambda \gg a$. A key issue to address is that Eq. (1) will only be valid under the assumption that the line density of current in the layer is constant, implying that the discontinuity in the tangential component of the magnetic field across the layer is constant as well. This requirement obviously contradicts the picture of clumping of currents outlined above. In the phenomenological model of Trubnikov[2] (where $\hat{H}$ is the Hilbert operator):

$$\rho_t' + (\rho v)_x' = 0, \quad v_t' + v v_x' = g_T \hat{H} \rho,$$
$$g_T = 2\pi I_1^2 / M_1 c^2 h, \quad (2)$$

the fundamental picture is entirely different. Here the thin current layer is a set of identical parallel wires—"elementary currents," each of which has a current I_1 and mass M_1 (per unit length); the value of ρ equals the product of the distance h between neighboring wires in equilibrium and the number of wires passing along the x axis per unit length (along the layer). These quantities also give meaning to the acceleration g_T that appears in Eq. (2) (the Trubnikov acceleration).

Systems (1) and (2) are not merely externally different, since they predict different dependences on wave number for the growth rate of the instability in the linear stage of its evolution:

$$\gamma_{BS} = k c_0, \quad \gamma_T = (k g_T)^{1/2}, \quad (3)$$

for Eqs. (1) and (2) respectively. This difference in the wavenumber dependence of the growth rate outwardly recalls what happens in the "wave-breaking" instability of a layer

of water,[6] which must be treated as either shallow or deep, depending on the ratio of the perturbation wave length to the layer thickness. In this case, the former situation has a growth rate proportional to the wave number, while in the second situation it is proportional to the square root of the wave number. In reality, this analogy is even deeper: the role of the parameter that separates long- and short-wavelength regions is played by the quantity $\varepsilon=\delta_0^2/a_0\lambda$.

3. In order to justify this assertion, let us follow Refs. 1 and 6 and use the equations of two-fluid hydrodynamics in the following form:

$$(na)_t'+(nav)_x'=0,$$

$$v_t'+vv_x'=-\frac{j_zB_y}{cn(m_i+m_e)}\equiv-\frac{|e|B_yu_j}{c(m_i+m_e)}. \tag{4}$$

Here the geometry of the layer is chosen so that current flows along the z axis; the coordinate y is transverse to the layer, and the coordinate x is measured along the layer. In this case, the usual notation is used, and in addition the plasma is treated as quasineutral: $n_e=n_i=n$, and the longitudinal velocities are assumed to be equal: $v_{xe}=v_{xi}=v$. For convenience we also introduce a current velocity

$$u_j=v_{zi}-v_{ze}=j_z/|e|n.$$

Let us calculate the value of this current velocity, taking into account the law of conservation of generalized momentum of ions and electrons which follows from the assumed uniformity of the layer with respect to coordinate z:

$$(mv_z+eA/c)_{e,i}=\text{const}, \tag{5}$$

where the quantity A is the z-component of the vector potential; the component of the magnetic field can be written in terms of it as follows:

$$B_x=\partial A/\partial y,\quad B_y=-\partial A/\partial x,\quad B_z=0.$$

Using Eq. (5), we obtain the z-component of the current velocity:

$$u_j=u_{j0}-|e|A/\mu c, \tag{6}$$

where $\mu=m_im_e/(m_i+m_e)$, and u_{j0} is the initial value, which is assumed to be constant. In Eq. (4), all the quantities are referred to the layer, i.e., for $y=0$. Outside the layer, the vector potential satisfies the following equation (neglecting displacement current):

$$\Delta A=(\partial^2/\partial x^2+\partial^2/\partial y^2)A=0, \tag{7}$$

which must be supplemented by the boundary condition

$$[\partial A/\partial y]_{y=0}=-(4\pi/c)j_za\equiv-(4\pi/c)|e|na(u_{j0}-|e|A/\mu c)_{y=0}, \tag{8}$$

where $[\partial A/\partial y]_{y=0}$ is the jump in $\partial A/\partial y$ transverse to the current layer. Following Refs. 1 and 2, we assume that the layer is immersed in vacuum; if this is not true, we must maintain a current on the right-hand sides of Eqs. (7) and (8) induced by the instability in the medium surrounding the layer.

We note that the formal solution to Eqs. (7) and (8) is the relation

$$B_y=-(2\pi/c)\hat{H}(j_za). \tag{9}$$

Relations of this kind were probably first introduced in similar problems involving currents in thin layers by Chukbar (see, e.g., Ref. 7).

It is further convenient to write the vector potential in the form of a sum

$$A=A_0+A_1,\quad A_0=-B_0|y|,\quad B_0=(2\pi/c)|e|n_0a_0u_{j0}, \tag{10}$$

which explicitly distinguishes the unperturbed field A_0. We now choose characteristic scales for the quantities λ, $\mu cu_{j0}/|e|, n_0, a_0$, i.e., the spatial coordinate, vector potential, ion density of the plasma, and layer thickness respectively. Then in dimensionless form condition (8) for the correction A_1 to the vector potential can be written as follows:

$$\varepsilon[\partial A_1/\partial y]_{y=0}=-(na(1-A_1)-1)_{y=0}, \tag{11}$$

where $\varepsilon=\delta_0^2/a_0\lambda$ is the parameter introduced above.

It is obvious that when ε is small, i.e., as assumed in Ref. 1, the left side of Eq. (11) may be treated as a small quantity, and to this accuracy we obtain from Eq. (11) the value of the potential at the layer:

$$A_1\approx 1-(na)^{-1}. \tag{12}$$

Using it in Eqs. (4), we can easily verify that Eq. (1) results. Obviously Eq. (12) is identically the condition for constancy of the line density of the current $j_za=\text{const}$, while the condition $\varepsilon\ll 1$ defines the shallow-water region, i.e., the constraint on the wavelengths of the perturbations $\lambda\gg\delta_0^2/a_0$.

In the opposite limit, when ε is large, we can expand in powers of ε^{-1} and neglect terms with the vector potential on the right side of Eq. (11), and accordingly in Eq. (8). It is easy to see that this is equivalent to the condition of constancy of the current velocity $u_j=\text{const}$. In this limit, it follows that the line current density equals approximately

$$j_za\approx|e|u_{j0}na, \tag{13}$$

and is proportional to the plasma line density of the current layer. Using Eq. (9) once more, we can convince ourselves that Eqs. (4) now reduce to

$$(na)_t'+(nav)_x'=0,$$

$$v_t'+vv_x'=g_H\hat{H}(na/n_0a_0), \tag{14}$$

which is equivalent to Eq. (2). Meanwhile, the phenomenological Trubnikov acceleration calculated from fluid mechanics is found to be

$$g_H=\frac{2\pi e^2(u_j^2na)_0}{(m_i+m_e)c^2}. \tag{15}$$

The condition $\varepsilon\gg 1$ combined with the condition $a_0\ll\lambda$ defines the short-wavelength range for the perturbations $a_0\ll\lambda\ll\delta_0^2/a_0$ (deep water in hydrodynamic terminology).

4. It is well known[1,6] that within the framework of model (1) the solutions typically have a tendency to break up the current layer in such a way that the line density decreases locally but not the current (!), which clearly contradicts the physical picture of the instability. The results obtained above

allow us to resolve this paradox, which has existed since the time of publication of Ref. 1. The fact is that in parameter ranges where the current layer breaks up, as the line density decreases a point is inevitably reached where the condition for applicability of Eqs. (1) is violated, i.e., the current density can no longer be treated as constant. According to Eq. (13), decreasing the line density leads to a proportional decrease in the current density, so that disruption of the current layer is necessarily accompanied by filamentation of the current in complete correspondence with the physics of the tearing instability. We will show that whenever Eqs. (14) are applicable, a strong collapse is inevitable with the appearance of current filaments that trap a finite current. Note that system (14) can be written in the equivalent Hamiltonian form:

$$\Psi_t' = -\delta H/\delta\rho, \quad \rho_t' = \delta H/\delta\Psi, \quad v = \Psi_x', \quad \rho = na/(na)_0, \tag{16}$$

where the ''Hamiltonian'' is

$$H = \int dx((\rho/2)[\Psi_x']^2 - (1/2)g_H\rho[\hat{K}\rho]). \tag{17}$$

Here the operator $\hat{K}$ is defined by the relation

$$\partial(\hat{K}\rho)/\partial x \equiv \hat{H}\rho.$$

Obviously, conservation laws apply for ''energy,'' ''momentum,'' and ''particle number'' respectively:

$$H = \text{const}, \quad P = \int dx(\rho\Psi_x') = \text{const},$$

$$N = \int dx(\rho) = \text{const}. \tag{18}$$

The scale transformation $x \to bx$, $\rho \to \rho/b$, $\Psi \to b\Psi$, which conserves the ''momentum'' P and ''particle number'' N, gives

$$H \to H - ([g_H/2\pi]N^2 \ln(1/b)), \tag{19}$$

and as the scale decreases $(b \to 0)$ the Hamiltonian can only decrease–a clear advantage!

In conclusion, we give an example of an exact solution in the form of an isolated collapsing filament:[8]

$$v = \xi(dL/dt), \quad \rho = (L_0/L)(1-\xi^2)^{1/2}, \quad \xi = x/L, \quad |x| \leqslant L, \tag{20}$$

where $L = L(t)$ is the half-width of the filament ($\rho \equiv 0$ for $|x| \geqslant L$), and L_0 is its initial value. The size of the filament is determined by the equation

$$Ld^2L/dt^2 = -g_H L_0,$$

according to which at time

$$t = t_c = [\pi L_0/(2g_H)]^{1/2}$$

this current filament ''collapses,'' trapping a finite current

$$I_f \sim \int_{-L}^{+L} \rho\, dx = \text{const}.$$

Elementary analysis shows that including the interaction with neighboring filaments located periodically in the current layer drags out the collapse, but does not eliminate it completely. Periodic systems of current filaments are obviously unstable against modulation with respect to position or value of the currents in the filaments, and this instability can also be described within the framework of Eqs. (14).

The authors are grateful to B. A. Trubnikov for discussing this work and also to the participants of the seminar of academic V. D. Shafranov for fruitful discussions.

*)E-mail: root@plasm.mephi.msk.su

[1] S. V. Bulanov and P. V. Sarasov, Fiz. Plazmy **4**, 746 (1978) [Sov. J. Plasma Phys. **4**, 418 (1978)].
[2] B. A. Trubnikov, Usp. Fiz. Nauk **160**, 167 (1990) [Sov. Phys. Usp. **33**, 87 (1990)].
[3] A. A. Galeev, in the *Handbook of Plasma Physics*, A. A. Galeev and R. N. Sudan (Eds.) (North-Holland, Amsterdam, 1984; Energoatomizdat, Moscow, 1984), Vol. 2, p. 331.
[4] B. B. Kadomtsev, Usp. Fiz. Nauk **151**, 3 (1987).
[5] S. V. Bulanov, G. I. Dudnikova, T. Zh. Esipenkov *et al.*, Fiz. Plazmy **22**, 867 (1996) [Plasma Phys. Rep. **22**, (1996)].
[6] S. K. Zhdanov and B. A. Trubnikov, *Quasi-Gaseous Unstable Media* (Nauka, Moscow, 1991) [in Russian].
[7] K. V. Chukbar, Fiz. Plazmy **19**, 1491 (1993) [Plasma Phys. Rep. **19**, 785 (1993)].
[8] S. K. Zhdanov and V. P. Vlasov, *Abst. Rep. 24th Zvenigorod Conf. on Plasma Physics and Thermonuclear Fusion*, Russia (1997), p. 86.

Translated by Frank J. Crowne

Generation of electromagnetic radiation in the motion of vortices in magnetically coupled superconducting films

V. A. Kozlov and A. V. Samokhvalov*)

Institute of Microstructure Physics, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia
(Submitted 19 June 1997)
Zh. Éksp. Teor. Fiz. **113**, 1319–1338 (April 1998)

We examine theoretically the generation of electromagnetic radiation in the relative motion of vortex lattices in magnetically coupled films in the dc transformer geometry. We establish the conditions under which the force of mutual pinning of the vortex lattices varies according to a harmonic law as a function of the relative displacement of the vortices in the films within a given range of magnetic field inductions. In this case the equation describing the viscous flow of vortex lattices in magnetically coupled films is the same as the equation of the resistively shunted Josephson junction model. We show that magnetically coupled superconductors exhibit the properties of a Josephson element without any restrictions on the geometrical size of such a system imposed by the coherence length ξ. The frequency f of the electromagnetic radiation generated by the relative motion of vortex lattices in magnetically coupled superconductors depends on the spatial period of the vortex lattices and the velocity of relative vortex motion, which means that the frequency of the radiation can be tuned by applying a magnetic field or a current. © *1998 American Institute of Physics.* [S1063-7761(98)01504-2]

1. INTRODUCTION

One manifestation of the time-dependent Josephson effect[1] in superconducting systems with weak links is the generation of narrow-band electromagnetic radiation. However, for the weak link to possess the properties of a Josephson junction, the characteristic geometrical size of the weak-link region must not be much larger than the coherence length ξ in the superconductor.[2,3] This condition strongly hinders the fabrication of reproducible Josephson junctions based on high-T_c superconductors, where the typical values of ξ are 3–10 Å. Hence it would be interesting to search for analogs of the Josephson effect in superconducting systems, in which such strong constraints on the size of the elements do no exist.

There is a close analogy between the time-dependent Josephson effect in systems with weak links and the resistive state of a superconductor in the flux-flow regime,[4] and a moving Abrikosov-vortex lattice[5] can be represented by an array of weak links.[6] Fiory[7] and Lee *et al.*[8] observed coherent vibrations of vortex lattices by the emergence of steps in the current–voltage characteristic of the superconducting microbridge generated by the rf component of an external current. For such quantum interference effects to exist the vortices must be pinned. Pinning guarantees that there is interaction between the ac component of the current from an external source and the natural oscillations of the supercurrent, which result from modulation of the mean vortex velocity.[9] Quantum interference becomes especially evident if coherent motion of a large number of vortices with an appreciable variable component of the velocity can be arranged.[7] This is achieved by using samples with a well-developed structure of pinning centers, e.g., by modulating the thickness of the superconducting film[10] or by creating a regular lattice of small (submicrometer) cylindrical microholes (antidots) in the film.[11]

The analogy between the resistive state of a superconducting microbridge and the time-dependent Josephson effect also manifests itself in the existence of intrinsic electromagnetic radiation generated by bridge structures in the flux-flow regime.[8,12,13] Such radiation is emitted if there is nonuniform motion of a vortex lattice, e.g., in the event of spatially nonuniform pinning. In the case of a weak random pinning potential, it is primarily the noise component that is emitted, with the nature of the emitted radiation determined largely by the nature and strength of the pinning.[14] The specific features of the rf response to the motion of an Abrikosov-vortex lattice in a periodic pinning potential generated by modulating the thickness of a superconducting film was observed by Martinoli *et al.*[15] However, the weakness of the link in the vortex lattice, the low pinning potential, and the difficulties encountered in monitoring the strength of this link made it impossible to generate appreciable radiative power.

Systems that are undoubtedly of interest as generators of electromagnetic radiation in which both the size of the coupling force between the vortex lattice and pinning potential and the spatial period of the pinning potential can be varied in a controllable manner. Obviously, in a periodic pinning potential such a coupling force is of a resonant nature; it is strongest if the period of the vortex lattice matches the period of the pinning potential.[15] The strength of the coupling force can be varied if there is simultaneous self-matching of the periods of the vortex lattice and the pinning potential, provided that the magnetic interaction between two identical

vortex lattices arranged in electrically insulated parallel superconducting films (as in a dc transformer[16]) is utilized. The periodic dependence of the interaction energy of vortex lattices in the two films on the relative displacement of these lattices is ensured by the existence of a self-matching periodic pinning potential,[17] and the necessary coupling force can easily be selected by changing the gap between the superconducting films. A variable component of the vortex velocity and hence a variable electric field and current are generated by the relative motion (slippage) of the vortex lattices in the two films.

The key issue in building a theory of magnetically coupled superconductors is the assumption[17] that a vortex lattice is displaced as a whole in the periodic pinning potential generated by another vortex lattice. The interaction energy between vortex lattices in magnetically coupled films was first calculated by Sherrill and Manson[18,19] for the case of Abrikosov vortices in thin films,[20] with the film thickness d much smaller than the magnetic-field penetration depth λ. These results were generalized to films of arbitrary thickness by Clem,[21,22] who also built a detailed theory describing the stripping of the vortex lattices in magnetically coupled films and did model calculations of the current–voltage characteristics of dc transformers. The theoretical ideas developed in Refs. 17, 21, and 22 were corroborated by detailed measurements of the current–voltage characteristics and the magnetic interaction between vortices in magnetically coupled films.[23,24]

In this paper we examine electromagnetic radiation due to the relative motion of vortex lattices in magnetically coupled films in the dc transformer geometry. In Sec. 2 we calculate the structure of the magnetic field of a square Abrikosov-vortex lattice formed by an external magnetic field in magnetically coupled superconducting films, and study the interaction between the vortex lattices in the two films. In Sec. 3 we write the equations of vortex motion in the periodic pinning potential generated by the magnetic interaction between the vortex lattices. The equations are similar to the equation of the resistively shunted Josephson junction model,[2] in which the relative displacement of the vortex lattices in the films acts as the phase difference φ. In Sec. 4 we study the rf response of magnetically coupled superconductors to the relative motion of vortex lattices in neighboring films. The frequency ω of the electromagnetic radiation depends on the spatial period of the vortex lattices, which means that it is possible to tune the frequency of the radiation by applying an external magnetic field. In Sec. 5 we discuss examples of Josephson systems based on magnetically coupled films. There we also compare the behavior of such systems with that of an ordinary Josephson junction.

2. MUTUAL VORTEX PINNING IN MAGNETICALLY COUPLED SUPERCONDUCTORS

We consider two parallel films of a type-II superconductor, of thicknesses d_1 and d_2, separated by an insulating layer of thickness d_3 (Fig. 1). We assume that generally the films are different, and each is characterized by a magnetic-field penetration depth λ_l and a coherence length ξ_l, where

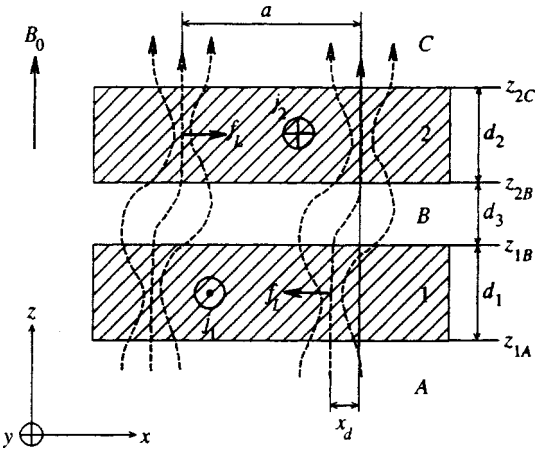


FIG. 1. Magnetically coupled superconducting films in an external magnetic field B_0. The structure of the magnetic lines is shown schematically for an arbitrary relative displacement x_d of the vortex lattices in the films.

the subscript l takes the values $l=1,2$ for the first and second films, respectively. For simplicity we assume that the external perpendicular magnetic field with induction $\mathbf{B}_0=B_0\mathbf{z}_0$ generates in each film a square Abrikosov-vortex lattice. The vortex lattice period a is determined by B_0 in accordance with the relation $B_0a^2=\Phi_0$, where $\Phi_0=\pi\hbar c/e$ is the quantum of magnetic flux, or fluxon. We also assume that the thickness d_3 of the insulating layer is so large that tunneling can be ignored ($d_3\gg\xi_l$). On the other hand, the attraction between vortices belonging to different films guarantees significant magnetic interaction between the vortex lattices. We choose a coordinate system (x,y,z) with x and y axes parallel to the basic vectors of a vortex lattice; the plane $z=0$ lies in the middle of the insulating layer, and the superconducting films occupy the regions $z_{1A}\leqslant z\leqslant z_{1B}$ and $z_{2B}\leqslant z\leqslant z_{2C}$, where $z_{1A}=-d_1-d_3/2$, $z_{2B}=-z_{1B}=d_3/2$, and $z_{2C}=d_2+d_3/2$. For definiteness we consider the case in which the external currents in the films are directed along one of the basic vectors of a vortex lattice: $\mathbf{j}_{1,2}\|\mathbf{y}_0$, and the current-induced flux flows along the x axis.

If there is no external current in the films, the Gibbs energy G of such a system is at its minimum when the axes of the vortices in the two films are opposite each other and there is no relative displacement x_d of the vortex lattices, i.e., $x_d=0$. Obviously, by virtue of translation invariance, the G vs. x_d dependence is a periodic function of the displacement x_d: $G(x_d+na)=G(x_d)$ for integer n. An external current of density $\mathbf{j}$ flowing in the films acts on unit length of each vortex line with a Lorentz force $\mathbf{f}_L=\Phi_0\mathbf{j}\times\mathbf{z}_0/c$, and the behavior of the vortex lattices is determined by the energy $G'=G-\Delta W$ (see Ref. 25), where ΔW is the work done by the current source in moving the vortices. The Lorentz force shifts the vortex lattices with respect to each other along the x axis, and the minimum in G' corresponds to an arrangement of vortices for which $x_d\neq 0$. If the external current density $\mathbf{j}$ is so high that the Lorentz force f_L exceeds the force of internal (volume and surface) pinning of the vortices in the films, f_p, as well as the maximum magnetic interaction force between the vortex lattices in the two films, $f_c^{\max}$,

then there is relative motion (slippage) of the vortex lattices, which leads to a periodic time-dependence of the vortex velocity. The corresponding dependence of the energy G' on the relative displacement x_d is a monotonically decreasing function with a periodic variation of the slope of the G' vs. x_d curve. Our qualitative treatment of the dependence of the vortex lattice energy in magnetically coupled films on the relative displacement x_d shows that the behavior of this system corresponds to that of a Josephson element.[26]

2.1. An Abrikosov-vortex lattice in magnetically coupled superconductors

We begin by calculating the magnetic field of a square lattice of Abrikosov vortices in magnetically coupled superconductors when there is relative displacement x_d of the lattices (Fig. 1). In the London approximation, which holds for superconductors with a large Ginzburg–Landau parameter $\kappa_l=\lambda_l/\xi_l\gg 1$, the distribution of the magnetic field $\mathbf{H}_l$ in a superconductor in each of the films is described by the London equation

$$\hat{L}_l(\mathbf{H}_l)\equiv\mathbf{H}_l+\lambda_l^2\nabla\times\nabla\times\mathbf{H}_l = \Phi_0\mathbf{z}_0\sum_{n,m=-\infty}^{+\infty}\delta(\mathbf{r}-\mathbf{r}_l(n,m)), \tag{1}$$

where the vector $\mathbf{r}_l(n,m)=(x_l+na)\mathbf{x}_0+ma\mathbf{y}_0$ determines the position of the vortex lines, and x_l specifies the displacement of the vortex lattice in relation to the zero position corresponding to the coaxial arrangement of vortices in the magnetically coupled films (n and m are arbitrary integers). We employ the fact that the London approximation is linear and write the desired general solution of the differential equation (1) as a linear combination of two solutions,

$$\mathbf{H}_l=\mathbf{H}_{lv}+\mathbf{H}_{ld}, \tag{2}$$

where the first term $\mathbf{H}_{lv}$ is a particular solution of the inhomogeneous equation (1),

$$\hat{L}_l(\mathbf{H}_{lv})=\Phi_0\mathbf{z}_0\sum_{n,m}\delta(\mathbf{r}-\mathbf{r}_l(n,m)), \tag{3}$$

and the second term $\mathbf{H}_{ld}$ is the general solution of the corresponding homogeneous London equation

$$\hat{L}_l(\mathbf{H}_{ld})=0, \tag{4}$$

which guarantees that the magnetic field components at the film surfaces are continuous.

We select the particular solution $\mathbf{H}_{lv}=H_{lv}\mathbf{z}_0$ of the inhomogeneous equation (3) in a form that describes the structure of the magnetic field of a square lattice of Abrikosov vortices in a massive superconductor. This solution can be written as a two-dimensional Fourier series in spatial harmonics:

$$H_{lv}=\frac{\Phi_0}{a^2}\sum_{n,m}S_l(n,m)\exp[iq_a(nx+my)-i\chi_l], \qquad q_a=\frac{2\pi}{a},\quad \chi_l=q_anx_l. \tag{5}$$

To broaden the range of applicability of the London approximation, we use the Clem model,[27] which provides a good description of a single Abrikosov vortex for an arbitrary value of the Ginzburg–Landau parameter κ. Representing the vortex lattice by a linear superposition of single vortices, we arrive at the following expression for the spatial Fourier spectrum $S_l(n,m)$:

$$S_l(n,m)=\frac{K_1(q_l\xi_l^v)}{q_l\lambda_lK_1(\xi_l^v/\lambda_l)}, \tag{6}$$

$$q_l\equiv q_l(n,m)=q_a\sqrt{n^2+m^2+\left(\frac{a}{2\pi\lambda_l}\right)^2}, \tag{7}$$

where the parameter ξ_l^v, which is the effective vortex-core size in the Clem model, satisfies the equation

$$\frac{\xi_l^v}{\sqrt{2}}=\xi_l\left[1-\frac{K_0^2(\xi_l^v/\lambda_l)}{K_1^2(\xi_l^v/\lambda_l)}\right]^{1/2},$$

and $K_{0,1}(\zeta)$ are modified Bessel functions. The fact that we use the linear superposition principle in calculating the lattice field of Abrikosov vortices means that there is no overlap of normal vortex cores, i.e., $\xi_l^v\ll a$. For superconductors with a large Ginzburg–Landau parameter, $\kappa_l\gg 1$, Eq. (6) can be simplified:

$$S_l(n,m)\simeq\left(\frac{1}{q_l\lambda_l}\right)^2. \tag{8}$$

The general solution of the homogeneous London equation (4),

$$\mathbf{H}_{ld}=(H_{ld}^x,H_{ld}^y,H_{ld}^z),$$

which guarantees that the appropriate boundary conditions at the film surfaces are met, will also be written as a two-dimensional Fourier series in spatial harmonics:

$$H_{ld}^\sigma=\frac{\Phi_0}{a^2}\sum_{n,m}[C_l^{\sigma+}e^{q_lz}+C_l^{\sigma-}e^{-q_lz}]\times\exp[iq_a(nx+my)-i\chi_l], \qquad \sigma=x,\ y,\ z. \tag{9}$$

Plugging (2), (5), and (9) into Eqs. (1) and (4) results in an algebraic relationship linking the unknown expansion coefficients $C_l^{\sigma\pm}$,

$$-q_anC_l^{x\pm}-q_amC_l^{y\pm}\pm iq_lC_l^{z\pm}=0, \tag{10}$$

and makes it possible to write

$$\hat{L}_l(\mathbf{H}_l)\equiv\frac{\Phi_0\mathbf{z}_0}{a^2}\sum_{n,m}\lambda_l^2q_l^2S_l(n,m)\times\exp[iq_a(nx+my)-i\chi_l]. \tag{11}$$

Everywhere outside the superconductor (i.e., $z<z_{1A}$, $z_{1B}<z<z_{2B}$, and $z>z_{2C}$) the magnetic field $\mathbf{H}_\beta$ satisfies Maxwell's equation

$$\nabla\times\mathbf{H}_\beta=0,\quad \beta=\begin{cases} C & \text{if } z>z_{2C},\\ B & \text{if } z_{1B}<z<z_{2B},\\ A & \text{if } z<z_{1A},\end{cases} \tag{12}$$

and is a potential field. The scalar potential U_β corresponding to this field,

$$\mathbf{H}_\beta=-\nabla U_\beta, \tag{13}$$

is a solution of Laplace's equation

$$\nabla^2 U_\beta=0. \tag{14}$$

We seek a solution of Eq. (14) in the form

$$U_\beta=-B_0 z+\frac{\Phi_0}{a}\sum_{n^2+m^2\neq 0}[C_\beta^+ e^{uz}+C_\beta^- e^{-uz}]\times\exp[iq_a(nx+my)-i\chi_\beta], \tag{15}$$

where $u=q_a\sqrt{n^2+m^2}$, $C_A^-=0$, $\chi_A\equiv\chi_1=q_a n x_1$, $\chi_B=0$, $C_C^+=0$, and $\chi_C\equiv\chi_2=q_a n x_2$.

It is convenient to write the boundary conditions that follow from the continuity of the magnetic field components at the film surfaces as follows:

$$z=z_{l\beta}\Rightarrow\mathbf{H}_l=\mathbf{H}_\beta,\ \text{where}\ l=1,2,\quad \beta=(A,B,C). \tag{16}$$

Substituting the Fourier expansions (5), (9), and (15) into the boundary conditions (16) and taking Eq. (10) into account, we arrive at a system of linear algebraic equations that make it possible to express the unknown coefficients $C_l^{\sigma\pm}(n,m)$ and $C_\beta^\pm(n,m)$ in terms of the spectral functions $S_l(n,m)$ (Eq. (6)) and the parameters of the problem. As a result, using Eqs. (5), (9), (13), and (15), we can find the desired distributions of the magnetic field.

The solutions obtained in this manner fully determine the structure of the magnetic field in magnetically coupled superconductors for an arbitrary relative displacement $x_d=x_2-x_1$ of the vortex lattices in the two films.

2.2. Gibbs energy of a vortex lattice in magnetically coupled superconductors

The behavior of a superconductor in an external magnetic field $\mathbf{B}_0$ is determined[28] by the Gibbs energy

$$G=F-\frac{1}{4\pi}\int dV(\mathbf{H}\cdot\mathbf{B}_0), \tag{17}$$

where F is the free energy of the system. In calculating the Gibbs energy per vortex in magnetically coupled superconductors, the integral must be evaluated within a unit cell of the vortex lattice, $S_c=(|x|,|y|\leqslant a/2)$:

$$\int dV=\int_{-\infty}^{+\infty}dz\int_{S_c}dx\,dy.$$

In the regions occupied by the superconductor ($z_{1A}\leqslant z\leqslant z_{1B}$ and $z_{2B}\leqslant z\leqslant z_{2C}$), the following expression for the free energy F_l corresponds to the London equation (1) (see Ref. 28):

$$F_l=\frac{1}{8\pi}\int_{V_l}dV\,[\mathbf{H}_l^2+(\lambda_l\nabla\times\mathbf{H}_l)^2], \tag{18}$$

where

$$V_l=\begin{cases}(x,y\in S_c,\quad z_{1A}\leqslant z\leqslant z_{1B}) & \text{if } l=1,\\ (x,y\in S_c,\quad z_{2B}\leqslant z\leqslant z_{2C}) & \text{if } l=2.\end{cases}$$

The energy of the magnetic field $\mathbf{H}_\beta$ outside the superconductor is determined in the usual way:

$$F_\beta=\frac{1}{8\pi}\int_{V_\beta}dV\,\mathbf{H}_\beta^2,$$

$$\text{where } V_\beta=\begin{cases}(x,y\in S_c,z>z_{2C}) & \text{if } \beta=C,\\ (x,y\in S_c,z_{1B}<z<z_{2B}) & \text{if } \beta=B,\\ (x,y\in S_c,z<z_{1A}) & \text{if } \beta=A.\end{cases} \tag{19}$$

Using Eqs. (1) and (13) and the boundary conditions (16), we see that the free energy

$$F=\sum_{l=1,2}F_l+\sum_{\beta=A,B,C}F_\beta \tag{20}$$

can be written

$$F=\frac{1}{8\pi}\Bigg\{\sum_{l=1,2}\int_{V_l}dV\,\mathbf{H}_l\hat{L}_l(\mathbf{H}_l)-\int_{S_c}d\mathbf{S}[U_A\hat{L}_1(\mathbf{H}_1)]_{z_{1A}}+\int_{S_c}d\mathbf{S}[U_B\hat{L}_1(\mathbf{H}_1)]_{z_{1B}}-\int_{S_c}d\mathbf{S}[U_B\hat{L}_2(\mathbf{H}_2)]_{z_{2B}}+\int_{S_c}d\mathbf{S}[U_C\hat{L}_2(\mathbf{H}_2)]_{z_{2C}}\Bigg\}, \tag{21}$$

where $d\mathbf{S}=dx\,dy\,\mathbf{z}_0$ is the surface area element.

The interaction energy G_c of Abrikosov-vortex lattices in magnetically coupled superconductors (or, in other words, the energy of mutual pinning of the vortex lattices) is the part of the Gibbs energy G (Eq. (17)) that depends on the relative displacement x_d of vortices in the two films:

$$G=G_0+G_c(x_d). \tag{22}$$

Obviously, the second term on the right-hand side of Eq. (17) contributes nothing to G_c since it contains no terms quadratic in the expansion coefficients S_l, $C_l^{\sigma\pm}$, and $C_\beta^\pm$. Substituting the Fourier expansions (5), (11), and (15) into (21) and dropping terms that are independent of x_d, we can write the interaction energy $G_c(x_d)$ per unit vortex cell S_c in terms of the spectral functions $S_l(n,m)$ and the Fourier expansion coefficients $C_l^{z\pm}(n,m)$:

$$G_c(x_d)=-\frac{\Phi_0^2}{8\pi a}\sum_{n^2+m^2\neq 0}[q_1S_1(C_1^{z+}e^{-\gamma_1}+C_1^{z-}e^{\mu_1})\times(1-e^{-\delta_1})+q_2S_2(C_2^{z+}e^{\mu_2}+C_2^{z-}e^{-\gamma_2})\times(1-e^{-\delta_2})]\frac{1}{u^2a}, \tag{23}$$

where $\delta_l=q_l d_l$, $\gamma_l=q_l d_3/2$, and $\mu_l=\delta_l+\gamma_l$. The coefficients $C_l^{z\pm}$ in (23) can be obtained by solving a system of linear algebraic equations (see Sec. 2.1) and can be expressed in terms of the spectral functions $S_l(n,m)$ (Eq. (6)) and the parameters of the problem. If in (23) we replace the

$C_l^{z\pm}$ by the corresponding solutions and collect the coefficients of like spatial harmonics, we obtain a Fourier series for the energy $G_c(x_d)$:

$$G_c(x_d)=-\frac{\Phi_0^2}{a}\sum_{n\geqslant 1} G_n \cos(q_a n x_d). \tag{24}$$

Here the parameter $x_d=x_2-x_1$ is the relative displacement of the vortex lattices in the films, and the amplitudes G_n of the spatial harmonics can be expressed in terms of the spectral functions $S_l(n,m)$:

$$G_n=\frac{1}{4\pi}\sum_{m=-\infty}^{\infty}\frac{P_1Q_2+P_2Q_1}{auD}, \tag{25}$$

where

$$P_l\equiv P_l(n,m)=q_lS_l(1-e^{-\delta_l})\left(1+\frac{q_l^-e^{-\delta_l}}{q_l^+}\right)e^{-\gamma_l},$$

$$Q_l\equiv Q_l(n,m)=S_l\left\{2q_l^-\left(1+\frac{q_l^-e^{-\delta_l}}{q_l^+}\right)\sinh\,\delta_l\right.$$

$$\left.+q_l^+\left[e^{\delta_l}-\left(\frac{q_l^-}{q_l^+}\right)^2e^{-\delta_l}\right](1-e^{-\delta_l})\right\}e^{-\mu_l}, \tag{26}$$

$$D\equiv D(n,m)=\left\{q_1^+q_2^+e^{\delta_3}\left[e^{\delta_1}-\left(\frac{q_1^-}{q_1^+}\right)^2e^{-\delta_1}\right]\right.$$

$$\times\left[e^{\delta_2}-\left(\frac{q_2^-}{q_2^+}\right)^2e^{-\delta_2}\right]$$

$$\left.-4q_1^-q_2^-e^{-\delta_3}\sinh\,\delta_1\sinh\,\delta_2\right\}e^{-\mu_1-\mu_2},$$

where $q_l^\pm=q_l\pm u$, and $\delta_3=ud_3$. The interaction force (per vortex) of vortex lattices in magnetically coupled films is

$$f_l=-\frac{dG_c}{dx_l} \tag{27}$$

and can be written as a Fourier series in the spatial harmonics of the relative displacement x_d:

$$f_1=-f_2=\frac{\Phi_0^2}{a^2}\sum_{n\geqslant 1}A_n\sin(q_anx_d), \tag{28}$$

$$A_n=\frac{n}{2}\sum_{m=-\infty}^{\infty}\frac{P_1Q_2+P_2Q_1}{auD}, \tag{29}$$

where $P_{1,2}$, $Q_{1,2}$, and D have been defined in (26). Equations (24)–(29) make it possible to calculate the mutual pinning energy $G_c(x_d)$ and the force $f_{1,2}(x_d)$ of the magnetic interaction of vortex lattices in magnetically coupled superconducting films.

Figure 2 depicts the dependence of the spatial-harmonic amplitudes A_n, calculated by Eqs. (26) and (29), on the induction B_0 of the external magnetic field for several values of the thicknesses d_1 and d_2 of the superconducting films and the thickness d_3 of the insulating layer. The amplitudes

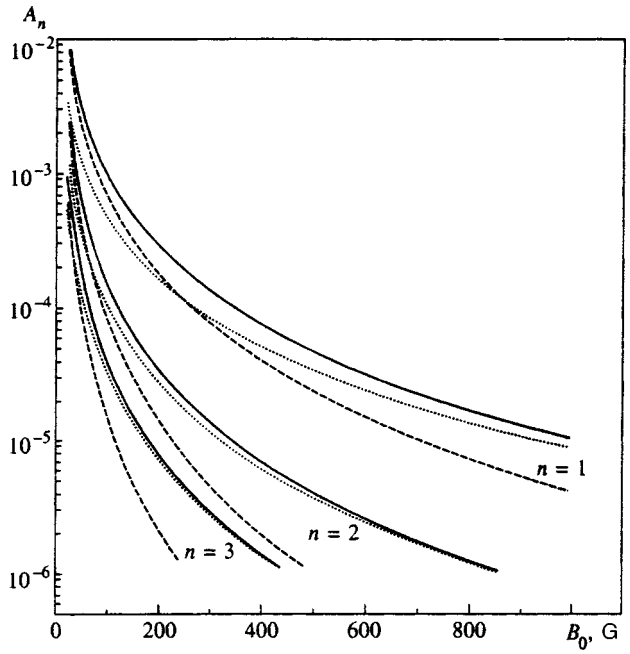


FIG. 2. Spatial-harmonic amplitudes A_n ($n=1-3$), calculated by Eqs. (26) and (29), as functions of the induction B_0 of the external magnetic field for several values of the thicknesses d_1 and d_2 of the superconducting films and the thickness d_3 of the insulating layer ($\lambda_{1,2}=2000$ Å and $\xi_{1,2}=20$ Å). The solid curves correspond to $d_{1,2}=1000$ Å and $d_3=100$ Å, the dashed curves to $d_{1,2}=1000$ Å and $d_3=300$ Å, and the dotted curves to $d_{1,2}=500$ Å and $d_3=100$ Å.

A_n rapidly decrease as the induction B_0 and the number n of the harmonic increase, so that harmonics with $n\geqslant 2$ in the expansion (28) are significant only in weak fields $B_0\sim H_{c1}$, where $H_{c1}=\Phi_0\ln\kappa/4\pi\lambda^2$ is the lower critical field in the London model.

If a vortex lattice is so dense that the distance a between vortices is at most of the order of the magnetic-field penetration depth $\lambda_{1,2}$, Eqs. (26) and (29) become much simpler. If in (29) we ignore, for $\lambda_{1,2}\gg a/2\pi$, the exponentially small terms and in (26) replace the spectral function $S_l(n,m)$ with the expression (8), which is valid for superconductors with a large Ginzburg–Landau parameter, $\kappa_{1,2}\gg 1$, we obtain a simple representation for the amplitudes A_n of the pinning force $f_{1,2}$:

$$A_n\simeq\frac{n}{2}\sum_{m=-\infty}^{\infty}\frac{e^{-ud_3}}{ua}\,\frac{1-e^{-ud_1}}{q_1^2\lambda_1^2}\,\frac{1-e^{-ud_2}}{q_2^2\lambda_2^2}, \tag{30}$$

where

$$u=\frac{2\pi\sqrt{n^2+m^2}}{a},\quad q_{1,2}=\frac{2\pi\sqrt{n^2+m^2+(a/2\pi\lambda_{1,2})^2}}{a}.$$

Clearly, the Fourier spectrum of A_n exponentially decays if n is larger than a certain value $n_m=a/2\pi d_3$. Thus, by selecting the proper value of the thickness d_3 of the insulating layer we can guarantee that the force of mutual pinning of vortex lattices, $f_{1,2}$, is harmonic over a broad range of film parameters,

$$f_{1,2}\simeq\pm f_{c1}\sin\,q_ax_d, \tag{31}$$

with the amplitude $f_{c1}=\Phi_0^2 A_1/a^2$ decreasing monotonically as B_0 increases.

3. EQUATION OF MOTION OF VORTICES IN MAGNETICALLY COUPLED SUPERCONDUCTORS

For each film we write the equation of motion of vortices in the periodic pinning potential generated by the interaction of vortex lattices in the two films. We assume that the dynamics of magnetic-flux motion is determined by the Lorentz force $\mathbf{f}_L=(j\Phi_0/c)\mathbf{x}_0$ per unit length of a vortex line, the force being generated by the interaction of the current **j** and a vortex. We also assume that the internal (volume and surface) pinning force f_p in both films is negligible and has no appreciable effect on vortex motion. A vortex moving with a velocity **v** is subject to a viscous friction force $\mathbf{f}_\eta=-\eta\mathbf{v}$ proportional to the velocity of the vortex. Under these assumptions, the equations of motion of vortex lattices along the x axis in magnetically coupled films can be written[17,21]

$$\begin{aligned} -\eta_1\dot{x}_1+\frac{\Phi_0}{c}j_1+\frac{f_1}{d_1}=0,\\ -\eta_2\dot{x}_2+\frac{\Phi_0}{c}j_2+\frac{f_2}{d_2}=0, \end{aligned} \tag{32}$$

where η_1 and η_2 are the phenomenological viscosity coefficients of the films, and the interaction forces f_1 and f_2 have been defined in (28). Introducing the phase difference $\varphi=q_a x_d$ determined by the relative displacement x_d of vortex lattices, we can easily transform (32) into the following equation, which describes the motion of vortices in magnetically coupled superconductors:

$$\alpha\dot{\varphi}+j_s(\varphi)=j. \tag{33}$$

The parameter $\alpha=ca\eta/2\pi\Phi_0$ decribes the dissipation that results from the relative motion of vortex lattices, and the function

$$j_s(\varphi)=\frac{cB_0}{d}\sum_{n\geqslant 1}A_n\sin n\varphi \tag{34}$$

determines the dependence of the current j_s flowing in the films without losses on the arrangement of the vortices in the two films. Here η, d, and j denote the reduced viscosity coefficient, the reduced film thickness, and the reduced external-current density, and can be expressed in terms of the parameters of the problem:

$$\eta=\frac{\eta_1\eta_2}{\eta_1+\eta_2},\quad d=\frac{d_1d_2(\eta_1+\eta_2)}{\eta_1d_1+\eta_2d_2},\quad j=\frac{j_2\eta_1-j_1\eta_2}{\eta_1+\eta_2}. \tag{35}$$

From now on we assume that the superconducting films are identical ($\lambda=\lambda_{1,2}$, $\xi=\xi_{1,2}$, $d_1=d_2$, and $\eta_1=\eta_2$) and that the currents in the films are equal in magnitude and opposite in direction ($j_2=-j_1$). In this case the reduced parameters specified in (35) are simply $d=d_{1,2}$ and $\eta=\eta_{1,2}/2$, and j coincides with the density of the external current in the superconducting films: $j=j_2=-j_1$.

The maximum current density

$$j_c=\max[j_s(\varphi)]=j_c(\varphi^c) \tag{36}$$

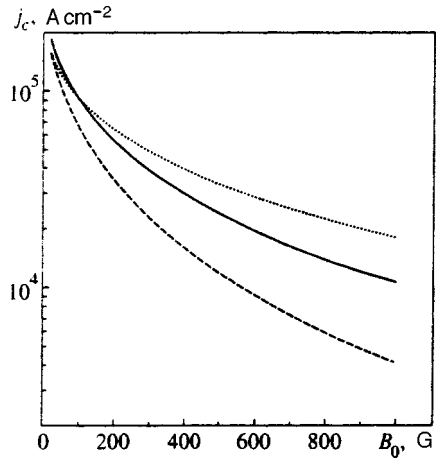


FIG. 3. Dependence of the critical current density j_c (see Eqs. (34), (36), and (37)) on the induction B_0 for several values of the film thickness d and the thickness d_3 of the insulating layer ($\lambda=2000$ Å and $\xi=20$ Å). The solid curve corresponds to $d=1000$ Å and $d_3=100$ Å, the dashed curve to $d=1000$ Å and $d_3=300$ Å, and the dotted curve to $d=500$ Å and $d_3=100$ Å.

that can flow in magnetically coupled films without dissipation in the geometry being discussed corresponds to a mutual displacement x_d^c of the vortex lattices in the two films for which the phase $\varphi^c=q_a x_d^c$ satisfies the equation

$$\left(\frac{dj_s}{d\varphi}\right)_{\varphi=\varphi^c}=\sum_{n\geqslant 0}nA_n\cos n\varphi^c=0. \tag{37}$$

Here j_c can be considered the critical current, and if currents higher than the critical value flow in the films, the vortex lattices become stripped. If the current j in the films is less than j_c, the mutual attraction of the vortices in the two films fixes the relative spatial arrangement of the vortex lattices and ensures nondissipative flow of the current. When $j>j_c$, there is relative motion of the vortex lattices in the two films, which leads to dissipation when such a current flows in the films. Note that the attraction of vortex lattices in magnetically coupled superconductors can be considered additional pinning of vortices, a factor that leads to an increase in the critical current as compared to the critical current j_p due to the intrinsic (internal and surface) pinning of vortices in the films. Such additional pinning can be detected in experiments and constitutes a problem in its own right.

Figure 3 depicts the dependence of the critical current density j_c (see Eqs. (34), (36), and (37)) on the induction B_0 for several values of the film thickness d and the thickness d_3 of the insulating layer. As B_0 increases, the vortex-lattice period a decreases, vortex overlap grows, and the distribution of the magnetic field in the films becomes more uniform. This leads to a decrease in the mutual vortex-lattice pinning force, and hence to a decrease in the critical current density j_c. Figure 4 depicts the dependence of the critical current density j_c on the film thickness d for several values of B_0 and the thickness d_3 of the insulating layer. The increase in j_c with d in the case of thin superconducting films ($d\ll\lambda$) can be explained by the decrease in the effective penetration depth $\lambda^d=\lambda^2/d$, which determines the characteristic spatial scale of the magnetic-field variations in a vortex.[20] As the

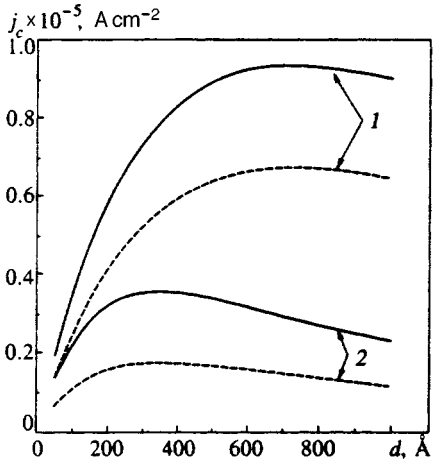


FIG. 4. Dependence of the critical current density j_c (see Eqs. (34), (36), and (37)) on the film thickness d for several values of the induction B_0 and the thickness d_3 of the insulating layer. Curves *1* correspond to $B_0 = 100$ G and curves *2*, to $B_0 = 500$ G ($\lambda = 2000$ Å and $\xi = 20$ Å); the solid curves correspond to $d_3 = 100$ Å and the dashed curves, to $d_3 = 300$ Å.

film thickness d becomes larger, a quasi-two-dimensional vortex transforms into an Abrikosov vortex, whose magnetic field changes significantly over the penetration depth λ. Here the relative effect of the finite film thickness d on the free energy $F_{1,2}$ of a vortex (Eq. (18)) weakens, and for $d > \lambda$ the perturbations introduced by the film surfaces are insignificant. In the latter case, however, the external current is distributed nonuniformly over the film thickness, and in writing the equations of vortex motion (32) one must bear in mind that the mean current density $j_{1,2}$ depends on the film thickness $d_{1,2}$, and furthermore, the axial lines of the vortices may be tilted.

For values of the external magnetic field B_0 at which the period of the vortex lattice created by the field is moderate ($a \leqslant \lambda$), the contribution of harmonics with $n \geqslant 2$ in the expansion (34) is insignificant (see Fig. 2). Hence, leaving only the term with $n = 1$ in (34), which corresponds to the fundamental harmonic with spatial frequency q_a, we obtain a simple sinusoidal dependence of the supercurrent j_s on the phase φ, corresponding to the harmonic law (31) for the pinning force,

$$j_{s1}(\varphi) \simeq j_{c1} \sin \varphi, \tag{38}$$

where the critical current density is

$$j_{c1} = \frac{cB_0 A_1}{d}. \tag{39}$$

If the insulating layer is not very thin ($d_3 \geqslant a/2\pi$), we can keep only the term with $m = 0$ in (30). If, in addition, for $d > a$ we ignore the fact that the thickness of the superconducting films is finite, them from (30) and (39) we can easily obtain

$$j_{c1} \simeq \frac{c\Phi_0 a^2 e^{-2\pi d_3/a}}{64\pi^5 \lambda^4 d}, \tag{40}$$

which is valid for superconducting films with a large Ginzburg–Landau parameter, $\kappa = \lambda/\xi \gg 1$. The equation of motion of vortices in magnetically coupled superconductors (Eq. (33)) corresponding to the sinusoidal dependence (38) of the supercurrent j_s on the phase φ coincides with the equation of the resistively shunted Josephson junction model:[29,30]

$$\alpha \dot{\varphi} + j_{c1} \sin \varphi = j, \tag{41}$$

where j_{c1} is given by (39), j is the reduced external current density in the films (see (35)), and the parameter α describes the dissipation that emerges in vortex motion. If the current j in the films is less than j_{c1}, Eq. (41) has a steady-state solution $\varphi^s = q_a x_d^s$, which determines the displacement of the vortex lattices in the two films relative to their equilibrium positions. In this case there is nondissipative current flow in the magnetically coupled films, which are in a mixed state, even when there is no intrinsic pinning of the vortices in the films.

4. RF RESPONSE OF MAGNETICALLY COUPLED SUPERCONDUCTORS

When the current j in a superconducting film exceeds the critical value j_c, the vortex lattices become stripped and there is slippage of the vortex lattices relative to one another. In view of the periodicity of the interaction potential, such motion has a variable velocity component and must lead to generation of electromagnetic radiation. Qualitatively, this motion can be considered the nonuniform motion of lattices consisting of magnetic dipoles. Note that the motion of vortices in the films generates an electric field $E = \dot{x}_d B_0/c$ with a variable component. We restrict our discussion to identical films with a dense vortex lattice ($a < \lambda$). This means that Eq. (41) suffices to describe the motion of vortices in magnetically coupled superconductors. Writing the viscosity coefficient $\eta = \Phi_0 B_0/\rho_f c^2$ in terms of the resistivity of the films in the flux-flow regime, $\rho_f = \rho_n B_0/H_{c2}$, where ρ_n is the resistivity of the films in the normal state, and $H_{c2} = \Phi_0/2\pi\xi^2$ is the upper critical field, we can obtain the following expression for the frequency of oscillation ω (see Ref. 2):

$$\omega = \frac{\sqrt{j^2 - j_{c1}^2}}{\alpha} = \omega_0 \sqrt{\left(\frac{j}{j_{c1}}\right)^2 - 1}, \tag{42}$$

$$\omega_0 = \frac{4\pi c j_{c1} \rho_f a}{\Phi_0} = \frac{4\pi c j_{c1} \rho_n}{a H_{c2}}. \tag{43}$$

The fact that the frequency ω of the electromagnetic radiation emitted by magnetically coupled films depends on the spatial lattice period a means that it is possible to tune the frequency by applying an external magnetic field B_0. If the external current j is substantially higher than the critical value j_{c1},

$$\omega \simeq \frac{4\pi c \rho_n}{H_{c2} \Phi_0^{1/2}} j \sqrt{B_0}. \tag{44}$$

By analogy with the well-known properties of Josephson junctions, the system of magnetically coupled superconductors is expected to be sensitive to an external variable field:[21]

because microwave radiation synchronizes the motion of the magnetic flux, Shapiro-type steps appear in the current–voltage characteristic.[29]

Let us discuss the relationship between the frequency ω and the constant voltage V_0 generated by the relative motion of vortex lattices. We take a superconducting bridge of length L and width W and place it in an external magnetic field B_0, so that a square lattice of $N\times M$ vortices is formed within a surface area of $L\times W$: $N=L/a$ and $M=W/a$, where a is the spatial period of the lattice. In the flux-flow regime, in the superconducting films there emerges an electric field E whose constant component $E_0\equiv\langle E\rangle_t$ ensures dissipation and is determined, in accordance with Maxwell's equation

$$\nabla\times\mathbf{E}=-\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t},$$

by the mean velocity v_0 of a vortex lattice in the films:

$$E_0=\frac{v_0B_0}{c}. \tag{45}$$

The electric field E_0 generates a constant voltage across the bridge,

$$V_0=E_0L=E_0aN. \tag{46}$$

After one period $T=2\pi/\omega$ the pattern of the mutual arrangement of the vortex lattices is repeated, with the vortex lattice in each of the magnetically coupled films being displaced with respect to its previous position by one half of the spatial period of the lattice, i.e., $a/2=v_0T$. This readily leads to a relationship between the applied voltage V_0 and the frequency ω,

$$V_0=\frac{\hbar}{2e}\frac{N}{2}\omega, \tag{47}$$

which differs from the well-known Josephson relation[1] by an additional factor $N/2$, where the integer N is equal to the number of vortices fitting into the bridge length L. Note that when only one vortex lattice is mobile while the other is immobile (e.g., because of strong intrinsic pinning), the factor $N/2$ in (47) must be replaced by N. The latter determines the number of fluxons exiting the contour in the course of one period.

The dc electrical resistance R_0 of the bridge due to the motion of the magnetic flux is $R_0=V_0/I_0$ at the operating point, where $I_0=jS$ is the total current flowing through the cross-sectional area of the film $S=Wd$. When I_0 is much greater than the critical current $I_c=j_cS$, the resistance R_0 approaches the characteristic resistance

$$R=\frac{\rho_fL}{S}, \tag{48}$$

which corresponds to a current-independent shunt resistance in the resistively shunted Josephson junction model.

A convenient parameter describing the operation of such a bridge is the characteristic voltage[31]

$$V_c=I_cR. \tag{49}$$

Note that an elementary chain of vortices moving transversely to the bridge involves a voltage $V_{c1}=V_c/N$, which is related to the characteristic frequency ω_0 through the Josephson relation (see Eq. (47)). At high currents ($I_0\gg I_c$) and under optimum matching conditions for the impedances of the bridge (R_c), and the load (R_L), i.e., $R_L\simeq R_c$, we can expect the rf power delivered to the load to be[32]

$$P_{\sim}\simeq\frac{RI_c^2}{8}. \tag{50}$$

5. DISCUSSION

We now compare the behavior of magnetically coupled superconductors in which there is relative slippage of the vortex lattice with that of an ordinary Josephson junction. As an example, we take an S–N–S junction with a resistance R_N and thickness d_N of the normal layer. The weak link in the S–N–S junction results from the proximity effect,[28] and the critical current in such a junction at temperature T is given by[3]

$$I_c^{SNS}=\frac{\pi\Delta^2}{4ek_BTR_N}\sum_{n\geqslant 0}\frac{8}{\pi^2(2n+1)^2}\frac{l_n}{\sinh l_n}, \tag{51}$$

where $l_n=(2n+1)^{1/2}d_N/\xi_N$, Δ is the energy gap in the superconductors, and ξ_N is the coherence length in the material of the normal layer. Clearly, Eq. (51) tells us that I_c^{SNS} can be high when the normal barrier is thin ($d_N\sim\xi_N$), while its resistance R_N is low. For $d_N\gg\xi_N$, the critical current I_c^{SNS} exponentially decreases as the normal layer becomes thicker. The low resistance R_N, which makes it difficult to match the S-N-S junction with other devices, can be increased by reducing the area of the junction. However, the smaller the junction area, the lower the radiated power.

When magnetically coupled superconductors are involved, the weak link is established by the magnetic interaction of vortex lattices in the superconducting films. Let us compare the expression for the critical current density j_{c1} in magnetically coupled superconducting films (Eq. (40)) with the expression (51) for the critical current of an S–N–S junction. One can easily see that $a/2\pi$ in (40) plays the same role as the parameter ξ_N for an S–N–S junction: the magnitude of the weak coupling between magnetically coupled superconductors decreases exponentially in magnetic fields B_0, for which the vortex lattice period a is of the order of the thickness d_3 of the insulating layer between the superconducting films. Since we have $a\gg\xi$ in fields $B_0\ll H_{c2}$, the requirements that the insulating layer in magnetically coupled superconductors must obey are not so stringent as they are for Josephson junctions.

The characteristic resistance R_0 (Eq. (48)) of a bridge of magnetically coupled superconductors does not explicitly depend on the gap thickness d_3. This makes it possible to match R_0 to the load impedance R_L by changing the length L and width W of the superconducting bridge without appreciably reducing the critical current $I_c=j_{c1}S$ and hence the radiated power $P_{\sim}$ (Eq. (50)).

For clarity, we rewrite the relationship (47) between the frequency $f=\omega/2\pi$ and the voltage V_0 across the bridge in the form

$$f=2\frac{2e}{h}V_1, \tag{52}$$

where $V_1=V_0/N$ is the voltage generated across an elementary chain of vortices moving transversely to the bridge, and the additional factor of 2 preceding the coefficient $2e/h \simeq 483.6\ \mathrm{MHz}/\mu\mathrm{V}$ reflects the obvious fact that for a given voltage V_1 the oscillating frequency f doubles because the vortices move toward one another.

We note one more interesting corollary of Eq. (52). For a bridge of given geometric dimensions $L\times W$, the dependence of the radiation frequency f on the external magnetic field B_0 appears in (52) via the relationship $N(B_0)=L(B_0/\Phi_0)^{1/2}$. Since N takes only integer values, the frequency f for a given voltage V_0 across the bridge can take only discrete values, i.e., if V_0 is fixed, changes in f induced by changes in B_0 can take place only in jumps: the jump to a neighboring step corresponds to a change in the number N of vortex chains by ± 1.

Expressing I_c and R_0 in terms of the dimensions of the bridge which is fabricated from magnetically coupled superconductors, we can write the expression (50) for the rf power in the form

$$P_{\sim}\simeq P_0(N\times M), \tag{53}$$

where $P_0=\rho_f j_c^2 a^2 d/8$ is the power radiated by a unit cell $(a\times a)$ containing only one vortex of a lattice. Note that such an elementary bridge, made from magnetically coupled superconductors containing only one vortex ($N=1$ and $M=1$), behaves in many respects like a single Josephson junction. Here the geometrical dimensions of such a bridge are determined not by the coherence length ξ but by $a=(\Phi_0/B_0)^{1/2}$, which simplifies the fabrication of such objects from high-T_c superconducting materials in comparison to the fabrication of ordinary Josephson junctions.

It is convenient to discuss the time-dependent Josephson effect in weak-link systems in terms of fluxons that penetrate the junction.[26] In an ordinary Josephson junction, it takes one period T for one fluxon Φ_0 to pass through the region occupied by the junction. This means that such a system operates at low power levels, and that raising the power requires synchronizing a large number of junctions.[32] Such synchronization constitutes a complicated problem, since the junctions are coupled only through the external circuit and there is no direct magnetic interaction of Josephson vortices in different junctions. In an $L\times W$ superconducting bridge based on magnetically coupled superconducting films, the external magnetic field B_0 generates a lattice of coupled vortices that consists of $N\times M$ unit cells. Equation (53) implies that it takes one period T for $N\times M$ vortices to re-interlock simultaneously, which is equivalent to synchronizing an array of $N\times M$ junctions. As a result, the rf power delivered to the load also grows by a factor of $N\times M$. Here the frequency f decreases in comparison to the Josephson frequency $f_J=V_0(2e/h)$ only by a factor of N for a fixed bias V_0 across the bridge. Such synchronous operation requires, however, that the vortex lattices move as a whole with respect to each other. This can be ensured if the interaction energy in each film, which determines the elastic properties of a vortex lattice,[25] is much greater than G_c (Eq. (23)), which characterizes the interaction of neighboring vortex lattices in magnetically coupled superconductors.

To illustrate the above results, we examine Josephson devices based on magnetically coupled superconducting films, and assume that the vortex lattices are counterpropagating. Using the parameters of the dc transformer discussed by Ekin *et al.*,[23,24] with the transformer fabricated from granular aluminum films ($\mathrm{Al/SiO_2/Al}$), i.e., $H_{c2}=50$ G, $\lambda=6000$ Å, $\rho_n=10\ \mu\Omega\cdot$cm, $d=750$ Å, and $d_3=120$ Å, we find that for an external magnetic field $B_0=10$ G ($a=1.4\ \mu$m) the potential of mutual pinning of the vortex lattices is $\Delta U_c\sim 10$ meV, which provides a critical current density $j_c\simeq 10^3\mathrm{A/cm^2}$. The critical current for a bridge with $L=0.7$ mm ($N=500$) and $W=20\ \mu$m proved to be $I_c=j_c W d=15\ \mu$A. Substituting these data into (43), (48), and (49), we can easily find the characteristic frequency $f_0=\omega_0/2\pi$, the resistance R_c, and the voltage V_c across a bridge based on such magnetically coupled superconducting films: $f_0\simeq 0.3$ GHz, $R_c\simeq 10\ \Omega$, and $V_c\simeq 0.15$ mV ($V_{c1}\simeq 0.3\ \mu$V). By virtue of (42), the frequency of the rf response in such a system, $f=\omega/2\pi$, depends on the excess of the current I_0 over the critical value I_c, i.e., on the choice of operating point. Assuming that the voltage V_0 across the bridge is equal to the characteristic value V_c, we find that $f=f_0$, which corresponds to the current I_0 in the films being equal to $\sqrt{2}I_c\simeq 21\ \mu$A, and to motion of the vortex lattices toward one another with a velocity $v_0=af/2\simeq 2\times 10^4$cm/s. Here Eq. (50) estimates the rf power $P_{\sim}$ delivered to a matched load, $R_L\simeq R_c$, to be $\sim 10^{-9}$ W.

Another example deals with magnetically coupled YBCO superconducting films with typical parameters[33] $\lambda=2000$ Å and $\rho_f=10\ \mu\Omega\cdot$cm and dimensions $d=500$ Å, $d_3=100$ Å, $L=0.7$ mm, and $W=20\ \mu$m. The mutual pinning potential ΔU_c of the vortex lattices in such magnetically coupled superconducting films in an external magnetic field $B_0=1000$ G ($a=1400$ Å and $N=5000$) reaches $\Delta U_c\sim 0.1$ eV. The critical current density and the critical current, determined by the interaction of the vortex lattices in the two films, are $j_c\simeq 2\times 10^4\mathrm{A/cm^2}$ and $I_c=0.2$ mA. The characteristic parameters of a bridge fabricated from such magnetically coupled superconducting films are $f_0\simeq 2.7$ GHz, $R_c\simeq 70\ \Omega$, and $V_c\simeq 14$ mV ($V_{c1}\simeq 2.8\ \mu$V). Assuming as we did in the previous example that the voltage V_0 across the bridge is equal to the characteristic value $V_0=V_c$, we find that the frequency f is equal to f_0 and that the current I_0 in the films is $\sqrt{2}I_c\simeq 0.28$ mA. The mean vortex velocity v_0 in the films is $v_0=af/2\simeq 2\times 10^4$cm/s, and an estimate of the emitted rf power $P_{\sim}$ (Eq. (50)) at the oscillating frequency yields $P_{\sim}\sim 0.35\ \mu$W.

Note that such high critical current densities, $j_c\sim 10^4-10^5\mathrm{A/cm^2}$, flowing in an $S-N-S$ junction with an area of $0.8\ \mu\mathrm{m}^2$ fabricated from high-T_c superconducting materials can be observed only at very low resistances of such a junction, $R_N\sim 0.1\ \Omega$ (see Ref. 34). However, the study of magnetically coupled films that are based on high-

T_c superconductors is hindered by the high values of the critical current density j_p (i.e., strong internal pinning of Abrikosov vortices in such films) characteristic of high-T_c superconducting materials. Granular high-T_c superconducting films, which have a low energy of internal pinning for the intergranular vortices, may prove to be promising high-$T_c$superconducting materials for studying the motion of vortices in magnetically coupled superconductors.[35]

Obviously, the moderate frequencies f at which magnetically coupled superconducting films radiate, result from the relatively low velocities of Abrikosov vortices in superconductors. However, the method of creating nonuniform motion of vortex structures discussed in this paper is universal and can be realized in long magnetically coupled Josephson junctions.[36] The characteristic velocities of Josephson vortices in flux-flow oscillators (FFO) reach 10^7-10^8cm/s at frequencies $f\sim 200-450$ GHz (see Ref. 37). The use of a pair of long Josephson junctions as an FFO between which there is additional magnetic coupling[38] make it possible to synchronize the oscillators with vortices moving in a single direction. The advantage of the regime suggested in this paper, with counterpropagating vortices, lies in the fact that the magnetic coupling between Josephson vortices makes it possible to increase the oscillating frequency and radiated power.

6. CONCLUSION

We have shown that when the vortex lattices in magnetically coupled superconducting films move relative to one another in the dc transformer geometry, the system generates electromagnetic radiation efficiently. The fact that there is magnetic coupling between identical vortex lattices guarantees self-consistency of the periods of the vortex lattices and pinning potential, and optimum coupling between them. The equation describing the viscous motion of the vortex lattices in magnetically coupled superconducting films coincides with the equation of the resistively shunted Josephson junction model, in which the relative displacement x_d of the vortex lattices in the films acts as the phase difference φ. The relatively large number of easily controlled parameters, i.e., the thickness of the superconducting films, the thickness of the insulating layer, and the temperature-dependent magnetic-field penetration depth, make it possible to a realize an operational mode in which the force of mutual pinning of the vortex lattice varies harmonically over a given range of external magnetic fields. This ensures a sinusoidal dependence of the supercurrent j_s on the phase φ in the equivalent resistively shunted Josephson junction model.

We find that magnetically coupled superconductors possess the properties of a Josephson element without any constraints imposed by the coherence length ξ on the spatial dimensions of the magnetically coupled system. The fact that there is an analogy between the motion of vortices in magnetically coupled superconductors and a Josephson junction implies that the system examined here is sensitive to external electromagnetic radiation: because microwave radiation synchronizes the motion of the magnetic flux, Shapiro-type steps appear in the current–voltage characteristic. The frequency of the electromagnetic radiation generated by the relative motion of vortex lattices in magnetically coupled superconductors depends on the spatial period of the vortex lattices and the velocity of relative vortex motion, which means that the frequency of the radiation can be tuned by applying a magnetic field or a current. When in magnetically coupled superconducting films there is relative motion of vortex lattices consisting of $N\times M$ unit cells, it takes one oscillation period for $N\times M$ vortices to re-interlock simultaneously, which is equivalent to synchronizing an array of $N\times M$ Josephson junctions.

The magnetic coupling of the vortex lattices in the neighboring films constitutes an additional pinning mechanism with a well-known and relatively simple pinning potential, which can be calculated analytically and compared directly with the experimental data. This opens new possibilities in establishing the nature of pinning, since it becomes possible to efficiently control the magnitude and spatial scale of the additional pinning potential.

The authors are grateful to A. A. Andronov and S. V. Gaponov for useful discussions, and to the Russian Fund for Fundamental Research (Project 96-02-19272). One of the authors (A.V.S.) is grateful to the International Soros Science Education Program (ISSEP) for financial support.

*)E-mail: samokh@ipm.sci-nnov.ru

[1] B. D. Josephson, Phys. Lett. **1**, 251 (1962).
[2] A. Barone and G. Paterno, *Physics and Applications of the Josephson Effect*, Wiley, New York (1982).
[3] K. K. Likharev, Rev. Mod. Phys. **51**, 101 (1979).
[4] B. B. Schwartz, Phys. Lett. **20**, 350 (1966).
[5] A. A. Abrikosov, Zh. Éksp. Teor. Fiz. **32**, 1442 (1957) [Sov. Phys. JETP **5**, 1174 (1957)].
[6] I. O. Kulik, Zh. Éksp. Teor. Fiz. **50**, 1617 (1966) [Sov. Phys. JETP **23**, 1077 (1966)].
[7] A. T. Fiory, Phys. Rev. Lett. **27**, 501 (1971).
[8] K. Lee, I. Iguchi, S. Lee *et al.*, Physica C **221**, 254 (1994).
[9] A. Schmid and W. Hanger, J. Low Temp. Phys. **11**, 667 (1973).
[10] P. Martinoli, O. Daldini, C. Leeman *et al.*, Solid State Commun. **17**, 205 (1975).
[11] E. Rosseel, M. Van Bael, M. Baert, R. Jonckheere, V. V. Moshchalkov, and Y. Bruynseraede, Phys. Rev. B **53**, R2983 (1996).
[12] L. É. Amatuni, A. A. Akhumyan, K. I. Konstantinyan, E. M. Martirosyan, and G. A. Ovsyannikov, JETP Lett. **50**, 385 (1989).
[13] J. Konopka and G. Jung, Europhys. Lett. **8**, 549 (1989).
[14] J. R. Clem, Phys. Rev. B **1**, 2140 (1970); Phys. Rep. **75**, 1 (1981).
[15] P. Martinoli, O. Daldini, C. Leemann, and B. Van den Brandt, Phys. Rev. Lett. **36**, 382 (1976).
[16] I. Giaever, Phys. Rev. Lett. **15**, 825 (1965); **16**, 460 (1966).
[17] P. E. Cladis, R. D. Parks, and J. M. Daniels, Phys. Rev. Lett. 21, 1521 (1968).
[18] M. D. Sherrill, Phys. Rev. B **7**, 1908 (1973).
[19] J. R. Manson and M. D. Sherrill, Phys. Rev. B **11**, 1066 (1975).
[20] J. Pearl, Appl. Phys. Lett. **5**, 65 (1964); J. Appl. Phys. **37**, 4139 (1966).
[21] J. R. Clem, Phys. Rev. B **9**, 898 (1974).
[22] J. R. Clem, Phys. Rev. B **12**, 1742 (1975).
[23] J. W. Ekin, B. Serin, and J. R. Clem, Phys. Rev. B **9**, 912 (1974).
[24] J. W. Ekin and J. R. Clem, Phys. Rev. B **12**, 1753 (1975).
[25] A. Campbell and J. Evetts, *Critical Currents in Superconductors*, Taylor & Francis, London (1972).
[26] K. K. Likharev, *Introduction to the Dynamics of Josephson Junctions* [in Russian], Nauka, Moscow (1985), p. 47.
[27] J. R. Clem, J. Low Temp. Phys. **18**, 427 (1975).
[28] A. A. Abrikosov, *Fundamentals of the Theory of Metals*, Elsevier, New York (1988).

[29] A. Gilabert, I. K. Schuller, V. V. Moshchalkov, and Y. Bruynseraede, Appl. Phys. Lett. **64**, 2885 (1994).
[30] V. A. Kozlov and A. V. Samokhvalov, Physica C **235–240**, 2011 (1994).
[31] K. K. Likharev and B. T. Ul'rikh, *Systems with Josephson Junctions* [in Russian], Moscow Univ. Press, Moscow (1978), p. 22.
[32] S. Han, B. Bi, W. Zhang, and L. E. Lukens, Appl. Phys. Lett. **64**, 1424 (1994).
[33] J. M. Roberts, B. Brown, B. A. Hermann, and J. Tate, Phys. Rev. B **49**, 6890 (1994).
[34] L. Antognazza, B. H. Moeckly, T. H. Geballe, and K. Char, Phys. Rev. B **52**, 4559 (1995).
[35] J. R. Clem, B. Bumble, S. I. Raider, W. J. Gallagher, and Y. C. Shih, Phys. Rev. B **35**, 6637 (1987).
[36] N. Grønbech-Jensen and M. R. Samuelsen, Phys. Rev. Lett. **74**, 170 (1995).
[37] V. P. Koshelets, A. Shchukin, and I. L. Lapytskaya, Phys. Rev. B **51**, 6536 (1995).
[38] N. Thyssen, H. Kohlstedt, and A. V. Ustinov, in *Applied Superconductivity Conference*, Pittsburgh (1996), p. 132.

Translated by Eugene Yankovsky

Temperature dependence of the heat capacity and saturation magnetization of dilute reentrant ferrimagnetic spinels

N. N. Efimova

Kharkov State University, 310077 Kharkov, Ukraine
(Submitted 18 July 1997)
Zh. Éksp. Teor. Fiz. **113**, 1339–1349 (April 1998)

Using model objects consisting of dilute reentrant spinels $Li_{0.5}Fe_{2.5-x}Ga_xO_4$ with $x=1.0-1.2$, this paper describes the temperature dependence of the magnetic contribution to the heat capacity, $C_m(T)$, at $H=0$ and temperatures exceeding the freezing temperature ($T_f \sim 10$ K) and, for $T \geqslant 4.2$ K, the low-field magnetization $\sigma_H(T)$ and the saturation magnetization $\sigma_s(T)$, as well as the magnetization isotherms $\sigma_T(H)$ in fields of up to 10 kOe. The specific features of the behavior of the overall characteristics of $C_m(T)$ and $\sigma_s(T)$ are found for the states of a frustrated ferrimagnetic that occur in the conditions considered here ($H=0$ and $T>T_f$ or $T \geqslant 4.2$ K and strong magnetic fields) and are discussed in connection with changes in the magnetic excitation spectrum caused by local breakdowns of collinear spin ordering and frustrations. © *1998 American Institute of Physics.* [S1063-7761(98)01604-7]

1. INTRODUCTION

This paper presents the results of a study of the magnetic and thermal properties of dilute two-sublattice ferrimagnetic spinels $Li_{0.5}Fe_{2.5-x}Ga_xO_4$ ($1.0 \leqslant x \leqslant 1.2$), which are included among Heisenberg magnets with short-range exchange. The $0.9 \leqslant x < 1.5$ interval on the $x-T$ phase diagram corresponds to the concentration region of reentrant states, where, as the temperature decreases, a paramagnetic–ferrimagnetic transition is initially observed at the Curie point T_C, and then a ferrimagnetic–ferrimagnetic spin-glass transition is observed at the freezing temperature $T_f < T_C$.[1,2] The latter is characterized by the existence at $T<T_f$ of long-range ferrimagnetic order (spontaneous magnetization $\sigma_s \neq 0$) with the typical properties of a spin glass. For $x=1.0-1.2$, the T_f values are about 10 K, and the ferrimagnetic spin-glass states break down in rather weak magnetic fields, $H<H_s$ (the technical saturation field of the ferrimagnet). The value of T_C decreases from 435 ($x=1.0$) to 325 K ($x=1.2$).

The problem of reentrant states in Heisenberg systems with short-range interaction (formation mechanisms, the presence of phase transitions with $0\ \text{K} < T < T_C$, magnetic structure) continues to attract the attention of investigators and, obviously, cannot be solved except in connection with ferrimagnetic states, which precede the ferrimagnetic spin glass in concentration and temperature ($T>T_f$).[3,4] Ferrimagnetic states of this kind are interesting in themselves. In various cross sections of $x-T-H$ parameter space, this can include a frustrated ferrimagnetic, ferrimagnetics with fluctuating exchange, with fluctuating exchange and stochastic magnetic structure, etc. The properties of such magnets, including the magnetic-excitation spectrum, can significantly differ from those inherent to homogeneous collinear structures.[5–8]

Guided by such considerations, we undertook studies of the concentration interval $x=1.0-1.2$, which corresponds to the beginning of the formation of ferrimagnetic spin-glass states on the $x-T$ phase diagram of Li–Ga spinels in the $T=0$ K cross section.[1] Particular attention was concentrated on the study of the ferrimagnetic region, namely, on the states that precede spin glass in temperature ($T_f<T<T_C$, $H=0$) or occur under the action of a magnetic field $H>H_s$ for $T \geqslant 4.2$ K, i.e., both for $T<T_f$ and for $T>T_f$. An advantage of the test objects considered here is that they correspond to the simulation concepts of Ref. 9, where the mechanism of the formation of spin-glass states (semispin-glass states, in the author's terminology) in dilute ferrimagnetic spinels with one sort of magnetic ions was theoretically considered.

According to Ref. 9, when dilution occurs in regions with an increased content of nonmagnetic ions (composition disorder), local breakdown of the collinearity, or canting, occurs. These canted spins partially polarize the collinear matrix, as a result of which more extended regions of local noncollinearity are formed. As a consequence of the competition of inter- and intrasublattice antiferromagnetic interactions, frustrated exchange coupling appears within these regions. At some concentration of nonmagnetic ions, these regions begin to overlap over the entire crystal. This corresponds to the formation of long-range transverse order of the spin glass with respect to spontaneous magnetization σ_s.

Such a mechanism for forming ferrimagnetic spin-glass states, unlike the conclusions of Gabay–Toulouse mean-field theory,[10] does not presuppose the formation of regular noncollinear structures in the region $T_f<T<T_C$. The state achieved in the finite temperature interval $T>T_f$ can be regarded as a state of a frustrated ferrimagnet with local destruction of collinear spin ordering in the form of regions of local noncollinearity. This paper is devoted to an experimental study of the thermal and magnetic properties of such a state, with a program that includes a study of the temperature dependence of the magnetic contribution to the heat capacity $C_m(T)$ and the saturation magnetization $\sigma_s(T)$, as well as

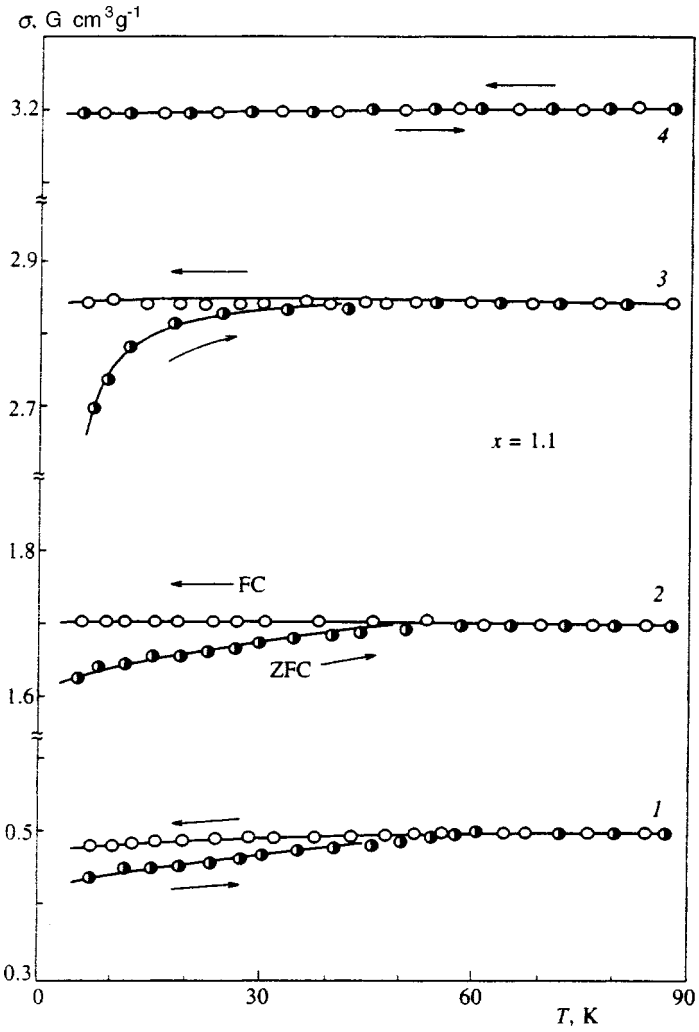


FIG. 1. Polytherms of the magnetization $\sigma_H(T)$ in ZFC regimes (preliminary cooling of the sample to 4.2 K in the absence of a field) and FC (cooling with $H \neq 0$) of the spinel $Li_{0.5}Fe_{1.4}Ga_{1.1}O_4$ for various values of the field: *1*—10 Oe, *2*—30 Oe, *3*—50 Oe, *4*—100 Oe.

the low-field polytherms of the magnetization $\sigma_H(T)$ and the isotherms of $\sigma_T(H)$ in fields up to 10 kOe.

2. SAMPLES AND MEASUREMENT TECHNIQUE

The studies were carried out on single-phase polycrystalline samples obtained by standard ceramic technology as in Refs. 1 and 2. The isotherms of the specific magnetization $\sigma_T(H)$ were measured at temperatures from 4.2 to 100 K in fields of up to 10 kOe on a ballistic magnetometer similar to that used in Refs. 1 and 2. The sensitivity of the apparatus was 10^{-3} G cm^{3}g^{-1}. The temperature was measured by means of a carbon thermometer and was kept constant to within ± 1 K during the measurements. Along with this, the $\sigma_H(T)$ dependences in fields of $H=5$ kOe, which exceeded the technical saturation field H_s, were directly measured with a temperature step of (3–5) K.

The method of measuring the heat capacity and the technique of distinguishing the magnetic contribution $C_m(T)$ are similar to what we used in Ref. 11. The low-temperature measurements from 2 to 22 K were also duplicated on the apparatus described in Ref. 12. The measurement error of heat capacity C was less than $\pm 1\%$. A carbon thermometer was also used to measure the temperature.

3. EXPERIMENTAL RESULTS AND THEIR DISCUSSION

3.1. Magnetic properties in weak ($H<H_s$) and strong ($H>H_s$) magnetic fields

First of all, it is expedient to consider the experimental results that can serve as evidence of the existence in weak magnetic fields at $T>T_f$ of inhomogeneities of the ferrimagnetic structure in the form of a region of local noncollinearity. Figures 1 and 2 show magnetization polytherms $\sigma_H(T)$ of samples with $x=1.1$ and 1.2, respectively. The notations ZFC and FC indicate different prehistory: ZFC denotes preliminary cooling of the samples to $T<T_f$ (in our case, to $T=4.2$ K) in the absence of a field, while FC denotes cooling with $H \neq 0$. As can be seen from these figures, the behavior of the magnetization polytherms is irreversible, $\sigma_{ZFC}(T,H) \neq \sigma_{FC}(T,H)$, over a wide region, $T>T_f \sim 10$ K, and becomes stronger as the temperature decreases, especially for $T<T_f$. As field H is increased in the high-temperature region, the irreversibility effects are suppressed, but they become more intense as $T \to T_f^+$: the difference $\Delta\sigma = \sigma_{FC} - \sigma_{ZFC}$ increases. Such behavior at low temperatures is typical of systems with spin-glass ordering.[1,2,13] For $x \leqslant 1.1$, the irreversibility effects manifest themselves in an identical way and disappear over the entire temperature

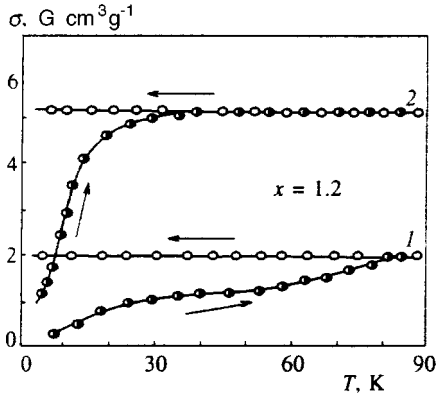


FIG. 2. Polytherms of the magnetization $\sigma_H(T)$ in ZFC and FC regimes of the spinel $Li_{0.5}Fe_{1.3}Ga_{1.2}O_4$ for various values of the field: *1*—50 Oe, *2*—100 Oe.

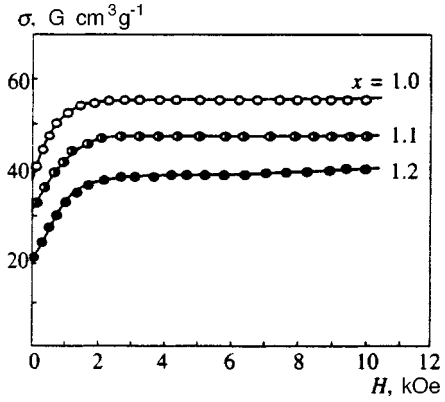


FIG. 3. Isotherms of the magnetization $\sigma_T(H)$ of the dilute spinels $Li_{0.5}Fe_{2.5-x}Ga_xO_4$ ($x=1.0-1.2$) for $T=4.2$ K.

range $T \geqslant 4.2$ K for $H \geqslant 100$ Oe. In contrast, the irreversibility is expressed more strongly for $x=1.2$ (see Fig. 2 and Ref. 1) and, as $T \to 0$, exists in fields up to several hundred oersted.

In the concentration region considered here, the behavior in small fields ($H<H_s$) must be determined by the domain structure. In particular, the most probable cause of irreversibility both for $T \to 0$ K and for higher temperatures is pinning of the domain boundaries. In this case, the pinning centers in general can consist of both inhomogeneities of the polycrystalline structure (inclusions, pores, local stresses, etc.) and inhomogeneities of the magnetic ordering caused by local breakdowns of spin collinearity.[14,15] The former mechanism causes irreversibility in ordinary magnetically soft ferromagnets: in weak magnetic fields, irreversibility exists at temperatures from 0 K to T_C, but the form of the $\sigma_{ZFC}(T)$ and $\sigma_{FC}(T)$ dependences can differ greatly.[14]

In comparison with this, irreversibility effects are expressed differently in Li–Ga spinels: they exist in far stronger magnetic fields, and a specific variation (of the spin-glass type) of the behavior of $\sigma_{ZFC}(T)$ is observed as the field increases and $T \to 0$ K. At the same time, their dependence on the concentration x of the nonmagnetic Ga^{3+} ions shows up clearly. In combination, all this is evidence of a second irreversibility mechanism, associated with the presence of regions of local noncollinearity. Note that the presence of these regions, which arise as pinning centers of the domain boundaries, can be used to explain the decrease of the real part of the dynamic susceptibility for $T<100$ K mentioned earlier.[2] Thus, the data obtained at low fields experimentally confirms the existence of regions of local noncollinearity theoretically predicted in Ref. 9 for the objects under consideration.

Figure 3 shows the magnetization isotherms $\sigma_T(H)$ of the test samples at $T=4.2$ K and $H \leqslant 10$ kOe. It can be seen that, for concentrations of $x=1.0$ and 1.1, technical saturation is reached at $H_s \sim 2$ kOe, and there is no para-process. This shows that local noncollinearity is broken down by the magnetic field, and a collinear ferrimagnetic state (or something very close to it) occurs for $H>H_s$. The noncollinear structure is more stable against the action of the field for the sample with $x=1.2$, as is evidenced by the presence of a para-process when $H>H_s$. This correlates with the irreversibility behavior, which is also expressed more strongly for this concentration.

3.2. Temperature dependences of the heat capacity $C_m(T)$

The temperature dependence of the magnetic contribution to the heat capacity $C_m(T)$ of Li–Ga spinels with $x=1.0-1.2$ for $T>T_f$, i.e., in the ferrimagnetic region, is shown in Fig. 4. For comparison, this figure also shows the data from Ref. 2 for a sample with $x=0.9$. For clarity, the experimental curves are shifted along the vertical axis (C_m) relative to each other. Assuming that the features of the behavior caused by the character of the ferrimagnetic states considered here can show up most clearly mainly in the temperature region close to T_f, we have initially restricted ourselves to the interval $10<T \leqslant 40$ K for all samples. Incidentally, we should point out that, for $T<T_f$ in the test objects, as in other reentrant magnets, $C_m(T) \sim T$.[11,13,16]

In analyzing the results given in Fig. 4, the following should be pointed out: The spin-wave law $C_m \sim T^{3/2}$, which is obeyed for unsubstituted Li spinels,[11] here clearly breaks down as a consequence of the formation of maxima. However, neglecting the maxima, the $C_m(T)$ curves as a whole maintain positive curvature (are concave), in which they differ from the similar dependence with higher Ga^{3+} concentrations,[11] as well as from those cited for other spin-glass systems, where the $C_m(T)$ curves have negative curvature (are convex).[13]

With regard to the latter, we should immediately point out that the transition from $C_m \sim T^{3/2}$ dependence (undiluted ferrimagnets) to convex curves (spin-glass systems), must naturally pass through certain intermediate stages that reflect the changes in the magnetic subsystem caused by a gradual increase of the concentration of nonmagnetic ions. The absence in the literature of detailed information on this question is evidence only that the region of the $x-T$ diagrams corresponding to the beginning of the formation of ferrimagnetic spin-glass states has so far been studied very little.

Taking into account that the concave shape of the $C_m(T)$ curves is caused by spin waves with a quadratic dispersion law, it was of interest to consider the concentration evolution

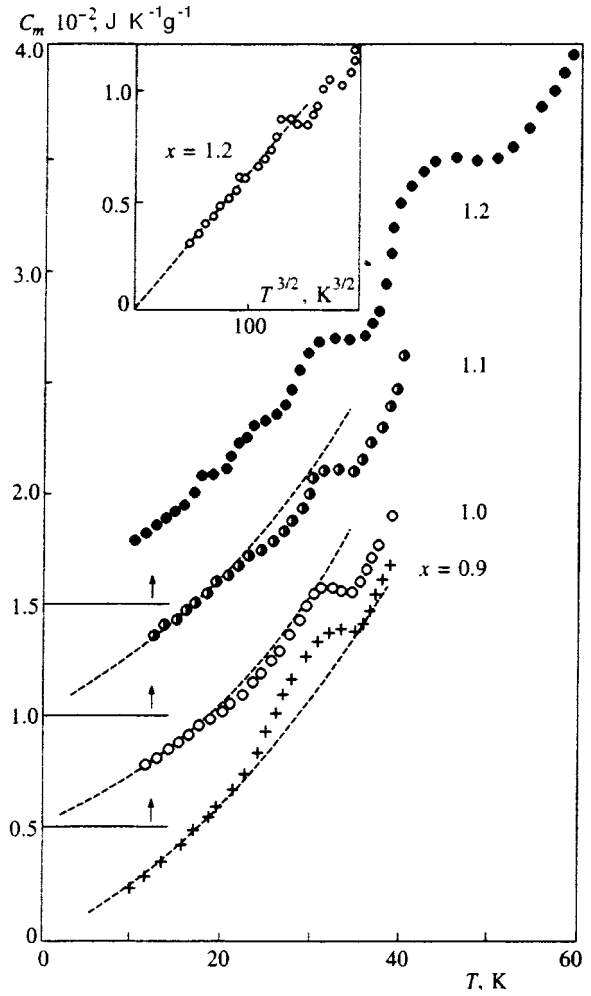


FIG. 4. Temperature dependences of the magnetic contribution to the heat capacity $C_m(T)$ of the dilute spinels $Li_{0.5}Fe_{2.5-x}Ga_xO_4$ ($x=1.0-1.2$ and $x=0.9$ from Ref. 2) for $H=0$. For $x=1.0-1.2$, the curves are shifted upwards along the vertical axis by (0.5, 1.0, and 1.5)$\times 10^2$ J K^{-1}g^{-1}, respectively. The inset shows the $C_m(T^{3/2})$ dependence for the sample with $x=1.2$. The dashed curves show extrapolations of $C_m(T)\sim T^{3/2}$ for $T\lesssim 20$ K.

of the $C_m(T)$ dependences from the standpoint of their deviation from the $T^{3/2}$ law. To do this, the experimental curves were replotted in $C_m-T^{3/2}$ coordinates. For illustration, one of these curves ($x=1.2$) is shown in the inset of Fig. 4. It was established by using this procedure that, for $x\geqslant 1.0$, the $T^{3/2}$ law is satisfied only in the temperature region directly adjacent to T_f, i.e., $T\lesssim 22$ K. At higher temperatures, despite the presence of maxima, the experimental $C_m(T)$ dependence rises less steeply than would follow from a $T^{3/2}$ law (dashed curves in Fig. 4). In contrast to this, for a sample with $x=0.9$, a maximum is superposed on the $T^{3/2}$ dependence.[2] As a whole, the behavior of the $C_m(T)$ temperature dependence for all samples with $x=0.9-1.2$ can obviously be interpreted as follows: In the temperature interval under consideration, the main type of excitation in the magnetic subsystem is spin waves, but the density of states in the spin-wave spectrum decreases as the degree of dilution and the temperature increase. Other contributions to the heat capacity also appear as temperature increases, because of which maxima appear on the $C_m(T)$ dependences.

The shape of these maxima, as well as the concentration variations of the course of the $C_m(T)$ dependences, clearly show that the detected features of the $C_m(T)$ behavior are not associated with a phase transition from a collinear to an ordered noncollinear ferrimagnetic structure. This is also supported by the fact that, for the temperatures corresponding to these maxima, no anomalies are observed in the behavior of the temperature dependences of the real part $\chi'(T)$ and imaginary part $\chi''(T)$ of the low-field dynamic susceptibility, whereas the latter has a characteristic maximum for $T=T_f$.[2,17]

At the same time, the appearance of maxima on the $C_m(T)$ dependences probably reflects the fact that additional degrees of freedom appear in the magnetic subsystem. Reference 9 expressed the assumption that this could be caused by regions of local noncollinearity. If the presence of these regions really causes the appearance of new modes in the excitation spectrum, because of which maxima are formed on the $C_m(T)$ dependences, it is clear that there must be several such maxima. This follows directly from the circumstance that the formation of such regions is a statistical process that is determined by the random placement of the nonmagnetic ions in the spinel lattice.[9] Taking into account the short-range character of the exchange, there is a good basis for assuming that, for the given x values, there exist several more probable types of regions of local noncollinearity (the degree of breakdown of the exchange, elongation, etc.). As a consequence, excitations with an identical nature, associated with different types of regions of local noncollinearity, must appear (and manifest themselves) at different temperatures.

The experimental data that have been obtained agree with such concepts. Actually, the maximum that appears for $x=0.9$ superposed on the $T^{3/2}$ dependence is maintained even at high concentrations of nonmagnetic ions, but its shape varies with increasing x. At the same time, the $C_m(T)$ behavior also varies in the low-temperature region—the $T^{3/2}$ law breaks down and, as is clearly seen from the $C_m(T)$ dependence for $x=1.2$, new maxima appear. By extending the temperature interval to 60 K for $x=1.2$, we convinced ourselves that the maxima exist even at higher temperatures.

It follows from the results of Sec. 3.1. that the regions of local noncollinearity are destroyed by a magnetic field. Therefore, studying the overall characteristics of $C_m(T)$ or $\sigma_s(T)$ in a magnetic field makes it possible to experimentally check the presence of coupling between the additional contribution to the heat capacity and regions of local noncollinearity. Keeping this in mind, we have also studied the temperature dependence of the magnetization in a field of $H=5$ kOe for all samples with $x\geqslant 1.0$.

3.3. Temperature dependence of the saturation magnetization

The temperature dependence of the magnetization of the test samples in a field of $H=5$ kOe at $T=(4.2-100)$ K are shown in Fig. 5. It can be seen from the results shown earlier in Sec. 3.1. (Fig. 3) that this magnetization can be regarded as saturation magnetization for $x=1.0$ and 1.1. Although the saturation is not complete at $H=5$ kOe for the sample with

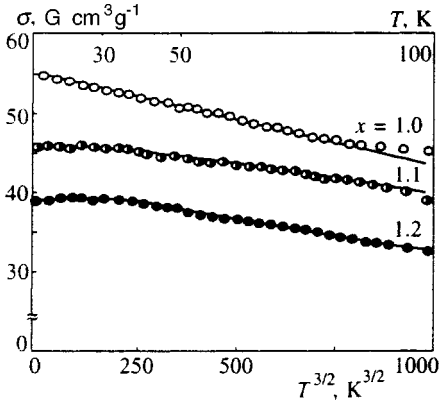


FIG. 5. Temperature dependences of the saturation magnetization $\sigma_s(T)$ of the dilute spinels $Li_{0.5}Fe_{2.5-x}Ga_xO_4$ ($x=1.0$–1.2) in $\sigma_s - T^{3/2}$ coordinates. The solid curves are calculated from Eq. (2). The values of A and Δ are given in the text.

$x=1.2$, technical saturation is achieved. Therefore, the term saturation magnetization (σ_s) will be used below for all the samples in just this sense.

As follows from the data of Fig. 5, unlike the $C_m(T)$ temperature dependence, there are no sharp changes in the temperature behavior for $\sigma_s(T)$ in each separate case. However, whereas the $C_m(T)$ behavior, in general, is of the same type for all the samples, appreciable concentration variations occur at low temperatures for $\sigma_s(T)$. This includes the appearance of a plateau for $T<30$ K for the $x=1.1$ case and a weakly expressed maximum for $x=1.2$ in the $T<50$ K region.

The uniform behavior of the $\sigma_s(T)$ dependences is evidence that additional excitations in the magnetic subsystem, manifesting themselves in the form of maxima on the $C_m(T)$ curves, either do not alter the z projection of the magnetization or are suppressed by the magnetic field because of the breakdown of their sources. If it is assumed that regions of local noncollinearity serve as such sources, both possibilities can take effect simultaneously. In particular, it was assumed in Ref. 9 that rigid rotation of these regions around the direction of $\sigma_s(T)$ can occur.

When the regions of local noncollinearity are completely destroyed by a magnetic field, i.e., when collinear ferrimagnetic structure is restored, the $\sigma_s(T)$ dependence must probably obey the spin-wave $T^{3/2}$ law. In order to check this, the experimental data for $\sigma_s(T)$ were also replotted in $\sigma_s - T^{3/2}$ coordinates, as is shown in Fig. 5.

It can be seen that the $T^{3/2}$ law is obeyed in the entire temperature range (4.2–100) K for a sample with $x=1.0$, whereas, for $x=1.1$ and 1.2, it is obeyed only in the comparatively high-temperature region, i.e., for $T>30$ K and $T>50$ K, respectively. Nevertheless, in our opinion, in combination with the data on the $C_m(T)$ behavior, this can support the hypothesis that the presence of regions of local noncollinearity for $H=0$ causes new degrees of freedom to appear in the magnetic subsystem, which make an additional contribution to the heat capacity.

The observed changes of the $\sigma_s(T)$ behavior at low temperatures for the cases $x=1.1$ and 1.2 do not contradict this, if local noncollinearity and frustrations are maintained in these samples and in the presence of a field $H>H_s$. As shown in Refs. 18–20, the $C_m(T)$ and $\sigma_H(T)$ behavior is determined in such a situation by the presence of a gap in the excitation spectrum, which has the form

$$\Delta=\mu(H-H_0), \tag{1}$$

where μ is the magnetic moment, H is the external field, and H_0 is the internal field. The presence of the term $-\mu H_0$, which, in contrast to the external field H, promotes the appearance of magnetic excitations, is caused by the presence of competition of the exchange and by frustrations.[19] Although the completely disordered state, i.e., a type of spin glass, was chosen as the initial state in the calculational model of Ref. 19, it changes in large fields to a frustrated ferrimagnetic state. Moreover, the magnetic excitations are regarded in Ref. 19 as noninteracting spin waves.

Starting from this, we attempted to approximate the experimental $\sigma_m(T)$ curves for $x=1.1$ and $x=1.2$ by a dependence of the form[21]

$$\sigma_s(T)=\sigma_{s0}[1-AT^{3/2}\exp(-\Delta/kT)], \tag{2}$$

where A, σ_{s0}, and Δ are parameters to be determined, and k is Boltzmann's constant.

We obtained the following values for the parameters by mathematically processing the experimental $\sigma_s(T)$ data: For $x=1.1$, we have $\sigma_{s0}=(46.7\pm0.7)$ G cm^{3}g^{-1}, $A=(12.0\pm0.5)\times10^{-5}$ K$^{-3/2}$, and $\Delta=(-15.2\pm0.1)$ K. For $x=1.2$, we have $\sigma_{s0}=(39.3\pm0.2)$ G cm^{3}g^{-1}, $A=(14.0\pm0.1)\times10^{-5}$ K$^{-3/2}$, and $\Delta=(-14.8\pm1.2)$ K. In both cases, the correlation coefficient is $R>0.99$. The calculated results are shown in Fig. 5 by solid curves.

The values obtained for the parameters of Eq. (2) are quite reasonable. For comparison, we should point out that, for $x=1.0$, the coefficient A in Bloch's law for $\sigma_s(T)$ equals $(18.1\pm1.2)\times10^{-5}$ K$^{-3/2}$. The negative sign of Δ is explained by the small value of the applied external field. Using as μ the magnetic moment of the Fe^{3+} ion, i.e., $5\mu_B$, at $H=5$ kOe we have $\mu H\simeq0.3$ K. Consequently, the gap is virtually completely determined by the internal field H_0. The numerical values of Δ are close to the corresponding freezing temperatures $T_f\sim10$ K. The same result ($|\Delta|\gtrsim T_f$) was obtained when the heat capacity $C_m(T)$ in large fields was investigated for samples of the (Eu–Sr)S spin-glass system.[18,20] There is thus a basis for assuming that the character of the $\sigma_s(T)$ temperature dependences at low temperatures in a field of $H=5$ kOe also reflects the specifics of the ferrimagnetic states that occur under these conditions, i.e., (T,H), in samples with $x=1.1$ and 1.2, namely, the presence of local noncollinearity and frustrations. For the $x=1.2$ case, their presence obviously follows from the results of Fig. 3 (see Sec. 3.1.). This apparently also occurs for the $x=1.1$ sample, although in a lesser degree. With the small susceptibility of the para-process, the absence of saturation in fields $H\leqslant10$ kOe can evidently by recorded only by a SQUID magnetometer.

4. CONCLUSION

The results of the study of the thermal and magnetic properties of model objects—the dilute reentrant ferrimagnetic spinels $Li_{0.5}Fe_{2.5-x}Ga_xO_4$ ($x=1.0-1.2$)—have confirmed the main concepts of the formation mechanism of ferrimagnetic spin-glass states in such systems, obtained theoretically.[9] The spin-glass state is preceded in temperature ($T>T_f$, $H=0$) by the frustrated ferrimagnet state, which is characterized by the presence of local breakdowns of collinear ferrimagnetic ordering—regions of local noncollinearity with frustrated exchange couplings retained within them. The ordered noncollinear structures whose existence for $T>T_f$ is predicted in the Gabay–Toulouse mean-field model[10] are absent here.

This model of the formation of ferrimagnetic spin-glass states is compared with the features detected in the behavior of the temperature dependence of the heat capacity, $C_m(T)$ at $H=0$, which as a whole deviate from the Bloch $T^{3/2}$ law and demonstrate the presence of maxima due to contributions not associated with spin waves. In the presence of a sufficiently strong magnetic field ($H>H_s$), which breaks down the ferrimagnetic spin-glass state and restores the collinear ferrimagnetic structure at all temperatures $T\geqslant 4.2$ K, the $T^{3/2}$ law is obeyed for saturation magnetization $\sigma_s(T)$ (the $x=1.0$ case), which confirms that the additional modes in the magnetic excitation spectrum are associated with regions of local noncollinearity.

If frustrations are maintained even in strong magnetic fields, then, according to the data that we obtained on the character of the $\sigma_s(T)$ dependence for samples with $x=1.1$ and 1.2 and the $C_m(T)$ data taken from Refs. 18–20 for the (Eu–Sr)S system, a gap of the form $\Delta=\mu(H-H_0)$ appears in the magnetic excitation spectrum. In this case, the excitations can be regarded as noninteracting spin waves, while the $-\mu H_0$ term that promotes their appearance is caused by the competition of the exchange interactions and by frustrations.

It is possible that the temperature dependence of $C_m(T)$ at $H=0$ ($x=1.1-1.2$) and of the saturation magnetization $\sigma_s(T)$ as $T\to 0$ K ($x=1.1$ and 1.2) detected in this work reflect certain common regularities in the variation of the character of the magnetic-excitation spectrum in frustrated ferro- and ferrimagnetic systems with short-range action. However, this question requires further study.

In conclusion, I consider it my pleasant duty to express my gratitude to V. A. Pervakov, S. R. Kufterina, and N. Yu. Tyutryumova for help in making the measurements.

[1] N. N. Efimova, Yu. A. Popkov, and N. V. Tkachenko, Zh. Éksp. Teor. Fiz. **90**, 1413 (1986) [Sov. Phys. JETP **63**, 827 (1986)]; Fiz. Nizk. Temp. **16**, 1565 (1990) [Sov. J. Low Temp. Phys. **16**, 881 (1990)].
[2] N. N. Efimova, Yu. A. Popkov, S. R. Kufterina, V. A. Pervakov, V. I. Ocharenko, and N. Yu. Tyutryumova, Fiz. Nizk. Temp. **20**, 546 (1994) [Low Temp. Phys. **20**, 431 (1994)].
[3] W. M. Saslow, and G. Papkep, Phys. Rev. Lett. **56**, 1074 (1986).
[4] J. R. Thomson, Hong Gud, D. Y. Ryan *et al*., Phys. Rev. B **45**, 3129 (1992).
[5] T. A. Kaplan, Phys. Rev. **109**, 782 (1958).
[6] I. Ya. Korenblit and E. F. Shender, Usp. Fiz. Nauk **126**, 233 (1978) [Sov. Phys. Usp. **21**, 832 (1978)].
[7] V. A. Ignatchenko and R. A. Iskhakov, Zh. Éksp. Teor. Fiz. **72**, 1005 (1977) [Sov. Phys. JETP **45**, 526 (1977)]; Izv. Akad. Nauk SSSR, Ser. Fiz. **44**, 1434 (1980).
[8] I. Ya. Korenblit and E. F. Shender, Phys. Rev. B **33**, 624 (1986).
[9] J. Villain, Z. Phys. B **33**, 31 (1979).
[10] M. Gabay and G. Toulouse, Phys. Rev. Lett. **47**, 201 (1981).
[11] N. N. Efimova, V. A. Pervakov, V. I. Ovcharenko, and N. Yu. Tyutyurmova, Fiz. Tverd. Tela **35**, 2838 (1993) [Phys. Solid State **35**, 1405 (1993)].
[12] É. A. Anders, B. Ya. Sukharevskiĭ, and I. A. Volchok, *Thermophysical Properties of Matter and Substances* (Amerind Publishing, New Delhi, 1974; Izd. Standartov, Moscow, 1976).
[13] K. Binder and A. P. Yuang, Rev. Mod. Phys. **58**, 801 (1986).
[14] L. V. Kirenskiĭ, A. I. Drokin, and D. A. Laptiĭ, *Temperature Magnetic Hysteresis of Ferrimagnets and Ferrites* (Novosibirsk, 1965) [in Russian].
[15] J. Hubsch and G. Gavoile, Phys. Rev. B **26**, 3816 (1982).
[16] N. N. Efimova, Yu. A. Popkov, and S. R. Kufterina, JETP Lett. **64**, 450 (1996).
[17] N. N. Efimova, Yu. A. Popkov, G. A. Takzeĭ, A. B. Surzhenko, and A. M. Dvoeglazov, Fiz. Tverd. Tela **36**, 490 (1994) [Phys. Solid State **36**, 271 (1994)].
[18] H. V. Löhneyesen, R. Berg, G. V. Lekomte, and W. Zinn, Phys. Rev. B **31**, 2920 (1985).
[19] U. Krey, J. Phys. (France) Lett. **46**, 845 (1985).
[20] J. Woznitza, H. V. Löhneysen, and U. Krey, Phys. Rev. B **33**, 3436 (1986).
[21] A. I. Akhiezer, V. G. Bar'yakhtar, and S. V. Peletminskiĭ, *Spin Waves* (North-Holland, Amsterdam, 1968; Moscow, Nauka, 1967).

Translated by W. J. Manthey

Kinetic theory of rotating molecule interaction with a solid surface

A. Yu. Pankov*) and V. D. Borman†)

Moscow State Engineering Physics Institute, 115409 Moscow, Russia

S. Yu. Krylov‡)

Institute of Chemical Physics, Russian Academy of Sciences, 117915 Moscow, Russia

(Submitted 23 July 1997)
Zh. Éksp. Teor. Fiz. **113**, 1350–1363 (April 1998)

A kinetic theory of interaction between molecules with rotational degrees of freedom and a solid surface for arbitrary ratios among the times of molecule rotation, flight through the region of surface forces, and relaxation of a molecular ensemble due to phonons has been developed. A kinetic equation for an ensemble of molecules residing in the field of surface forces has been derived from the equation for the one-particle distribution function of molecules by averaging it along the dynamical trajectories in the region of surface force action. A simple analytic expression for the probability of trapping a molecule with rotational degrees of freedom has been obtained. Experimental data on rotational cooling and rotational polarization of desorbed molecules are discussed. © *1998 American Institute of Physics.* [S1063-7761(98)01704-1]

1. INTRODUCTION

In recent years considerable attention has been focused on the dynamics of multiatomic molecules interacting with solid surfaces. Features of the adsorption dynamics show up in distributions of molecules over various states affected by their interaction with a surface. Using laser spectroscopy techniques, which are sensitive to rotational states of a molecule, researchers have obtained abundant information about the distribution of molecules scattered in the thermal regime.[1–5] Effects of rotational polarization and the ''rotational rainbow''[1–3] in scattered molecules, translational cooling, rotational cooling, and polarization of desorbed molecules[3–5] have been observed. Features of the distribution over rotational states of molecules during their interaction with surfaces manifest themselves in transport processes in rarefied gas—in particular in light-induced drift[6] and the influence of the magnetic field on heat and particle flows.[7,8]

Effects connected with the nonequilibrium distribution of molecules interacting with a surface can be described in a natural way by the kinetic theory. Techniques based on the microscopic approach have been developed thus far only for monatomic particles.[9–11] As for molecules with internal degrees of freedom, the theory runs into considerable difficulties. Thus, in polyatomic molecules the interaction potential with a surface strongly depends on the angular orientation of the molecule.[12] For thermal molecules, the times of rotation and flight through the region of surface forces (10^{-13}–10^{-12} s) are comparable to the relaxation time of the molecular ensemble due to phonons in the solid.[11] Progress in describing surface scattering of rotating molecules has been achieved only in computer simulations (molecular dynamics[12,13]). Some features of the distribution of fast scattered molecules have been satisfactorily described by dynamical scattering theory.[14,15]

In the reported work we have developed a consistent kinetic theory of rotating molecules interacting with a solid surface (Sec. 2). It is based on the theory[11] proposed earlier for monatomic particles and is a generalization of the latter theory to molecules with rotational degrees of freedom. A fundamentally new feature of our theory, which allows us to describe the distribution of molecules residing in a complex multidimensional field, is the introduction of new variables defining molecular trajectories in the mean surface field instead of traditional canonically conjugate variables. Below we purpose a kinetic equation obtained by averaging the exact equation for a one-particle distribution function along various dynamical trajectories. This equation contains the loss and gain terms describing the arrival of molecules from gas (molecular beam) to the interaction region, and their departure back to the gas phase as a result of molecule motion along the trajectories above the surface potential well, and the collisional integral due to phonons, which leads, in particular, to transitions of molecules to bound (adsorbed) states.

Solution of the kinetic equation for an ensemble of surface particles allows one to determine the distribution function of molecules departing from the surface (a surface boundary condition for the gas distribution function). With the scattering integral in the relaxation-time approximation, we obtain (Sec. 3) a simple analytic expression for the probability of trapping a molecule with rotational degrees of freedom, and provide a microscopic justification for the previously suggested phenomenological description of certain effects.[16] On the basis of our results, we discuss general features of translational and rotational cooling of desorbed molecules, and rotational polarization of desorbed and scattered particles.

2. KINETIC EQUATION

We start our consideration with the potential V of interaction between a molecule and a solid, which depends on the

coordinates $\mathbf{r}_j$ of all atoms in the solid, the coordinates x, y, and z of the molecule (the z-axis is normal to the surface, x and y are coordinates in the plane parallel to the surface), and other generalized coordinates Ω. The coordinates Ω specify molecular orientation angles with respect to the surface; for a linear molecule, $\Omega=\{\vartheta,\phi\}$, i.e., the polar and azimuthal angles of the molecule, and for a more complex shape of the molecule $\Omega=\{\vartheta,\varphi,\psi\}$, i.e., the three Euler's angles. Let us denote by $\{\alpha_k\}$ the set of generalized coordinates and by $\{\beta_k\}$ the set of corresponding generalized momenta, $k=1,\ldots,K$; K is the number of molecular degrees of freedom.

Following the conventional technique of constructing kinetic equations for a system of $N+1$ particles,[17,18] let us express the classical kinetic equation for the one-particle distribution function $F(t,\alpha,\beta)$ of molecules interacting with an equilibrium solid surface in the form

$$\frac{\partial F}{\partial t}+\sum_{k=1}^{K}\left(\dot{\alpha}_k\frac{\partial F}{\partial \alpha_k}+\dot{\beta}_k\frac{\partial F}{\partial \beta_k}\right)=I[F], \tag{1}$$

$$\dot{\alpha}_k=\frac{\partial H_M}{\partial \beta_k},\quad \dot{\beta}_k=-\frac{\partial H_M}{\partial \alpha_k}, \tag{2}$$

where $H_M=H_M(\{\alpha_k\},\{\beta_k\})$ is the Hamiltonian of a molecule, which equals the sum of the rotational and translational kinetic energies, and the potential energy $V^{(1)}(\{\alpha_k\})$ averaged over equilibrium states of the solid.[18] The right-hand side of Eq. (1) is due to the residual part of the interaction potential, $V^{(2)}=V-V^{(1)}$, which depends on thermal displacements $\delta\mathbf{r}_j=\mathbf{r}_j-\mathbf{R}_j$ of particles with respect to their lattice sites $\mathbf{R}_j$. With due account of $V^{(2)}$, which varies in time owing to thermal oscillations of surface atoms, the energy of a molecule moving in the region of surface forces can change due to exchange with the solid. This exchange, obviously, can be described in terms of creation and annihilation of phonons; in other words, $I[F]$ can be treated as a collisional integral between molecules and phonons. The interaction among molecules is neglected, since the surface coverage of adsorbed molecules is assumed to be small.

Terms on the left of Eq. (1) are due to both the explicit time dependence of the distribution function and the change in F during motion in potential $V^{(1)}$. The potential is shown in Fig. 1. Depending on the sign of the total energy E of the molecule (the sum of the kinetic and potential energies), motion along z is finite for $E<0$ and semi-infinite for $E>0$. The scattering integral takes into account transitions between states of the same sign of E, and transitions between states of opposite sign, $E>0$ and $E<0$.

Equation (1) contains detailed information about the variation of the molecular distribution function in the region of surface forces. In recent experiments, however, mean values of parameters which are integrals of the motion outside the region of interaction with the surface are measured.[1] In determining the distribution function of molecules moving away from the surface after interaction, we will follow the approach of Ref. 11 and develop a rough approximation by averaging Eq. (1) over the range of surface forces. To this end, we first transform from variables $\{\alpha,\beta\}$ to new independent ones. In this capacity, let us use variables defining molecular trajectories in the potential $V^{(1)}$ and positions of molecules on their trajectories.

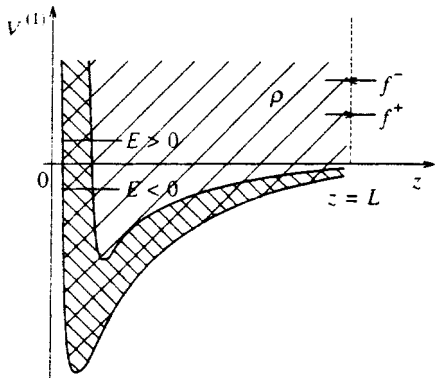


FIG. 1. Shape of potential $V^{(1)}$ describing interaction of a molecule with a surface. Values of $V^{(1)}$ between the two curves correspond to various molecular orientations with respect to the surface and different positions of molecules' centers of mass in the surface plane; f^-, f^+, and ρ are distribution functions of molecules moving towards the surface, away from the surface, and of the ensemble of molecules in the surface region $0<z<L$, respectively.

By virtue of the equations of motion (2), all generalized coordinates and momenta at an arbitrary moment are uniquely determined by their initial values. Let us draw the plane $z=L$, which bounds the region of surface forces (Fig. 1), and select the initial value $z=\tilde{z}$: $\tilde{z}=L$ for $E>0$ and arbitrary $\tilde{z}$ in the region of surface forces for $E<0$. The initial values of all variables $\{\alpha,\beta\}$ except $z=\tilde{z}$ will be denoted by γ ($\gamma=\{\gamma_i\}$, $i=1,\ldots,2K-1$). If we solve the equations of motion with initial conditions $\{\alpha,\beta\}|_{t=0}=\{\tilde{z},\gamma\}$ and express the time in terms of the coordinate z, the variables $\{\alpha,\beta\}$ can be expressed in terms of variables $\{z,\gamma\}$:

$$\alpha_1=z,\quad \alpha_2=A_1^{(\xi)}(\gamma_1,\ldots,\gamma_{2K-1},z),\ldots,$$
$$\alpha_K=A_{K-1}^{(\xi)}(\gamma_1,\ldots,\gamma_{2K-1},z),$$
$$\beta_1=B_1^{(\xi)}(\gamma_1,\ldots,\gamma_{2K-1},z),\ldots,$$
$$\beta_K=B_K^{(\xi)}(\gamma_1,\ldots,\gamma_{2K-1},z).$$

In this case, the values of the inverse functions

$$\Psi_1^{(\xi)}(\alpha_1,\ldots,\alpha_K,\beta_1,\ldots,\beta_K)=\gamma_1,\ldots,\Psi_{2K-1}^{(\xi)}$$
$$\times(\alpha_1,\ldots,\alpha_K,\beta_1,\ldots,\beta_K)=\gamma_{2K-1}$$

are constant with time and can be considered independent integrals of the motion.[19] Any fixed set of variables γ specifies a molecular trajectory in the potential $V^{(1)}$, whereas the coordinate z and superscript ξ ($\xi=1,\ldots,\Xi(\gamma)$) specify the molecule's position on its trajectory. The superscripts ξ are introduced because each trajectory has at least two branches of the function relating $\{\alpha,\beta\}$ to $\{z,\gamma\}$, which correspond to the motion of a molecule towards the surface ($\beta_1\equiv p_z<0$ for $\xi=1$) and away from it ($p_z>0$ for $\xi=2$). The number of branches Ξ equals two, however, only in the simplest case, when the potential $V^{(1)}$ depends only on z. If the potential also depends on other coordinates, there can be trajectories

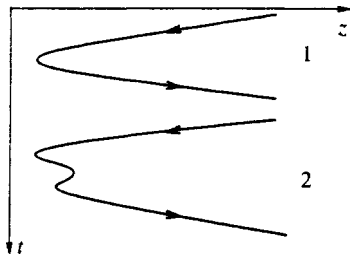


FIG. 2. Trajectories for $E>0$ in potential $V^{(1)}$: 1) for $\Xi=2$; 2) for $\Xi>2$.

with $\Xi>2$. Such trajectories are due to the existence of additional turning points in the potential $V^{(1)}$ (when a molecule passes through a turning point, ξ increases by one; see Fig. 2) because of the exchange of kinetic energy between translation along the normal to the surface and other degrees of freedom.[1)] In particular, if the potential is a function of z and molecular rotation angles, this is the exchange between the translational and rotational degrees of freedom. If the potential depends on coordinates x and y in the plane of the surface, the correlations between normal and tangential motion play a similar role. For trajectories with $E>0$, Ξ can be assumed to be finite, since the motion along the normal to the surface is semi-finite. Without loss of generality, we assume for simplicity that trajectories with $E<0$ are closed, so Ξ is finite even when $E<0$.

Since the transformation between variables $\{\alpha,\beta\}$ and $\{z,\gamma\}$ is one-to-one for all given ξ, any function of $\{\alpha,\beta\}$ can be reduced to a set of functions of $\{z,\gamma\}$ labeled by ξ. The element of the phase space in variables $\{z,\gamma\}$ takes the form

$$d\alpha\, d\beta=|J^{(\xi)}(z,\gamma)|dz\, d\gamma, \tag{3}$$

where $d\alpha$, $d\beta$, and $d\gamma$ denote products of differentials of the corresponding variables (see Appendix), and $J^{(\xi)}(z,\gamma)$ is the Jacobian, of the transformation, which can be interpreted as the density of states in terms of the new phase variables. Irrespective of the number of molecular degrees of freedom,

$$|J^{(\xi)}(z,\gamma)|=|v_z^{(\xi)}(z,\gamma)|^{-1}|\tilde{v}_z|. \tag{4}$$

Here $\tilde{v}_z$ is the initial velocity on trajectory γ at $z=\tilde{z}$ ($\tilde{z}=L$ for $E>0$). Equation (4) can be derived by differentiating the left- and right-hand sides of Eq. (3), with respect to time with due account of Liouville's theorem about conservation of phase space, and the fact that at $z=\tilde{z}$ the variables $\{\alpha,\beta\}$ are equal to their initial values $\{\tilde{z},\gamma\}$.

Replacing the canonically-conjugate variables $\{\alpha,\beta\}$ in Eq. (1) with the integrals of the motion $\{\gamma\}$ in potential $V^{(1)}$ and the coordinate z, and noting that $\dot{\gamma}_i=0$, we obtain

$$\frac{\partial F_\xi(t,z,\gamma)}{\partial t}+v_z^{(\xi)}(z,\gamma)\frac{\partial F_\xi(t,z,\gamma)}{\partial z}=I_\xi[F_{\xi'}]. \tag{5}$$

It is clear from (5) that the kinetic equation can be simplified considerably in variables $\{z,\gamma\}$ since the sum of $2K$ terms on the left-hand side of Eq. (1) corresponds to just one term in Eq. (5) (the second). It describes variations in the one-particle distribution function due to the motion of molecules along trajectories.

We make the natural assumption that the distribution function is continuous along trajectories, and supplement Eq. (5) with boundary conditions at the turning points $z_0^{(\xi)}$ (Fig. 2):

$$F_\xi(t,z=z_0^{(\xi)},\gamma)=F_{\xi+1}(t,z=z_0^{(\xi)},\gamma),$$

and for $E<0$ we add the condition

$$F_1(t,z=z_0^{(1)},\gamma)=F_\Xi(t,z=z_0^{(1)},\gamma),$$

which reflects the periodicity of motion along closed trajectories.

Equation (5) will be averaged below along trajectories in the region of surface forces. Before averaging the equations, we improve upon the set of variables, which will then conveniently specify trajectories. Note first that when $E>0$, the set of variables γ, which is defined at $z=L$, can be selected either at the point of entry to the surface region ($\tilde{v}_z<0$), or at the point of egress ($\tilde{v}_z>0$). We denote these sets of variables by γ^- and γ^+, respectively. In describing effects associated with the distribution of molecules after their interaction with the surface, the set γ^+ is the natural one.

Second, one can determine trajectories by using instead of $2K-1$ initial values of coordinates and momenta $2K-1$ independent combinations of these values. For $E>0$, variables γ can be chosen so that a subset of them, denoted by Γ, coincide at $z=L$ with the constants of the motion of the molecule outside the region of surface forces. For a linear molecule, these are the variables $\Gamma=\{\mathbf{v},\mathbf{J}\}$, i.e., the components of the velocity and angular momentum vectors. The rest of the variables, $\gamma_\sim$, specify the initial values of quantities that change during free motion of the molecule. For a linear molecule, these are the initial values of x and y in the plane of the surface and the molecular axis orientation angles in a plane perpendicular to $\mathbf{J}$. The representation of variables γ in the form $\gamma=\{\Gamma,\gamma_\sim\}$ for linear and more complex molecules is considered in greater detail in the Appendix. For $E<0$ we choose variables γ in the same way as for $E>0$ (this is of no further significance).

Note that information about variables $\gamma_\sim$ at $z=L$ is lost as a molecule moves from the region of surface forces to a detector. This is due to the fact that the phase variables, with initial values $\gamma_\sim$, change within certain limits, and these parameters run over their ranges many times during the molecular motion. The equations below will be averaged over $\gamma_\sim$. Note that this averaging is quite similar to averaging over rapidly varying phases in the kinetic theory of a gas with rotational degrees of freedom.[21]

Let us average Eq. (5) along trajectories in the surface region $0<z<L$. Multiplying Eq. (5) by

$$|J^{(\xi)}(z,\gamma)|\left(\omega^{-1}\int d\gamma_\sim\sum_{\xi=1}^{\Xi}\int_0^L dz|J^{(\xi)}(z,\gamma)|\right)^{-1},$$

then integrating over z, summing over ξ with due account of Eq. (4), and averaging over $\gamma_\sim$, we obtain

$$\frac{\partial \rho}{\partial t}+\theta(E)t_{fl}^{-1}(f^{+}-\hat{A}f^{-})=\langle I\rangle. \tag{6}$$

Here $\theta(E)=1$ for $E>0$ and 0 for $E<0$, and f^{+} and f^{-} are distribution functions of molecules moving towards the surface and away from the surface outside the interaction region:

$$f^{+}\equiv f^{+}(t,\Gamma^{+})=\omega^{-1}\int d\gamma_{\sim}^{+}F_{\xi=\Xi}(t,z=L,\gamma^{+}),$$
$$f^{-}\equiv f^{-}(t,\Gamma^{-})=\omega^{-1}\int d\gamma_{\sim}^{-}F_{\xi=1}(t,z=L,\gamma^{-}). \tag{7}$$

The function

$$\rho\equiv\rho(t,\Gamma)=\frac{\omega^{-1}\int d\gamma_{\sim}\sum_{\xi=1}^{\Xi}\int_{0}^{L}dz|J^{(\xi)}(z,\gamma)|F_{\xi}(t,z,\gamma)}{\omega^{-1}\int d\gamma_{\sim}\sum_{\xi=1}^{\Xi}\int_{0}^{L}dz|J^{(\xi)}(z,\gamma)|} \tag{8}$$

is the averaged distribution function of molecules in the surface region $0<z<L$. It describes the distribution of molecules over states Γ which are specified by the set of parameters that do not vary when a molecule moves in free space; ω is the phase space corresponding to variables $\gamma_{\sim}$ (see Appendix). By virtue of Eq. (8), the function ρ in equilibrium equals the usual Boltzmann distribution function $\rho_{0}(E)\propto\exp(-E/T)$, and since the original distribution function F is normalized to the total number of molecules,

$$\int d\alpha\, d\beta F=N,$$

the distribution function ρ is normalized to the number of molecules N_{s} in the surface region $0<z<L$,

$$N_{s}=\int d\Gamma g(\Gamma)\rho(t,\Gamma),$$
$$g(\Gamma)=\int d\gamma_{\sim}\sum_{\xi=1}^{\Xi}\int_{0}^{L}dz|J^{(\xi)}(z,\gamma)|. \tag{9}$$

The function $g(\Gamma)$ has the sense of a density of states in the variables Γ. The operator $\hat{A}$,

$$\hat{A}f^{-}\equiv f^{-}(t,\Gamma^{+})=\omega^{-1}\int d\gamma_{\sim}^{+}F_{\xi=1}(t,z=L,\gamma^{+}), \tag{10}$$

is introduced because the distribution function of molecules incident on the surface, which depends in a natural way on the variables γ^{-} of particles arriving at the surface region, can be expressed in terms of variables γ^{+} of particles leaving the surface region using an inverse transform performed along the trajectories in potential $V^{(1)}$.

Equation (6) also contains the quantity t_{fl}, which has the sense of an averaged time of flight through the surface region $0<z<L$ along the trajectory γ. The average is taken over the x, y, and the orientation angles, which vary during molecular motion in free space:

$$t_{fl}\equiv t_{fl}(\Gamma)=\omega^{-1}\int d\gamma_{\sim}\, t_{fl}(\gamma), \tag{11}$$

where

$$t_{fl}(\gamma)=\sum_{\xi=1}^{\Xi}\int_{0}^{L}\frac{dz}{|v_{z}^{(\xi)}(z,\gamma)|}. \tag{12}$$

The mean collisional integral $\langle I\rangle$ in Eq. (6) is determined by an expression whose only difference from Eq. (8) is that F_{ξ} is replaced with I_{ξ} (see Eq. (5)).

Note that Eq. (6) is strictly correct. In order to derive a closed equation for the distribution function ρ, let us use, as in Ref. 11, a rough model of molecular interaction with the surface. We assume that the distribution f^{+} of particles leaving the interaction region is determined by the average distribution ρ of surface particles[11]:

$$f^{+}=\theta(E)\rho. \tag{13}$$

Note that Eq. (13) is reasonable for $t>t_{fl}$. Using Eq. (13), let us rewrite Eq. (6) as

$$\frac{\partial \rho}{\partial t}=\theta(E)t_{fl}^{-1}\hat{A}f^{-}-\theta(E)t_{fl}^{-1}\rho+\tau^{-1}(\rho_{0}-\rho). \tag{14}$$

This is the sought-for approximate kinetic equation for molecules with internal degrees of freedom. The first two terms on the right-hand side of Eq. (14) describe the arrival of molecules from the gas phase (molecular beam) at the surface region and their departure to the gas phase. According to Eq. (14), particles come and go during the time of flight (averaged over the ''unmeasured'' variables $\gamma_{\sim}$) through the surface force region along trajectories with $E>0$. Unlike the corresponding terms on the left-hand side of Eq. (6), the gain and loss terms in Eq. (14) do not change sign under time reversal, and generate additional entropy (as a source of dissipation supplementing the collisional integral) in the ensemble of surface particles. Their dissipative property is due to Eq. (13), which breaks the symmetry of the left-hand side of Eq. (6) under time reversal. Note that the structure of Eq. (14) is similar to that of a previously suggested kinetic equation (Eq. (8) in Ref. 11) describing monatomic particle interactions with a surface. Unlike the mentioned equation, which was derived for the case of a potential that was a function of only the z coordinate, Eq. (14) allows the interaction potential to depend on an arbitrary number of generalized molecular coordinates.

In deriving Eq. (14), we have expressed the collsional integral in the relaxation time approximation, i.e.,

$$\langle I\rangle=\tau^{-1}(\rho_{0}-\rho). \tag{15}$$

In our previous work[11,16] we showed that the τ-approximation provides a satisfactory description of some effects observed in transport through a rarefied gas and scattering of molecular beams on surfaces. The rate τ^{-1} introduced in Eq. (14) describes relaxation due to phonons of the distribution function ρ averaged over the surface layer. Note that in real molecule/surface systems, the rate of collisions with phonons strongly depends on z, since the potential $V^{(2)}$ responsible for this interaction vanishes outside the range of

the surface field. Thus, τ^{-1} has the sense of the collision rate averaged over the surface region. By assuming also (as usual[10,22]) that an average probability of transitions between molecular trajectories due to interaction with the solid has been determined, the relaxation rate can be expressed as

$$\tau^{-1}\equiv\tau^{-1}(\Gamma)=\sum_{\Gamma'}\dot{W}(\Gamma\to\Gamma'),$$
$$\dot{W}(\Gamma\to\Gamma')=\omega^{-1}\int d\gamma_{\sim}\int d\gamma'_{\sim}\dot{w}(\gamma\to\gamma'), \tag{16}$$

where $\dot{w}(\gamma\to\gamma')$ and $\dot{W}(\Gamma\to\Gamma')$ are the rates of transitions out of trajectory γ to γ' and from state Γ to Γ'.

3. RESULTS OF CALCULATIONS IN τ-APPROXIMATION

3.1. As follows from Eq. (14), the distribution function of particles in the region of surface forces takes shape in a time of order t_{fl} and τ; this rate is of order 10^{-13}–10^{-12} s in realistic systems with physically adsorbed particles,[11] at times much greater than 10^{-12} s, the time derivative can be neglected in comparison with the terms on the right-hand side of Eq. (14). In this quasi-stationary regime, we derive from Eq. (14)

$$\rho=\begin{cases}S\rho_0+(1-S)\hat{A}f^-, & E>0,\\ \rho_0, & E<0,\end{cases} \tag{17}$$

$$S\equiv S(\Gamma)=\left(1+\frac{\tau(\Gamma)}{t_{fl}(\Gamma)}\right)^{-1}. \tag{18}$$

According to Eq. (17), molecules in bound states ($E<0$) are in equilibrium, whereas molecules in unbound states separate into two groups. Particles in the first group are those whose relaxation process has been completed, while the distribution of the rest depends on the distribution function f^- of particles incident on the surface, and is determined by its transform as the particles pass through the region of surface forces, in accordance with Eq. (10). Equations (17) and (18) indicate that S has the sense of a relaxation probability.

3.2. Let us derive equations that describe the transport of particles in the system. By integrating Eqs. (14) and (13) over states Γ with weighting functions $g(\Gamma)$ and $g(\Gamma)t_{fl}^{-1}(\Gamma)$, respectively (see Eq. (9)), using Eqs. (4) and (17), we obtain

$$\frac{dN_s}{dt}=S^-G^--S_0(t_a^{(0)})^{-1}N_s, \tag{19}$$

$$G^+=(1-S^-)G^-+S_0(t_a^{(0)})^{-1}N_s. \tag{20}$$

Here

$$G^{\mp}=\omega\int d\Gamma|v_z|f^{\mp}(\Gamma)$$

are fluxes of molecules moving towards the surface and away from it outside the interaction region, and factors S_0 and S^- in Eqs. (19) and (20) are given by the relaxation probability averaged in various ways:

$$S_0=\left(\int d\Gamma\,\theta(E)|v_z|\rho_0(E)\right)^{-1}\int d\Gamma\,\theta(E)|v_z|S(\Gamma)\rho_0(E). \tag{21}$$

The difference between Eq. (21) and the expression for S^- is that the distribution function $\rho_0(E)$ in the latter is replaced with the distribution $f^-(\Gamma)$ of molecules incident on the surface. The quantity $t_a^{(0)}$, which has the sense, as follows from Eq. (20), of the mean particle lifetime at the surface in equilibrium (in this case $G^+=G^-=(t_a^{(0)})^{-1}N_s$), is given by

$$(t_a^{(0)})^{-1}=\left(\int d\Gamma g(\Gamma)\rho_0(E)\right)^{-1}\times\int d\Gamma g(\Gamma)\theta(E)t_{fl}^{-1}(\Gamma)\rho_0(E). \tag{22}$$

According to this formula, the equilibrium escape rate is determined by the transit time through the region of surface forces along the trajectories of $E>0$ averaged over the entire ensemble of particles. It follows from Eq. (22) that in the case of a deep potential well (when $V_m\gg T$), $t_a^{(0)}$ is given by the Frenkel–Arrhenius formula $t_a^{(0)}=t_0\exp(V_m/T)$, and in the special case of a potential that depends only on z, the coefficient of the exponential equals the oscillation period near the well minimum:

$$t_0\approx m^{1/2}(d^2V^{(1)}/dz^2)^{-1/2},$$

where m is the molecule mass.

Equations (19) and (20) are formally identical to the equations of phenomenological adsorption theory. The adsorption time, in accordance with Eq. (20), is expressed as $t_a=t_a^{(0)}/S_0$, and this expression depends (see Eqs. (21) and (18)) on the ratio between the characteristic relaxation time τ and time of flight τ_{fl}. In the limiting cases we have

$$t_a=\begin{cases}(\tau/t_{fl})t_a^{(0)}, & \tau/t_{fl}\gg 1,\\ t_a^{(0)}, & \tau/t_{fl}\ll 1.\end{cases}$$

Note that the expression for the lifetime at $\tau/t_{fl}\gg 1$ is in agreement with the Kramers theory in the case of weak friction.[23] Although our formula applies to the case of intermediate friction ($\tau/t_{fl}\leqslant 1$), our result is different from the Kramers one in the limit $\tau/t_{fl}\to 0$, since the latter yields an infinitely growing lifetime as the friction increases. The results are different because the limiting case of strong friction cannot be described when the kinetic equation is averaged over z, and approximation (13) obviously ignores the fact that the particle mobility drops to zero if the friction is strong. It is known that the number of particles in this case can be obtained using Smoluchowski's equation. Therefore our model of particle interaction with a surface is limited to the case of moderate friction. The distribution shape, however, which is the only interesting result for us, is described adequately in both limiting cases, regardless of the number of escaping particles. Indeed, in accordance with Eqs. (17) and (18), we have $f^+=\rho=\hat{A}f^-$ for $\tau/t_{fl}\gg 1$, whereas for $\tau/t_{fl}\ll 1$ we have $f^+=\rho=\rho_0$. Note also that the strong friction regime, as is well known, cannot be realized in interaction of real molecules with a surface.[10] So, the ratio τ/t_{fl} cannot be arbitrarily small, since the relaxation time due to

phonons obviously has a lower limit determined by the oscillation period of atoms in the solid ($\sim 10^{-13}$ s), while the time of flight through the interaction region several Å thick is about 10^{-13}–10^{-12} s.

3.3. By substituting Eq. (17) into (13), let us determine the distribution functions of molecules moving away from the surface at times much longer than t_{fl} and τ:

$$f^+ = S\rho_0 + (1-S)\hat{A}f^-. \tag{23}$$

When applied to scattering of a molecular beam on a surface, Eq. (23) relates the distribution function f^+ of molecules moving from the surface to the distribution function f^- of molecules incident on the surface outside the interaction region. The distribution functions ρ_0 and f^+ are normalized in accordance with Eqs. (19) and (20), which describe the particle transport. Therefore Eq. (23) means that ρ_0 can change in a time of about the adsorption time. Thus, for scattering on a surface that was initially clean (not coated with an adsorbate), we can obtain by solving Eqs. (19) and (20) that $N_s \simeq 0$, $G^+ \simeq (1-S^-)G^-$ for $t \ll t_a$ and $N_s \simeq S^- G^- t_a$, $G^+ \simeq G^-$ for $t \gg t_a$. Hence, the first term of Eq. (23) can be neglected for $t \ll t_a$, but it becomes important at times comparable to the adsorption time. In another special case, when the arrival of particles is zero, Eq. (23) reduces to a simple expression for the distribution function of desorbed molecules:

$$f^+_{\rm des} = c(t)S(\Gamma)\exp(-E/T_s),$$
$$c(t) = \text{const}\cdot\exp(-t/t_a). \tag{24}$$

An important point is that Eq. (23) can be used as a boundary condition for the Boltzmann equation in describing nonequilibrium processes in the gaseous phase. The thickness of the surface-force region (1–5 Å) is usually small in comparison with the molecule free path, and the times t_{fl} and τ are much shorter than the characteristic relaxation times in gases. Therefore Eq. (23) can be used as a boundary condition corresponding to a quasi-stationary solution of Eq. (14) for the surface phase, where the number of particles is variable. In the steady state, which takes place under the condition $t \gg t_a$, the distribution functions f^+ and ρ_0 can be normalized (see Eqs. (19) and (20)) by equating the flux of molecules incident on and reflected from a surface.

Distributions like that given by Eq. (23) can be interpreted by distinguishing two channels of molecules interaction with a surface. Some of the molecules comes to equilibrium with the surface during transitions (accompanying the absorption and emission of phonons in the solid) between different states in the surface field, and others elastically scatter on their trajectories ($E>0$) without exchanging energy with the solid. The simplicity of our result is obviously due to the τ-approximation, which considerably simplifies the collisional integral.

There is another important point. Models with two components in distribution functions of molecules reflected from a surface have been used in the literature since Maxwell (the specular-diffuse boundary condition in rarefied gas dynamics[24,25]). In recent experiments with molecular beams, the two-component structure of the distribution of scattered molecules was actually detected in various systems.[26–32] The observed diffusion component is usually attributed to trapping of molecules in the potential well. This statement seems to be only partially true. Really, the diffuse component is important in the case of a deep potential well provided that adsorbed states are occupied by molecules. The function $S(\Gamma)$ in Eqs. (23) and (24), however, has the sense of the relaxation probability, and trapping of molecules in bound states is unnecessary since the rate τ^{-1} for molecular relaxation with $E>0$ can be associated, using Eq. (16), with transitions to both bound states ($E'<0$) and states above the potential well ($E'>0$). Another feature of our theory is that the presence of trajectories with $\Xi>2$ at $E>0$ (or short-lived collisional complexes as they are known) does not provide an alternative ''trapping'' channel, as suggested in some publications,[15] but only leads to an increase in the relaxation probability S. The latter is due to the longer transit time (12) for ''longer'' trajectories with $\Xi>2$. Note that this factor was mentioned in the discussion in Ref. 20 on a qualitative level.

3.4. The dependence of the ''trapping'' probability on parameters which determine the molecule state manifests itself in various effects related to the shape of the distribution of molecules interacting with a surface. They include, in particular, translational cooling of desorbed molecules[28,30,33] and deviation of the angular distribution of desorbed molecules from the cosine law,[33] as well as rotational cooling and rotational polarization of desorbed molecules, and rotational polarization of scattered particles.[3] These effects show up in the following way. The mean measured translational and rotational kinetic energies of desorbed molecules are less than their equilibrium values. The angular distribution of escaping particles has an excessive fraction of particles desorbed at large angles with respect to the normal. The angular momentum vectors **J** of desorbed molecules are aligned predominantly with the normal to the surface (this effect is traditionally termed helicopter-like rotational alignment of molecules[3]), whereas the vectors **J** of scattered molecules are aligned predominantly perpendicular to the normal (cartwheel alignment).

Before discussing the rules governing the observed effects, note that the ''trapping'' probability S can be calculated using Eqs. (11), (12), and (16) for t_{fl} and τ. Leaving aside calculations more sophisticated than the semiphenomenological theory under consideration, we limit our discussion to the qualitative interpretation based on the phenomenological analysis given above.

Usually the interaction potential depends weakly on the coordinates in the surface plane.[12] The parameters t_{fl}, τ, and S for linear molecules and a perfectly smooth surface depend largely on only two variables, namely v_z and J_t, where J_t is the projection of **J** on the surface. The effects of translational cooling and ''non-cosine'' desorption can be ascribed, in accordance with Eq. (24), to the relaxation probability, which decreases with the velocity component v_z. The decrease in S with increasing v_z is easily understandable, since the time of flight (12) obviously decreases with increasing velocity. Given Eq. (18) for the relaxation (trapping) probability, some experimental data were satisfactory described.[11,34]

The rotational cooling and predominance of the helicopter-like alignment of desorbed molecules can be attributed to the decreasing function $S(J_t)$. Moreover, the decrease in S with increasing J_t enables us to understand (see Eq. (23)) why the cartwheel-like polarization is predominant in the scattering channel. We previously,[16] analyzed the effects due to a rotational polarization of molecules. We assumed that the interaction parameters t_{fl} and τ were functions of the rotational state. These functions were derived from experimental data in order to predict results of other experiments. It turned out that, using Eq. (18) for the relaxation probability, we could give a consistent description of experimental data[35,36] concerning rotational cooling and polarization of desorbed molecules and also relate the temperature dependence of the rotational polarization in a very rarefied gas, obtained previously in studies of the effect of magnetic field on the heat flow in a rarefied polyatomic gas between two surfaces,[37] to the temperature dependence of the energy accommodation coefficient.

We acknowledge helpful discussions of this work with L. A. Maksimov.

APPENDIX A.

Here we give details concerning the description of a molecule's rotational motion in a potential that depends on the angles Ω that define the orientation of its principal axes.

1. For a diatomic (linear) molecule $\Omega=\{\vartheta,\phi\}$ are the polar and azimuthal angles of the molecule with respect to the surface. For a more complex molecule $\Omega=\{\vartheta,\varphi,\psi\}$ are the three Euler angles. We define them as follows. The x-, y-, and z-axes are fixed and attached to the surface, and x'-, y'-, and z'-axes are attached to the rotating molecule and aligned with the principal axes of the inertia tensor. The Euler angles ϑ,φ,ψ define, accordingly, the orientation of the z'-axis with respect to z, and the orientation of the line of nodes ζ which points in the direction of the vector product $\mathbf{n}_z\times\mathbf{n}_{z'}$, with respect to the x- and x'-axes; ϑ and $\phi=\varphi-\pi/2$ are the polar and azimuthal angles of the z'-axis with respect to x, y, and z. The definition of angles and axes is the same as in Fig. 16.2 of Ref. 38.

The angular momenta canonically conjugate to the Euler angles ϑ, φ, and ψ are projections of vector **J** on the ζ-, z-, and z'-axes.[39] The variable canonically conjugate to the angles ϑ and ϕ of a linear molecule are the projections J_ζ and J_z, respectively.

The element of phase space in the canonically conjugate variables has the form

$$d\Upsilon=d\Upsilon_{\rm tr}d\Upsilon_{\rm rot}, \tag{A1}$$

$$d\Upsilon_{\rm tr}=dx\ dy\ dz\ dp_x dp_y dp_z,$$

$$d\Upsilon_{\rm rot}=d\vartheta\ d\phi\ dJ_\zeta dJ_z \tag{A2}$$

for a linear molecule and

$$d\Upsilon_{\rm rot}=\sin^2\vartheta\ d\vartheta\ d\varphi\ d\psi\ dJ_\zeta dJ_z dJ_{z'} \tag{A3}$$

for a more complex molecule. The factor $\sin^2\vartheta$ is due to the angle between the fixed z-axis and rotating z'-axis.

2. Let us illustrate the transformation from variables γ to $\{\Gamma,\gamma_\sim\}$ taking as an example the states with $E>0$. The superscript "+" will be omitted in variables describing trajectories. For a linear molecule $\Gamma=\{\mathbf{p},\mathbf{J}\}$ are the momentum and angular momentum vectors of the rotating molecule, $\gamma_\sim=\{\varphi',x,y\}$, where φ' is the angle defining the projection of the molecular axis on the plane perpendicular to **J**. In this case

$$d\gamma=d\Gamma d\gamma_\sim, \tag{A4}$$

$$d\Gamma=dp_z dp_x dp_y J\ dJ\ dO_J,\quad d\gamma_\sim=d\varphi' dx\ dy;$$

the phase volume corresponding to variables $\gamma_\sim$ is equal to $\omega=2\pi s$, where s is the surface area element.

For a more complex molecule, $\{\Gamma,\gamma_\sim\}=\{\mathbf{p},\mathbf{J},\vartheta',\varphi',\psi',x,y\}$, where ϑ', φ', and ψ' are the Euler angles defining the axis aligned with **J**. In particular, for a symmetric top, $\Gamma=\{\mathbf{p},\mathbf{J},\vartheta'\}$ (angle θ' is constant because the molecule axis precesses around **J** when moving freely) and $\gamma_\sim=\{\varphi',\psi',x,y\}$. For a spherical top, $\Gamma=\{\mathbf{p},\mathbf{J}\vartheta',\psi'\}$ and $\gamma_\sim=\{\varphi',x,y\}$. In the new variables

$$d\gamma=dp_z dp_x dp_y J^2 dJ dO_J\ \sin\vartheta' d\vartheta' d\varphi' d\psi' dx\ dy. \tag{A5}$$

In the general case $d\gamma=d\Gamma d\gamma_\sim$, where the separation into $d\Gamma$ and $d\gamma$ has been given above. The phase volume $\omega=(2\pi)^3 s$ for a symmetric top and $\omega=2\pi s$ for a spherical one.

*)E-mail: pankov_a@yahoo.com
†)E-mail: borman@bvd.mephi.ru
‡)E-mail: krylov@redline.ru

1)When $E>0$, $\Xi>2$ may be associated with the existence of short-lived collisional complexes; see, for example, Ref. 20.

[1] J. A. Barker and D. J. Auerbach, Surf. Sci. Rep. **4**, 1 (1985).
[2] A. W. Kleyn and T. C. M. Horn, Phys. Rep. **199**, 192 (1991).
[3] D. C. Jacobs, K. W. Kolasinski, S. F. Shane, and R. N. Zare, J. Chem. Phys. **91**, 3182 (1989).
[4] M. C. Lin and G. Ertl, Ann. Rev. Phys. Chem. **37**, 587 (1986).
[5] D. S. King and R. R. Cavanagh, Adv. Chem. Phys. **76**, 45 (1989).
[6] R. W. M. Hoogeveen, R. J. C. Spreeuw, and L. J. F. Hermans, Phys. Rev. Lett. **59**, 447 (1987).
[7] V. D. Borman, L. A. Maksimov, B. I. Nikolaev, and V. I. Troyan, Zh. Éksp. Teor. Fiz. **64**, 526 (1973) [Sov. Phys. JETP **37**, 269 (1973)].
[8] J. J. G. M. Van der Tol, L. J. F. Hermans, S. Yu. Krylov, and J. J. M. Beenakker, Physica A **131**, 545 (1985).
[9] W. Brenig, in *Springer Ser. Surface Sci.* (1986), Vol. 8, p. 19.
[10] H. J. Kreuzer and Z. W. Gortel, *Physisorption Kinetics*, Springer, Berlin (1986).
[11] V. D. Borman, S. Yu. Krylov, A. V. Prosyanov, Zh. Éksp. Teor. Fiz. **94**(10), 271 (1988) [Sov. Phys. JETP **67**, 2110 (1988)].
[12] C. W. Muhlhausen, L. R. Williams, and J. C. Tully, J. Chem. Phys. **83**, 2594 (1985).
[13] D. C. Jacobs and R. N. Zare, J. Chem. Phys. **91**, 3196 (1989).
[14] T. Brunner, R. Brako, and W. Brenig, Phys. Rev. A **35**, 5266 (1987).
[15] C. Haug, W. Brenig, and T. Brunner, Surf. Sci. **265**, 56 (1992).
[16] V. D. Borman, S. Yu. Krylov, L. J. F. Hermans, and A. Yu. Pankov, in *Rarefied Gas Dynamics*, A. E. Beylich (ed.), VCH, Weinheim (1991), p. 1434.
[17] W. L. Schaich, J. Chem. Phys. **60**, 1087 (1974).
[18] C. Heer, *Statistical Mechanics, Kinetic Theory and Stochasic Processes*, Academic Press, New York (1972).
[19] L. D. Landau and E. M. Lifshitz, *Mechanics*, Pergamon Press, New York (1974).
[20] R. B. Gerber, Chem. Rev. **87**, 29 (1987).

[21] Yu. Kagan and A. M. Afanas'ev, Zh. Éksp. Teor. Fiz. **41**, 1536 (1961) [Sov. Phys. JETP **14**, 1096 (1961)].
[22] L. A. Maksimov and A. V. Il'in, in *Lectures on Physical Kinetics*, MIPT, Moscow (1974), p. 33.
[23] H. A. Kramers, Physica **7**, 284 (1940).
[24] J. C. Maxwell, *The Scientific Papers of James Clerk Maxwell*, W. D. Niven (ed.), Dover, New York (1972), Vol. 2, p. 706.
[25] J. Ferziger and H. Kaper, *Mathematical Theory of Transport Processes in Gases*, North-Holland, Amsterdam (1972).
[26] J. E. Hurst, C. A. Becker, J. P. Cowin *et al.*, Phys. Rev. Lett. **43**, 1175 (1979).
[27] K. C. Janda, J. E. Hurst, C. A. Becker *et al.* J. Chem. Phys. **72**, 2403 (1980).
[28] W. L. Guthrie, T.-H. Lin, S. T. Ceyer, and G. A. Somorjai, J. Chem. Phys. **76**, 6398 (1982).
[29] K. C. Janda, J. E. Hurst, J. P. Cowin *et al.*, Surf. Sci. **130**, 395 (1983).
[30] F. Budde, A. Modl, A. V. Hamza *et al.*, Surf. Sci. **192**, 507 (1987).
[31] Y. Matsuo, H. Vach, M. Châtelet *et al.* J. Chem. Phys. **93**, 4368 (1990).
[32] J. L. W. Siders and G. O. Sitz, J. Chem. Phys. **101**, 6264 (1994).
[33] J. E. Hurst, L. Wharton, K. C. Janda, and D. J. Auerbach, J. Chem. Phys. **83**, 1376 (1985).
[34] A. Yu. Pankov, V. D. Borman, and S. Yu. Krylov, in *Europhysics Conference Abstracts, ECOSS 16*, Geneva (1996), Vol. 20F, WeAP50.
[35] M. Asscher, W. L. Guthrie, T.-H. Lin, and G. A. Somorjai, J. Chem. Phys. **78**, 6992 (1983).
[36] D. C. Jacobs, K. W. Kolasinski, R. J. Madix, and R. N. Zare, J. Chem. Phys. **87**, 5038 (1987).
[37] V. D. Borman and D. I. Talyzin, in *Proceedings of 9th All-Union Conference on Rarefied Gas Dynamics* (Sverdlovsk, 1987) [in Russian], Izdatel'stvo Ural'skogo Universiteta, Sverdlovsk (1988), Vol. 2, p. 67.
[38] I. I. Ol'khovskiĭ, *Course of Theoretical Mechnanics for Physicists* [in Russian], Nauka, Moscow (1970), p. 144.
[39] I. I. Ol'khovskiĭ, Yu. G. Pavlenko, and L. S. Kuz'menkov, *Problems in Theoretical Mechanics for Physicists* [in Russian], Moscow State University, Moscow (1977), p. 58.

Translation provided by the Russian Editorial office.

Synchronization of high-frequency vibrations of slipping phase centers in a tin whisker under microwave radiation

V. I. Kuznetsov and V. A. Tulin*)

Institute of Microelectronics Technology and High-Purity Materials, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

(Submitted 25 September 1997)
Zh. Éksp. Teor. Fiz. **113**, 1364–1375 (April 1998)

Current–voltage characteristics of a system with a variable number of slipping phase centers resulting from phase separation in a tin whisker under external microwave field with a frequency $\Omega/2\pi \cong 35$–45 GHz have been studied experimentally. Emergence and disappearance of steps with zero slope in a whisker's current–voltage characteristic at $U_{m/n}=(m/n)U_\Omega$, where m and n are integers and U_Ω is determined by Josephson's formula $\hbar\Omega=2eU_\Omega$, have been investigated. Microwave field generated by slipping phase centers is nonharmonic, and the system of slipping phase centers permits synchronization of internal oscillations at a microwave frequency by an external field with a frequency which is the n-th harmonic of internal oscillations. The estimated microwave power generated by a whisker is 10^{-8} W. Stimulation of superconductivity in a current-carrying whisker has been detected. © *1998 American Institute of Physics.* [S1063-7761(98)01804-6]

1. INTRODUCTION

Microwave generation in a Josephson junction (a weak-coupling element in a superconducting circuit) under a dc voltage has attracted researchers' attention since the time when the ac Josephson effect was discovered. The simple structure of the experimental device and easy control of the generated frequency are the most attractive features of the effect. The frequency generated by the junction is determined by the formula

$$\omega = 2Ue/\hbar,$$

where U is the voltage drop across the junction, e is the electron charge, and $\hbar$ is the Planck constant. The disadvantages of these devices are their low output and difficulties in matching the superconducting circuits containing Josephson junctions to the microwave circuits. Attempts have been made to overcome these difficulties using circuits of short junctions.[1–4]

Josephson junctions have a typical linear size in the direction perpendicular to the supercurrent density vector, namely the Josephson penetration depth λ_j. If the junction dimension in the direction perpendicular to the supercurrent satisfies the condition $d<\lambda_j$ (a short junction), the phase variation is uniform over the junction volume, and one has a single source of microwave radiation. In the case of a network of short synchronized junctions, it seems possible to derive a high microwave output close to the sum of powers generated by each element.

A long uniform superconducting channel with slipping phase centers can be classified with such systems. Phase slipping centers occur in resistive states of a long narrow channel carrying a constant current at a temperature close to the superconducting transition ($I>I_c$, $T<T_c$).[5,6] Real structures in which phase slipping centers have been detected are thin films with width w and single-crystal wires (whiskers) with diameter d smaller than the superconductor coherence length ξ. From the viewpoint of experimenters dealing with superconducting channels, whiskers (thin crystalline wires) are preferable because their uniformity over the length is higher. But thin films have some advantages when applications are concerned, since their dimensions are directly controlled during their manufacture. On the other hand, microscopic inhomogeneities due to fabrication technologies can lead to considerable degradation of parameters of slipping phase centers, and a thin film may behave like a system of weak superconducting bounds localized along the narrow film.

An isolated slipping phase center is an nonstationary, inhomogeneous entity ''localized in the space'' and containing an internal region with a size of about ξ where the superconducting order parameter oscillates at the Josephson frequency $\omega=2eU_\omega/\hbar$. At temperatures near the transition point, the voltage averaged over the oscillation period, U_ω, in the slipping phase center is due to penetration of a nonuniform longitudinal electric field into the outer region of the center through a distance of about l_E (the electric field penetration range), and the electric resistance of each phase slipping center is $R_0=2\rho_N l_E/S$, where ρ_N is the material resistivity in the normal state and S is the channel cross section.[5–8] At the moment when the absolute value of the order parameter vanishes, the phase difference over the center jumps by 2π. Current–voltage characteristics (CVC) of such superconducting channels contain a set of sloped linear sections corresponding to resistances $R_n=nR_0$ (where n is an integer) connected by sections of curves with current jumps. Extrapolations of these linear sections cross the current axis at approximately the same point I_0 (an excess current).[5–7,9]

Although the number of publications dedicated to slipping phase centers is fairly large,[5,6] the dynamics of systems

with slipping phase centers has been studied insufficiently.[7,8,10–12] The reversed ac Josephson effect under external electromagnetic radiation was detected in thin tin films at a frequency of 10 GHz[7] and in single-crystal wires (whiskers) at frequencies of up to 900 MHz.[13–15] In both these cases, a CVC contains, in addition to sloping steps, a fundamental step with a zero slope at voltage U_Ω in the region of parameters corresponding to one phase slipping center and associated with high-frequency oscillations of the order parameter in the center, and ''weak'' steps at $U_{m/n}=(m/n)U_\Omega$, where U_Ω is the voltage corresponding to the external field frequency and m and n are integers. Ivlev and Kopnin[16] analyzed the ac Josephson effect in terms of the microscopic theory. The pattern of various zero-slope steps at different direct currents and microwave frequencies in systems with variable numbers of slipping phase centers has not been investigated in full. A current-carrying whisker under an electromagnetic field with a frequency higher than 900 MHz has never been studied.

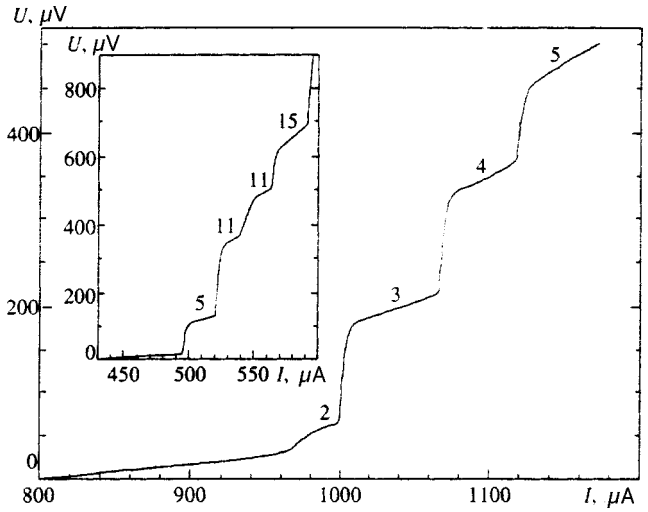


FIG. 1. CVC of the Sn3 whisker ($R_0\approx0.19\ \Omega$, $T_c\approx3.71$ K, $R_{300}/R_{4.2}\approx73$) without irradiation by an external microwave field at $T\approx3.56$ K. The insert shows the CVC of the Sn2 sample ($R_0\approx0.21\ \Omega$, $T_c\approx3.72$ K, $d\approx0.8\ \mu$m, $R_{300}/R_{4.2}\approx50$) without irradiation at $T\approx3.63$ K.

2. SAMPLES AND EXPERIMENTAL DETAILS

In the reported work, we have studied the effect of microwave fields with frequencies ranging between 35 and 45 GHz on CVCs of tin whiskers in the regime when several slipping phase centers exist in a sample at voltages of order of U_Ω. In previous experiments[7,13–15] the parameters T and $\Omega/2\pi$ were selected so that the mean voltage U_ω across one center, which determined the frequency of proper high-frequency oscillations, $\omega=2Ue/\hbar$, could be tuned to U_Ω, i.e., the frequency ω of internal oscillations should be equal to that of applied microwave field. We have used higher microwave frequencies $\Omega/2\pi$ and temperatures at a greater distance from T_c than Tidecks et al.[13–15] so that to satisfy the condition $U_\omega=U_\Omega/n$, i.e., $n\omega=\Omega$ $(n>1)$ when a sample contained several slipping phase centers at voltages about U_Ω. It follows from the microscopic theory[16] that this is the condition under which induced steps on a CVC are generated at voltages U_Ω/n. The presence of such steps means that the radiatition generated by the system of slipping phase centers is nonharmonic. Given the higher uniformity of whiskers over their lengths and smaller number of structural defects than in films, they are preferable for such experiments. Moreover, zero-slope steps on a CVC of an irradiated whisker[13–15] are considerably wider than in narrow films.[7] In many experiments (see for example Ref. 17), low-frequency oscillations instead of high-frequences oscillations were detected in narrow films. Whiskers grown from thin tin films deposited on silicon substrates had diameters $d=0.2$–0.8 μm, lengths of about 1 mm, resistance ratio $R_{300}/R_{4.2}<100$, and $T_c\approx3.7$ K. A whisker was set across a 300-μm gap in a thin tin film about 1000 Å thick. A whisker was attached to electrodes by electrostatic forces at the initial moment, then, apparently, by the Van der Waals forces. It is not easy to remove a whisker from the substrate surface. The heat-sinking conditions, probably, were fairly good because the greater part of the sample was in contact with the polished substrate surface, therefore measured CVCs did not exhibit a notable hysteresis in the studied temperature range, unlike CVCs reported in Refs. 13–15. CVCs were measured using either the two-terminal configuration (this was possible because T_c of films was higher than that of whiskers) or the four-terminal configuration. The substrate supporting the whisker was placed in a copper waveguide and insulated from environment by a superconducting lead shield. The curves of the critical current and resistance versus temperature for the case of a single slipping phase center at $T_c-T<10$ mK had shapes typical of whiskers: $I_c\sim(1-T/T_c)^{3/2}$, $R_0\sim(1-T/T_c)^{-1/4}$.[13]

3. EXPERIMENTAL RESULTS

Current–voltage characteristics of all samples are piecewise linear, i.e., they are composed of linear sections connected by nonlinear sections with larger slopes. Figure 1 shows the examples of CVCs of superconducting whiskers with microwave radiation off. The initial parts of the whisker CVCs without microwaves are also shown in Figs. 2–4 by dashed lines. The numbers near the linear sections of the whisker CVCs indicate the ratios between their resistances and that of a single phase slipping center, R_0. The latter parameter was determined as the largest common divisor of differential resistance values of all linear CVC sections and compared to an estimate derived from the size and resistivity of the whisker. The CVC linear sections are connected by nonlinear sections, which are reproducible and reversible in the range of studied frequencies. Note that in most experiments, the initial CVC sections at $I>I_c$ (curves (a) in Figs. 2 and 3) without radiation are nonlinear, and the first reproducible linear sections correspond to states with several slipping phase centers (the linear section $3R_0$ on curve (a) in Fig. 2 and $5R_0$ on curve (a) of Fig. 3). In earlier experiments,[13–15] the states with one slipping phase center could be regularly produced. In contrast to those experiments, where the temperature difference T_c-T was less than 10 mK, we measured CVCs mostly at temperatures 70–160 mK below T_c. In this case, states with several slipping phase centers were

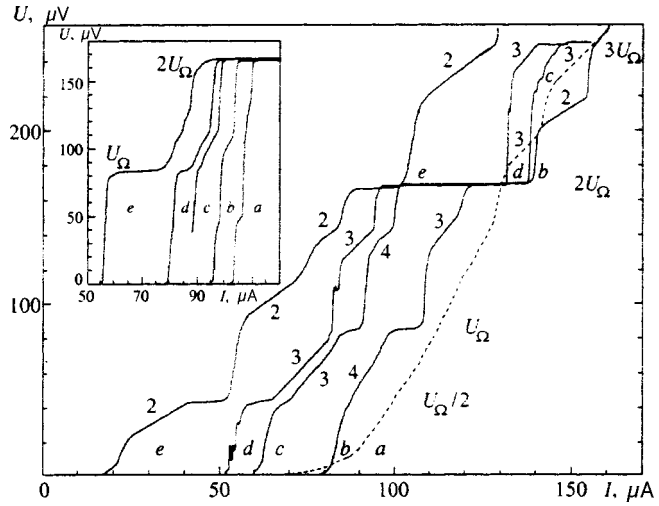


FIG. 2. Set of CVCs of the Sn1 sample ($R_0=0.79$–$0.63\ \Omega$, $T_c\approx3.69$ K, $d\approx0.3\ \mu$m, $R_{300}/R_{4.2}\approx20$) at different powers of external microwave irradiation at frequency $\Omega/2\pi=40.62$ GHz and $T\approx3.62$ K: (a) 70 dB (dashed line); (b) 36 dB; (c) 31 dB, (d) 30 dB, (e) 28.6 dB. The insert shows low-current sections of CVCs of the Sn1 sample at approximately equal parameters in another cycle of measurements: (a) 32.6 dB; (b) 30.6 dB; (c) 29 dB; (d) 28 dB; (e) 25.2 dB.

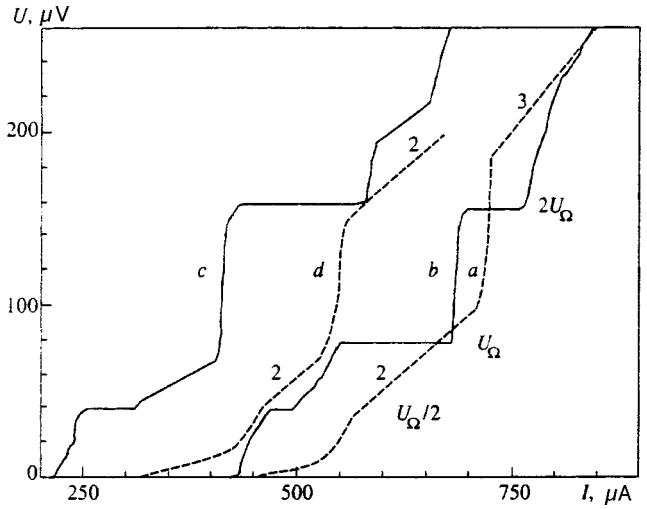


FIG. 4. CVC of the Sn3 whisker ($R_0=0.21\ \Omega$, $T_c\approx3.71$ K, $R_{300}/R_{4.2}\approx73$) at different microwave powers at frequency $\Omega/2\pi=37.5$ GHz at $T\approx3.58$ K (curves a, b, and c) and $T\approx3.63$ K (curve d); (a) 70 dB (dashed line); (b) 19.5 dB; (c) 12 dB; (d) 70 dB (dashed line).

stable at notably larger temperature differences T_c-T. This can be seen by comparing the CVC shown in Fig. 1 with the CVC in the insert to this graph. Moreover, the linear sections with the same resistance (such as $3R_0$ sections on curve (a) in Fig. 2 for Sn1 and $2R_0$ in curve (d) in Fig. 4 for Sn3) separated by voltage jumps were recorded many times. On the basis of these observations, we have come to the conclusion that, in spite of some complications in interpreting CVCs of our whiskers, they are superconducting channels with slipping phase centers at appropriate temperatures and transport currents.

When samples are exposed to microwave radiation, their CVCs contain, in addition to linear sloping sections due to the presence of certain numbers of slipping phase centers, steps with a zero slope at voltages $U_{m/n}=(m/n)U_\Omega$. At low microwave powers, the channel critical current was higher, i.e., stimulation of superconductivity theoretically described by Eliashberg[18] took place. Instead of the emergence of the zero-slope step first at U_Ω,[13–15] we observed the sequential appearance of steps at $2U_\Omega$, U_Ω, $3U_\Omega$, and $U_\Omega/2$ for a sample Sn1 (Fig. 2), and in Sn2 (Fig. 3) we first observed a step at $U_\Omega/2$ and then at U_Ω (not shown in the graph). At lower temperatures the unusual shapes of the CVCs at zero radiation intensity with linear sections of the same slope (curve (d) in Fig. 4) or the lowest linear sections corresponding to several slipping phase centers were replaced by more common CVC shapes. The sequence of microwave-induced steps in whiskers' CVCs emerging with increasing microwave power also became more like the usual sequence at lower temperatures, namely, the step at U_Ω was detected first, then the step at $2U_\Omega$, and at still higher microwave power at $U_\Omega/2$ (Fig. 4). The curve became similar to those given in Refs. 13–15. As the microwave power increased, the sloping linear sections due to the slipping phase centers became more pronounced on CVCs (Figs. 2 and 3).

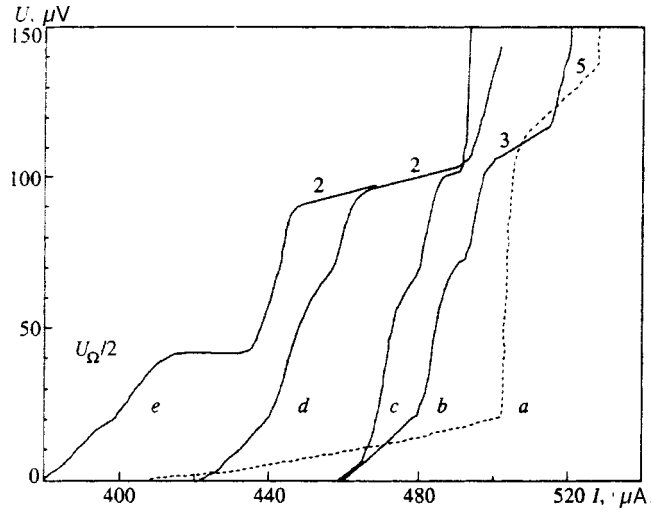


FIG. 3. Set of CVCs of the Sn2 sample ($R_0=0.23$–0.18Ω, $T_c\approx3.72$ K, $d\approx0.8\ \mu$m, $R_{300}/R_{4.2}\approx50$) at different powers of microwaves at frequency $\Omega/2\pi=40.62$ GHz at $T\approx3.63$ K: (a) 70 dB (dashed line); (b) 30.2 dB; (c) 30 dB; (d) 28.2 dB; (e) 26.1 dB.

Steps with zero slope emerge on linear sections of CVCs, which either exist in the samples not exposed to microwaves or appear in the samples irradiated by the microwave field. For example, the step of zero slope on the CVC of the Sn1 whisker at U_Ω (insert to Fig. 2) appears after the emergence of a linear section on the curve, its growth, and the shift of its lower edge to the required voltage (curves (a), (b), and (c) in the insert to Fig. 2). As soon as the edge of the linear section achieves U_Ω, a zero-slope step is produced (curve (d), and its width increases with the microwave power curve (e)). The steps at $U_\Omega/2$ (curves (c), (d), and (e) in Fig. 2) emerge in a similar manner. The step at $3U_\Omega$ (curve (c) in Fig. 2) appears when the sloping linear section with differential resistance $3R_0$ extends to this region. A zero-slope step can disappear at a higher microwave power (for ex-

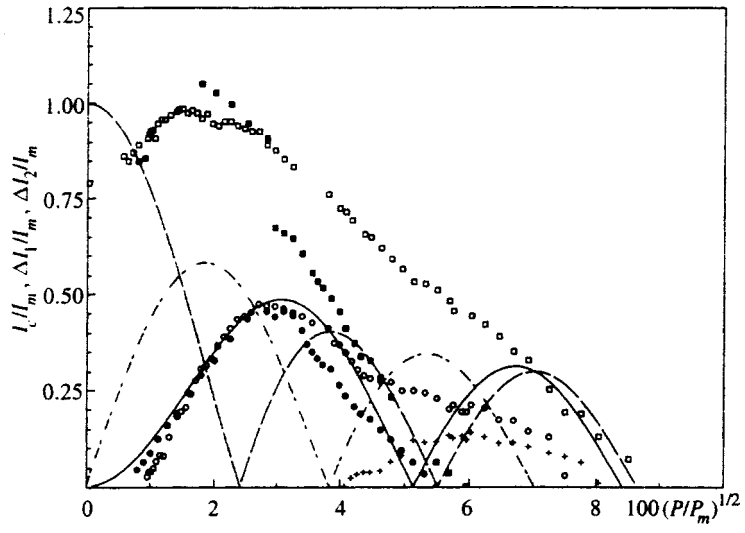


FIG. 5. Normalized critical current I_c/I_m versus the relative amplitude of external microwave field at frequency $\Omega/2\pi=40.62$ GHz at $T\approx3.62$ K for the Sn1 whisker in two different cycles of measurements (full squares are the data of the first cycle and empty squares correspond to the second cycle), $I_m\approx107$ μA. Normalized widths of zero-slope steps on a CVC as functions of relative microwave field amplitude at voltage U_Ω (crosses plot data of the second cycle) and at voltage $2U_\Omega$ (full circles are the data of the first cycle and empty circles correspond to the second cycle). The dashed, dash-dotted, and solid lines show absolute values of Bessel functions $J_0(x)$, $J_1(x)$, and $J_2(x)$, respectively; $x=100(P/P_m)^{1/2}$, P is the power, and P_m is the maximal output of the microwave generator.

ample, the steps at $3U_\Omega$, U_Ω, and $U_\Omega/2$) when the upper edge of the linear section shifts below the respective voltage, and a vertical CVC section moves to this region. In this case, the differential resistance of the linear section can have a jump (curves (d) and (e) in Fig. 2), namely, the linear section at about $3U_\Omega$ changed its factor from 3 to 2. Thus, a linear section on a CVC of a sample with or without microwave pumping at $U_{m/n}$ is a necessary condition for formation of a zero-slope step, i.e., for the existence of the required number of slipping phase centers in the sample.

By tuning the incident microwave frequency Ω, we could detect zero-slope steps not observed previously when voltage $U_{m/n}$ coincided with a linear section of a CVC recorded without irradiation.

Sloping linear sections in a CVC of a whisker containing a certain number of slipping phase centers and exposed to microwaves of a fixed power could decrease their resistance factor with respect to the resistance of an isolated slipping phase center if the direct transport current increased (see curve (b) in Fig. 2). The resistance factor could also remain unchanged (curve (e) in Fig. 2, section 2). An increase in the incident microwave power could cause, in addition to suppression of both the critical and excess current at a fixed voltage, a switch-over to a linear section with a lower differential resistance. On the curves in Fig. 2, the resistance factor dropped from four to two, and in Fig. 3 from five to two. The CVCs of the Sn2 whisker (Fig. 3) initially contained a linear section with resistance $5R_0$ at voltages above U_Ω, and under microwave irradiation this parameter dropped to $3R_0$ and then $2R_0$. At higher microwave powers the length of the $2R_0$ section increased at a constant resistance factor. Note that R_0 could vary under microwave radiation within 20%. Thus, microwaves not only produce horizontal steps on CVCs, but also strongly affect CVCs of tin whiskers.

We have also measured the widths of microwave induced steps as functions of the incident power over the interval of their existence. The experimental dependencies of current-normalized widths of zero-slope steps at voltages U_Ω and $2U_\Omega$, and of the critical current for a sample Sn1 obtained in different measurement cycles at approximately equal parameters as functions of the relative microwave amplitude are given in Fig. 5. The graph also shows as an illustration the absolute values of Bessel functions of order 0, 1, and 2 ($J_0(x)$, $J_1(x)$, and $J_2(x)$) although we believe that the experimental curves are not directly related to these functions. Note the main features of the curves in Fig. 5. (1) The microwave stimulation of superconductivity led to an increase in the critical current of about 20%. (2) The zero-slope step at $2U_\Omega$ emerged at a lower microwave power and had the maximum width of about $0.5I_c$. (3) The step at U_Ω observed in the second cycle of measurements (it was too small in the first cycle and its width is not shown in Fig. 5) appeared at a higher microwave power, and in its presence the width of the $2U_\Omega$ step and the critical current as functions of the microwave field amplitude changed considerably. In this case the critical current and width of the $2U_\Omega$ step vanished at a notably higher microwave field amplitude than in the first cycle. (4) Induced U_Ω and $2U_\Omega$ steps appeared at a finite microwave power, i.e., there is a certain threshold microwave power needed for formation of these steps. This threshold is related to the extension of the linear CVC sections to voltages U_Ω and $2U_\Omega$. (5) There is only one interval of the microwave field amplitude on which the critical current and CVC steps exist. No oscillations have been detected on the curves of critical current and step width.

In studying step widths as functions of the microwave power, we recorded (in several cases) nonmonotonic curves with relatively narrow down-peaks against the background of wide bell-shaped curves.

4. DISCUSSION OF RESULTS

The current–voltage characteristic of a uniform superconducting channel, which is our model for a whisker, depends on its length. In the case of a short whisker section through which current is fed, $l\approx l_E$, the presence of one slipping phase center allows the sample to conduct a current higher than the critical value. If $l\gg l_E$, the exponentially decaying parameters of slipping phase centers have little effect on the channel properties, therefore it should contain several slipping phase centers, whose number is determined by the channel length. In our samples, the condition $l\gg l_E$

was satisfied ($l \approx (10-20) l_E$), therefore we assume that several slipping phase centers were necessary to conduct a current slightly higher than the critical value. In the process of generation of the required number of centers, the instantaneous number of centers can be unstable and variable in both time and space.

The CVCs of our samples have piecewise linear shapes with sections characterized by differential resistance $R = nR_0$, where n is an integer. These sections correspond to definite numbers of slipping phase centers, which can be derived from the sample sizes. In addition, there are the nonlinear sections on which the number of centers is probably unstable and varies with time. A dedicated investigation is needed to verify this hypothesis. The linear sections of CVCs of the superconducting channel in the simplified model[7] are described by the formula $U = nR_0(I - I_0)$. The excess current I_0 is usually related to the average superconducting component of the total current. This formula is not universal for all linear sections, because $I_0 \neq \text{const}$ for all groups of linear sections.[9] The CVCs of our samples contain neighboring linear sections with equal n but different I_0.

Microwave irradiation of our samples has a dual effect on their CVCs. The first effect is the generation of constant voltage steps, which was the main subject of the reported study. The second effect is the change in the number of slipping phase centers under microwave radiation and stabilization of CVC sections with definite numbers of these centers. This shows up in the extension of linear sections and transformation of some nonlinear CVC portion to linear.

The existence of constant-voltage steps under microwave radiation indicates that there are currents of microwave frequencies with spectral components $\omega = 2enU_\omega/\hbar$ ($n = 1,2,3, \ldots$) in the regions of slipping phase centers. When the external frequency equals that of one of these harmonics, several centers are synchronized, which shows up in the form of constant-voltage steps at $U_m = mU_\omega$, where m is the number of slipping phase centers, $U_\omega = \hbar\Omega/2en$, and Ω is the external radiation frequency. As a result, steps can occur at $U_{m/n} = (m/n)\hbar\Omega/2e$ if this voltage coincides with an inherent or microwave-induced linear section of CVC with a definite (integral) number of slipping phase centers.

Unfortunately, it is difficult to determine m and n with certainty using CVCs. Linear sections from which the number of centers could be exactly determined could be seen near constant-voltage steps only at certain values of parameters. We believe that the step at $U_\Omega/2$ in Fig. 2 is due to the synchronization of two slipping phase centers by the fourth harmonic of proper oscillations, i.e., $U_\Omega/2 \to 2U_\Omega/4$, similarly $U_\Omega \to 4U_\Omega/4$, $2U_\Omega \to 4U_\Omega/2$, $3U_\Omega \to 6U_\Omega/2$, the step at $4U_\Omega \to 8U_\Omega/2$ is not shown; in Fig. 4 $U_\Omega/2 \to 2U_\Omega/4$, $U_\Omega \to 2U_\Omega/2$, and $2U_\Omega \to 4U_\Omega/2$; in Fig. 3, $U_\Omega/2 \to 2U_\Omega/4$; in Fig. 6, $U_\Omega/2 \to 3U_\Omega/6$, $2U_\Omega/3 \to 4U_\Omega/6$, and $U_\Omega \to 6U_\Omega/6$. At other values of parameters this sample demonstrated steps at $5U_\Omega/6 \to 5U_\Omega/6$ and $U_\Omega/3 \to 2U_\Omega/6$ (not shown in the graphs of this paper).

Thus, at certain positions of these voltages in the whisker CVCs, microwave field synchronizes oscillations of the order parameter in all slipping phase centers present in a sample, which results in constant voltage drops across isolated centers and across the entire sample. States with synchronized slipping phase centers under microwave radiation emerge predominantly at corresponding locations in the CVCs. Other CVC sections may correspond to states in which some slipping phase centers are synchronized by external field and the rest are not. This conjecture allows us to interpret the drop in the differential resistance of linear sections (and the behavior of the differential resistance in general) when the current increases under microwave radiation. The existence of neighboring sloping steps with equal resistance but different excess current can also be interpreted in these terms. A similar effect without microwave radiation can be attributed to a different but, in a sense, similar phenomenon. So-called Fiske steps[19] were detected in experiments with tunneling Josephson junctions when the frequency generated by the junction was locked to the resonant frequency of the structural cavity in the experimental device. In this case, constant-voltage steps determined by the Josephson formula with the resonant cavity frequency could be seen on CVCs. The gap in the tin film on which the whisker was mounted could act as a structural resonator. The length of this gap was about 5 mm, and, given the silicon substrate dielectric constant ($\varepsilon \approx 12$), we have a resonant frequency in the studied microwave band. In this case, a section with a constant voltage due to synchronization of a group of slipping phase centers (Fiske step) can occur. The centers whose oscillations are not locked to the resonant frequency should demonstrate a linear behavior. As a result, the CVC of the sample should have a linear section with the resistance corresponding to the number of unlocked centers, which is smaller than the total number. The question why horizontal steps have not been observed remains unanswered. Doubts in this interpretation could be eliminated by directly measuring microwaves generated in the sample.

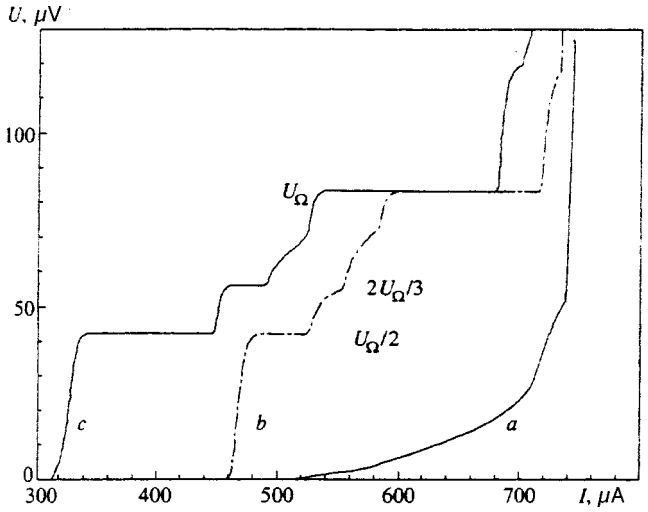


FIG. 6. Low-current sections of CVCs of the Sn2 sample at different microwave field powers at frequency $\Omega/2\pi = 40.62$ GHz in the second cycle of measurements at $T \approx 3.60$ K: (a) 70 dB; (b) 24.49 dB; (c) 22 dB.

Figure 5 shows the widths of constant-voltage steps as functions of the microwave field amplitude in relative units.

The maximal width of these steps allows us to estimate the microwave power generated by the whisker:

$$P \approx (\Delta I)^2 m R_0,$$

where ΔI is the step width in terms of current. Hence, $P \approx 10^{-8}$ W.

The microwave generation in the slipping phase centers can be interpreted in terms of the order parameter versus time, which vanishes at some moment and then increases to some value. At the moment when the order parameter is zero, the difference between the phases on different sides of the slipping phase center drops by 2π. It would be interesting to estimate the times of these processes and compare their reciprocal values with the frequencies of the order parameter oscillations and external radiation. The most important parameter is the time τ_Δ in which the order parameter recovers. When τ_Δ is much longer than the order parameter oscillation period determined by the Josephson formula, both the mean and instantaneous absolute values of the order parameter within the center are much smaller than the equilibrium value in other regions of the superconducting channel. If τ_Δ is comparable to or smaller than the period of the order parameter oscillations, the instantaneous value of the gap in the slipping phase center can be large and comparable to the gap in the surrounding regions. The spectra of normal excitations in slipping phase centers should be notably different in these two cases, which can lead to differences in some electrical properties of slipping phase centers. Since the energy relaxation time of current carriers in tin is 3×10^{-10} s, the first case is realized in the microwave frequency band.

The behavior of the step width is determined by two factors. The first is the width of the step against the background of an infinite linear CVC section with a definite number of slipping phase centers as a function of the microwave field amplitude. The second is the limitation of the constant-voltage step by the length of the CVC linear sloping section, whose positions, as follows from experimental data, are also functions of the microwave power. A change in the number of slipping phase centers breaks the initial synchronization condition, and the system can switch to either a totally unsynchronized state, or a partially synchronized state, or fully synchronized state at a different harmonic and with a different number of slipping phase centers (for example, the zero-slope step at $2U_\Omega$ in Fig. 2 can be due to synchronization of four centers by the second harmonic or six centers by the third harmonic). Given these two effects, we could not determine the constant-voltage step width as a function of the microwave amplitude unambiguously and compare it to the theoretical model. The existence of the microwave power threshold at which induced steps appear and the absence of oscillations in both the zero-slope step width at $U_{m/n}$ and critical current as functions of the microwave field amplitude are due to a definite number of slipping phase centers required at these voltages.

5. CONCLUSION

In the reported work, we have studied the effect of microwave radiation on current–voltage characteristics of whiskers with submicron diameters. Such whiskers can serve as microwave oscillators at frequencies of up to 40 GHz with an output of about 10^{-8} W. The spectrum of generated waves contains many harmonics, and the generation occurs on CVC sections with stable numbers of slipping phase centers. Features of CVCs of our samples under microwave radiation are determined by changes in the number of slipping phase centers and the synchronization degree of generation in these centers.

The work was supported by the *Superconductivity* subprogram of the *Physics of Condensed State* program sponsored by the Russian government (Project No. 95021), and by the *Physics of Solid-State Nanostructures* program (Project No. 1-084/4).

*)E-mail: tulin@ipmt-hpm.ac.ru

[1] D. W. Palmer and J. E. Mercereau, Appl. Phys. Lett. **25**, 467 (1974).
[2] M. Octavio and W. J. Skocpol, J. Appl. Phys. **50**, 3505 (1979).
[3] L. É. Amatuni, V. N. Gubankov, A. V. Zaitsev, and G. A. Ovsyannikov, Zh. Éksp. Teor. Fiz. **83**, 1851 (1982) [Sov. Phys. JETP **56**, 1070 (1982)].
[4] L. É. Amatuni, V. N. Gubankov, and G. A. Ovsyannikov, Fiz. Nizkikh Temp. **9**, 939 (1983) [Sov. J. Low Temp. Phys. **9**, 484 (1983)].
[5] B. I. Ivlev and N. B. Kopnin, Usp. Fiz. Nauk **142**, 435 (1984) [Sov. Phys. Usp. **27**, 206 (1984)].
[6] R. Tidecks, *Current-Induced Nonequilibrium Phenomena in Quasi-One-Dimensional Superconductors*, in *Springer Tracts in Modern Physics*, Vol. 121, Springer (1990).
[7] W. J. Skocpol, M. R. Beasley, and M. Tinkham, J. Low Temp. Phys. **16**, 145 (1974).
[8] S. M. Gol'berg, N. B. Kopnin, and M. I. Tribel'skii, Zh. Éksp. Teor. Fiz. **94**(4), 289 (1988) [Sov. Phys. JETP **67**, 812 (1988)].
[9] J. Meyer and G. Minnigerode, Phys. Lett. A **38**, 529 (1972).
[10] J. D. Meyer and R. Tidecks, Solid State Commun. **24**, 639 (1977).
[11] M. Tinkham, J. Low Temp. Phys. **35**, 147 (1979).
[12] X. Yang and R. Tidecks, Z. Phys. B **83**, 113 (1991).
[13] R. Tidecks and G. von Minnigerode, Phys. Status Solidi A **52**, 421 (1979).
[14] R. Tidecks and G. Slama, Z. Phys. B **37**, 103 (1980).
[15] B. Damaschke and R. Tidecks, Z. Phys. B **77**, 17 (1989).
[16] B. I. Ivlev and N. B. Kopnin, Solid State Commun. **41**, 107 (1982).
[17] G. E. Churilov, V. M. Dmitriev, and V. N. Svetlov, Fiz. Nizk. Temp. **9**, 495 (1983) [Sov. J. Low Temp. Phys. **9** 250 (1983)].
[18] G. M. Éliashberg, JETP Lett. **11**, 114 (1970).
[19] M. D. Fiske, Rev. Mod. Phys. **36**, 221 (1964).

Translation was provided by the Russian Editorial office.

Conductance of a quantum wire in a longitudinal magnetic field

V. A. Geĭler*), V. A. Margulis, and L. I. Filina

N. P. Ogaryov Mordovian State University, 430000 Saransk, Russia
(Submitted 2 July 1997)
Zh. Éksp. Teor. Fiz. **113**, 1376–1396 (April 1998)

We examine the ballistic conductance of a quantum wire in a parallel magnetic field in the presence of several impurities and derive analytic expressions for the transmission coefficient and the conductance in such a system. We show that scattering by impurities leads to a number of sharp peaks near the thresholds of the conductance quantization steps. The number of such peaks is determined by the distance between the impurities and the energy of the scattered particle. We also study the conductivity of a quantum wire in the region where the transport mechanism is diffusive. The conductivity is examined for the case in which charge carriers are scattered by randomly distributed point impurities. We study Shubnikov–de Haas oscillations in such a system. The general oscillation pattern consists of broad minima separated by irregularly spaced sharp peaks of the burst type. © *1998 American Institute of Physics.* [S1063-7761(98)01904-0]

1. INTRODUCTION

The conductance of an electron gas in quasi-one-dimensional nanostructures has attracted growing attention in view of the fact that systems of this type exhibit interesting physical effects, such as the Aharonov–Bohm effect, the weak localization effect, and the quantization of conductance. A magnetic field **B** applied along the axis of a quantum wire enhances lateral geometric confinement,[1,2] so that by varying **B** we can alter the effective geometrical size of the system, and hence the functional dependence of the conductance on the field **B**. In particular, by varying **B** we can change the parameters of the conductance quantization steps.[3,4] Studying the conductance of a quantum wire in a lateral magnetic field makes it possible to investigate such important characteristics of electrons as the parameters of the confining potential and the subband energies in the energy spectrum.[5,6]

Two fundamentally distinct cases are possible in studies of the conductance of quantum wires. If the wire's radius and length are less than the electron mean free path, the conductance is ballistic and can be expressed in terms of appropriate transmission coefficients.[7–9] Otherwise, the conductance is diffusive and can be described by an appropriate transport equation.

Elastic scattering by impurities plays an important role in such systems.[9–20] In particular, as shown in Ref. 4, scattering by a single point impurity in a ballistic quasi-one-dimensional nanostructure in the presence of a longitudinal magnetic field **B** leads to the emergence of narrow and high peaks to the left of the threshold of the conductance quantization steps. As the Landauer–Büttiker formula[7–9] implies, the wire conductance in the ballistic regime is described by the sum of the electron transmission coefficients over the various scattering channels. These coefficients change in the presence of an impurity due to the possibility of reflection of the incident electron wave, which depends on the parameters of the scattering center. The effect may be even stronger if there are several impurities. General considerations suggest that since each impurity can reflect the incident electron wave, variations in the relative positions of the impurities can have a strong effect on the shape of the conductance curve $G(E)$. Moreover, due to multiple reflections, the number of peaks in the G vs. E curve may be high even for two impurities. In the diffusive regime, scattering by impurities leads to oscillations of the Shubnikov–de Haas type in the dependence of the conductivity σ on the field **B**.

The goal of our investigation is to study the conductance of a quantum wire in a longitudinal magnetic field when carriers are scattered by point impurities in both the ballistic regime and the diffusive transport regime.

Various models of the confining potential have been used to describe theoretically the one-electron states in a quantum wire: an infinitely long waveguide with constant cross section,[21,22] saddle-point potentials for constrictions in quantum channels,[7–11] and a symmetric quadratic potential.[23–25] A more rigorous approach to the shape of the confining potential requires self-consistent solution of the Poisson equation and the Schrödinger equation. Numerical solutions of these equations yield an almost parabolic potential for channels in a two-dimensional gas with, however, a flat lower boundary (the lower part being cut off[26]). Since such a shape is very close to parabolic, a harmonic potential is quite realistic from an experimental standpoint.[23] The convenience of the harmonic potential for theoretical studies of quasi-one-dimensional microstructures in a longitudinal magnetic field is explained by the fact that such a potential yields explicit formulas for the spectral characteristics and the scattering parameters.

To describe one-electron states in a quantum wire we use a symmetric confining potential $V(x,y)=m^*\omega_0^2(x^2+y^2)/2$, where m^* is the effective electron mass and ω_0 is the characteristic confining-potential frequency, which is related to the effective radius l_0 of the wire by l_0

$=\sqrt{\hbar/m^*\omega_0}$. For one-electron states that are unperturbed by impurities and are in a longitudinal magnetic field $\mathbf{B}\|z$, the Hamiltonian of our model is

$$H_0=\frac{1}{2m^*}\left(\mathbf{p}-\frac{e}{c}\mathbf{A}\right)^2+\frac{m^*\omega_0^2}{2}(x^2+y^2). \tag{1}$$

We select a symmetric gauge for the vector potential of the magnetic field, $\mathbf{A}(\mathbf{r})$, i.e., $\mathbf{A}=(-yB/2,\ xB/2,0)$. Then the spectrum of the Hamiltonian (1) can be written

$$E_{mnp}=\frac{\hbar\omega_c}{2}m+\frac{\hbar\Omega}{2}(2n+|m|+1)+\frac{p^2}{2m^*},$$

and the corresponding wave functions in cylindrical coordinates ρ, φ, and z are

$$\Psi^0_{mnp}=\exp\left(\frac{ipz}{\hbar}\right)\frac{\exp(im\varphi)}{\sqrt{2\pi}}R_{mn}(\rho), \tag{2}$$

where $m\in\mathrm{Z}$, $n\in\mathrm{N}$, and $p=p_z$ is the projection of momentum on the z axis. Here

$$\begin{aligned}R_{mn}(\rho)&=C_{mn}\rho^{|m|}\exp\left(-\frac{\rho^2}{4\ell^2}\right)L_n^{|m|}\left(\frac{\rho^2}{2\ell^2}\right),\\ C_{mn}&=\frac{1}{\ell^{1+|m|}}\left[\frac{n!}{2^{|m|}(n+|m|)!}\right]^{1/2},\quad \ell=\sqrt{\frac{\hbar}{m^*\Omega}},\end{aligned} \tag{3}$$

$L_n^{|m|}(x)$ are generalized Laguerre polynomials, $\Omega=\sqrt{\omega_c^2+4\omega_0^2}$, and ω_c is the cyclotron frequency.

In this paper we study electron transport in a quantum wire with allowance for elastic scattering by a system of potentials of short-range impurities. We model the potential of short-range impurities as a sum of point potentials[27–29]:

$$V(\mathbf{r})=\sum_j a_j\delta(\mathbf{r}-\mathbf{r}_j)[1+(\mathbf{r}-\mathbf{r}_j)\cdot\nabla], \tag{4}$$

where a_j are the coupling constants of the pseudopotential, and $\mathbf{r}_j$ are the radius vectors of the impurity centers.

From (3) we see that the unperturbed wave functions decrease exponentially with distance from the conducting channel axis (as ρ increases). The characteristic size of the wave function in a plane perpendicular to the magnetic field is ℓ, while the characteristic size of the microstructure is obviously $\sqrt{\hbar/m^*\omega_0}=\ell_0$. Below we study only the important case in which ℓ is several times smaller than ℓ_0. This occurs in the presence of strong magnetic quantization and when the microstructure is pinched.[9]

The restriction on the magnitude of the magnetic field $\mathbf{B}$ can be lifted if all impurities are near the conducting channel axis. We know of no experimental work in which such an impurity distribution is employed. However, recent novel technology employing a scanning tunneling microscope makes it possible to manipulate impurity atoms with the required precision.

2. SCATTERING BY A SYSTEM OF POINT IMPURITIES

Note that the use of potentials of type (4) in 2D and 3D requires a special approach.[30] In particular, perturbation theory techniques do not work in this case.[27] However, if Kreĭn's formula for the resolvent of the Hamiltonian H_0 perturbed by such a potential is employed, we can derive exact analytic formulas for the Green's function of such a Hamiltonian and the wave function.

Using (2), we find the Green's function of the operator H_0 in the form

$$\begin{aligned}G_E^0(\mathbf{r},\mathbf{r}')=&\frac{m^*}{2\pi i\hbar}\sum_{n=0}^{\infty}\sum_{m=-\infty}^{\infty}\exp[im(\varphi-\varphi')]\\ &\times R_{mn}(\rho)R_{mn}(\rho')\frac{\exp(ip_{mn}|z-z'|/\hbar)}{p_{mn}},\end{aligned} \tag{5}$$

where

$$p_{mn}=\left[2m^*\left(E-\frac{\hbar\omega_c}{2}m-\frac{\hbar\Omega}{2}(2n+|m|+1)\right)\right]^{1/2}.$$

We now examine the Green's function and the state of a scattered particle for the operator H_0 perturbed by a single-point potential concentrated at point $\mathbf{r}_0$. As shown, e.g., in Ref. 30, a convenient way to study operators of this type is to employ the approach based on Kreĭn's formula for resolvents. According to this formula, the Green's function of the Hamiltonian $H=H_0+V$ has the form

$$\begin{aligned}G_E(\mathbf{r},\mathbf{r}')=&G_E^0(\mathbf{r},\mathbf{r}')-[Q(E,\rho_0)+a^{-1}]^{-1}\\ &\times G_E^0(\mathbf{r},\mathbf{r}_0)G_E^0(\mathbf{r}_0,\mathbf{r}').\end{aligned} \tag{6}$$

Here $Q(E,\rho_0)$ is Kreĭn's $\mathcal{Q}$-function, which is defined, to within a constant, by

$$Q(E,\rho_0)=\lim_{\rho\to\rho_0}[G_E^0(\mathbf{r},\mathbf{r}_0)-G_{E_0}^0(\mathbf{r},\mathbf{r}_0)], \tag{7}$$

where E_0 is some fixed value of E, and a is related to the scattering length λ by $a=2\pi\hbar^2\lambda/m^*$. Combining (5) and (7), we have

$$\begin{aligned}Q(E,\rho_0)=&\frac{\sqrt{m^*}}{2\sqrt{2}\pi i\hbar}\sum_{n=0}^{\infty}\sum_{m=-\infty}^{\infty}R_{mn}^2(\rho_0)\\ &\times\left[\left(E-\frac{\hbar\omega_c}{2}m-\frac{\hbar\Omega}{2}(2n+|m|+1)\right)^{-1/2}\right.\\ &\left.-i\left(\frac{\hbar\omega_c}{2}m+\frac{\hbar\Omega}{2}(2n+|m|+1)\right)^{-1/2}\right]+C,\end{aligned} \tag{8}$$

where the constant C remains to be determined.

Now suppose that $\Psi^0(\mathbf{r})$ is a delocalized state of H_0. By virtue of Eq. (6), the corresponding state $\Psi(\mathbf{r})$ of the operator H is

$$\Psi(\mathbf{r})=\Psi^0(\mathbf{r})-[Q(E,\rho_0)+a^{-1}]^{-1}\Psi^0(\mathbf{r}_0)G_E^0(\mathbf{r},\mathbf{r}_0). \tag{9}$$

We see from (9) that the electron scattering amplitude in a quantum wire then takes the form

$$t(E,\mathbf{r}_0)=[Q(E,\rho_0)+a^{-1}]^{-1}. \tag{10}$$

The constant C in (7) and (8) can easily be found by passing to the limits $\rho\to0$ and $E\to0$. The result is

$$C=\frac{m^*}{2\pi\hbar^2}\left(\frac{m^*\Omega}{2\hbar}\right)^{1/2}\zeta\left(\frac{1}{2},\frac{1}{2}\right). \tag{11}$$

Here $\zeta(s,\sigma)$ is the generalized Riemann zeta function. Note that by passing to the limit $\omega_0\to 0$ in (8) we obtain

$$Q(E,0)=\alpha a^{-1}\zeta\left(\frac{1}{2},\frac{1}{2}-\frac{E}{\hbar\omega_c}\right),$$

where $\alpha=\lambda/\sqrt{2l_B}$, with $l_B=\sqrt{\hbar/m^*\omega_c}$ the magnetic length. For the scattering amplitude in this limit we have

$$t(E)=\frac{a}{1+\alpha\zeta(1/2,1/2-E/\hbar\omega_c)}. \tag{12}$$

The right-hand side of Eq. (12) coincides with the expression in Ref. 28 for the scattering amplitude of an electron in a magnetic field, scattered by a point potential at $\mathbf{r}_0=0$. Note the difference between Eqs. (12) and (10), which becomes important in our subsequent investigation. Equation (10) contains the distance ρ_0 from the wire's axis to the point at which the impurity is located. Such an inhomogeneity in Kreĭn's $\mathcal{Q}$-function results, in the present case, from the lack of invariance of the Hamiltonian H_0 under the group of magnetic translations.

We now examine the Green's function of the operator H_0 perturbed by a system of identical impurities. The Green's function of this Hamiltonian can be obtained in explicit form (see Refs. 31 and 32):

$$G_E(\mathbf{r},\mathbf{r}')=G_E^0(\mathbf{r},\mathbf{r}')-\sum_{i,j=1}^{N_i}[Q(E)+a^{-1}]_{ij}^{-1}\times G_E^0(\mathbf{r},\mathbf{r}_i)G_E^0(\mathbf{r}_j,\mathbf{r}'). \tag{13}$$

Here Kreĭn's matrix $Q_{ij}(E)$ has the form

$$Q_{ij}(E)=\begin{cases}G_E^0(\mathbf{r}_i,\mathbf{r}_j), & i\neq j,\\ Q(E,\rho_j), & i=j,\end{cases} \tag{14}$$

and $(a^{-1})_{ij}=a^{-1}\delta_{ij}$. In Eq. (14), $Q(E,\rho_j)$ is given by (8) with ρ_0 replaced by ρ_j, where ρ_j is the polar radius of the jth impurity. Combining (13) and (14), we can determine the scattering operator $T(E)$. The Green's function of H can be expressed[27] in terms of the Green's function G_E^0 and the operator $T(E)$:

$$G_E(\mathbf{r},\mathbf{r}')=G_E^0(\mathbf{r},r')-G_E^0(\mathbf{r},\mathbf{r}')T(E)G_E^0(\mathbf{r}',\mathbf{r}). \tag{15}$$

Combining (13) and (15), we find that

$$T(E)=\sum_{i,j}[Q(E)+a^{-1}]_{ij}^{-1}|\delta(\mathbf{r}-\mathbf{r}_i)\rangle\langle\delta(\mathbf{r}-\mathbf{r}_j)|. \tag{16}$$

Using this expression for the scattering operator, we can easily write the scattering matrix $T_{\beta\beta'}(E)$ in the form

$$T_{\beta\beta'}(E)=\sum_{i,j}[Q(E)+a^{-1}]_{ij}^{-1}\Psi_\beta^0(\mathbf{r}_i)\Psi_{\beta'}^{0\,*}(\mathbf{r}_j). \tag{17}$$

According to what was said in the Introduction, the impurities that can effectively scatter are those near the axis of the conducting channel. The results of Ref. 4 imply that for $\rho_j\ll\ell$, corrections in ρ_j to the elements of Kreĭn's $\mathcal{Q}$-matrix are of order $O(\rho_j^2/\ell^2)$. We will subsequently see that the scattering parameters for the Hamiltonian H can be expressed in terms of these elements; neglecting corrections of order $O(\rho_j^2/\ell^2)$, we assume from now on that $\rho_j=0$. In this case, by virtue of (3), the only electrons scattered are those with angular momentum $m=0$.

Equation (13) implies that delocalized states $\Psi(\mathbf{r})$ of the Hamiltonian H_0+V with energy E have the form

$$\Psi(\mathbf{r})=\Psi_0(\mathbf{r})-\sum_{i,j}[Q(E)+a^{-1}]_{ij}^{-1}\Psi_0(\mathbf{r}_i)G_E^0(\mathbf{r},\mathbf{r}_j), \tag{18}$$

where Ψ_0 is the delocalized wave function of the unperturbed Hamiltonian H_0 with the same energy. Consider the transition, due to the scatterers, from the state $(0,n_0)$ with energy $E=\hbar\Omega(n_0+1/2)+p^2/2m^*$ to the state $(0,n_1)$ with the same energy. Suppose that the wave that propagates in the mode $(0,n_0)$ is

$$\Psi_{0n_0p_0}^0=\frac{1}{\sqrt{2\pi}}\exp\left(\frac{ip_{0n_0}z}{\hbar}\right)R_{0n_0}(\rho). \tag{19}$$

According to (18), the corresponding state of the Hamiltonian H with the same energy E is

$$\Psi(\mathbf{r})=\frac{\exp(ip_{0n_0}z/\hbar)}{\sqrt{2\pi}}R_{0n_0}(\rho)-\frac{1}{\sqrt{2\pi}}\frac{m^*}{2\pi i\hbar\ell^2}\times\sum_{n=0}^{\infty}\frac{R_{0n}(\rho)}{p_{0n}}\sum_{j,k}[Q(E)+a^{-1}]_{jk}^{-1}\times\exp\left[i\left(p_{0n_0}z_j+\frac{p_{0n}|z-z_k|}{\hbar}\right)\right], \tag{20}$$

where we have allowed for the fact that $R_{0n}(0)=\ell^{-1}$ for all values of n.

3. TRANSMISSION COEFFICIENT AND CONDUCTANCE IN THE BALLISTIC REGIME

Using the results obtained for the wave functions in Sec. 2, we can find the transmission coefficients and the conductance in the ballistic regime.

When $z<-\max|z_k|$, the coefficient $\varphi^{(-)}$ of $R_{0n_0}(\rho)$ in (20) has the form

$$\varphi^{(-)}=\frac{\exp(ip_{0n_0}z/\hbar)}{\sqrt{2\pi}}-\frac{m^*}{2\pi i\hbar\ell^2p_{0n_0}}\left(\sum_{j,k}[Q(E)+a^{-1}]_{jk}^{-1}\times\exp\frac{ip_{0n_0}(z_j+z_k)}{\hbar}\right)\frac{\exp(-ip_{0n_0}z/\hbar)}{\sqrt{2\pi}}, \tag{21}$$

and when $z>\max|z_k|$, the coefficient $\varphi^{(+)}$ of $R_{0n_1}(\rho)$ is

$$\varphi^{(+)}=\frac{\exp(ip_{0n_0}z/\hbar)}{\sqrt{2\pi}}\delta_{n_0n_1}-\frac{m^*}{2\pi i\hbar\ell^2 p_{0n_1}}\times\left(\sum_{j,k}[Q(E)+a^{-1}]^{-1}_{jk}\exp\frac{i(p_{0n_0}z_j-p_{0n_1}z_k)}{\hbar}\right)\times\frac{\exp(ip_{0n_1}z/\hbar)}{\sqrt{2\pi}}. \tag{22}$$

Comparing $\varphi^{(+)}$ and $\varphi^{(-)}$, we see that the partial transmission coefficient $T_{n_0\to n_1}(E)$ is

$$T_{n_0\to n_1}(E)=\left|\delta_{n_0n_1}-\frac{m^*}{2\pi i\hbar\ell^2}\times\frac{\sum_{j,k}[Q(E)+a^{-1}]^{-1}_{jk}\exp[i(p_{0n_0}z_j-p_{0n_1}z_k)/\hbar]}{[p^2_{0n_0}+2m^*\hbar\Omega(n_0-n_1+1/2)]^{1/2}}\right|^2. \tag{23}$$

Following the Landauer–Büttiker formalism,[7–9] we can find the conductance of the microstructure:

$$G(E)=\frac{2e^2}{h}\sum_{n_0,n=0}^{N}T_{n_0\to n}(E), \tag{24}$$

where the integer N satisfies $E=\hbar\Omega(N+1/2+\delta)$, $0<\delta<1$. Since it is difficult to use the general formulas (23) and (24) in an analytic investigation due to the complexity of the expression for Kreĭn's $\mathcal{Q}$-matrix, we consider the various limits. First we obtain convenient estimates for the off-diagonal elements of this matrix. The diagonal elements of Kreĭn's $\mathcal{Q}$-matrix are equal to $Q_0(E)+a^{-1}$.

At $\rho_j=\rho_k=0$ the off-diagonal elements of Kreĭn's $\mathcal{Q}$-matrix are

$$G_{jk}\equiv G^0_E(\mathbf{r}_j,\mathbf{r}_k)=\frac{m^*}{2\pi i\hbar\ell^2}\sum_0^\infty\frac{1}{p_{0n}}\exp\frac{ip_{0n}|z_j-z_k|}{\hbar}. \tag{25}$$

Using the method developed in Refs. 28 and 29, we can transform the sum in (25) into an integral:

$$G_{jk}=\frac{m^*}{2\pi\hbar^2|z_j-z_k|}\exp\left[-i\sqrt{\frac{2E}{\hbar\Omega}-1}\left|\frac{z_j-z_k}{\ell}\right|\right]+\frac{m^*}{\sqrt{2}\pi^{3/2}\hbar^2\ell}\int_0^\infty\left(\frac{1}{2}+\frac{1}{2}\coth\frac{t^2}{2}-\frac{1}{t^2}\right)\times\exp\left[-\frac{t^2}{2}\left(1-\frac{2E}{\hbar\Omega}\right)-\frac{1}{2t^2}\left(\frac{z_j-z_k}{\ell}\right)^2\right]dt. \tag{26}$$

This formula proves to be convenient in analyzing the various limits. Indeed, suppose that $|z_j-z_k|/\ell\ll 1$, which corresponds to small distances between impurities. Then Eq. (26) yields the asymptotic behavior

$$G_{jk}=\frac{m^*}{2\pi\hbar^2|z_j-z_k|}\left[1+O\left(\left|\frac{z_j-z_k}{\ell}\right|\right)\right]. \tag{27}$$

If the distances between the impurities are large, i.e., $|z_j-z_k|/\ell\gg 1$, Eq. (26) yields

$$G_{jk}=-\frac{m^*i}{2\pi\hbar^2\ell\sqrt{2E/\hbar\Omega-1}}\times\exp\left[-i\sqrt{\frac{2E}{\hbar\Omega}-1}\frac{|z_j-z_k|}{\ell}\right]\times\left[1+O\left(\frac{|z_j-z_k|}{\ell}\right)\right]. \tag{28}$$

Equation (27) shows that when the distances between the impurities are small, the matrix elements G_{jk} are large, while Eq. (28) shows that when these distances are large, $|G_{jk}|$ is independent of the coordinates of the impurities. In the latter case we can estimate the ratio of the off-diagonal elements of Kreĭn's matrix to the diagonal elements. Let us introduce the scattering length $\lambda=m^*a/2\pi\hbar^2$ and the constant $\gamma=|\lambda|/\sqrt{2}\ell$. In all realistic situations $\gamma\ll 1$, so that we have

$$\left|\frac{G_{jk}}{Q_0(E)+a^{-1}}\right|=\frac{\sqrt{2}}{\sqrt{2E/\hbar\Omega-1}}\times\frac{\gamma}{|1+\gamma\zeta(1/2,1/2-E/\hbar\Omega)|}. \tag{29}$$

Clearly, for $E\gg\hbar\Omega/2$ we have from (29)

$$\left|\frac{G_{jk}}{Q_0(E)+a^{-1}}\right|\ll 1,\quad j\neq k.$$

Discarding the small off-diagonal terms in the matrix $[Q(E)+a^{-1}]^{-1}$, we obtain a convenient estimate for the transmission coefficient:

$$T_{n_0\to n_1}(E)\cong\left|\delta_{n_0n_1}-\frac{m^*}{2\pi i\hbar\ell^2}\times\frac{[Q_0(E)+a^{-1}]^{-1}\sum_j\exp[i(p_{0n_0}-p_{0n_1})z_j/\hbar]}{[p^2_{0n_0}+2m^*\hbar\Omega(n_0-n_1+1/2)]^{1/2}}\right|^2. \tag{30}$$

Note that when there are only two impurities in the conducting channel and they are near the wire's axis, we can simplify Eq. (30). We position one impurity at the origin and introduce the quantity $\Delta=|z_1-z_2|$, the distance between the impurities. Then we can write

$$T_{n_0\to n_1}=|\delta_{n_0n_1}+i\gamma(N+\delta+n_1)^{-1/2}\times[1+\gamma\zeta(1/2,-N-\delta)](1+\cos\beta+i\sin\beta)|^2, \tag{31}$$

where $\beta=(p_{0n_0}-p_{0n_1})\Delta/\hbar$.

4. SCATTERING BY TWO POINT IMPURITIES

We now discuss elastic scattering by the potential of two impurities in greater detail. What is interesting here is that the problem makes it possible to carry out a fairly simple analytic and numerical analysis of the expression for the microstructure's conductance. In addition, it can be expected that this case will reveal the main features of scattering by several point potentials.

Using the shift formula for the Riemann zeta function, we separate the real and imaginary parts of this function:[33]

$$\text{Re } \zeta=\text{Re } \zeta\left(\frac{1}{2},-N-\delta\right)=\zeta\left(\frac{1}{2},1-\delta\right), \tag{32}$$

$$\text{Im } \zeta=\text{Im } \zeta\left(\frac{1}{2},-N-\delta\right)=\sum_{n=0}^{N}(N+\delta-n)^{-1/2}.$$

To estimate $\zeta(1/2,1-\delta)$ we use the results of Ref. 31, whereupon

$$\zeta\left(\frac{1}{2},1-\delta\right)\cong(1-\delta)^{-1/2}+\frac{1}{2}(2-\delta)^{-1/2}-2\sqrt{2-\delta}+\frac{1}{24}(2-\delta)^{-3/2}. \tag{33}$$

Combining (32) and (31), we obtain

$$T_{n_0\to n}=\delta_{n_0n}\left\{1+\frac{4\gamma^2\ \text{Im } \zeta}{\sqrt{N+\delta-n}[(1+2\gamma\ \text{Re } \zeta)^2+(2\gamma\ \text{Im } \zeta)^2]}\right\}+\frac{\gamma^2(N+\delta-n)^{-1}}{[1+\gamma\ \text{Re } \zeta(1+\cos\beta)-\gamma\ \text{Im } \zeta\ \sin\beta]^2+[\gamma\ \text{Im } \zeta(1+\cos\beta)+\gamma\ \text{Re } \zeta\ \sin\beta]^2}. \tag{34}$$

Let us compare this with the expression for the partial transmission coefficient in the case of scattering by a single impurity (obtain in Ref. 4). The coefficient of δ_{n_0n} in (34) (the expression in braces) differs from the corresponding coefficient in Ref. 4 in that γ is replaced by 2γ, which means that the coupling constant in (4) has doubled. This term in $T_{n_0\to n}$ does not contain the distance between the impurities, and its structure is the same as for scattering by a single impurity. The second term in the sum for $T_{n_0\to n}$ contains the distance between the impurities, and in the limit $\Delta\to 0$, it simply turns into the analogous corresponding term in the transmission coefficient for the case involving a single impurity. Both terms are proportional γ^2 and, generally speaking, are small because $\gamma\ll 1$. However, at values of the electron energy E at which $1+2\gamma\ \text{Re } \zeta=0$, the first term of those proportional to γ^2 ceases to be small, since in (34) the small factor γ^2 in the numerator cancels the same factor in the denominator, and both terms in braces become quantities of the same order. In the second term, which is proportional to γ^2, the same situation occurs at $1+2\gamma[(1+\cos\beta)\text{Re } \zeta-\sin\beta\ \text{Im } \zeta]=0$. While in the former case there is only one such point E at which the above situation occurs, in the latter there can be many, and their location on the E axis depends on the distance Δ between impurities. As shown in Ref. 4, at energy values satisfying the condition $1+2\gamma\ \text{Re } \zeta=0$, a sharp maximum emerges at the conductance quantization steps near the threshold of each step. We expect that in the cases considered here, there can be several maxima of this kind.

We now consider the conductance $G(E)$ of a quantum wire for the important case where in which the distance Δ between the two impurities is large, $\Delta/\lambda\gg 1$.

Using Eqs. (24) and (34), we obtain

$$\frac{G}{G_0}=\sum_{n=0}^{N}\left\{1+\frac{4\gamma^2\ \text{Im } \zeta}{\sqrt{N+\delta-n}[(1+2\gamma\ \text{Re } \zeta)^2+(2\gamma\ \text{Im } \zeta)^2]}\right\}+\gamma^2\sum_{n,n_0=0}^{N}\frac{(N+\delta-n)^{-1}}{[1+\gamma\ \text{Re } \zeta(1+\cos\beta)-\gamma\ \text{Im } \zeta\ \sin\beta]^2+\gamma^2[(1+\cos\beta)\text{Im } \zeta+\text{Re } \zeta\ \sin\beta]^2}, \tag{35}$$

where $G_0=2e^2/h$ is the conductance quantum. Calculating the sums in (35) in the same way as in Ref. 4, we obtain that

$$\frac{G}{G_0}=N+1+\frac{(2\gamma\ \mathrm{Im}\ \zeta)^2}{(1+2\gamma\ \mathrm{Re}\ \zeta)^2+(2\gamma\ \mathrm{Im}\ \zeta)^2}+\gamma^2\sum_{n_0,n=0}^{N}\frac{(N+\delta-n)^{-1}}{1+2\gamma[(1+\cos\beta)\mathrm{Re}\ \zeta-\sin\beta\ \mathrm{Im}\ \zeta]+2\gamma^2(1+\cos\beta)[(\mathrm{Re}\ \zeta)^2+(\mathrm{Im}\ \zeta)^2]}. \tag{36}$$

Clearly, the conductance in (36) consists of three terms: $G/G_0=G_1+G_2+G_3$. The first, $G_1=1+[x]$, where $[x]$ is the integer part of $x=(E/\hbar\Omega-1/2)$, is independent of scattering and produces conductance quantization steps of width $\hbar\Omega$ and height equal to the conductance quantum G_0. The second term, G_2, is given by

$$G_2=(2\gamma)^2\left|\zeta\left(\frac{1}{2},-x\right)-\zeta\left(\frac{1}{2},1-\{x\}\right)\right|^2\times\left[1+4\gamma\zeta\left(\frac{1}{2},1-\{x\}\right)+4\gamma^2\left|\zeta\left(\frac{1}{2},-x\right)\right|^2\right]^{-1}, \tag{37}$$

where $\{x\}$ is the fractional part of x. Finally, the third term is

$$G_3=\gamma^2\sum_{n,n_0=0}^{\infty}(x-n)^{-1}\left\{1+2\gamma\left[(1+\cos\beta)\zeta\left(\frac{1}{2},1-\{x\}\right)-\sin\beta\left|\zeta\left(\frac{1}{2},-x\right)-\zeta\left(\frac{1}{2},1-\{x\}\right)\right|\right]+2\gamma^2(1+\cos\beta)\times\left|\zeta\left(\frac{1}{2},-x\right)\right|^2\right\}^{-1}. \tag{38}$$

The last two terms reflect the deviation of the shape of the quantization steps from ideal. Both terms, G_2 and G_3, are due to scattering by impurities. The G vs. E curves are depicted in Figs. 1–3, with the details of one of the peaks at the step threshold depicted separately. The diagrams representing $G(\Delta)$ are depicted in Figs. 4–6, which show how conductance quantization breaks down in the case of scattering by two point impurities. The number of resonance peaks in the G vs. E curves depends strongly on the distance between impurities and the number of the conductance step, i.e., the number of peaks increases with Δ and N. All peaks are located near a step threshold, which is due to the aforementioned fact that $G_2(E)$ and $G_3(E)$ are small everywhere except in the vicinity of the threshold. The height of these peaks decreases as the distance from the step threshold increases, the reason obviously being that as E increases, we move farther away from the points of resonance, i.e., the sum $G_2(E)+G_3(E)$ gets small as we move away from the main peak.

We also note that the peaks in $G_2(E)$ and $G_3(E)$ are located at different points in the vicinity of a step threshold and that the peak in $G_2(E)$ is much lower than the main peak in $G_3(E)$, since $G_3^{\max}(E)/G_2^{\max}(E)\sim N$. Indeed, the second term in $G_2(E)$ in the neighborhood of the resonance point determined by the equality $1+2\gamma\ \mathrm{Re}\ \zeta=0$ is of order unity. It is clear why the resonance in $G_2(E)$ is determined by such an equality: when this condition is met, the small factor γ^2 in the numerator in (36) cancels the same factor in the denominator. A similar situation occurs for $G_3(E)$ if the following condition is met:

$$1+2\gamma[(1+\cos\beta)\mathrm{Re}\ \zeta-\sin\beta\ \mathrm{Im}\ \zeta]=0. \tag{39}$$

The factors Re ζ and Im ζ in (39) behave differently, i.e., Im ζ is large for $\delta\ll 1$, while Re ζ is large for $\delta\approx 1$. Equations (32) and (33) clearly demonstrate this. On the other hand, because of the smallness of γ, for condition (39) to be met at least one of these factors must be large. Clearly,

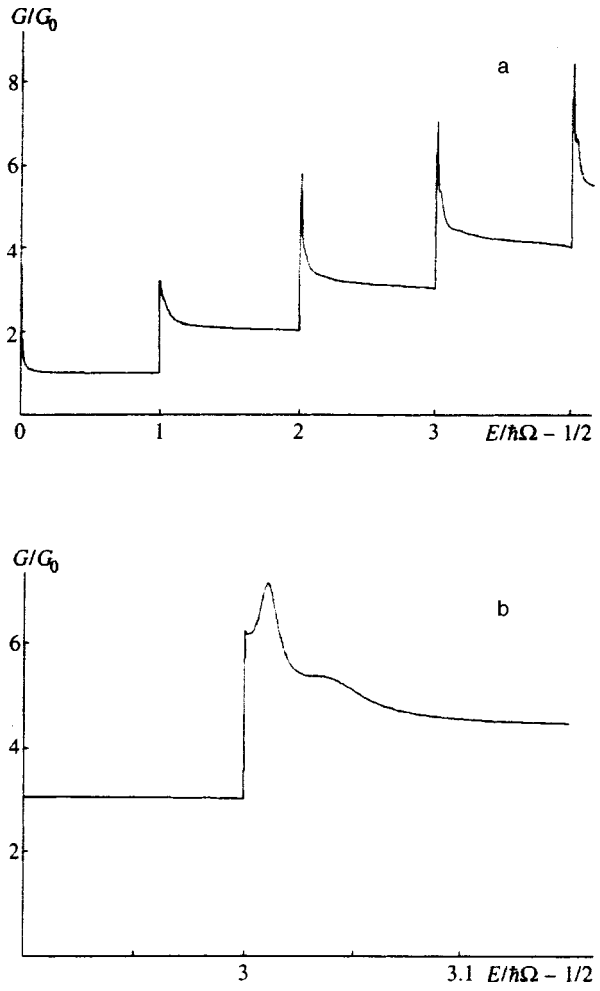


FIG. 1. (a) Energy dependence of the conductance of a microstructure, $G(E)$, for the case of scattering by two impurities with $l_0=2.5\times10^{-6}$ cm, $l=1.1\times10^{-6}$ cm, and $\Delta=2\times10^{-5}$ cm. (b) Detailed structure of the resonance peaks.

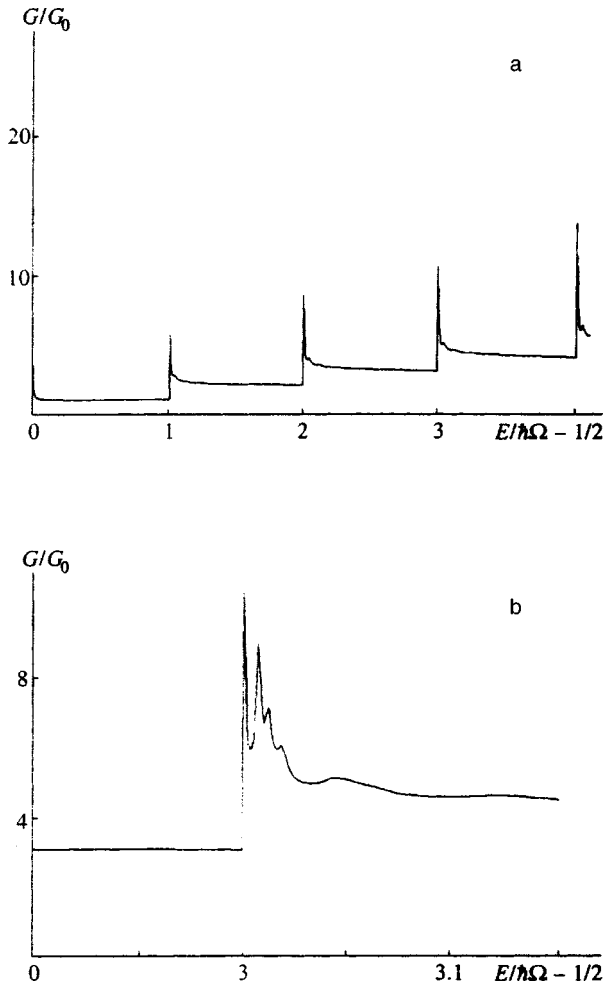


FIG. 2. a) Energy dependence of the conductance of a microstructure, $G(E)$, under the same conditions as in Fig. 1a but with a different distance between impurities, $\Delta = 5\times 10^{-5}$ cm. b) Detailed structure of the peaks.

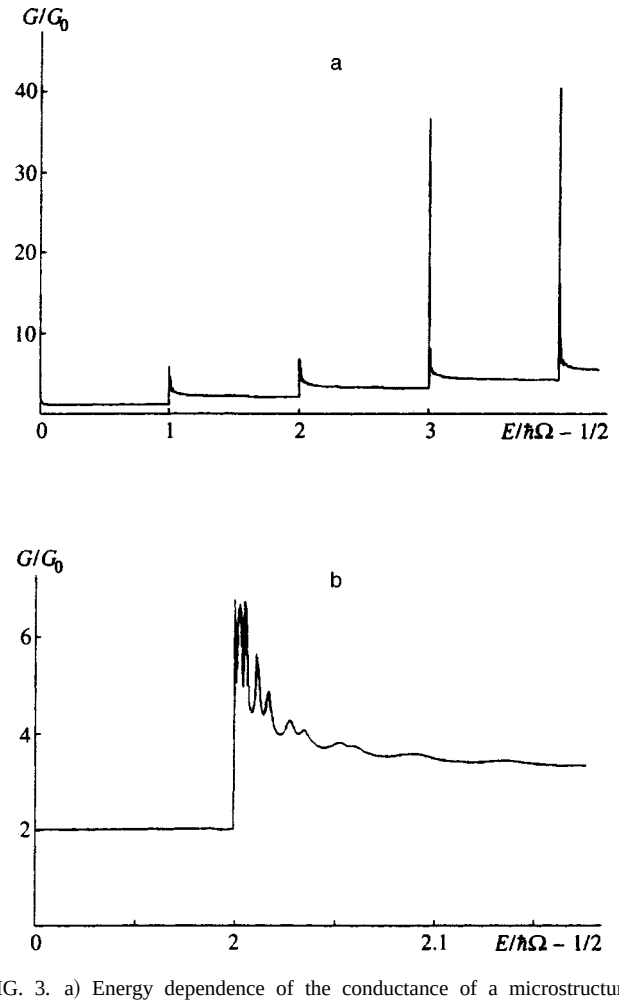


FIG. 3. a) Energy dependence of the conductance of a microstructure, $G(E)$, under the same conditions as in Figs. 1a and 2a but with $\Delta = 10^{-4}$ cm. b) Detailed structure of the peaks.

the condition (39) is met only if $\delta \ll 1$. Then near the threshold (39) yields the estimate $\sin\beta\,\mathrm{Im}\,\zeta \simeq 1/2\gamma$, and since $\mathrm{Im}\,\zeta \simeq 1/\sqrt{\delta}$ for $\delta \ll 1$, we have $\sin\beta \simeq \sqrt{\delta}/2\gamma$. Hence $0 < \delta < 4\gamma^2$ near a resonance. Using what was said earlier, from (36) we find that in the neighborhood of a resonance

$$G_3(E) \simeq \frac{1}{2}(N+1)\sum_{n=0}^{N} \frac{(N+\delta-n)^{-1}}{1+\sqrt{1-\delta/4\gamma^2}} \frac{1}{(\mathrm{Im}\,\zeta)^2}. \tag{40}$$

Then for $\delta \ll 1$ we obtain

$$G_3(E) \simeq \frac{N+1}{2}\left(\frac{1}{1+\sqrt{1-\delta/4\gamma^2}}\right). \tag{41}$$

The factor in parentheses ranges from 1/2 to 1, so that $(N+1)/4 < G_3(E) < (N+1)/2$. Clearly, for $N \gg 1$ we have $G_3(E) \gg G_2(E)$, while for $N \sim 1$ and $\delta \ll 1$ all three terms ($G_1(E)$, $G_2(E)$, and $G_3(E)$) are of the same order.

5. CONDUCTANCE OF A QUANTUM WIRE IN THE DIFFUSIVE REGIME

We now discuss the conductance of a quantum wire in the case in which the system is in the diffusive regime. To find the probability of an electron transition from the initial state $|\beta\rangle = |m,n,p\rangle$ to the final state $|\beta'\rangle = |m',n',p'\rangle$, we use the Lippmann equation[34]

$$W_{\beta\beta'} = \frac{2\pi}{\hbar}|T_{\beta\beta'}|^2 \delta(\varepsilon_\beta - \varepsilon_{\beta'}). \tag{42}$$

Introducing the scattering amplitude $t_{\beta\beta'}(E)$, we obtain

$$W_{\beta\beta'} = \frac{2\pi}{\hbar V^2}|t_{\beta\beta'}(E)|^2 \delta(\varepsilon_\beta - \varepsilon_{\beta'}), \tag{43}$$

where $T_{\beta\beta'}(E) = V^{-1} t_{\beta\beta'}(E)$, with V the normalization volume.

The electron motion along the wire's axis in a longitudinal magnetic field (the z axis) is semiclassical, so that we can use the Boltzmann transport equation to find the nonequilibrium distribution function for the electrons. The transport equation linearized in the electric field $\mathscr{E}$ parallel to the z axis has the form

$$\frac{\partial f_0}{\partial \varepsilon_\beta}\frac{e\mathscr{E}p}{m^*} = I^{ei}_{\beta\beta'}(f), \tag{44}$$

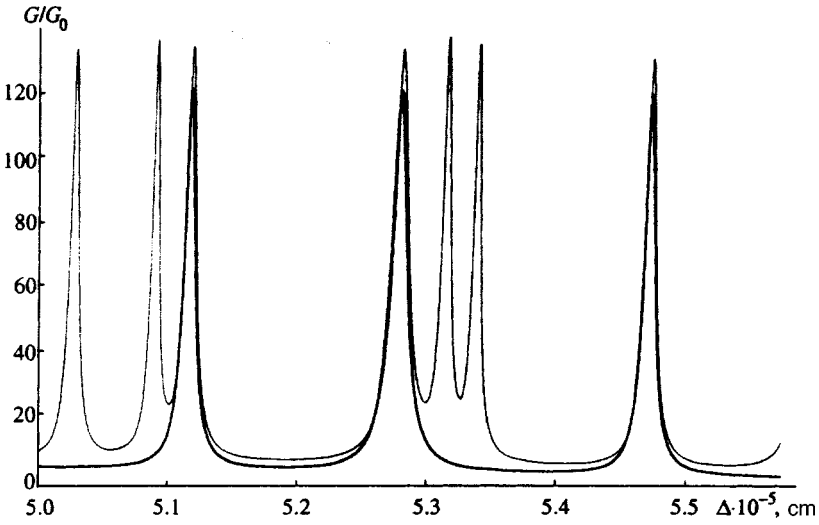


FIG. 4. G vs. Δ at $l_0=2.5\times10^{-6}$ cm and $l=1.1\times10^{-6}$ cm. The heavy curve corresponds to $E/\hbar\Omega-1/2=2$, and the light curve to $E/\hbar\Omega-1/2=4$.

where $I^{ei}_{\beta\beta'}(f)$ is the collision integral, and $f_0(\varepsilon)$ and $f(\varepsilon)$ are, respectively, the equilibrium and nonequilibrium distribution functions. The collision integral is given by the well-known formula

$$I^{ei}_{\beta\beta'}=\sum_{\beta'} W_{\beta\beta'}[f(\varepsilon_\beta)-f(\varepsilon_{\beta'})]. \tag{45}$$

In the relaxation-time approximation, the nonequilibrium distribution function is

$$f(\varepsilon_\beta)=f_0(\varepsilon_\beta)-\frac{e\mathscr{E}p}{m^*}\tau(\varepsilon)\frac{\partial f_0}{\partial\varepsilon}. \tag{46}$$

Combining (45) and (46), we obtain the following expression for the relaxation time $\tau^{-1}(\varepsilon_\beta)=\Sigma_{\beta'}W_{\beta\beta'}$:

$$\tau^{-1}(\varepsilon_\beta)=\frac{2\pi}{\hbar V^2}\sum_{\beta'} t_{\beta\beta'}(E_{\beta'})\delta(\varepsilon_\beta-\varepsilon_{\beta'}). \tag{47}$$

The longitudinal conductivity is

$$\sigma_{zz}=-\frac{e^2}{Vm^{*2}}\sum_\beta p^2\frac{\partial f_0}{\partial\varepsilon_\beta}\tau(\varepsilon_\beta). \tag{48}$$

Then the general formula for σ_{zz} becomes

$$\sigma_{zz}=-\frac{e^2\hbar^2}{2\pi m^{*2}}\sum_\beta p^2\frac{\partial f_0}{\partial\varepsilon_\beta}\Bigg[\sum_{\beta'}|t_{\beta\beta'}(\varepsilon_\beta)|^2\times\delta(\varepsilon_\beta-\varepsilon_{\beta'})\Bigg]^{-1}. \tag{49}$$

Here

$$t_{\beta\beta'}(\varepsilon_\beta)=V\sum_{i,j}[Q(\varepsilon_\beta)+a^{-1}]^{-1}_{ij}\Psi^{0*}_\beta(\mathbf{r}_i)\Psi^0_{\beta'}(\mathbf{r}_j), \tag{50}$$

with the matrix in the denominator given by (14).

The equation for the scattering amplitude is extremely awkward, which means that the general expression (49) is extremely difficult to analyze, so that below we consider various limiting cases.

We examine the important case in which geometric confinement is weaker than magnetic. This situation is realized when the radius ρ of the wire is much greater than l_B. Here the characteristic scale of variation of the wave function (2)

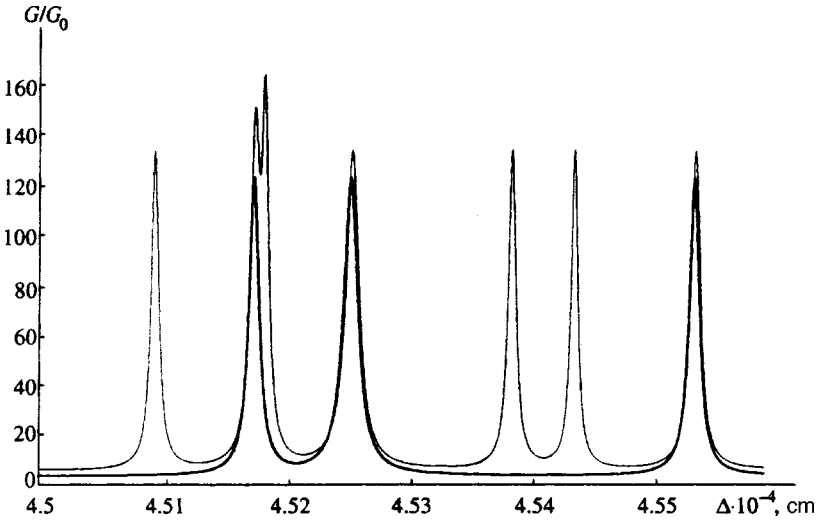


FIG. 5. The same dependence as in Fig. 4 but for a different range of Δ.

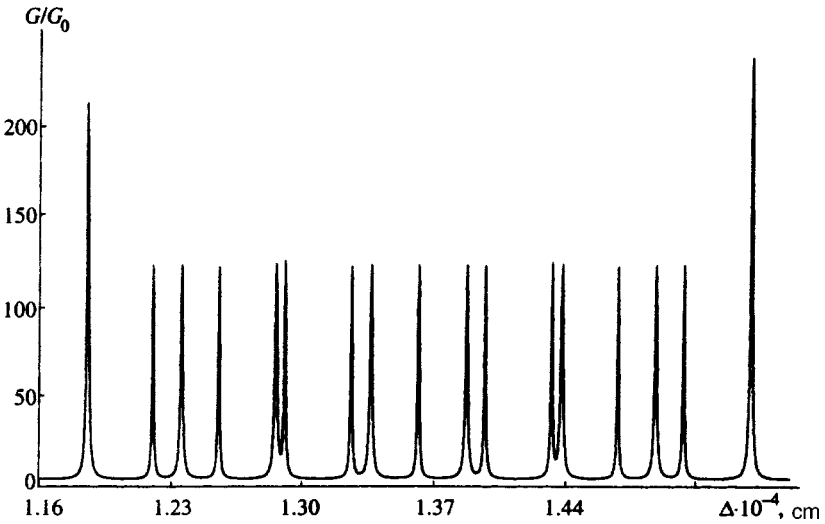


FIG. 6. G vs. Δ at $l_0=2.5\times10^{-6}$ cm, $l=1.1\times10^{-6}$ cm, and $E/\hbar\Omega-1/2=4$. The distance Δ between the impurities is much greater than in Fig. 4.

in the geometric-confinement plane is of order l, the hybrid length, and $l\ll\rho$. In view of this, the main contribution to the current is provided by electrons moving along the wire's axis, while the contribution of the electrons at a distance from the axis is exponentially small ($\propto\exp(-\rho^2/2l^2)$). Hence we can set ρ_0 to zero in the expression (8) for $Q(E,\rho_0)$.

In this case Kreĭn's $\mathcal{Q}$-function is homogeneous, and we can average (42) over the positions of the randomly distributed impurities. If the impurity concentration is low, i.e., if for a characteristic momentum p of the electron gas we can write $p\gg\hbar(n_i)^{1/3}$, where n_i is the impurity concentration, we can linearize the electron transition probability (43) in the impurity concentration, with the result that $W_{\beta\beta'}=n_iVW^0_{\beta\beta'}$, where $W^0_{\beta\beta'}$ is the probability of elastic scattering by a single impurity at the origin,

$$W^0_{\beta\beta'}=\frac{2\pi}{\hbar V^2}\frac{a^2}{1+\alpha\zeta(1/2,1/2-\varepsilon_\beta/\hbar\Omega)}\delta(\varepsilon_\beta-\varepsilon_{\beta'}). \tag{51}$$

Bearing all this in mind, we can find $\tau(E)$ using Eq. (47):

$$\tau^{-1}(\varepsilon)=\frac{n_iL}{\hbar^2V}|t(\varepsilon)|^2\frac{m^*}{\sqrt{2m^*\hbar\Omega}}\sum_n\left(\frac{\varepsilon}{\hbar\Omega}-n-\frac{1}{2}\right)^{-1/2}, \tag{52}$$

where

$$t(\varepsilon)=\frac{a}{1+\alpha\zeta(1/2,1/2-\varepsilon/\hbar\Omega)}, \tag{53}$$

and L is the length of the wire.

Combining (49), (52), and (53), we find the general expression for the conductivity in the present case:

$$\sigma_{zz}=-\frac{e^2\hbar^3}{\pi l^2m^{*2}n_i}\int_{\hbar\Omega/2}^{\infty}d\varepsilon\,\frac{\partial f_0}{\partial\varepsilon}|t(\varepsilon)|^{-2}\times\frac{\sum_n(\varepsilon/\hbar\Omega-n-1/2)^{1/2}}{\sum_n(\varepsilon/\hbar\Omega-n-1/2)^{-1/2}}. \tag{54}$$

We first examine a nondegenerate electron gas in the ultraquantum limit. The equilibrium distribution function for electrons with a vanishing projection of angular momentum on the wire's axis is

$$f_0(\varepsilon_\beta)=A\,\exp\left\{-\frac{1}{T}\left[\frac{p^2}{2m^*}+\hbar\Omega\left(n+\frac{1}{2}\right)\right]\right\}, \tag{55}$$

where the normalization constant is

$$A=\frac{8\pi\hbar^2n_e}{m^*\Omega\sqrt{2\pi m^*T}}\sinh\frac{\hbar\Omega}{2T},$$

with n_e the electron concentration. In the ultraquantum limit ($\hbar\Omega\gg T$), combining (54) and (55) yields

$$\sigma_{zz}=A\frac{2e^2\hbar^3TV}{\pi^3m^{*2}n_i\hbar\Omega a^2}\tau(\alpha)\exp\left(-\frac{\hbar\Omega}{2T}\right), \tag{56}$$

where

$$\tau(\alpha)=\int_0^\infty xe^{-x}\left|1+\alpha\zeta\left(\frac{1}{2},\frac{1}{2}-\frac{xT}{\hbar\Omega}\right)\right|^2dx. \tag{57}$$

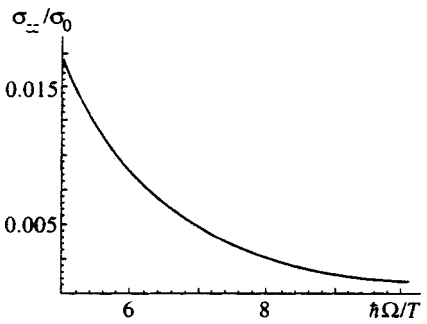


FIG. 7. Dependence of the longitudinal conductivity of a nondegenerate electron gas in the ultraquantum limit $\hbar\Omega\gg T$ with $\sigma_0=4e^2\hbar^4n_e\times[\pi^3a^2n^{*2}n_i\sqrt{2\pi m^*T}]^{-1}$.

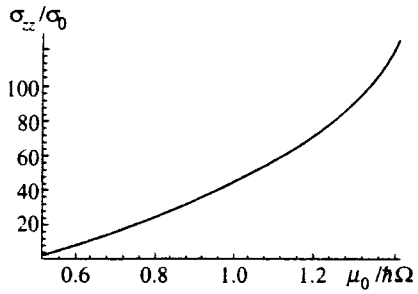


FIG. 8. Dependence of the longitudinal conductivity on $\mu_0/\hbar\Omega$ in the ultraquantum limit.

To estimate the integral in (57) we note that $x \leqslant 1$ is the region in which the integrand varies substantially, which implies that $xT/\hbar\Omega \ll 1$. Then for the Riemann zeta function we can use the asymptotic formula[33]

$$\zeta\left(\frac{1}{2},\delta\right) \simeq \frac{1}{\sqrt{\delta}} + \frac{1}{2}\,\frac{1}{\sqrt{1+\delta}}. \tag{58}$$

Combining (57) and (58), we obtain

$$\tau(\alpha) \simeq 1 + \alpha^2 \frac{\hbar\Omega}{T}. \tag{59}$$

The final expression for this limit is

$$\sigma_{zz} \simeq \frac{4e^2\hbar^3 n_e T}{\pi^3\sqrt{2\pi m^* T}\, m^{*2} n_i \Omega a^2} \times \exp\left(-\frac{\hbar\Omega}{2T}\right)\left(1+\alpha^2\frac{\hbar\Omega}{T}\right). \tag{60}$$

The diagram representing this dependence is depicted in Fig. 7.

We now turn to the case of a degenerate gas in the ultraquantum limit ($n=0$, $T=0$). If we bear in mind the asymptotic behavior of the zeta function, Eq. (54) yields

$$\sigma_{zz} = \frac{2e^2\hbar^3\alpha^2(\mu_0/\hbar\Omega - 1/2)L}{\pi^3 a^2 m^{*2} N_i}(u^2+v^2), \tag{61}$$

where

$$u = \frac{1}{\sqrt{3/2-\mu_0/\hbar\Omega}} + \frac{1}{2\sqrt{5/2-\mu_0/\hbar\Omega}} - 2\sqrt{\frac{5}{2}-\frac{\mu_0}{\hbar\Omega}} + \frac{1}{24}\left(\frac{5}{2}-\frac{\mu_0}{\hbar\Omega}\right)^{-3/2} + \alpha^{-1},$$
$$v = \left(\frac{\mu_0}{\hbar\Omega}-\frac{1}{2}\right)^{-1/2}, \quad N_i = Vn_i, \tag{62}$$

and $0.5\hbar\Omega < \mu_0 < 1.5\hbar\Omega$, with μ_0 the Fermi level. The diagram representing (62) is depicted in Fig. 8.

Finally, we consider the Shubnikov–de Haas oscillations in the longitudinal conductivity of a quantum wire. Estimating the ratio of the sums in (54) via the Poisson formula by a method similar to the one used in Ref. 33, we easily obtain

$$\sigma_{zz} = \frac{8e^2\hbar^2\mu}{3\pi^3 m^{*2} n_i \Omega a^2} \times \frac{1-\sqrt{\hbar\Omega/\mu}\sum_{k=1}^{\infty}(-1)^k A_k \cos(2\pi k\mu/\hbar\Omega - \pi/4)}{\alpha^2\hbar\Omega/\mu + 4\left|1+\alpha\zeta(1/2,1/2-\mu/\hbar\Omega)\right|^{-2}}, \tag{63}$$

where μ is the chemical potential of the electron gas in the wire, and

$$A_k = \frac{2\pi^2 kT}{\hbar\Omega\,\sinh(2\pi^2 kT/\hbar\Omega)}\,\frac{1}{\sqrt{2k}}. \tag{64}$$

The curve representing the dependence of σ_{zz}/σ_0 on the ratio ω_c/ω_0, where

$$\sigma_0 = \frac{8e^2\hbar^2\mu}{3\pi^3 a^2 m^{*2}\omega_0 n_i},$$

is depicted in Fig. 9.

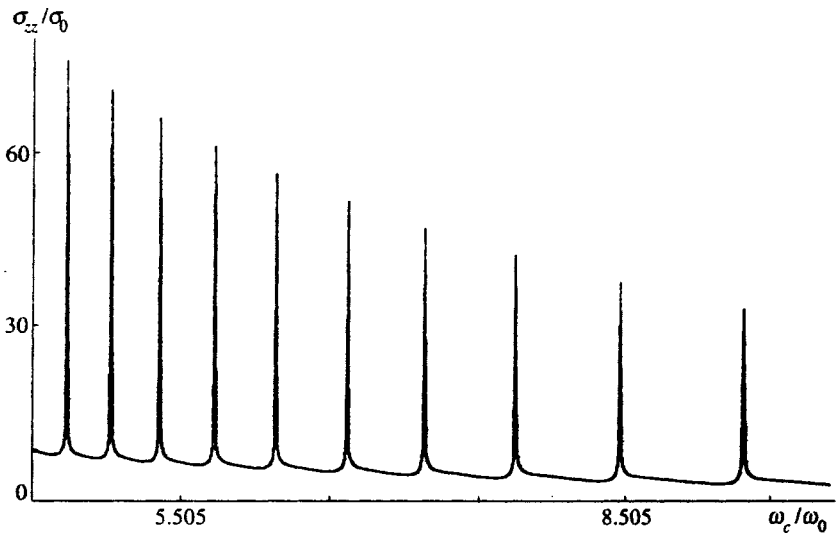


FIG. 9. Oscillations of the longitudinal conductivity of a quantum wire with variations of the magnetic field at $\mu/T = 100$ and $\mu/\hbar\omega_0 = 100$.

6. DISCUSSION

For a more detailed analysis of the resonance peaks in the conductance we consider the formula (38) used for estimates. This formula implies that $G_3(E)$ has a maximum at $\delta \simeq 4\gamma^2$. Then from the estimates in Sec. 4 it follows that $\sin\beta \simeq 1$, which means that $p_{0n_0} - p_{0n} \simeq (\hbar/\Delta)(\pi/2 + 2\pi k)$, with $k = 0,1,2,\ldots$. This implies that

$$\sqrt{N - n_0 + \delta} - \sqrt{N - n + \delta} \simeq \frac{\ell}{\sqrt{2}\Delta}\left(\frac{\pi}{2} + 2\pi k\right). \tag{65}$$

The condition (65) shows which of the partial transmission coefficients $T_{n_0 \to n}$ yield a resonance contribution to the conductance. Suppose that at an energy E (N and $\delta = 4\gamma^2$ are fixed) the condition (65) is met for certain values of n_0 and n. Then $T_{n_0 \to n}$ provides a resonance contribution to the third term, $G_3(E)$. For all other transmission coefficients $T_{n_0' \to n'}$ the quantity $\sqrt{N - n_0' + \delta} - \sqrt{N - n' + \delta}$ can obviously only satisfy (65) at values of δ not equal to $4\gamma^2$. The corresponding resonance contributions of these coefficients are smaller.

Note that the peaks from these coefficients can lie either to the right or left of the resonance point on the E axis. Clearly, the height of these peaks decreases as we move away from the point of resonance to the right. Such behavior of $G_3(E)$ is due to the dependence of β on n_0 and n, and differs considerably from the behavior of $G_2(E)$, where all transmission coefficients provide a resonance contribution at the same point. The behavior of $G(E)$ corresponds to diagrams built according to (35) and depicted in Figs. 1a–3a. The detailed structure of $G(E)$ is depicted in Figs. 1b–3b.

Let us roughly estimate the number N_0 of peaks in the vicinity of a step threshold, where $\delta \ll 1$. Equation (65) implies that $N_0 \sim \sqrt{2N}\Delta/\pi\ell$. This estimate shows that N_0 increases with energy in proportion to $\sqrt{E/\hbar\Omega}$ and the distance between the impurities. A comparison of Figs. 1b–3b shows that the above conclusion agrees with the behavior of $G(E)$ near the step threshold.

Figures 4–6 depict the G vs. Δ dependence for different values of E. We see that as $E/\hbar\Omega$ increases, new peaks appear, but the old peaks do not change position, and change shape negligibly. Such behavior of the resonance peaks can be explained by the fact that each peak originates from one of the transmission coefficients, and the contribution of all the other peaks to the region where one of them is large is insignificant. Moreover, the diagrams imply that the number of peaks in $G(\Delta)$ increases with Δ in any neighborhood of the step threshold.

We now estimate the parameters of the peaks. We start with the behavior of the second conductance term, $G_2(E)$. The term has a maximum at a point $\delta_0 \ll 1$ such that $\gamma\zeta(1/2, 1-\delta_0) = -1$. To the left of the step threshold, where δ approaches unity, $G_2(E)$ is of order $1-\delta$ and hence is extremely small, which means that the peak begins at the point where $\delta = 0$, i.e., at the step threshold, and since $\delta_0 \ll 1$, rises almost vertically to the resonance point, where $\delta = \delta_0$. The declining part of the peak can be approximately described by a formula that follows from (37) for $\delta_0 < \delta \ll 1$. Then

$$G_2(E) \simeq \frac{1}{1 + \delta\gamma^{-2}}. \tag{66}$$

Let us now find the halfwidth $\Gamma(E)$ and the height of the resonance peaks in $G_2(E)$. The maxima in $G_2(E)$ are shifted by δ_0 to the right of the threshold of each step and have an amplitude $\Delta G \simeq 1$. We wish to find the value δ_1 at which $G_2(\delta_1) = G_2(\delta_0)/2$. An estimate of δ_1 by (66) yields γ^2, with the result that $\Gamma(E) \sim \gamma^2\hbar\Omega$. Thus, the halfwidths of all peaks in $G_2(E)$ are the same, do not depend on E, and are extremely small.

Next we examine the behavior of the third term in the conductance, $G_3(E)$, near a resonances. Since as noted earlier, the peaks in $G_3(E)$ are due to a resonance in the corresponding transmission coefficient $T_{n_0 \to n}$, below we examine the behavior of such coefficients. To the left of the step threshold, where $\delta \simeq 1$, the transmission coefficients $T_{n_0 \to n}$ are of order $(N+1-n)(1-\delta)/2(1+\cos\beta)$ and hence are small for all n_0 and n with the exception of certain value at which $\beta = \pi/2 + 2\pi k$ (these values of β correspond to a resonance, as mentioned earlier). The line of reasoning used in studying $G_2(E)$ shows that the increasing parts are almost vertical (to the left of a resonance point). To the right of a resonance point the decreasing parts of the peaks are approximately described by

$$T_{n_0 \to n} \simeq \frac{(N + \delta - n)^{-1}}{1 + \cos\beta + \delta\gamma^{-2}}. \tag{67}$$

Let us now find the halfwidth $\Gamma(E)$. Clearly, in this case we also have $\Gamma(E) \sim \gamma^2\hbar\Omega$. Since as noted earlier, only one resonance term contributes significantly in the vicinity of each peak in $G_3(E)$, this is the estimate of the halfwidth in $G_3(E)$. Thus, all peaks in $G_3(E)$ are located to the right of the step threshold, and their halfwidths are extremely small and depend neither on the energy E nor on the distance Δ between the impurities.

The relationships of Secs. 3–5 were obtained on the assumption that there is strong magnetic quantization, i.e., the hybrid length is much less than the effective radius of the wire. However, the above results concerning the corrections to Kreĭn's $\mathcal{O}$-function caused by deviations of impurities from the wire's axis imply that changes induced by these corrections in $G_2 + G_3$ are higher-order corrections $\sim \gamma^2(\rho/\ell_0)^2$. Hence the deviations of impurities from the microstructure axis have no significant effect on the position and shape of the resonance peaks in $G(E)$.

Equation (63) suggests that the σ_{zz} vs. B dependence is of an oscillatory nature. These oscillations appear when the Fermi level intersects a hybrid-oscillatory level with energy $\hbar\Omega(n+1/2)$. Note that since the Ω vs. B dependence is nonlinear, the peaks of the curve in Fig. 9 are nonuniformly spaced.

Now let us discuss the nature of the oscillation maxima near the points where $\mu = \hbar\Omega(n+1/2)$. The dependence of the scattering amplitude $t(\mu)$ on the magnetic field near the points where $\mu = \hbar\Omega(n+1/2)$ has a strong effect on the conductivity, i.e., at these points the factor $|1 + \alpha\zeta(1/2, 1/2 - \mu/\hbar\Omega)|^{-2}$ vanishes, so that the denomina-

tor of the second fraction in (63) becomes a small quantity of order $\alpha^2\hbar\Omega/\mu$. The quantity α is of order 0.1 for $B\sim 10$ T. This means that at the maxima the values of σ_{zz} increase by a factor of 100 for $\mu/\hbar\Omega\sim 10$. The overall oscillation pattern is one of broad minima separated by nonequidistant sharp peaks of the burst type.

Finally, we estimate the region where the results of Sec. 5 are valid. The transport mechanism is diffusive rather than ballistic if the length and radius of the wire meet certain conditions: $\rho,\ L>\lambda$, where λ is the mean free path in the wire, i.e., for typical values of $\lambda\sim 10^{-5}$ cm, ρ must be greater than 10^{-5} cm. For magnetic quantization to be much greater than size quantization, ρ must be much greater than l_B, which is achieved if $B>1.5$ T, i.e., under ordinary conditions, where magnetic quantization must be taken into account. Thus, in the region where electron transport is diffusive and under ordinary conditions for the magnitude of the quantizing magnetic field, magnetic quantization is stronger than size quantization, so that the results of the present section correspond to the situation with a quantum wire.

The present work was supported by the Russian Fund for Fundamental Research (Grant No. 96-01-00074), the Russian Ministry of Education, and the Universities of Russia Program.

*)E-mail: root@mathan.mordovia.su

[1] J. K. Jain and S. A. Kivelson, Phys. Rev. Lett. **60**, 1542 (1988).
[2] M. Ya. Azbel, Phys. Rev. B **43**, 2435 (1991).
[3] B. J. Van Wees, L. P. Kouwenhoven, H. Van Houten, C. W. J. Beenakker, J. E. Mooij, C. T. Foxon, and J. J. Harris, Phys. Rev. B **25**, 3625 (1988).
[4] V. A. Geĭler and V. A. Margulis, Zh. Éksp. Teor. Fiz. **111**, 2215 (1997) [JETP **84**, 1209 (1997)].
[5] J. F. Weis and K. F. Berggen, Phys. Rev. B **40**, 1325 (1989).
[6] D. A. Wharam, U. Ekenberg, M. Pepper, D. G. Hasko, H. Ahmed, J. E. F. Frost, D. A. Ritchie, D. C. Peacock, and G. A. C. Jones, Phys. Rev. B **39**, 6289 (1989).
[7] M. Büttiker, Phys. Rev. B **41**, 7906 (1990).
[8] H. A. Fertig and B. I. Halperin, Phys. Rev. B **36**, 7969 (1987).
[9] Y. B. Levinson, M. I. Lubin, and E. V. Sukhorukov, Phys. Rev. B **45**, 11 936 (1992).
[10] Y. B. Levinson, M. I. Lubin, and E. V. Sukhorukov, JETP Lett. **54**, 401 (1991).
[11] M. I. Lubin, JETP Lett. **57**, 361 (1993).
[12] D. H. Gobden, N. K. Patel, M. Pepper, D. A. Ritchie, J. E. F. Frost, and G. A. C. Jones, Phys. Rev. B **42**, 1938 (1991).
[13] A. B. Fowler, G. L. Timp, J. J. Wainer, and R. A. Webb, Phys. Rev. Lett. **57**, 138 (1986).
[14] T. E. Kopley, P. L. McEuen, and R. G. Wheller, Phys. Rev. Lett. **61**, 1654 (1988).
[15] P. L. McEuen, B. W. Alphenaur, R. G. Wheller, and R. N. Sacks, Surf. Sci. **312**, 229 (1990).
[16] S. J. Bending and M. R. Beasley, Phys. Rev. Lett. **55**, 324 (1985).
[17] Y. Xu, A. Matsuda, and M. R. Beasley, Phys. Rev. B **42**, 1492 (1990).
[18] M. W. Dellow, P. H. Beton, C. J. G. M. Landerak, T. J. Foster, P. C. Main, L. Eaves, M. Henini, S. P. Beaumont, and C. D. W. Wilkinson, Phys. Rev. Lett. **68**, 1754 (1992).
[19] A. K. Geim, P. C. Main, N. La Scala Jr., L. Eaves, T. J. Foster, P. H. Foster, P. H. Beton, J. W. Sakai, F. W. Sheard, M. Henini, G. Hill, and M. A. Pate, Phys. Rev. Lett. **72**, 2061 (1994).
[20] A. K. Geim, T. J. Foster, A. Nogaret, N. Mori, P. J. McDonnell, N. La Scala Jr., P. C. Main, and L. Eaves, Phys. Rev. B **50**, 8074 (1994).
[21] C. S. Chu and R. S. Sorbello, Phys. Rev. B **40**, 5941 (1989).
[22] P. F. Bagwell, Phys. Rev. B **41**, 10 354 (1990).
[23] T. Martin and S. Feng, Phys. Rev. Lett. **64**, 1971 (1990).
[24] S. B. Kaplan and A. C. Warren, Phys. Rev. B **34**, 1346 (1986).
[25] C. T. Liu, K. Nakamura, D. C. Tsui, K. Ismail, D. A. Antoniadis, and H. I. Smith, Appl. Phys. Lett. **55** (2), 168 (1989).
[26] C. W. J. Beenakker and H. Van Houten, in *Solid State Physics: Advances in Research and Applications*, H. Ehrenreich and D. Turnbull (eds.), Vol. 44, Academic Press, New York (1991), p. 83.
[27] A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions, and Decay in Nonrelativistic Quantum Mechanics* [in Russian], 2nd ed., Nauka, Moscow (1971).
[28] Yu. N. Demkov and G. F. Drukarev, Zh. Éksp. Teor. Fiz. **49**, 257 (1965) [Sov. Phys. JETP **22**, 182 (1966)].
[29] Yu. N. Demkov and V. N. Ostrovskiĭ, *The Method of Zero-Radius Potentials in Atomic Physics* [in Russian], Leningrad Univ. Press, Leningrad (1975).
[30] V. A. Geĭler, V. A. Margulis, and I. I. Chuchaev, JETP Lett. **58**, 648 (1993).
[31] V. A. Geĭler and V. A. Margulis, Teor. Mat. Fiz. **58**, 461 (1984).
[32] V. A. Geĭler and V. A. Margulis, Teor. Mat. Fiz. **61**, 140 (1984).
[33] V. A. Geĭler, V. A. Margulis, I. V. Chudaev, and I. I. Chuchaev, Zh. Éksp. Teor. Fiz. **107**, 187 (1995) [JETP **80**, 100 (1995)].
[34] B. A. Lippmann, Phys. Rev. **15**, 11 (1965).

Translated by Eugene Yankovsky

Spectroscopy of quasiparticle excitations of superconducting bismuth cuprate at high pressures

V. M. Svistunov*), V. Yu. Tarenkov, and A. I. D'yachenko

A. A. Galkin Donetsk Physicotechnical Institute, Ukrainian National Academy of Sciences, 340114 Donetsk, Ukraine

R. Aoku

College of Industrial Technology, Amagasaki Nishi-Koya, Japan

H. Murakami and M. Shirai

Osaka University, Japan

(Submitted 5 August 1997)
Zh. Éksp. Teor. Fiz. **113**, 1397–1410 (April 1998)

We study the energy spectrum of Bi2223 ($Bi_{1.6}Pb_{0.4}Sr_{1.8}Ca_{2.2}Cu_3O_x$) at high hydrostatic pressures by Andreev- and tunneling-spectroscopy methods. We determine the gap anisotropy in the basal ab plane and find the following values for the parameters $\Delta(\varphi)$: $\Delta_{\max}=42$ mV, and $\Delta_{\min}=19.5$ mV ($T_c=110$ K and $dT_c/dP=0.16$ K/kbar). We detect an increase in the ratio $R=2\Delta_{\max}/kT_c$ with pressure P; for Bi2223 cuprate, $dR/dP\approx 0.017$ kbar^{-1}. In the phonon-frequency region we detect a ''softening,'' due to pressure, of the high-frequency part of the phonon spectrum corresponding to ''breathing'' modes of oxygen, as well as other optical modes of Cu–O. The characteristic frequencies of the spectrum for $\hbar\Omega>60$ mV are found to decrease, with increasing pressure, at a rate $d\ \ln(\hbar\Omega)/dP\approx -6.5\pm 0.5\times 10^{-3}$ kbar^{-1}. This result explains the observed increase in the ratio $2\Delta/kT_c(P)$ in the model of strong electron–phonon interaction. © *1998 American Institute of Physics.* [S1063-7761(98)02004-6]

1. INTRODUCTION

It is impossible to decipher the pairing mechanisms in high-T_c superconductors without analyzing the excitation spectrum of the superconductor over a broad energy range. Junction methods are the main source of such information, especially the tunneling effect in S–I–S structures.[1] In junctions with direct conductivity of S–c–S and S–c–N type, the spectroscopic characteristics of the superconductor are observed in the Andreev-reflection regime[2] (S stands for a superconductor, c for a constriction, N for a normal metal, and I for an insulator). In this paper we use both junction regimes to study the effect of pressure on the gap and phonon spectra of a bismuth metal oxide, with each regime providing complementary information.

For isotropic superconductors the theory of electron–phonon interaction yields a unique relation between the ratio $2\Delta/kT_c$ and the phonon spectrum: as the phonon frequencies increases, $2\Delta/kT_c$ must decrease, while softening of the phonon frequencies must lead to an increase in this ratio.[3] This conclusion is fully corroborated by tunneling experiments at high pressure, which serve as an additional argument in favor of the phonon mechanism of superconductivity.[4,5] Here the gap anisotropy of low-temperature superconductors plays no important role, since it is rapidly suppressed by elastic scattering by impurities.[1]

The situation is different for high-T_c superconductors. A characteristic feature of such superconductors is strong gap anisotropy,[6] which accounts for the anomalous value of the ratio, $2\Delta/kT_c\geqslant 7$, and probably ensures the high value of T_c (see Ref. 7). Hence complete information about the mechanism of high-T_c superconductivity can only be obtained if we take gap anisotropy into account. For instance, to determine $\Delta(\varphi)$ of cuprates from tunneling-spectroscopy studies, junctions must be created in each crystallographic direction, which technologically is extremely difficult.[8] At the same time, the anisotropic energy gap in a superconductor and the changes in the gap produced by high pressure can be deduced from Andreev-reflection measurements.[9] The experiment can involve a single S–c–N junction and is therefore more promising.

As a result of Andreev reflection from the N–S interface, there is inversion of electrons into holes, with a hole reflected in the same direction in which the respective electron had been moving. This leads to a doubling of the junction conductance $G(V)$ at $V=0$, i.e., $G(0)/G(eV\gg\Delta)\approx 2$ (see Refs. 10 and 11). Such behavior in the reflection is retained even for large electron incidence angles. As a result, if an NS microjunction is prepared on a single crystal, Andreev-reflected electrons carry information about the value of $\Delta(\mathbf{k})$ for all directions of the wave vector $\mathbf{k}$ (see Ref. 11). Thus, measuring the conductance of a single point NS junction makes it possible to completely reconstruct the anisotropy $\Delta(\mathbf{k})$ of the energy gap of the superconductor.

In our research we used three types of junction, each with its own advantages in studies of electron–phonon inter-

actions in high-T_c superconductors. Andreev junctions of the S–c–N type make it possible to find gap anisotropy of cuprates in the ab plane and the reaction of gap anisotropy to pressure. In S–c–S microjunctions no anisotropy effects were observed, but the junctions proved suitable for the elastic Andreev spectroscopy of phonons at high pressure. Finally, tunnel measurements were found to yield exhaustive information about phonons but cannot be used to measure the size of the energy gap and gap anisotropy to high accuracy.

2. GAP ANISOTROPY

Information about gap anisotropy and the phonon spectrum was obtained from the characteristics of S–c–N and S–c–S microjunctions and tunnel S–I–S structures. The objects of investigation were bismuth cuprates ($Bi_{1.6}Pb_{0.4}Sr_{1.8}Ca_{2.2}Cu_3O_x$) (95% of the Bi2223 phase, $T_c = 110$ K) prepared by the solid-phase synthesis method from chemically pure oxides. Bismuth ceramic plates $1\times0.1\times0.01$ cm^3 were prepared by subjecting a powder of the Bi2223 phase to a pressure of 30–40 kbar between two steel anvils. For the bandage supports of the powder being pressed we used copper wires with a diameter $D=0.1$–0.2 mm fastened to the anvil surface parallel to each other. As a result of pressure the powder was compacted into dense plane-parallel plates with a thickness $d\leqslant0.1$ mm. The ceramic samples with current and potential leads manufactured from a silver paste were then annealed at $T=845$ °C.

The method of fabrication of S–c–N microjunctions was based on pressing a fine silver powder into a ceramic material. The silver powder with particles of diameter $d=3$ μm was spread over the surface of the ceramic plate and covered with a drop of glue. Then the sample was placed in a hydrostatic press and subjected to a pressure of up to 15 kbar. The high mobility of the silver powder and hydrostatic pressing resulted in the powder being pressed into the surface of the cermet, and a stable metallic junction formed. The outer diameter of such a junction was 1–1.5 mm, while the area of the spectroscopic S–c–N microjunction estimated by Sharvin's formula (see Ref. 2) is much smaller, about 10^{-11}–10^{-12} cm^2. A pure S–c–N junction is probably produced when microscopic cracks are formed in the thin microcrystals of the ceramic plate under shear deformations that appear in the pressing process and the simultaneous influx of the silver powder into the cracks. Then, probably, the Sharvin junction is formed at the fracture of a microcrystal along the (001) or (010) plane. Note that what is measured in this case are the characteristics of fairly perfect microcrystals of the ceramic material. This is suggested by the high repeatability of the Andreev spectra for different junctions.

In addition to normal-metal–superconductor structures we manufactured and studied S–I–S junctions of the break junction type. Here is a brief description of the method used in manufacturing such junctions. A thin ceramic plate with current and potential leads was placed on a support manufactured from a textolite foil. Then the plate was covered by a thick, $d\sim0.5$ mm, layer of varnish. After polymerization of the varnish the substrate was bent until the ceramic plate broke, and this moment was monitored by the change in the resistance of the plate. The resistance of the break junctions fabricated at room temperature was either metallic or tunnel in nature. The stability of the characteristics of the samples made it possible to study the samples at low temperatures and high pressures. For spectroscopic measurements we used a high-pressure chamber of the piston–cylinder type,[12] which contained not only the sample but also pressure and temperature pickups. To measure the characteristics we used a standard four-probe circuit. The differential conductance curves were recorded by a circuit with a highly constant modulating voltage.

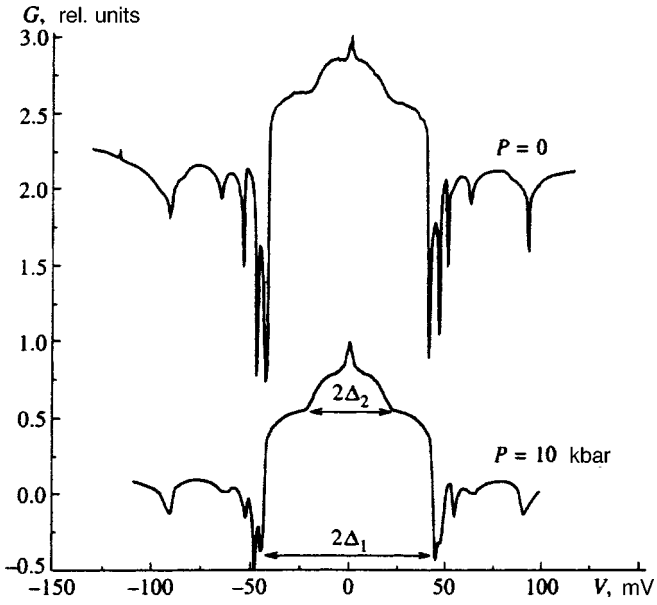


FIG. 1. The characteristic of the Bi2223–Ag junction (S–c–N) demonstrating the singularities of the anisotropic energy gap at $eV=2\Delta_1$ and $eV=2\Delta_2$ at $P=0$ and 10 kbar.

The characteristic spectrum (i.e., the voltage dependence of $G(V)=dI/dV$) for a Bi2223–Ag microjunction is depicted in Fig. 1. The main features of the spectrum are easily reproducible for junctions that differ greatly in their conductance. Note that the position of the local minima in the dynamic conductance (the resistance peaks) at $eV=\Delta_1=42\pm0.5$ meV ($P=0$) corresponds to the size of the of the superconducting energy gap $\Delta_{ab}=40$–50 meV of Bi2223 cuprates previously found in tunneling measurements in the ab plane.[6,8,13] We obtained close values of Δ_{ab}(Bi) from tunneling S–I–S spectra. And although, as we will see below, the peaks in the tunneling conductance at $eV\approx\Delta$ are clearly visible, their considerable smearing made it impossible to determine quantitatively the rate of change of the parameter Δ under pressure. This drawback does not exist for Andreev NS microjunctions (Fig. 1).

In a series of experiments with microjunctions whose resistance differed considerably, the singularities depicted in Fig. 1 at $eV=\Delta_1,\Delta_2$ were found to occupy the same positions, which suggests that the nature of these singularities is spectroscopic. Another indication of the relation of these singularities to the superconductivities of cuprates is the temperature dependence of $\Delta(T)$. For instance, the Δ_1 vs. T

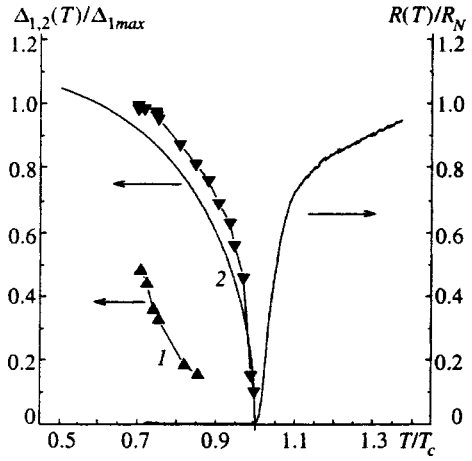


FIG. 2. Temperature curves for Δ_2 (curve *1*) and Δ_1 (curve *2*); the solid curves represent $\Delta(T)_{\rm BCS}$ and $R(T)$, the superconducting transition of the ceramic material being studied.

curve for Bi2223–Ag microjunctions depicted in Fig. 2 is similar to the well-known BCS $\Delta(T)$ curve for the energy gap of a superconductor. On the hand, the singularity for the small gap Δ_2 rapidly flattened out with increasing temperature and had a quite different (non-BCS)nature. Such temperature curves for Δ_1 and Δ_2 of bismuth cuprate were discussed by Pashitskiĭ and Pentegov[14] and was observed in the experiments of Ma *et al.*[15] involving IR-spectroscopy with high angular resolution.

The characteristic features of the resulting spectra (Fig. 1) are sharp dips in dynamic conductance whose positions are related to the energy gap Δ_1. We believe that these dips are a manifestation of an effect accompanying Andreev reflection and due to the high degree of locality of the feed current of a Sharvin microjunction. Earlier it was noted that in cuprates, as well as in ordinary superconductors[16] (including superconductors with heavy fermions[17]), the dynamic resistance of N–S junctions under a bias voltage $V=V_d$ may exhibit peaks, with V_d related to Δ by $V_d \approx \Delta/e$. The temperature dependence of a peak at V_d was found to follow the temperature dependence of $\Delta(T)$.

The resistance peak in the dynamic characteristics of metallic N–S junctions results from nonequilibrium processes at quasiparticle energies $2eV=2eV_d \sim 2\Delta$ (Ref. 18) or at high measuring current densities.[16,17] What is important here is that irrespective of the specific mechanism of singularity formulation in the spectrum at $V=V_d$, the relationship of this singularity to the size Δ of the energy gap of the superconductor yields the convenient possibility of quantitatively following the changes in $\Delta(\varphi)$ induced by hydrostatic pressure.

By analyzing the experimental curves, we found that the break at $eV=\Delta_2$ (Fig. 1) is probably a reflection of the gap anisotropy in cuprates in the ab plane. The singularity at $eV=\Delta_2$ cannot be related to the energy gap in the c-direction, since in this direction the Fermi velocity V_{Fs} in metal oxides is of order 10^6 cm/s, which is much less than the Fermi velocity in silver, $V_{Fn} \sim 10^8$ cm/s. Hence the effectiveness of Andreev reflection in the direction of the c axis must drop considerably (by a factor of approximately V_{Fn}/V_{Fs}).[10] The effect of inhomogeneity of the order parameter due to structural defects or impurities can also be ruled out, since the value of Δ_2 was reproduced in measurements involving a large number of samples. The narrow conductance peak near the zero-bias potential is often observed in experiments with N–S junctions, but its nature has yet to be explained (see, e.g., Ref. 19).

As the pressure grows, $\Delta_1=\Delta_{\max(ab)}$ increases and $\Delta_2=\Delta_{\min}(ab)$ decreases, i.e., the gap anisotropy $a=\Delta_{\max}/\Delta_{\min}$ gets stronger, and for Bi2223 the gap-anisotropy increase rate was found to be $da/dP=0.003\pm0.001$ kbar^{-1}. The critical temperature increase rate was found to be $dT_c/dP=16$ K/kbar, which yields the rate of change of the ratio $R=2\Delta_{\max(ab)}/kT_c$, namely $dR/dP=0.017\pm0.005$ kbar^{-1}.

To get a feeling for the nature of the observed curve, we calculated the conductivity $\sigma(V)$ of a normal-metal–anisotropic-superconductor microjunction. For simplicity, we assumed that anisotropy of the energy gap Δ in the ab plane is the most important and that the Blonder–Tinkham–Klapwijk parameter Z is small.[10] Then

$$G(V)\approx 1+\frac{2}{\pi}\int_0^{\pi/2} d\varphi |A(V,\varphi)|^2,$$

$$A=\frac{E+i\Gamma-\{(E+i\Gamma)^2-\Delta^2\}^{1/2}}{\Delta}, \qquad (1)$$

where $\Delta=|\Delta(\varphi)|$, and the Dines parameter Γ accounts for the smearing of the singularity in the Gor'kov anomalous Green's function. The dependence of the parameter Δ on the azimuthal angle φ in the ab plane was approximated by the formula

$$\Delta(\varphi)=\Delta_0\left(1+\sum_k C_k \cos(2\varphi k)\right), \qquad (2)$$

which makes it possible to allow for s- and d-pairing effects simultaneously. The parameters C_k in (2) were selected so that the results provided by (1) would reasonably fit the experimental $G(V)$ dependence (Fig. 1). Equation (1) for calculating the function $G(V)$ reflected the main features of the experimental curve (Fig. 1) for $eV\leqslant 2\Delta$ (with the exception of the peak at small bias voltages), which made it possible to relate the singularities of the observed spectrum of the NS junction to the anisotropy of the energy gap $\Delta(\varphi)$ of the superconductor.

It is currently believed that the values of T_c and $R=2\Delta/kT_c$ in cuprates are high as a result of substantial anisotropy of the energy gap in the ab plane.[6,7] The observed increase in the ratio R and in gap anisotropy at high pressure agrees with these ideas. Still, this is not an argument in favor of the nonphonon mechanism of superconductivity, since appreciable gap anisotropy and d-wave superconductivity may also be possible for a modified electron–phonon pairing mechanism (see, e.g., Ref. 20).

3. PHONON SPECTRUM

The unusual symmetry of the order parameter in high-T_c is corroborated by numerous experiments, but still electron–

phonon interactions play an important role in pairing.[13,21–23] The phonon spectrum of cuprates extends to 100 mV and, as previous research has shown,[2,13,21] there is strong coupling between electrons and high-energy phonons. In the present paper this is corroborated by experiments in elastic spectroscopy of the electron–phonon interaction in S–c–S microjunctions of the Andreev type and in tunnel S–I–S samples.

Tunnel investigations of ordinary superconductors at high pressure demonstrate beyond any doubt that the deviation of the ratio $R=2\Delta/kT_c$ from the universal BCS value $R=3.53$ is due to strong electron–phonon interaction.[24] In anisotropic high-T_c superconductors, there is no universal relation between the maximum energy gap $2\Delta_{\max}$ and the transition temperature T_c. However, one can expect an appreciable contribution from electron–phonon interactions to the ratio $R=2\Delta_{\max}/kT_c$ for the nonphonon pairing mechanism. As in the case of ordinary superconductors, this contribution shows up at high pressure, since the phonon frequencies contribute the most to variations of the ratio $R(P)$ with pressure (these frequencies comprise the lower part of the bosonic spectrum of cuprates). The greatest contribution to superconductivity is provided by phonons with large vectors $q\sim\pi/a$, where a is the lattice constant. These are the phonons detected by junction methods and neutron spectroscopy methods.[1,2] Hence the information about $\omega(P)$ obtained via these methods reflects the essence of the electron–phonon interaction in superconductors more accurately than the Raman spectroscopy method, which determines the phonon frequencies $\omega_{\mathbf{q}}$ only at $q=0$, i.e., at the center of the Brillouin zone.

The elastic Andreev spectroscopy of bismuth metal oxide involved using S–c–S junctions, which made it possible to simultaneously determine the ratio $2\Delta/kT_c$ and the shift of phonon frequencies $\omega(P)$ for different pressures. The feasibility of such spectroscopy in Andreev-type junctions is based on the fact that for strong electron–phonon interaction, the dynamic conductance G ''feels'' the frequency dependence of the complex-valued gap function $\Delta(\omega)$ (see Ref. 25):

$$G=\frac{dI}{dV}=\frac{1}{R_N}\left\{1+\left|\frac{\Delta(\omega)}{\hbar\omega+[(\hbar\omega)^2-\Delta^2(\omega)]^{1/2}}\right|^2\right\}_{\omega=eV/\hbar}, \tag{3}$$

where R_N is the resistance of the junction in the normal state. From (3) it follows that the derivative dG/dV of S–c–N and S–c–S junctions reflects the behavior of $\Delta(\omega)$ and hence the phonon spectrum of the superconductor. For S–c–S junctions the singularities in the conductance G due to strong electron–phonon interaction ought to show up at biases $eV=\hbar\omega_i+2\Delta$, where ω_i are the characteristic frequencies of the phonon density of states $F(\omega)$.

We measured break-junction S–c–S junctions whose metallic nature was monitored both by the temperature dependence of their conductance and by the weak reaction of $G(0)$ of the junctions to pressure. At low temperatures $T<T_c$, the current–voltage characteristics of S–c–S samples had excess current, which is a true sign of junctions with direct conductance.[2]

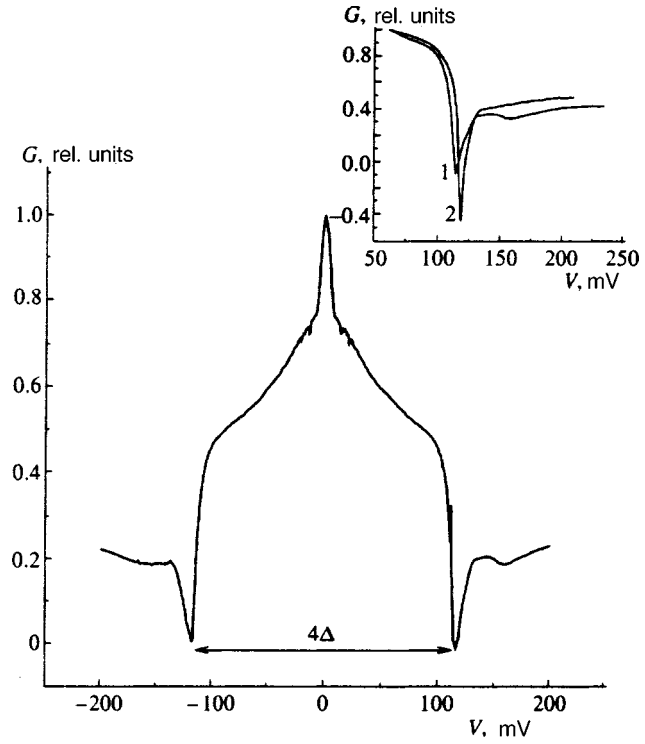


FIG. 3. Effect of gap singularities in the conductance of the Bi2223–Bi2223 microjunction (S–c–S). The inset illustrates the effect of hydrostatic pressure on the position of the singularity at $eV=2\Delta$ in the spectrum: curve *1* corresponds to a pressure $P=0$, and curve *2* to a pressure $P=10$ kbar.

Figure 3 depicts $G(V)$ for an S–c–S junction, demonstrating the emergence of an energy gap at $eV=2\Delta=114$ meV ($T=77$ K). The observed discrepancy in the values of the energy gap parameter for Bi2223 obtained from measurements of an S–c–N junction ($\Delta=42$ mV) and an S–c–S junction ($\Delta=57$ mV) may be related to partial suppression of the order parameter and the S–N interface due to the proximity effect. As in the case of an S–c–N junction, the temperature dependence $\Delta(T)$ of an S–c–S junction corresponds to the BCS curve. Note that here there is no singularity corresponding to the small energy gap Δ_2, which is probably due to the special features of multiple Andreev reflection of an electron between two anisotropic superconductors.[18]

At bias voltages eV higher than 2Δ, the conductance of the S–c–S was found to exhibit singularities, which can be interpreted as a reflection of the phonon structure of the metal oxide being investigated (Fig. 4). The spectroscopic nature of these curves appears to be proved by the fact that for different junctions, either with tunneling conduction or with direct conduction, the arrangement of the singularities in the conductance measured from the sum of the energy gaps, 2Δ, was the same. To establish the nature of the phonon structure in the spectra of Andreev-type junctions, we calculated the junction conductance with Eq. (3), where the complex-valued gap parameter for cuprates, $\Delta(\omega)$, was determined from the solution of the Eliashberg equations for the phonon density of states $F(\omega)$ of Bi2223 (see Ref. 26). The value of the electron–phonon coupling constant $\alpha^2(\omega)$ in the function $g(\omega)=\alpha^2(\omega)F(\omega)$ was found by the method

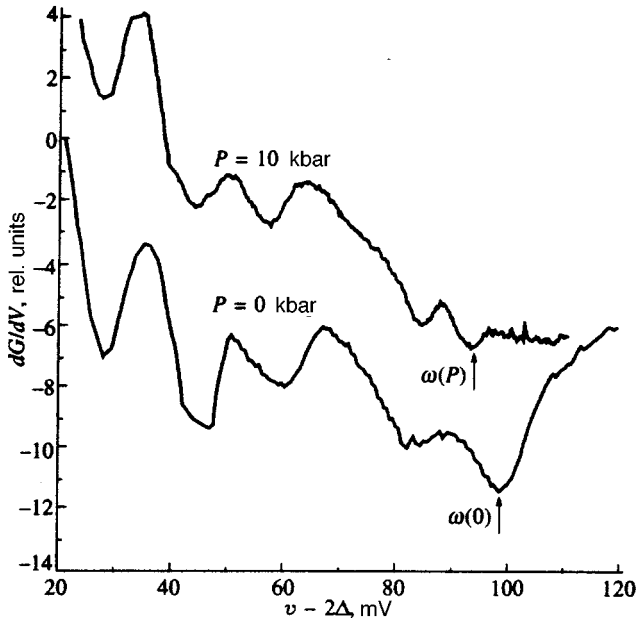


FIG. 4. The spectrum of dG/dV in S–c–S junctions of the Andreev type at zero and finite pressure. The arrows indicate singularities of the curves that proved to be pressure-sensitive.

developed in Ref. 27, while the electron–phonon coupling constant λ was found by requiring that the calculated transition temperature match the experimental value of T_c. At $\lambda \approx 3.3$ and the value 0.1 for the Coulomb coupling constant $\mu^* = 0.1$, the calculated value of T_c was found to be 110 K (the corresponding isotropic energy gap Δ_0 was found to be 22 meV). We see (Fig. 5) that the general structures of the observed and calculated features of dG/dV agree. The coincident locations and similarity of the phonon features in the Andreev (Fig. 4) and tunneling (Fig. 6) spectra make it possible to ignore the possible distortions in the conductance of S–c–S junctions related to inelastic processes.[2]

The effect of 10 kbar on the structure of the second derivative of current of one of the Andreev S–c–S junctions is shown in Fig. 4. The initial region of the phonon spectrum of Bi2223 changes little under pressure, $d\ln(\omega)/dP = (1-1.5) \times 10^{-3}$ kbar^{-1} (this agrees with the Raman spectroscopy data in Refs. 21 and 28). The most significant changes take place in the high-frequency part of the spectrum at $\hbar\omega = 70-95$ meV, where pressure is found to considerably reduce the phonon energy at a rate $d\ln(\omega)/dP = -6 \times 10^{-3}$ kbar^{-1}.

Although the results of elastic Andreev spectroscopy at high pressures can easily be reproduced, this method has not gained wide acceptance, in contrast to tunneling spectroscopy. This fact stimulated similar tunnel investigations. Comparative analysis of the effect of high pressure on the phonon spectrum of the metal oxide conducted in conjunction with the data obtained by the Andreev and tunneling spectroscopy methods greatly increases the credibility of the results.

For the tunnel current in isotropic superconductors we can write[1]

$$G(V) = \left.\frac{dI}{dV}\right|_{eV=\hbar\omega} \sim N(\omega) = \mathrm{Re}\,\frac{\hbar\omega}{[(\hbar\omega)^2 - \Delta^2(\omega)]^{1/2}}, \tag{4}$$

where the complex-valued energy gap $\Delta(\omega)$ of the superconductors has singularities at frequencies $\hbar\omega = \Delta_0 + \hbar\Omega$ corresponding to virtual phonons with an energy $\hbar\Omega$. Here Δ_0 is the BCS gap, and $N(\omega)$ is the tunnel density of states. In the derivative of the tunnel conductance, dG/dV, the phonon structure manifests itself in the form of dips, whose positions give the phonon frequencies to high accuracy.[1] For S–I–S junctions, such singularities appear at biases $eV_i = 2\Delta_0 + \hbar\Omega_i$. As shown in Ref. 29, the condition $eV_i = 2\langle\Delta\rangle + \hbar\Omega_i$ (where $\langle\Delta\rangle$ is an ''average'' gap) remains valid for strongly anisotropic superconductors (Bi2223 is

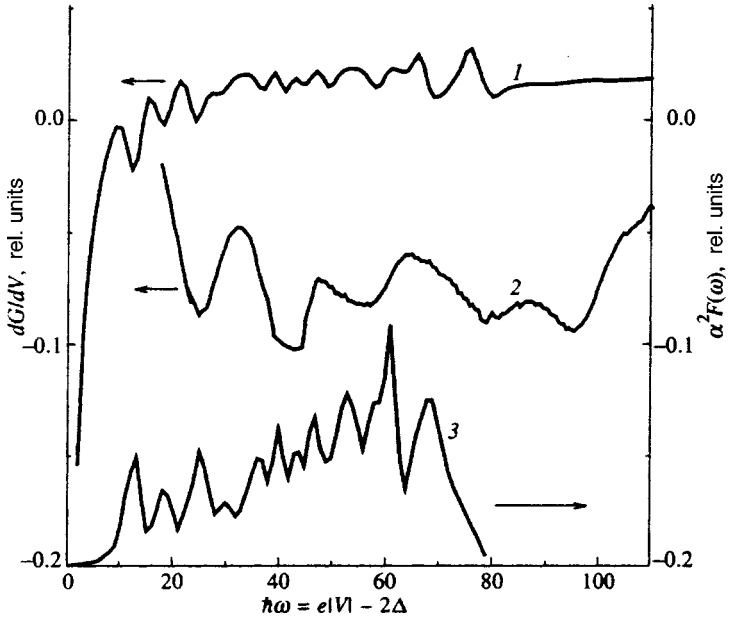


FIG. 5. Comparative characteristics of the experimental and calculated curves for a bismuth metal oxide: curve *1* represents dG/dV reconstructed from $\alpha^2F(\omega)$, curve *2* represents the experimental behavior of dG/dV, and curve *3* represents the behavior of the density of states $F(\omega)$ for a bismuth metal oxide.[26]

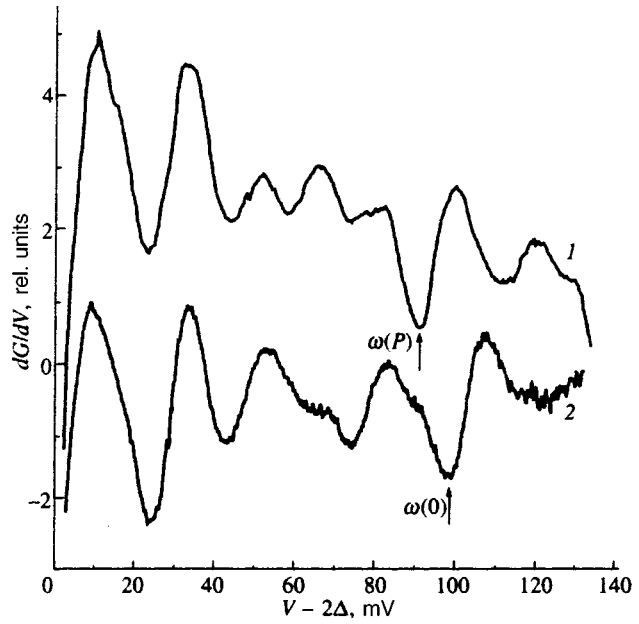


FIG. 6. dG/dV of a tunnel Bi2223–Bi2223 break junction at $P=0$ (curve *2*) and 10 kbar (curve *1*).

such a superconductor). This makes it possible to determine the energies of the characteristic phonon frequencies Ω_i of an anisotropic superconductor by measuring the tunnel dG/dV spectra.

The inset in Fig. 7 depicts the behavior of $G=dI/dV$ for an S–I–S junction with an energy gap at $eV=2\Delta=75$ meV. Similar values of 2Δ for the bismuth metal oxide with $T_c=110$ K have been obtained by Kane and Ng.[8] The observed smearing of the gap characteristics is inherent in tunnel high-T_c superconductors[13] and emerge in high-quality junctions because of substantial anisotropy in $\Delta(\mathbf{k})$. However, this anisotropy does not prevent the phonon structure in the tunnel spectrum from being detected, since the characteristic values of phonon frequencies are linked to a certain average value $\langle\Delta\rangle$ (see Ref. 29).

To establish the extent to which the spectrum of the tunnel sample reflects the bulk properties of the ceramic material, we measured the temperature dependence $G(T)$ junction conductance at zero bias. By comparing this behavior with $R(T)$ for the transition of the ceramic material to the superconducting state (Fig. 7), we found that the point at which the resistance $R(T)$ disappears is essentially the same as the break in $G(T)$ for the junction, which corresponds to the point at which the energy gap of the superconductor begins. In this case the spectroscopic characteristics of the junction are determined by the macroscopic properties of the sample.

At bias voltages eV higher than 2Δ, the derivative of the junction conductance $dG/dV=d^2I/dV^2$ exhibits features related to the phonon structure of the metal oxide (Fig. 6). Comparison of the dG/dV spectra of a tunnel S–I–S break junction and Andreev S–c–S microjunctions shows (Figs. 4 and 6) that the curves coincide in both the number and location of the features (if units of measurement along the horizontal axes are the same, $V-2\Delta$). This points to the same nature of the two spectra. According to Ref. 26, the phonon spectrum of Bi2223 cuts off in the vicinity of 80–90 meV, while the tunnel characteristics and the Andreev-reflection spectra demonstrate the existence of a feature at $eV-2\Delta$

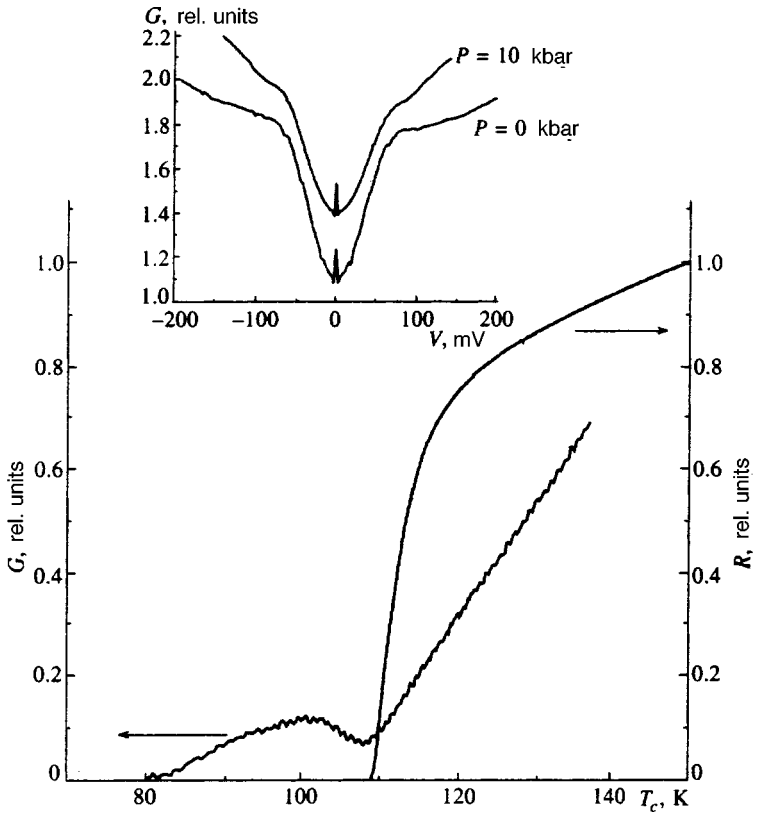


FIG. 7. Temperature dependence of the conductance G of a tunnel Bi2223–Bi2223 junction at zero bias, and $R(T)$, the superconducting transition of the ceramic material being studied. The inset illustrates the effect of the energy gap Δ on the conductance of the tunnel junction at $P=0$ and 10 kbar.

$=90-95$ meV. Note that for tunnel S–I–S and Andreev S–c–S junctions, results with coinciding (in location) characteristic phonon frequencies were obtained despite the fact that for the S–C–S junction the parameter $2\Delta=2\Delta_{\max}=114$ meV is much larger than in tunnel junctions ($2\Delta=70-84$ meV).

Applying hydrostatic pressure up to 10 kbar introduced no significant changes in the tunnel value of the gap, which is probably due to the large smearing of the average tunnel density of states (inset in Fig. 7). Note that the change in location of the beginning of the tunnel gap (Fig. 7) induced by applying pressure is essentially the same as the pressure dependence of the transition temperature $T_c(P)$ of the junction edges, at a rate $d\ \ln T_c/dP=1.4\times10^{-3}$ kbar^{-1}.

Figure 6 depicts the effect of pressure on the features of the spectrum of Bi2223. As in the case of Andreev-type junctions, the high-frequency part of the spectrum near $eV\sim90$ meV proved to be the most sensitive to pressure. This part is probably related to the vibrational modes of oxygen and its environment, and in particular to ''breathing'' modes. As the pressure rises, the rate of shift of the mode ω depicted in Fig. 6 by an arrow is negative and amounts to $d\ \ln(\omega)/dP\approx-6.5\pm0.5\times10^{-3}$ kbar^{-1}. At the same time, the lower part of the phonon spectrum ($\hbar\omega<40$ meV) shifts at a rate $d\ \ln(\omega)/dP\approx1\times10^{-3}$ kbar^{-1} toward higher frequencies. These data fully agree with the above results for Andreev spectra.

Our results concerning the variation of the characteristic phonon frequencies with pressure make it possible to calculate the effect of pressure on the ratio $2\Delta/kT_c$ in cuprates in the strong electron–phonon interaction approximation. The calculation were done according to the method described above, where the variation of the function $F(\omega)$ with pressure was found with allowance for the experimental values of $\omega(P)$. The calculated variation of the ratio $R=2\Delta/kT_c$ under pressure was found to be $\delta=[R(P)-R(0)]/R(0)=0.018$, which is very close to the experimental value $\delta=0.017$ and the value $\delta=0.02$ obtained via the Geĭlikman–Kresin formula[3]

$$\frac{2\Delta_{\max}}{kT_c}=C\left(1+5.3\left[\frac{T_c}{\omega_0}\right]^2\ln\left[\frac{\omega_0}{T_c}\right]\right) \tag{5}$$

with the characteristic phonon frequency ω_0 of the breathing mode being equal to 75 meV. Here $C(\Delta)$ is a constant whose value is determined by the gap anisotropy function $\Delta(\mathbf{k})$ (see, e.g., Ref. 30). In the isotropic BCS theory, $C=3.53$.

4. CONCLUSIONS

Our experimental results indicate that in the Bi2223 cuprate, the high-frequency phonons, related to optical vibrations of oxygen atoms and its environment, undergo the greatest change when pressure is applied. These frequencies correspond to energies of 70–95 MeV and, in particular, to the breathing mode at $\hbar\omega=70-75$ meV. The acoustic and optical modes of the phonon spectrum with energies $\hbar\omega<40$ meV shift toward higher frequencies at a rate close to the values provided by Raman spectroscopy data.[21,28]

A characteristic feature of ordinary electron–phonon interactions in superconductors is an increase in the frequency of the phonon modes, ω_q, with increasing pressure.[24] A rare exception here is lanthanum, for which softening of the low-frequency part of the spectrum is observed. In our case we detected a softening of the high-frequency phonon modes with wave vectors at the edge of the Brillouin zone. A similar effect of softening of the edge LO phonon modes under doping was observed in YBCO, LaSrCuO, and BaKBiO by employing the inelastic neutron scattering method. In $La_{2-x}Sr_xCuO_4$, the energy $(\pi,0)$ of the phonon mode was found to decrease from 79 meV at $x=0$ to 70 meV at $x=0.15$ (Ref. 31); a similar mode in $YBa_2Cu_3O_{7-x}$ was found to soften from 77 meV at $x=1$ ($T_c=0$) to 58 meV at $x=0$ ($T_c=92$ K) (Ref. 22); in $Bi_2Sr_2(Ca_{1-x}Y_x)Cu_2O_8$ significant softening of the ion modes was observed as x varied from 1 to 0 ($T_c=82$ K) (Ref. 32). In all these cases the increase in the transition temperature of cuprates was accompanied by a decrease (softening) in the high-frequency modes of the phonon spectrum. We observed a similar effect: an increase in T_c accompanied by an increase in pressure is accompanied by a decrease in the energy of the upper frequencies of the phonon spectrum. Hence it is possible that the microscopic reasons for the softening of the phonon modes under pressure and under concentration variations are the same.

The superconducting properties of cuprates manifest themselves at the edge of the Mott metal–insulator transition.[33] In view of the hypothesis concerning the universal nature of this transition, the reaction of the critical temperature T_c to pressure P and concentration c of impurities is determined by a function of a dimensionless parameter x characterizing the closeness of the system to the phase transition point, at which $x=x_c$. In this region, the difference $x-x_c=P/P_c+C/C_c$, where P_c and C_c are parameters characterizing the effect of pressure and impurity concentration on the correlation length.

Recently the significant softening of the edge optical phonon modes near a metal–insulator junction has been computer-simulated in the two-band model of the Peierls–Hubbard Hamiltonian.[34] The results show that the reason for such softening may be electron–electron correlations, which make it profitable for the energy of the high-frequency LO oxygen modes to decrease.

We note, finally, that the observed decrease in the frequencies of the upper part of the phonon spectrum explains the increase in the ratio $2\Delta_{\max}/kT_c(P)$ irrespective of the nature of the effect. Actually, agreement with the experimental results is achieved if we use a simple generalization of the Geĭlikman–Kresin equation (5), where $C=C(\mathbf{k})$ is a constant determined by the anisotropic pairing mechanism in cuprates in the weak coupling limit.[29] The parameter Δ in (5) is equal to the maximum value of $\Delta(\mathbf{k})$, or $\Delta_{\max}$. The characteristic phonon frequency ω_0 is the breathing mode, whose energy $\hbar\omega_0$ decreases under pressure at a rate $d\ \ln(\omega)/dP=-6\times10^{-3}$ kbar^{-1}. This means that essentially the entire observed increase in the ratio $2\Delta_{\max}/kT_c$ under pressure is due to the strong electron–phonon interaction.

One of the authors (V. M. S.) is grateful to Dr. C. Lobb

at the Center for Superconductivity Research, University of Maryland, for fruitful discussions of the problems touched on in this paper.

The work was supported by a grant from the Telecommunications Advancement Organization of Japan.

*)E-mail: svistuno@hsts.dipt.donetsk.ua

[1] E. L. Wolf, *Principles of Electron Tunneling Spectroscopy*, Oxford Univ. Press, New York (1985).
[2] I. K. Yanson, Phys. Scr. **23**, 88 (1988); I. K. Yanson, Fiz. Nizk. Temp. **17**, 275 (1991) [Sov. J. Low Temp. Phys. **17**, 143 (1991)].
[3] B. T. Geĭlikman and V. Z. Kresin, Fiz. Tverd. Tela (Leningrad) **7**, 3294 (1966) [Sov. Phys. Solid State **7**, 2659 (1966)].
[4] N. V. Zavaritskiĭ, E. S. Itskevich, and A. N. Voronovskiĭ, JETP Lett. **7**, 211 (1968); Zh. Éksp. Teor. Fiz. **60**, 1408 (1971) [Sov. Phys. JETP **33**, 762 (1971)].
[5] A. A. Galkin and V. M. Svistunov, Phys. Status Solidi **26**, K55 (1968).
[6] D. J. Van Harlingen, Rev. Mod. Phys. **67**, 515 (1995).
[7] M. T. Béal-Monod and K. Maki, Phys. Rev. B **53**, 5775 (1996).
[8] J. Kane and K.-W. Ng, Phys. Rev. B **53**, 2819 (1996).
[9] A. F. Andreev, Zh. Éksp. Teor. Fiz. **46**, 1823 (1964) [Sov. Phys. JETP **19**, 1228 (1964)].
[10] G. E. Blonder, M. Tinkham, and T. M. Klapwijk, Phys. Rev. B **25**, 4515 (1982).
[11] Y. Tanaka and S. Kashiwaya, Phys. Rev. Lett. **74**, 3451 (1995).
[12] E. S. Itskevich, Fiz. Tekh. Éksper. No. 4, 148 (1963).
[13] V. M. Svistunov, M. A. Belogolovskiĭ, and A. I. Khachaturov, Usp. Fiz. Nauk **163**, No. 2, 61 (1993) [Phys. Usp. **36**, No. 2, 65 (1993)]; J. R. Kirtley, Int. J. Mod. Phys. **4**, 201 (1990).
[14] É. A. Pashitskiĭ and V. I. Pentegov, JETP Lett. **63**, 583 (1996).
[15] Jian Ma, G. Quitmann, R. J. Kelley *et al.*, Physica C **235–240**, 1875 (1994).
[16] Peng Xiong, Gang Xiao, and R. B. Laibowitz, Phys. Rev. Lett. **71**, 1907 (1994).
[17] Y. De Wilde, J. Heil, A. G. M. Jansen, P. Wyder, R. Deltour, W. Assmus, A. Menovsky, W. Sun, and L. Taillefer, Phys. Rev. Lett. **72**, 2278 (1994).
[18] M. Octavio, M. Tinkham, G. E. Blonder, and T. M. Klapwijk, Phys. Rev. B **27**, 6739 (1983); B. A. Aminov, B. Aschermann, M. A. Hein, F. Hill, M. Lorenz, G. Müller, and H. Peil, Phys. Rev. B **52**, 13 631 (1995).
[19] C. W. Beenakker, J. A. Melsen, and P. W. Broower, Phys. Rev. B **51**, 13 883 (1995); Y. Takagaki and H. Takayanagi, Phys. Rev. B **53**, 14 530 (1996); Y. V. Nazarov and T. H. Stoof, Phys. Rev. Lett. **76**, 823 (1996).
[20] H.-B. Schuttler, K. Yonemitsu, and J. Zhong, J. Supercond. **8**, 555 (1995).
[21] A. P. Litvinchuk, C. Thompson, and M. Cardona, in *Physical Properties of High-Temperature Superconductors*, D. M. Ginsberg (ed.), World Scientific, Singapore (1994), p. 375.
[22] L. Pintschovius and W. Reichardt, in *Physical Properties of High-Temperature Superconductors IV*, D. M. Ginsberg (ed.), World Scientific, Singapore (1994), p. 295.
[23] A. E. Karakozov, E. G. Maksimov, and A. A. Mikhaĭlovskiĭ, Zh. Éksp. Teor. Fiz. **102**, 132 (1992) [Sov. J. Low Temp. Phys. **75**, 70 (1992)].
[24] V. M. Svistunov, M. A. Belogolovskiĭ, and O. I. Chernyak, Usp. Fiz. Nauk **151**, 31 (1987) [Sov. Phys. Usp. **30**, 1 (1987)].
[25] A. I. Omel'yanchuk, S. I. Beloborod'ko, and I. O. Kulik, Fiz. Nizk. Temp. **14**, 1142 (1988) [Sov. J. Low Temp. Phys. **14**, 630 (1988)].
[26] S. Mase and T. Yasuda, Solid State Commun. **68**, 655 (1988).
[27] R. Aoki, H. Murakami, and T. Kita, Physica C **235–240**, 1891 (1994).
[28] J. S. Schilling and S. Klotz, *Physical Properties of High-Temperature Superconductors*, Vol. III, D. M. Ginsberg (ed.), World Scientific, Singapore (1989); H. Takahashi and N. Mori, in *Studies of High-Temperature Superconductors*, Vol. 16, A. V. Narlikar (ed.), Nova Science Publishers, New York (1995), p. 1.
[29] A. I. D'yachenko and V. M. Svistunov, Fiz. Nizk. Temp. **22**, 547 (1996) [Low Temp. Phys. **22**, 421 (1996)].
[30] K. Langfeld and E. Frey, Phys. Rev. B **48**, 4176 (1993).
[31] R. J. McQueeney, T. Egami, G. Shirane, and Y. Endoh, Phys. Rev. B **54**, R9689 (1996).
[32] B. Renker, F. Compf, D. Ewert, P. Adelmann, H. Schmidt, E. Gering, and H. Mutka, Z. Phys. B **77**, 65 (1989).
[33] T. Schneider and H. Keller, Int. J. Mod. Phys. B **8**, 487 (1993).
[34] S. Ishihara, T. Egami, and M. Tachiki, Phys. Rev. B **55**, 3163 (1997).

Translated by Eugene Yankovsky

Ballistic transport and spin–orbit interaction of two-dimensional electrons on a cylindrical surface

L. I. Magarill,*) D. A. Romanov, and A. V. Chaplik

Institute of Semiconductor Physics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted 11 August 1997)
Zh. Éksp. Teor. Fiz. **113**, 1411–1428 (April 1998)

The components of the ballistic magnetoconductance tensor of a two-dimensional electron gas placed on a cylindrical sector are calculated for various geometries. For a quasiclassical system a method is proposed for finding the conductance based only on the Bohr–Sommerfeld quantization condition and not requiring a knowledge of the matrix elements of the velocity. The effect of curvature of the surface on the spin–orbit interaction in a two-dimensional electron gas is investigated. As examples, the microwave absorption and longitudinal conductance of a hollow cylindrical wire are calculated, and also the conductance of a cylindrical sector. There are qualitative differences from planar systems, in particular the relative sign of the curvature and the spin–orbit coupling constant becomes important. [S1063-7761(98)02104-0]

1. INTRODUCTION

Nonplanar two-dimensional electron systems have attracted interest for a number of years. Experiments have been performed on faceted surfaces of GaAs/AlGaAs structures containing a two-dimensional electron gas.[1,2] In this case, the gas occupies a region consisting of planar segments oriented differently in space (in the experiment, relative to an external magnetic field). Recently it has become possible[3] by using special techniques (lift-off) to separate a thin layer of the GaAs/AlGaAs heterojunction together with the two-dimensional electron gas and bend it up to angles on the order of 180°. A number of works have theoretically examined the spectrum of a curved two-dimensional electron gas,[4–7] its plasma modes,[8] magnetotransport,[5,9] etc.

Going over to the model of low-dimensional electrons on a curved manifold (a surface or contour) is not a trivial procedure. In classical mechanics the problem reduces simply to introducing couplings which reduce the effective number of degrees of freedom, and therefore from the ''very outset'' can make use of curvilinear coordinates. In the quantum problem we must speak of waveguides, one or two characteristic dimensions of which become much smaller than all the remaining lengths of the problem, and the wave equation must be transformed in the spirit of the adiabatic approximation (the degrees of freedom separate into fast and slow). In this case an adiabatic potential of geometrical origin arises. As far as we know, this question was first discussed by da Kosta.[10] We showed in Ref. 11 for the case of an electric ring that the result depends on the method of ''one-dimensionalization.'' Thus, the system preserves its memory of the shape of the n-dimensional waveguide, or more generally, of the potential limiting the motion of the n-dimensional particle, from which in the limit an $(n-1)$-dimensional wave equation arises (this circumstance was overlooked in Ref. 10).

The situation is, however, significantly simplified for a surface of constant curvature (sphere, circular cylinder). In this case, it is possible at once to write down the two-dimensional Laplace operator in the corresponding coordinates since the adiabatic potential arising in the transition from the three-dimensional problem reduces to a constant. In the present work we consider the spectrum and ballistic conductance of a two-dimensional electron gas on the surface of a cylindrical sector. Such a system is the closest to that realized experimentally by the lift-off method.[3] The main effect consists of an effective variation of the normal component of the magnetic field, which is all that the two-dimensional electron ''sees.'' The Landau levels split, and this leads to the observed changes in the ballistic transport in comparison with a planar structure.

Less trivial is the effect of curvature on the spin-orbit component of the total Hamiltonian of the two-dimensional particle. The mathematical reason for this lies, naturally, in the behavior of spinors in a curved space. In the present paper we discuss the physical consequences of the effect of curvature on the spin–orbit coupling of two-dimensional electrons located on the surface of a circular cylinder or cylindrical sector.

2. ELECTRON SPECTRUM IN CROSSED FIELDS AND THE HALL CONDUCTIVITY (INTERNAL STATES)

The investigated system consists of a sector $-\varphi_0 \leqslant \varphi \leqslant \varphi_0$ of a circular cylinder of radius R in a uniform magnetic field **B** acting in the plane perpendicular to the axis of the cylinder (we choose the z axis to coincide with this axis). We restrict the discussion to the case in which the magnetic field is directed along the x axis which passes through the top of the sector (see Fig. 1). An electric field F transverse to **B**, directed along the arc $[-\varphi_0, \varphi_0]$, is also applied to the system (it is assumed that the poles of the battery are connected

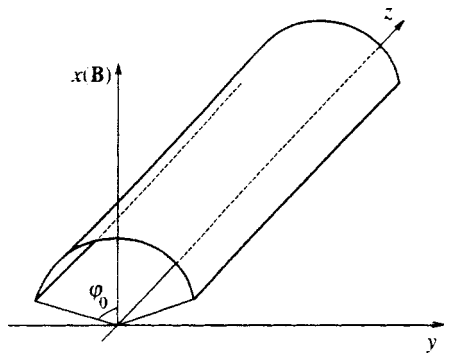


FIG. 1. Cylindrical sector in a magnetic field.

to the end-faces forming the sector $\varphi=\pm\varphi_0$). We choose the vector potential of the magnetic field $\mathbf{B}=(B,0,0)$ in the form $\mathbf{A}=(0,0,By)$ and use cylindrical coordinates $(\rho=R,\varphi,z)$.

Separating out the motion along the z axis in the form of a plane wave $[\Psi=\psi(\varphi)\exp(ip_z z)/\sqrt{L}$, where L is length of the system], we arrive at the one-dimensional Schrödinger equation (we neglect the electron spin)

$$-\frac{1}{2mR^2}\frac{\partial^2\psi}{\partial\varphi^2}+[eFR\varphi+U_{\rm eff}(\varphi;p_z)]\psi=E\psi, \tag{1}$$

where $U_{\rm eff}(\varphi;p_z)=m\Omega^2R^2(\sin\varphi+p_z l_B^2/R)^2/2$ is the effective potential energy; $\Omega=eB/mc$ is the cyclotron frequency, and $l_B=\sqrt{c/eB}$ is the magnetic length. We further assume $\varphi_0<\pi/2$ and we set $\hbar=1$. First we examine the energy spectrum of the electrons with $|p_z|<k_0\equiv R\sin\varphi_0/l_B^2$. This inequality means that the suspension point of the Landau oscillator lies inside the sector (''internal'' states). In a strong magnetic field $l_B\ll R$, it is possible to expand the potential in Eq. (1) near its minimum $\varphi_k=\arcsin(p_z l_B^2/R)$. In this (parabolic) approximation Eq. (1) is easily solved, and we find the following expression for the energy:

$$E_\beta(F)=\Omega_k\left(n+\frac{1}{2}\right)-eFR\,\arcsin\left(\frac{p_z l_B^2}{R}\right)-\frac{e^2F^2}{2m\Omega_k^2}, \tag{2}$$

$$\Omega_k=\Omega\sqrt{1-p_z^2l_B^4/R^2},\quad \beta=(n,p_z),\quad n=0,1,2,\ldots. \tag{3}$$

This result is quite transparent: we obtain the Landau spectrum in crossed electric and magnetic fields with local value of the normal component of the magnetic field $B_k=B\cos\varphi_k$.

Knowing the energy spectrum, it is not hard to find the Hall current j_z in the ballistic regime. Toward this end it is necessary, following the method expounded in the classic work of Adams and Holstein,[12] to sum up the velocity of the Landau state $v_\beta(F)=\partial E_\beta(F)/\partial p_z$ over all levels weighted by the equilibrium Fermi distribution function $f(E_\beta^{(0)})$:

$$j_z=-\frac{2e}{2R\varphi_0 L}\sum_\beta v_\beta(F)f(E_\beta^{(0)}), \tag{4}$$

where $E_\beta^{(0)}\equiv E_\beta(0)$, the factor 2 takes spin degeneracy into account, and $2R\varphi_0$ is the arc length of the sector. Substituting Eq. (2) into Eq. (4), we find for the contribution of the internal states to the Hall conductance $G_H\equiv\sigma_{z\varphi}=j_z/F$:

$$G_H=\frac{G_0 l_B^2}{2R\varphi_0}\sum_n\int_{-k_0}^{k_0}dp_z\,\frac{f(E_n^{(0)}(p_z))}{\sqrt{1-p_z^2l_B^4/R^2}} \tag{5}$$

($G_0=2e^2/h$ is the conductance quantum). Expression (5) can also be found from Kubo's linear formula. Indeed, in the absence of scattering the Kubo formula yields the following expression for G_H:

$$G_H=-\frac{2e^2}{\varphi_0 L}\sum_{\beta'\neq\beta}\frac{(\hat v_z)_{\beta'\beta}\varphi_{\beta\beta'}f(E_\beta^{(0)})}{E_{\beta,\beta'}^{(0)}}. \tag{6}$$

Here $\hat v_z=\hat p_z/m+\Omega R\sin\varphi$ is the operator of the z component of the velocity, $E_{\beta,\beta'}^{(0)}=E_\beta^{(0)}-E_{\beta'}^{(0)}$; and $\beta'\neq\beta$ means that $n_{\beta'}\neq n_\beta$ (the matrix elements of the operators $\hat v_z$ and φ are diagonal in p_z). The sum over β' in Eq. (6) is proportional to the correction of first-order perturbation theory (as the perturbation we consider the interaction with the electric field $eRF\varphi$) to the diagonal matrix element of the operator $\hat v_z$. This allows us to prove the equivalence of expressions (6) and (5).

In the ultraquantum limit (when only the lower Landau band is populated, i.e., $n=0$, which is possible for $\cos\varphi_0>1/3$) we find from Eq. (5) for $T=0$

$$G_H=G_0\left[1-\frac{\arcsin(\sin\varphi_0-\varphi_0\nu/2)}{\varphi_0}\right], \tag{7}$$

where $\nu=2\pi N_s l_B^2$ is the filling factor and N_s is the surface carrier density. For $\varphi_0\ll 1$ we have from Eq. (7)

$$G_H=\frac{N_s ec}{B}+\frac{G_0\varphi_0^2}{12}\left[1-\left(1-\frac{\nu}{2}\right)^3\right]. \tag{8}$$

The second term on the right in Eq. (8) is the correction to the curvature.

If more than one Landau band is filled, it is necessary to differentiate cases of overlapping and non-overlapping bands. In general, for $T=0$ Eq. (5) yields the following expression for G_H:

$$G_H=G_0\sum_n\left\{1-\frac{\arcsin(\sqrt{1-(E_F/\varepsilon_n)^2})}{\varphi_0}\,\theta(\varepsilon_n^2-E_F^2)\right\}\times\theta\left(1-\frac{\arcsin(\sqrt{1-(E_F/\varepsilon_n)^2})}{\varphi_0}\,\theta(\varepsilon_n^2-E_F^2)\right), \tag{9}$$

where E_F is the Fermi energy, $\varepsilon_n=\Omega(n+1/2)$, and $\theta(x)$ is the Heaviside step function. Similarly, it is possible to find an expression for the surface concentration as a function of E_F:

$$N_s=\frac{1}{\pi l_B^2\varphi_0}\sum_n\left\{\sin\varphi_0-\sqrt{1-\left(\frac{E_F}{\varepsilon_n}\right)^2}\,\theta(\varepsilon_n^2-E_F^2)\right\}\times\theta\left(\sin\varphi_0-\sqrt{1-\left(\frac{E_F}{\varepsilon_n}\right)^2}\,\theta(\varepsilon_n^2-E_F^2)\right). \tag{10}$$

In order to find $G_H(B;N_s)$, it is necessary to express E_F with the help of Eq. (10) in terms of N_s and substitute this result into expression (9). In general, this program cannot be carried out analytically (in contrast to a planar, unbounded two-dimensional gas, where $G_H=G_0\nu/2\equiv N_s ec/B$). The situa-

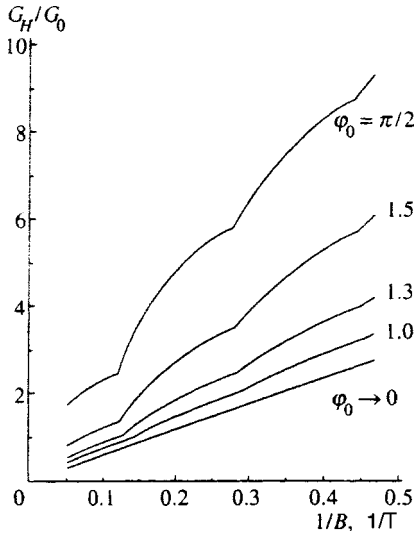


FIG. 2. Magnetic-field dependence of the Hall conductance for a cylindrical sector. The straight line corresponds to a flat strip.

tion simplifies somewhat when the Fermi level is found in the group of non-intersecting bands. The bands with numbers $l=0,1,...,\ n_0+1$ are separated from one another by gaps if

$$\frac{n_0+1/2}{n_0+3/2}\leqslant\cos\ \varphi_0\leqslant\frac{n_0+3/2}{n_0+5/2}. \tag{11}$$

The remaining bands with $l\geqslant n_0+1$ overlap. Conditions (11) follow from the inequalities $\max(E^{(0)}_{n_0+1})\geqslant\min(E^{(0)}_{n_0+2})$ and $\max(E^{(0)}_{n_0})\leqslant\min(E^{(0)}_{n_0+1})$. Thus, if φ_0 satisfies inequalities (11) and $E_F\leqslant\min(E^{(0)}_{n_0+2})\equiv\Omega\cos\varphi_0(n_0+5/2)$, then we have the case where the Fermi level lies in the group of non-overlapping bands. In this case, m bands are filled (from the zeroth to the $(m-1)$-th) and the mth band is partially filled. Then from Eq. (10) we find $\sqrt{1-(E_F/\varepsilon_m)^2}=(m+1)\sin\varphi_0-\nu\varphi_0/2$, and correspondingly for G_H[1)]

$$G_H=G_0\left\{m+1-\frac{\arcsin((m+1)\sin\ \varphi_0-\varphi_0\nu/2)}{\varphi_0}\right\}. \tag{12}$$

The correction to the curvature for small φ_0 in this case has the form $(m\varphi_0\ll 1)$

$$G_H-\frac{N_s ec}{B}\approx\frac{G_0\varphi_0^2}{12}\left[(m+1)-\left(m+1-\frac{\nu}{2}\right)^3\right]. \tag{13}$$

The filling here is such that $m<\nu/2<m+1$. For arbitrary bending angles, the system of equations (9) and (10) was solved numerically. The dependence of the Hall conductance on the magnetic field for fixed concentration N_s for sectors with different angles is plotted in Fig. 2. Instead of the straight-line dependence characterizing planar systems ($\varphi_0=0$), the B-dependence of G_H for a curved surface has cusps at values of the magnetic field corresponding to tangency of the Fermi level with the tops of the Landau bands $E^{(0)}_n(p_z=0)=\varepsilon_n$.

3. ACCOUNT OF BOUNDARY STATES

The foregoing treatment pertains obviously to the case when the parameter $\alpha=(l_B/R\varphi_0)^2$ is small. To take account of boundary states it is necessary to solve the Schrödinger equation for arbitrary α, and also to lift the restriction $|p_z|<k_0$. The Hall conductance in the linear approximation in F can be found if the quantity $\partial E_\beta(F)/\partial F|_{F=0}$ is known. Indeed, integrating by parts with respect to p_z in the general formula (4), we arrive at the following expression:

$$G_H=\frac{e}{2\pi R\varphi_0}\sum_n\int_{-\infty}^{\infty}dp_z v^{(0)}_\beta\left.\frac{\partial E_\beta(F)}{\partial F}\right|_{F=0}\frac{\partial f(E^{(0)}_\beta)}{\partial\mu}, \tag{14}$$

$$v^{(0)}_\beta=\frac{\partial E^{(0)}_\beta}{\partial p_z}, \tag{15}$$

where μ is the chemical potential. Note that the formula of type (14) for the contribution of only the internal states (considered in the previous section) contains in addition a term outside the integral, arising from the finiteness of the integration region over p_z:

$$G^{(\mathrm{int})}_H=\frac{e}{2\pi R\varphi_0}\sum_n\left\{\int_{-k_0}^{k_0}dp_z v^{(0)}_n(p_z)\left.\frac{\partial E_n(p_z;F)}{\partial F}\right|_{F=0}\frac{\partial f(E^{(0)}_n(p_z))}{\partial\mu}+2\left.\frac{\partial E_n(k_0;F)}{\partial F}\right|_{F=0}f(E^{(0)}_n(k_0))\right\}. \tag{16}$$

Employing the spectrum (2), it is hence easy to obtain again expression (5).

In what follows, we limit the discussion to the case in which the temperature is equal to zero. In this case $\partial f/\partial\mu\to\delta(E^{(0)}_n(p_z)-E_F)$ [where $E_F=\mu(T=0)$ is the Fermi energy] and integration over p_z gives

$$G_H=\frac{e}{2\pi R\varphi_0}\sum_{n,i}\ \mathrm{sign}(v^{(0)}_n(k_{n,i}(E_F)))\left.\frac{\partial E_n(k_{n,i}(E_F);F)}{\partial F}\right|_{F=0}, \tag{17}$$

where $k_{n,i}(E_F)$ are the roots of the equation

$$E^{(0)}_n(p_z)=E_F. \tag{18}$$

Let us calculate the quantities $v^{(0)}_n$ and $\partial E_n/\partial F|_{F=0}$ entering into expression (17), in the quasiclassical approximation. The energy spectrum is found from the Bohr–Sommerfeld quantization rule:

$$R\int_{\varphi_1}^{\varphi_2}d\varphi\sqrt{2m[E-U_{\mathrm{eff}}(\varphi;p_z)-eFR\varphi]}=\pi(n+\gamma). \tag{19}$$

Here $\varphi_1=\max(\varphi_l,-\varphi_0)$, $\varphi_2=\min(\varphi_r,\varphi_0)$, where $\varphi_{l,r}$ are the turning points [roots of the integrand in Eq. (19)]. We assume that at the boundaries of the sector (at the points $\pm\varphi_0$) there is an infinitely high barrier—a ''hard wall.'' The value

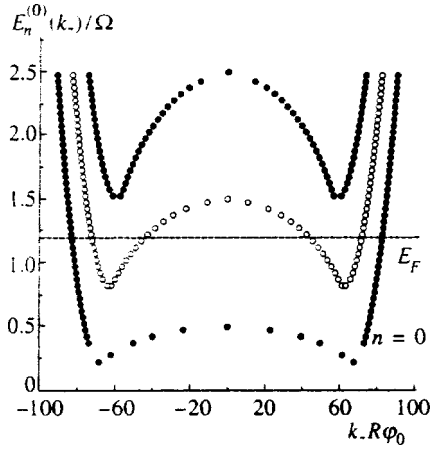


FIG. 3. Energy spectrum for a sector with $\varphi_0=1.3$; the first three Landau zones.

of the parameter γ depends on the positions of the turning points: $\gamma=1/2$ if the turning points lie inside the sector, i.e., $\varphi_r<\varphi_0$ and $\varphi_l>-\varphi_0$; $\gamma=3/4$ if for the given energy E the particle reaches one of the boundaries of the sample, and $\gamma=1$ for the ''passing'' particles, when the turning points lie at $\varphi=-\varphi_0$ and $\varphi=\varphi_0$. There is a quite complicated procedure, proposed by Langer,[13] allowing one to smooth out the jumps in γ. Our numerical calculation shows that even for $n=0$ (when the effect of the phase of γ is maximum) the difference between the results of Langer's method and calculations for $\gamma=\text{const}=1/2$ does not exceed 10%. On the other hand, the jumps in γ at the hard wall are obviously not due to curvature. Therefore, being interested mainly in the effects of curvature, in what follows we will, for simplicity, write the Bohr–Sommerfeld conditions with constant γ equal to 1/2. For illustration, Fig. 3 plots the dispersion curves $E_n^{(0)}(p_z)$ for a sector with $\varphi_0=1.3$ for $d=0.01$, obtained numerically in the quasiclassical approximation.

Differentiating Eq. (19) with respect to F, we find

$$\left.\frac{\partial E_n}{\partial F}\right|_{F=0}=\frac{\epsilon R I^{(3)}(E_n^{(0)}(p_z),p_z)}{I^{(2)}(E_n^{(0)}(p_z),p_z)}. \tag{20}$$

Similarly, differentiating Eq. (19) with respect to p_z for $F=0$, we arrive at an expression for $v_n^{(0)}(p_z)$:

$$v_n^{(0)}(p_z)=\Omega R \times\frac{I^{(1)}(E_n^{(0)}(p_z),p_z)+(p_z l_B^2/R)I^{(2)}(E_n^{(0)}(p_z),p_z)}{I^{(2)}(E_n^{(0)}(p_z),p_z)}. \tag{21}$$

In expressions (20) and (21) we have introduced the functions $I^{(s)}$:

$$I^{(s)}(E,k)=\int_{\varphi_1}^{\varphi_2}d\varphi\,\frac{R^{(s)}(\varphi)}{\sqrt{E-U_{\text{eff}}(\varphi;k)}}, \tag{22}$$

where $R^{(1)}=\sin\varphi$, $R^{(2)}=1$, and $R^{(3)}=\varphi$.

Substituting Eqs. (20) and (21) into Eq. (17), we finally find for the Hall conductance

$$G_H=\frac{G_0}{2\varphi_0}\sum_{n,i}\ \text{sign}\left(I^{(1)}(E_F,k_{n,i})+\frac{k_{n,i}l_B^2}{R}I^{(2)}\times(E_F,k_{n,i})\right)\frac{I^{(3)}(E_F,k_{n,i})}{I^{(2)}(E_F,k_{n,i})}. \tag{23}$$

Setting $E=E_F$ in the Bohr–Sommerfeld condition for $F=0$, we find numerically the quantities $k_{n,i}(E_F)$ and thereupon the conductance from Eq. (23). The proposed method allows one to avoid calculating velocity matrix elements (as is required in the Kubo formulism), which in turn requires that we know the wave functions. Instead, with the proposed method it is sufficient to use just the Bohr–Sommerfeld equation defining the energy levels and to calculate the integrals $I^{(s)}$.

Results of numerical calculations of the dependence of the Hall conductance on the Fermi energy for a sector with $\varphi_0=1$ are shown in Fig. 4a. For comparison, Fig. 4b plots the analogous curves for a flat strip of width d ($R\to\infty$, $\varphi_0\to 0$, $2R\varphi_0=d$). The solid curves in both figures correspond to the case $\alpha\to 0$. In this limit the edge states are unimportant, and the problem admits the analytical solution given by expressions (9). Here the role of curvature is especially clear: the magnetic field acting on the two-dimensional electrons becomes effectively inhomogeneous. As a result, instead of vertical steps, as in the case of a plane wave, a cylindrical sector is characterized by smoothed steps in $G_H(E_F)$.

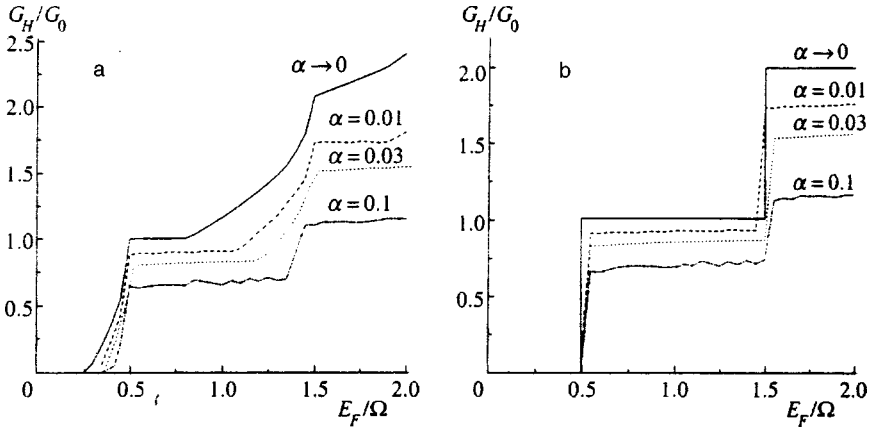


FIG. 4. Curves of the dependence of the Hall conductance on the Fermi energy for different values of the parameter α; a) cylindrical sector with $\varphi_0=1$, b) flat strip of the same width.

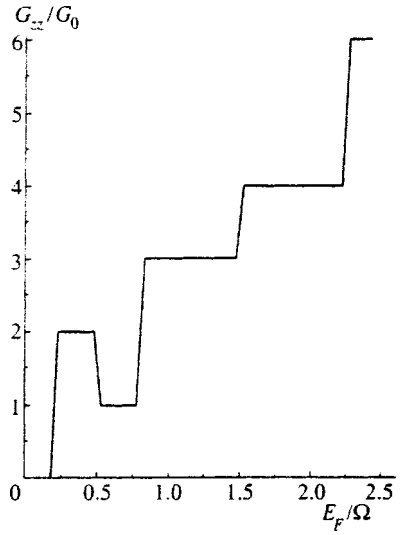


FIG. 5. Dependence of the longitudinal conductance G_{zz} on the Fermi energy for a cylindrical sector with $\varphi_0=1.3$.

4. LONGITUDINAL BALLISTIC CONDUCTANCE

To calculate the longitudinal conductance in the z direction, G_{zz}, we make use of the Buttiker–Landauer approach (see, e.g., Ref. 14). Neglecting scattering, we take all transmission coefficients to be equal to zero or unity (depending on the position of the Fermi level). Then $G_{zz}(E_F)$ is determined by the number of roots $k_{n,i}$ of Eq. (18):

$$G_{zz}=\frac{G_0}{2}\sum_{n,i}\theta(v_n^{(0)}(k_{n,i}(E_F))). \tag{24}$$

Thanks to the nonmonotonic dependence of $E_n^{(0)}(p_z)$ (see Fig. 3), the number of roots of Eq. (18) for some values of E_F is doubled in comparison with the flat strip. As a result, G_{zz} depends nonmonotonically on E_F (see Fig. 5).

Finally, let us consider a quantum constriction curved along the direction of the current: a narrow strip cut out of the cylindrical sector transverse to those forming it (an ''arched bridge''). The external magnetic field is perpendicular to the sample at its upper point. The situation here is qualitatively different since in no direction now are there any solutions in the form of traveling waves. We will assume (as is often done in problems of quantum constrictions) that motion along the z axis is bounded by a parabolic potential $m\omega^2z^2/2$, and we carry out the procedure of ''adiabatization,'' setting z up as the fast variable and φ as the slow variable. In the vector-potential gauge adopted above the Hamiltonian of the problem has the form

$$H=-\frac{1}{2mR^2}\frac{\partial^2}{\partial\varphi^2}+\frac{1}{2m}\left(-i\frac{\partial}{\partial z}+\frac{R}{l_B^2}\sin\varphi\right)^2+\frac{m\omega^2z^2}{2}. \tag{25}$$

Writing the wave function in the form

$$\Psi(z,\varphi)=\sum_n\psi_n(z;\varphi)\chi_n(\varphi)\exp(im\Omega Rz\sin\varphi), \tag{26}$$

where $\psi_n(z;\varphi)$ are the oscillator eigenfunctions (φ is a parameter), we arrive in the usual way at the equation for $\chi_n(\varphi)$ containing the adiabatic potential:

$$\left[-\frac{1}{2mR^2}\frac{\partial^2}{\partial\varphi^2}+\sqrt{\omega^2+\Omega^2\cos^2\varphi}\left(n+\frac{1}{2}\right)\right]\chi_n=E\chi_n. \tag{27}$$

In the spirit of the adiabatic approximation, the nondiagonal terms in n describing transitions between the adiabatic terms have been dropped in Eq. (27).

Now, to find the ballistic conductance $G_{\varphi\varphi}$, we need to find the transmission coefficients T_n of the barrier figuring in Eq. (27). In the quasiclassical approximation the coefficients T_n are substantially different from zero and unity only for energies near the barrier maximum for $\varphi=0$. Expanding the adiabatic potential in the vicinity of this point (cf. Ref. 15), we arrive at Campbell's parabolic barrier problem (see Ref. 16). Hence it is easy to find the form of the steps of the conductance:

$$G_{\varphi\varphi}^{(n)}=G_0\left[n+\frac{1}{1+\exp(-A)}\right],\qquad A=\frac{2\pi[E_F-\tilde{\omega}(n+1/2)]R}{\tilde{\omega}\tilde{l}\sqrt{n+1/2}}, \tag{28}$$

where $\tilde{\omega}=\sqrt{\Omega^2+\omega^2}$ and $\tilde{l}^2=1/m\tilde{\omega}$. The step width grows with magnetic field as $\sqrt{\tilde{\omega}}$.

5. SPIN–ORBIT INTERACTION ON A CYLINDRICAL SURFACE

The spin–orbit interaction of two-dimensional electrons is described by two contributions to the effective spin–orbital Hamiltonian. One of them, known as the Rashba model, is written in the invariant form[17,18]

$$\hat{V}_1=\gamma_1\hat{\boldsymbol{\sigma}}\times\hat{\mathbf{p}}\cdot\mathbf{n}, \tag{29}$$

where $\hat{\sigma}_i$ and $\mathbf{p}$ are respectively the Pauli matrices and the two-dimensional momentum operator, $\mathbf{n}$ is the surface normal, and γ_1 is the spin–orbit coupling constant. The contribution $\hat{V}_1$ is nonzero only for structures asymmetric in the normal direction (a typical example is a triangular quantum well). In this sense we say (in the two-dimensional limit) that $\gamma_1\neq0$ for an oriented surface on which the two directions of $\mathbf{n}$ are non-equivalent.

The second contribution arises from the terms in the bulk Hamiltonian that are cubic in the momentum (first considered by Dresselhaus[19] after allowing for quantization in the normal direction (the z axis). It can be written down explicitly only for a certain choice of the coordinate axes:[18]

$$\hat{V}_2=\gamma_2(\hat{\sigma}_x\hat{p}_x-\hat{\sigma}_y\hat{p}_y). \tag{30}$$

The constant γ_2 is also nonzero for a symmetric quantum well. The relative role of $\hat{V}_1$ and $\hat{V}_2$ depends on the width of the well (and of course on the material), but, strictly speaking, both terms must be taken into account simultaneously.

For planar two-dimensional systems the Hamiltonians $\hat{V}_1$ and $\hat{V}_2$ are unitarily equivalent (see Ref. 18) and their

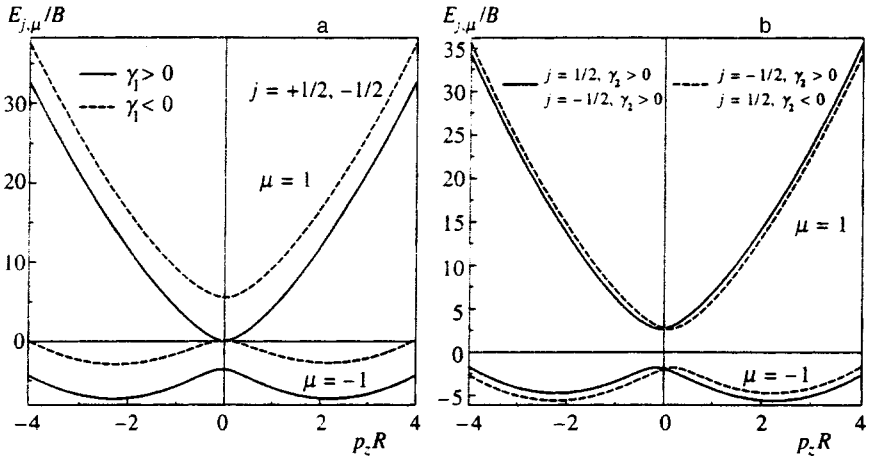


FIG. 6. Energy spectrum of two-dimensional electrons located on the surface of a cylinder: a) Rashba model, b) Dresselhaus model. The lower subbands are shown ($j=\pm 1/2$, $\mu=\pm 1$).

energy spectra are identical and depend only on their absolute values $|\gamma_1|$ and $|\gamma_2|$. It turns out that in the case of a curved surface all these statements lose their validity.

First of all, let us consider a hollow circular cylinder, for which both models admit an exact analytical solution. It follows from Eq. (29) that the Hamiltonian of two-dimensional electrons located on the surface of a cylinder has, in the Rashba model, the form

$$\hat{H}^{(R)}=\frac{\hat{p}_z^2+\hat{p}_\varphi^2}{2m}+\gamma_1(\hat{\sigma}_z\hat{p}_\varphi-\hat{\Sigma}\hat{p}_z), \tag{31}$$

$$\hat{\Sigma}=\begin{bmatrix} 0 & -ie^{-i\varphi} \\ ie^{i\varphi} & 0 \end{bmatrix}, \tag{32}$$

where we assume that the radius of the cylinder R is much larger than the lattice constant and we use cylindrical coordinates with the z axis as the axis of the cylinder. Here $\hat{p}_z$ is the longitudinal momentum operator and $\hat{p}_\varphi = -i(1/R)\partial/\partial\varphi$. The Hamiltonian (31) leads to the spectrum (see Ref. 9)

$$E_{j,\mu}^{(R)}(k;\Lambda_1)=B\left[k^2+j^2+\frac{(1-2\Lambda_1)}{4}+\mu\sqrt{j^2(\Lambda_1-1)^2+k^2\Lambda_1^2}\right], \tag{33}$$

where $B=1/2mR^2$, $k=p_zR$, j is the projection of the total angular momentum on the cylindrical axis (it has a half-integer value), $\Lambda_1=2m\gamma_1R$, and $\mu=\pm 1$ is a quantum number labeling the two branches of the spin-split dispersion law of each subband j.

A similar solution with the same structure of the wave function

$$\Psi(\varphi)=\frac{\exp(ip_zz)}{\sqrt{L}}\begin{pmatrix} e^{i(j-1/2)\varphi}\psi^{(1)} \\ e^{i(j+1/2)\varphi}\psi^{(2)} \end{pmatrix} \tag{34}$$

($\psi^{1,2}$ are constants independent of φ) is also possible for the Dresselhaus model. The Hamiltonian in this model is given by

$$\hat{H}^{(D)}=\frac{\hat{p}_z^2+\hat{p}_\varphi^2}{2m}+\gamma_2[\{\hat{\Sigma},\hat{p}_\varphi\}-\hat{\sigma}_z\hat{p}_z], \tag{35}$$

where the braces denote the symmetrized operator product. Solving the Schrödinger equation for the Hamiltonian (35) with periodic (in φ) boundary conditions, we obtain for the spectrum

$$E_{j,\mu}^{(D)}(k;\Lambda_2)=B\left[k^2+j^2+\frac{1}{4}+\mu\sqrt{j^2\Lambda_2^2+(k\Lambda_2+j)^2}\right]. \tag{36}$$

Here $\Lambda_2=2m\gamma_2R$.

It is clear from Eqs. (33) and (36) that the following symmetry properties hold for the energy spectrum:

$$E_{j,\mu}^{(R)}(k;\Lambda_1)=E_{j,\mu}^{(R)}(-k;\Lambda_1)=E_{-j,\mu}^{(R)}(k;\Lambda_1),$$
$$E_{j,\mu}^{(R)}(k;\Lambda_1)\neq E_{j,\mu}^{(R)}(k;-\Lambda_1), \tag{37}$$
$$E_{j,\mu}^{(D)}(k;\Lambda_2)=E_{j,\mu}^{(D)}(-k;-\Lambda_2)$$
$$=E_{-j,\mu}^{(D)}(-k;\Lambda_2)=E_{-j,\mu}^{(D)}(k;-\Lambda_2). \tag{38}$$

Properties (37) and (38) have the result that all thermodynamic and kinetic characteristics of the system that contain sums over states depend on the relative sign of the curvature and of γ_1 in the Rashba model and are independent of the sign of γ_2 in the Dresselhaus model. The contribution $\hat{V}_1$ can be said to be the cause of differences in the spectra and other physical characteristics of concave and convex cylindrical systems fabricated from a GaAs/AlGaAs heterojunction. Figures 6a and 6b plot the dependences $E_{j,\mu}^{(R)}(k)$ and $E_{j,\mu}^{(D)}(k)$ for the case of the heterostructure GaAs/AlGaAs. For the parameters $\Lambda_{1,2}$ we have chosen the value 4.6, which corresponds to $\gamma_1=\gamma_2=2.5\times 10^{-10}$ eV·cm (see the estimates in Refs. 17 and 18) and $R=1\,\mu m$.

It is not hard to find the energy spectrum of the two-dimensional electrons on the surface of a cylinder for simultaneous action of both mechanisms of the spin–orbit interaction. The expression for it has the form

$$E_{j,\mu}(k;\Lambda_1)=B\left\{k^2+j^2+\frac{(1-2\Lambda_1)}{4}+\mu\sqrt{[j(\Lambda_1-1)-k\Lambda_2]^2+[\Lambda_1k-\Lambda_2j]^2}\right\}. \tag{39}$$

Hence it follows that the fundamental possibility exists of separating the contributions $\hat{V}_1$ and $\hat{V}_2$ when measuring the same thermodynamic (kinetic) property for systems differing only in their sign of curvature: for example, a GaAs/AlGaAs heterojunction bent such that the two-dimensional electron gas is found either on the inner or outer side of the cylindrical surface.

The longitudinal conductance of the hollow cylinder G for Fermi energies E_F near the origin of the spectrum depends substantially on the spin–orbit interaction. We studied the shape of the conductance steps in the Rashba model in an earlier paper (Ref. 9). The number of intersection points of the dispersion curve with the Fermi level (this number determines the height of the conductance steps) varies within the limits of the lower subbands ($j=\pm 1/2$, $\mu=\mp 1$) as the Fermi level is increased, in the sequence 8–4–8. In the Dresselhaus model, as can be seen from Fig. 6b, the corresponding sequence looks like 4–8–4–8, which leads to a different shape of $G_{zz}(E_F)$. It must be emphasized once again that $G_{zz}(E_F)$ in this model, in contrast to the Rashba model, does not depend on the sign of γ_2.

Observable consequences of this effect of the dependence of the electron dispersion law on the sign of γ_1 include features of the microwave absorption spectrum of a hollow cylindrical wire. The normalized solutions of the Schrödinger equation with the Hamiltonian (31) have the form (34), where

$$\psi_{j^+}^{(1)}=\psi_{j^-}^{(2)}=iA_jC_j/\Lambda_1 k, \quad \psi_{j^-}^{(1)}=\psi_{j^+}^{(2)}=A_j,$$
$$A_j=|\Lambda_1 k|/\sqrt{4\pi D_j C_j}, \quad C_j=D_j+j(\Lambda_1-1),$$
$$D_j(p_z)=\sqrt{j^2(\Lambda_1-1)^2+k^2\Lambda_1^2}.$$

The absorbed power of an electromagnetic wave of frequency ω due to the interaction with the electron, is given by the following expression (per unit length of the cylinder):

$$Q(\omega)=\frac{2\pi\omega}{L}\sum_{\beta',\beta}|\langle\beta'|H_{\rm int}|\beta\rangle|^2\delta(E_{\beta'}-E_\beta-\omega)\times[f(E_{\beta'})-f(E_\beta)]. \tag{40}$$

In Eq. (40), $\hat{H}_{\rm int}=e\hat{\mathbf{v}}\cdot\mathbf{A}_0/2c$ is the interaction operator, $\mathbf{A}=\mathrm{Re}(\mathbf{A}_0\exp(-i\omega t))$ is the vector potential of the electromagnetic wave, $\hat{\mathbf{v}}$ is the velocity operator, $\beta=(p_z,j,\mu)$ is the set of quantum numbers characterizing the state of the electron, and $f(E_\beta)$ is the Fermi distribution function.

To start with, let us consider the interaction of the electrons with an electromagnetic wave polarized in the z direction. In this case we have $\langle\beta'|H_{\rm int}|\beta\rangle\propto\langle\beta'|v_z|\beta\rangle$. For the operator $\hat{v}_z$ we have from Eq. (31) $\hat{v}_z=\hat{p}_z/m-\gamma_1\hat{\Sigma}$. The nonzero matrix elements of the operator $\hat{v}_z$ are diagonal in p_z and j. Thus, for a z-polarized wave absorption takes place only for transitions with a change in μ (with ''spin flip''). The corresponding matrix elements are equal to

$$\langle p_z,j,+|v_z|p_z,j,-\rangle=\langle p_z,j,-|v_z|p_z,j,+\rangle^*=\frac{i\gamma_1(\Lambda_1-1)j}{D_j(p_z)}. \tag{41}$$

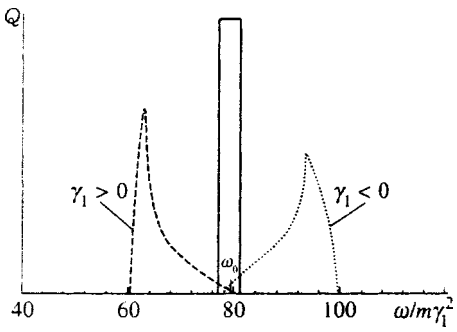


FIG. 7. Absorption of a z-polarized electromagnetic wave: $R=1$ μm, $E_F=5$ meV.

For reasonable values of the cylinder radius R and typical values of the surface concentration N_s the ratio E_F/B is large (e.g., for a GaAs/AlGaAs heterojunction with $N_s=2\times10^{11}$ cm^{-2} and $R=1$ μm we have $E_F/B\approx1.6\times10^4$). This makes it possible to calculate the absorption quasiclassically, i.e., to replace the sum over j in Eq. (40) by an integral

$$\sum_j \rightarrow|R|\int_{-\infty}^{\infty}dp_y.$$

After integrating over $|\mathbf{p}|$ ($\mathbf{p}=(p_y,p_z)$) using the δ-function, we arrive at the result

$$Q^{(z)}(\omega)=\frac{e^2E_0^2(\Lambda_1-1)^2 2\pi|R|}{8\pi\Lambda_1^2}\int_0^1 dx\,\frac{x^2}{\sqrt{1-x^2}}\,\frac{1}{q^4(x)}\times\left[f\left(\frac{\omega^2}{8m\gamma_1^2q^2(x)}-\frac{\omega}{2}\right)-f\left(\frac{\omega^2}{8m\gamma_1^2q^2(x)}+\frac{\omega}{2}\right)\right]. \tag{42}$$

In expression (42) we have introduced the function $q(x)=\sqrt{1-(2\Lambda_1-1)x^2/\Lambda_1^2}$; E_0 is the amplitude of the electric field of the electromagnetic wave.

Figure 7 presents an example of the dependence $Q^{(z)}(\omega)$ given by expression (42) for $\gamma_1=2.5\times10^{-10}$ eV·cm (the GaAs/AlGaAs heterostructure, Ref. 18), $R=1$ μm and $T=0$ K. The central peak of width $4m\gamma_1^2$ corresponds to the planar system ($R\rightarrow\infty$). Curvature leads to a substantial change in the absorption line. The absorption maximum is shifted toward the blue (red) for negative (positive) values of $\gamma_1 R$. The departure of the absorption maximum from its position for a planar system $\omega_0=2|\gamma_1|mv_F$ is of the order of v_F/R, where $v_F=\sqrt{2E_F/m}$ is the Fermi velocity. Let us turn our attention now to the strong asymmetry in the shape of the absorption lines for positive and negative values of $\gamma_1 R$.

Somewhat more involved calculations are required for the case of a circularly polarized electromagnetic wave propagating along the axis of the cylinder. The selection rules for this case are $\Delta p_z=0$, $\Delta j=\pm1$, $\Delta\mu=0,\pm2$. Under the same condition that we used above ($B\ll E_F$), we find for the absorption power due to spin-flip transitions ($\Delta\mu=\pm2$)

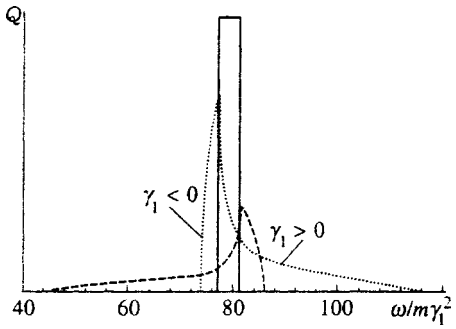


FIG. 8. Absorption of a circularly polarized electromagnetic wave: $R = 1\ \mu$m, $E_F = 5$ meV.

$$Q^{(c)}(\omega) = \frac{e^2 E_0^2 (\Lambda_1 - 1)^2 2\pi R}{16\pi\Lambda_1^2} \times \sum_{\pm} \int_{-1}^{1} dx \sqrt{1-x^2}\, \frac{\theta(\pm g_{\pm}(x))}{q^2(x) g_{\pm}^2(x)} \times \left[f\left(\frac{\omega^2}{8m\gamma_1^2 g_{\pm}^2(x)} \mp \frac{\omega q(x)}{2g_{\pm}(x)} \right) - f\left(\frac{\omega^2}{8m\gamma_1^2 g_{\pm}^2(x)} \mp \frac{\omega q(x)}{2g_{\pm}(x)} + \omega \right) \right], \tag{43}$$

where $g_{\pm}(x) = q(x) \pm x/|\Lambda_1|$, and $\theta(x)$ is the Heaviside step function. Figure 8 plots the dependence of $Q^{(c)}(\omega)$ for the same parameters as for $Q^{(z)}(\omega)$ in Fig. 7. In comparison with the z-polarization case the absorption maximum is shifted in the opposite direction and the absorption line is roughly two times wider.

If the double inequality $1 \ll |\Lambda_1| \ll v_F/\gamma_1$ is satisfied, expressions (42) and (43) can both be represented in simple analytical form ($\Delta\omega = \omega - \omega_0$):

$$Q^{(z)}(\omega) \propto \sqrt{\frac{-\Delta\omega R\ \mathrm{sign}(\gamma_1)/v_F}{1 + \Delta\omega R\ \mathrm{sign}(\gamma_1)/v_F}},$$

$$Q^{(c)}(\omega) \propto \sum_{\pm} \sqrt{\frac{1 - [\mathrm{sign}(\gamma_1)/2 \pm \sqrt{1/4 - \Delta\omega R\ \mathrm{sign}(\gamma_1)/v_F}]^2}{1/4 - \Delta\omega R\ \mathrm{sign}(\gamma_1)/v_F}}. \tag{44}$$

The frequency ω is restricted to the interval ensuring real values of the radicals.

It is technologically simpler to fabricate a curved two-dimensional system in the form of a cylindrical sector $\rho = R, -\varphi_0/2 \leqslant \varphi \leqslant \varphi_0/2, -\infty \leqslant z \leqslant \infty$. Such samples have already been obtained[3] by separating off a thin layer containing two-dimensional electrons from a GaAs/AlGaAs heterostructure by the ''lift-off'' technique.

The corresponding calculations are quite cumbersome. The transverse part of the electron wave function obeys the one-dimensional Schrödinger equation

$$\frac{\partial^2 \boldsymbol{\psi}}{\partial\varphi^2} + \frac{2mR^2}{\hbar^2} (\tilde{E} - \hat{V}_{SO}) \boldsymbol{\psi} = 0. \tag{45}$$

Here ψ is a two-component spinor and $\tilde{E} = E - p_z^2/2m$ is the transverse energy. A very important difference from the hollow cylinder case discussed above is that in the present case zero, and not periodic, boundary conditions are imposed on the wave function ψ:

$$\boldsymbol{\psi}(\varphi = \pm\varphi_0) = 0. \tag{46}$$

The spinor ψ can be represented in the form

$$\psi(\varphi) = \begin{pmatrix} e^{i(\lambda_i - 1/2)\varphi} \psi^{(1)} \\ e^{i(\lambda_i + 1/2)\varphi} \psi^{(2)} \end{pmatrix} \tag{47}$$

($\psi^{(1,2)}$ are constants which are independent of φ). By virtue of the absence now of axial symmetry the quantities λ_i ($i = 1,2,3,4$) are not half-integers (as in the case of a hollow cylinder), but are given by expressions which follow from Eqs. (45) and (47): $\lambda_{1,2} = \pm q_+ R$ and $\lambda_{3,4} = \pm q_- R$, where

$$q_{\pm} = \sqrt{(\beta_1 - \xi)^2 + 2m\tilde{E} + \beta_1^2 \mp 2\sqrt{(\beta_1 - \xi)^2 (2m\tilde{E} + \beta_1^2) + p_z^2 \beta_1^2}} \tag{48}$$

($\beta_1 = m\gamma_1$, $\xi = 1/2R$).

The general solution of Eq. (45) is a superposition of four wave functions of the form (47) (with different λ_i):

$$\boldsymbol{\psi}(\varphi) = \sum_{i=1}^{4} A_i \psi_{\lambda_i}(\varphi).$$

Subjecting these solutions to boundary conditions (46) and making use of the connection between $\psi_{\lambda_i}^{(1)}$ and $\psi_{\lambda_i}^{(2)}$, which is given by Eq. (45), after some quite cumbersome transformations we arrive at the following equation for the electron energy:

$$\begin{aligned} &(f_+ f_- - p_z^2) \sin^2\left((q_+ + q_-)\, \frac{a}{2} \right) \\ &\quad = (f_+ - f_-)^2 p_z^2\, \sin^2\left((q_+ - q_-)\, \frac{a}{2} \right), \\ &f_{\pm} = \frac{2m\tilde{E} + \beta_1^2 - (q_{\pm} \pm |\beta_1 - \xi|)^2}{2|\beta_1|}, \end{aligned} \tag{49}$$

where $a = 2R\varphi_0$ is the width of the sector. This equation does not have an analytical solution. However, it can be at once seen that in the distinction to the case of a hollow cylinder degeneracy of the spectrum for $p_z = 0$ is conserved.

We can carry out the same treatment for the case of a flat strip of two-dimensional electron gas. The resulting dispersion equation has the same form as (49), but the quantities $q_{\pm}$ and $f_{\pm}$ are given by different expressions [the superscript (0) corresponds to the planar case]:

$$\begin{aligned} q_{\pm}^{(0)} &= \sqrt{(\sqrt{2mE^{(0)} + \beta_1^2} \mp |\beta_1|)^2 - p_z^2}, \\ f_{\pm}^{(0)} &= \pm(\sqrt{2mE^{(0)} + \beta_1^2} - q_{\pm}^{(0)}) - |\beta_1|. \end{aligned} \tag{50}$$

Comparing these two dispersion equations, it is not hard to convince oneself that the following relation holds between the transverse energies of a curved and a flat strip [$\tilde{E}(p_z; \beta_1, \xi)$ and $\tilde{E}^{(0)}(p_z; \beta_1)$]:

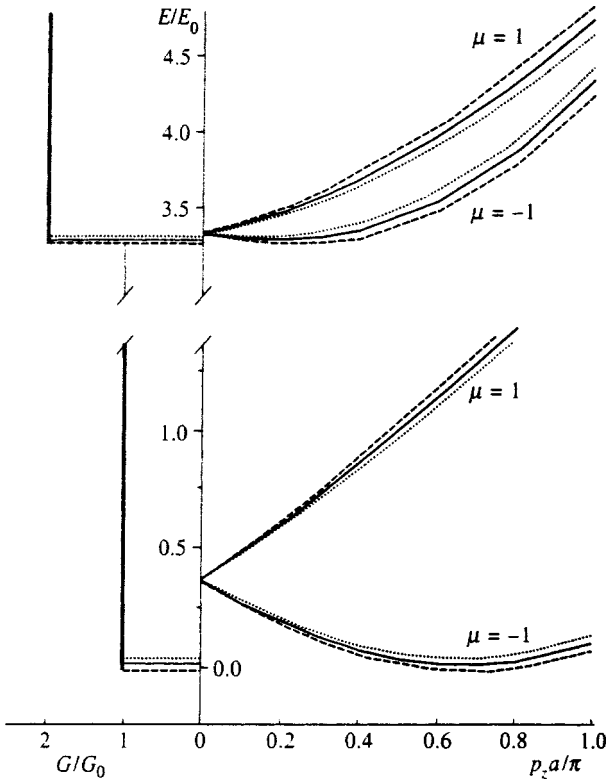


FIG. 9. Right: energy spectrum (first two subbands) of a curved strip: dashed line—$\gamma_1>0$, dotted line—$\gamma_1<0$, solid line—flat strip. Left: first two corresponding steps in the dependence of the longitudinal conductance on the Fermi energy; $E_0=\hbar^2/2ma^2$, $G_0=2e^2/h$ is the conductance quantum.

$$\widetilde{E}_{l,\mu}(p_z;\beta_1,\xi)=\widetilde{E}^{(0)}_{l,\mu}(p_z|\beta_1|/|\beta-\xi|,\beta_1-\xi) \tag{51}$$

($l=1,2,\ldots$ is the index of the subband).

It follows from this expression that the spectrum of a curved strip remains unchanged only for simultaneous change of sign of γ_1 and of the curvature $1/R$. Thus, curved two-dimensional electron systems are sensitive to the sign of the spin–orbit interaction. Results of numerical solution of Eq. (49) for an InAs/GaSb heterostructure ($\gamma_1=0.9\times10^{-9}$ eV·cm (Ref. 20), $R=1$ μm, and strip width a equal to 0.3 μm) are shown in Fig. 9. The figure shows the first two spin-split subbands of the energy spectrum of a curved strip (cylindrical sector). The spectrum of a flat strip is also shown for comparison. Also shown are steps of the longitudinal conductance, whose edges are determined by the minima of the dispersion curves.

In the Dresselhaus model the equation determining the value of λ_i is not biquadratic, as in the case of the Rashba model, but fully quartic. Therefore it is not possible to obtain a relation of the type (51) and the problem must be solved numerically from the very start. As has already been said, it is clear that the dependence $G(E_F)$ will not be sensitive to the sign of γ_2.

To summarize, we have considered some problems of ballistic magnetotransport of two-dimensional electrons on a curved surface. Depending on the geometry of the experiment, the effective inhomogeneity of the magnetic field leads either to a smearing out of the sharp edges of the conductance steps (as functions of the Fermi energy) or to a non-monotonic dependence of $G(E_F)$. Cusps corresponding to the extrema of the Landau bands arise in the magnetic field dependence of the Hall conductance for fixed electron concentration.

The spin–orbit interaction of two-dimensional electrons undergoes a more significant change in the transition to a curved surface. In particular, for an oriented surface it becomes possible to experimentally determine the sign of the spin–orbit coupling constant in the Rashba model.

ACKNOWLEDGMENTS

We are grateful to Yu. A. Bychkov, G. E. Pikus, and É. I. Rashba for helpful discussions of some aspects of the work. This work was carried out with the support of the Russian Fund for Fundamental Research (Project No. 96-02-19058a), the program ''Physics of Solid-State Heterostructures,'' and also grants from INTAS (Grant No. 95-0657) and the Volkswagen Foundation (Grant No. 1/71162).

*)E-mail: levm@isp.nsc.ru

1)We take this opportunity to point out a typographical error in Ref. 9: at the end of p. 422, $m\varphi_0$ should read $m\sin\varphi_0$.

[1] M. L. Leadbeater, C. L. Foden, T. M. Burke, J. H. Burroughes, M. P. Grimshaw, D. A. Ritchie, L. L. Wang, and M. Pepper, J. Phys.: Condens. Matter **7**, L307 (1995).
[2] M. L. Leadbeater, C. L. Foden, J. H. Burroughes, M. Pepper, T. M. Burke, L. L. Wang, M. P. Grimshaw, and D. A. Ritchie, Phys. Rev. B **52**, 8629 (1995).
[3] A. Lorke, A. Wixforth, and J. P. Kotthaus, private communication.
[4] C. L. Foden, M. L. Leadbeater, J. H. Burroughes, and M. Pepper, J. Phys.: Condens. Matter **6**, L127 (1994).
[5] C. L. Foden, M. L. Leadbeater, and M. Pepper, Phys. Rev. B **52**, 8646 (1995).
[6] H. Aoki and H. Suezawa, Phys. Rev. A **46**, R1163 (1992).
[7] V. M. Nabutovskiĭ and D. A. Romanov, Zh. Éksp. Teor. Fiz. **90**, 232 (1986) [Sov. Phys. JETP **63**, 133 (1986)].
[8] V. V. Rotkin and R. A. Suris, Fiz. Tverd. Tela **36**, 3569 (1994) [Phys. Solid State **36**, 1899 (1994)].
[9] L. I. Magarill, D. A. Romanov, and A. V. Chaplik, JETP Lett. **64**, 460 (1996); A. V. Chaplik, D. A. Romanov, and L. I. Magarill, submitted to S. & M. (1997).
[10] R. C. T. da Kosta, Phys. Rev. A **23**, 1982 (1981).
[11] L. I. Magarill, D. A. Romanov, and A. V. Chaplik, Zh. Éksp. Teor. Fiz. **110**, 669 (1996) [JETP **83**, 361 (1996)].
[12] E. N. Adams and T. D. Holstein, J. Phys. Chem. Solids **10**, 254 (1959).
[13] R. E. Langer, Phys. Rev. **51**, 669 (1937).
[14] C. W. J. Beenakker and H. van Houten, edited by H. Ehrenreich and D. Turnbull, in *Solid State Physics* (Academic Press, Boston, 1991), Vol. 44, p. 83.
[15] L. I. Glazman, G. B. Lesovik, D. E. Khmel'nitskiĭ, and R. I. Shekhter, JETP Lett. **48**, 238 (1988).
[16] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed. (Pergamon Press, Oxford, 1980).
[17] Yu. A. Bychkov and É. I. Rashba, JETP Lett. **39**, 78 (1984).
[18] E. I. Rashba and V. I. Sheka, in *Landau Level Spectroscopy*, edited by G. Landwehr and E. I. Rashba (Elsevier, 1991), p. 178.
[19] G. Dresselhaus, Phys. Rev. **100**, 580 (1955).
[20] J. Luo, H. Munekata, F. F. Fang, and P. J. Stiles, Phys. Rev. B **41**, 7685 (1990).

Translated by Paul F. Schippnick

Weak localization in semiconductor structures with strong spin–orbit coupling

N. S. Averkiev, L. E. Golub,*) and G. E. Pikus

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted 30 September 1997)
Zh. Éksp. Teor. Fiz. **113**, 1429–1445 (April 1998)

A theory of weak localization is constructed for p-type semiconductor structures with a complex Γ_8 valence band. An equation for the Cooperon is obtained and solved in the case when spin relaxation cannot be treated as a perturbation. The anomalous magnetoresistance is calculated in bulk samples as a function of the external deformation and in quantum wells as a function of the doping level. The results of the theory are represented in a form that allows direct comparison with experiment. © *1998 American Institute of Physics.* [S1063-7761(98)02204-5]

1. INTRODUCTION

The phenomenon of weak localization consists in quantum interference of waves propagating along the same trajectory but in opposite directions. One of the most striking consequences of this phenomenon is the anomalous change in the resistance in classically weak magnetic fields. The reason for this is that when waves propagate in opposite directions along the same path in a magnetic field an additional phase difference arises that is proportional to the magnetic flux through the area enclosed by the path. As a result, the initial interference is destroyed and the anomalous contribution to the conductivity is decreased.

Besides a magnetic field, inelastic processes and spin relaxation also destroy interference. Here effects associated with spin relaxation depend significantly on the total angular momentum of the two waves. Thus, in the absence of magnetic impurities, only states with nonzero total angular momentum are susceptible to spin relaxation. A theory taking these facts into account and explaining anomalous magnetoresistance in metals and metal films was developed in Refs. 1 and 2, and for two-dimensional carriers in semiconductor heterostructures in Refs. 2–5. These works assumed that the spin relaxation times can be comparable with the dephasing time of the wave function, but both these times are much longer than the momentum relaxation time. In these papers the spin–orbit interaction, which leads to spin relaxation, was treated as a perturbation.

However, it is well known that in III–V semiconductors, Si, Ge, and heterostructures based on them the valence band is formed as a result of a strong spin–orbit interaction, and the total angular momentum is coupled with the quasimomentum of the particle. As a result, the spin and momentum relaxation times are of the same order of magnitude and, consequently, the methods for calculating the magnetoresistance used in Refs. 1–5 are inapplicable for these systems.

The goal of the present work is to create a theory of weak localization leading to anomalous magnetoresistance in semiconductor structures with a strong spin–orbit interaction. We will consider non-deformed and deformed bulk semiconductors of p-type with a complex Γ_8 valence band. For quantum-well structures we calculate the dependence of the magnetoresistance on the carrier concentration. All calculations are carried out in the single-particle approximation, whose domain of applicability was indicated in Ref. 2. In the present paper odd terms in the wave vector in the spectrum are not taken into account, since the spin relaxation time due to them exceeds the momentum relaxation time for not too large a deformation or not too narrow wells.

2. HOLE SPECTRUM AND WAVE FUNCTIONS

In a bulk cubic semiconductor the energy level in a Γ_8 valence band at the quasimomentum $\mathbf{k}=0$ is quadruply degenerate. In the spherical approximation, which we use here, these four states are classified according to the projection of the total angular momentum $J=3/2$. For $\mathbf{k}\neq 0$ the states are double degenerate and are characterized by the projection of the angular momentum in the $\mathbf{k}$ direction, where states with projections differing only in sign have the same energy: for heavy holes $\mathbf{J}\cdot\mathbf{k}/k=\pm 3/2$, and for light holes $\mathbf{J}\cdot\mathbf{k}/k=\pm 1/2$. Taking uniaxial deformation into account, the hole wave functions can be written in the form[6]

$$\hat{\Psi}_{\alpha\mathbf{k}}=e^{i\mathbf{k}\cdot\mathbf{r}}\hat{F}_{\alpha}(\mathbf{k}), \tag{1}$$

where $\alpha=h1$, $h2$ and $l1$, $l2$ label the states of the heavy and light holes, respectively, and $\hat{F}_{\alpha}$ are four-component column vectors in the basis of Bloch functions of the top of the valence band. For uniaxial deformation along the (100) axis the hole energies are

$$E_{l,h}=Ak^2\pm\sqrt{(Bk^2)^2+b\varepsilon B(3k_{\parallel}^2-k^2)+(b\varepsilon)^2}, \tag{2}$$

$$E_{h1}=E_{h2}=E_h, \quad E_{l1}=E_{l2}=E_l,$$

where A and B are the band parameters determining the effective masses m_h and m_l of the heavy and light holes, $\varepsilon=\varepsilon_{zz}+(\varepsilon_{xx}+\varepsilon_{yy})/2$ is the relative deformation, b is the constant of the deformation potential, and the symbols $\parallel$ and $\perp$ here and below denote vector projections on the deformation axis and on the plane perpendicular to it. Formula (2) is written in the spherical approximation, where the band constant is $D=B\sqrt{3}$. For uniaxial deformation along the (111)

axis, b is replaced by $d/\sqrt{3}$. For $b=d/\sqrt{3}$ formula (2) is valid for any direction of uniaxial deformation. Functions $\hat{F}_\alpha$ corresponding to the same energy can be chosen so that they are related to one another by time inversion. The volume of the sample is assumed to be equal to unity.

We examine scattering by the short-range potential

$$V(\mathbf{r})=V_0\delta(\mathbf{r}).$$

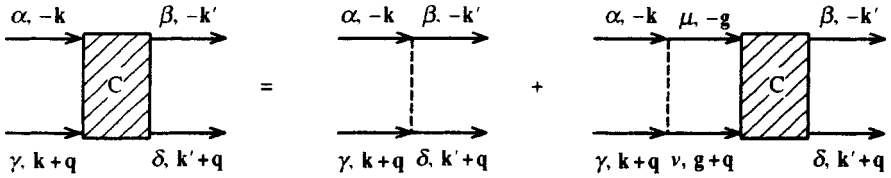
As can be seen from Eq. (1), the corresponding scattering matrix element

$$V_{\alpha\beta}(\mathbf{k},\mathbf{k}')=\langle\hat{\Psi}_{\alpha\mathbf{k}}|V(\mathbf{r})|\hat{\Psi}_{\beta\mathbf{k}'}\rangle=V_0\hat{F}^{\dagger}_{\alpha}(\mathbf{k})\hat{F}_{\beta}(\mathbf{k}') \qquad (3)$$

depends on both the initial and final quasimomenta of the hole. It follows from Eq. (3) that transitions are possible in which not only the quasimomentum but also the projection of the angular momentum changes. This means that the spin relaxation times can be of the same order of magnitude as the momentum relaxation time. This constitutes the main difference between the effects of weak localization in a complex band and the case when the spin–orbit interaction is weak.

3. EQUATIONS FOR THE COOPERON

As was shown in Ref. 7, the main quantum correction to the conductivity arises when ''fan'' diagrams are taken into account. Such diagrams describe the interference arising as a result of multiple backscattering. The amplitude of this interference (a Cooperon) is determined by the sum of the ladder diagrams at small total momentum $\mathbf{q}$:



The corresponding integral equation for the Cooperon $C^{\alpha\beta}_{\gamma\delta}(\mathbf{k},\mathbf{k}',\mathbf{q})$, averaged over the uncorrelated impurity distribution, has the form

$$C^{\alpha\beta}_{\gamma\delta}(\mathbf{k},\mathbf{k}',\mathbf{q})=\mathcal{N}V_{\alpha\beta}(-\mathbf{k},-\mathbf{k}')V_{\gamma\delta}(\mathbf{k}+\mathbf{q},\mathbf{k}'+\mathbf{q})+\mathcal{N}\sum_{\mu\nu}\int\frac{d^zg}{(2\pi)^z}V_{\alpha\mu}(-\mathbf{k},-\mathbf{g})V_{\gamma\nu}(\mathbf{k}+\mathbf{q},\mathbf{g}+\mathbf{q})C^{\mu\beta}_{\nu\delta}(\mathbf{g},\mathbf{k}',\mathbf{q})G^R_\nu(\mathbf{g}+\mathbf{q})G^A_\mu(-\mathbf{g}). \qquad (4)$$

where z is the dimensionality of the space, $\mathcal{N}$ is the impurity concentration, and $G^{A,R}_\nu$ are the advanced and retarded Green's functions for holes of type ν:

$$G^{A,R}_\nu(\mathbf{k})=\frac{1}{E_F-E_\nu(\mathbf{k})\pm i\hbar/2\tau_\nu(\mathbf{k})\pm i\hbar/2\tau^{(\nu)}_\varphi(\mathbf{k})}. \qquad (5)$$

Here E_F is the Fermi energy, determined by the total concentration of light and heavy holes, and $\tau_\mu(\mathbf{k})$ is the total relaxation time of the state $|\mu,\mathbf{k}\rangle$:

$$\frac{1}{\tau_\mu(\mathbf{k})}=\mathcal{N}\frac{2\pi}{\hbar}\sum_\nu\int\frac{d^zk'}{(2\pi)^z}|V_{\mu\nu}(\mathbf{k},\mathbf{k}')|^2\delta[E_F-E_\nu(\mathbf{k}')]. \qquad (6)$$

and $\tau^{(\nu)}_\varphi(\mathbf{k})$ is the phase relaxation time of the wave function of a hole of type ν. The quantities E_F and $\tau_\nu(\mathbf{k})$ depend on the applied deformation, but the following relations hold:

$$\tau_{h1}=\tau_{h2}=\tau_h,\quad \tau_{l1}=\tau_{l2}=\tau_l.$$

In the absence of a deformation τ_ν does not depend on the direction of $\mathbf{k}$ or on $\tau_l=\tau_h$. Note that the proposed theory is valid for $E_F\tau_\nu/\hbar\gg 1$ (Ref. 7).

As is well known,[7] a divergence arises in the integration over $\mathbf{g}$ on the right-hand side of Eq. (4) in the limit $q\to 0$. This means that $\mathbf{q}$ should be retained in Eq. (4) only in $G^R_\nu(\mathbf{g}+\mathbf{q})$. Expanding $E_\nu(\mathbf{g}+\mathbf{q})$ to second order in $\mathbf{q}$ and then integrating over $E_\nu(\mathbf{g})$, we obtain the equation

$$C^{\alpha\beta}_{\gamma\delta}(\mathbf{k},\mathbf{k}',\mathbf{q})=\mathcal{N}V_{\alpha\beta}(-\mathbf{k},-\mathbf{k}')V_{\gamma\delta}(\mathbf{k},\mathbf{k}')+\mathcal{N}\sum_{\mu\nu}\zeta(\mu,\nu)\int d\Omega_{\mathbf{g}}\frac{2\pi\tau_\nu(\mathbf{g})N_\nu(\mathbf{g})}{\hbar}\times V_{\alpha\mu}(-\mathbf{k},-\mathbf{g})V_{\gamma\nu}(\mathbf{k},\mathbf{g})C^{\mu\beta}_{\nu\delta}(\mathbf{g},\mathbf{k}',\mathbf{q})\times\left\{1-i\mathbf{v}^{(\nu)}(\mathbf{g})\mathbf{q}\tau_\nu(\mathbf{g})-[\mathbf{v}^{(\nu)}(\mathbf{g})\times\mathbf{q}\tau_\nu(\mathbf{g})]^2-\frac{\tau_\nu(\mathbf{g})}{\tau^{(\nu)}_\varphi(\mathbf{g})}\right\}, \qquad (7)$$

$$\mathbf{v}^{(\nu)}(\mathbf{g})=\frac{1}{\hbar}\frac{\partial E_\nu(\mathbf{g})}{\partial\mathbf{g}}.$$

The quantity $N_\nu(\mathbf{g})$ has the meaning of a density of states of holes of type ν on the Fermi surface per unit solid angle $d\Omega_{\mathbf{g}}$:

$$N_\nu(\mathbf{g})=\frac{g_\nu^{z-1}}{(2\pi)^z}\left|\frac{\partial E_\nu}{\partial g}\right|^{-1}_{g=g_\nu},$$

where g_ν is the absolute value of the quasimomentum of a hole of type ν on the Fermi surface, determined from the equation $E_\nu(\mathbf{g})=E_F$. Here the total relaxation time is also expressed in terms of $N_\nu(\mathbf{g})$:

$$\frac{1}{\tau_\mu(\mathbf{k})}=\mathcal{N}\frac{2\pi}{\hbar}\sum_\nu\int d\Omega_{\mathbf{g}}|V_{\mu\nu}(\mathbf{k},\mathbf{g})|^2N_\nu(\mathbf{g}). \qquad (8)$$

The factor $\zeta(\mu,\nu)$ allows for the fact that the pole contribution to the Cooperon arises only for $E_\mu(\mathbf{k})=E_\nu(\mathbf{k})$ and is equal to unity if both indices correspond to heavy holes or to light holes, and is otherwise equal to zero. As a consequence, the index ν takes the same value for N_ν, τ_ν, $\mathbf{v}^{(\nu)}$, and $\tau_\varphi^{(\nu)}$ inside the integral in Eq. (7). For $E_\mu(\mathbf{k})\neq E_\nu(\mathbf{k})$ the corrections to expression (7) are $\sim(m_h/m_l-1)^{-1}\hbar/(E_F\tau_\nu)\sim\hbar/(E_F\tau_\nu)\ll 1$.

In general, Eq. (7) can be solved as follows. To start with, consider the homogeneous integral equation with the same kernel as Eq. (7), but where the small quantities $\mathbf{v}^{(\nu)}(\mathbf{g})\mathbf{q}\tau_\nu(\mathbf{g})$, $[\mathbf{v}^{(\nu)}(\mathbf{g})\mathbf{q}\tau_\nu(\mathbf{g})]^2$, and $\tau_\nu(\mathbf{g})/\tau_\varphi^{(\nu)}(\mathbf{g})$ have been discarded:

$$\lambda_i\mathscr{A}_{i\gamma}^{\alpha}(\mathbf{k})=\mathscr{N}\sum_{\mu\nu}\zeta(\mu,\nu)\int d\Omega_{\mathbf{g}}\,\frac{2\pi\tau_\nu(\mathbf{g})N_\nu(\mathbf{g})}{\hbar}\times V_{\alpha\mu}(-\mathbf{k},-\mathbf{g})V_{\gamma\nu}(\mathbf{k},\mathbf{g})\mathscr{A}_{i\nu}^{\mu}(\mathbf{g}). \tag{9}$$

In this equation λ_i are the eigenvalues and $\mathscr{A}_i$ is a set of eigenfunctions, which can be chosen to be orthonormal:

$$\sum_{\alpha\gamma}\zeta(\alpha,\gamma)\int d\Omega_{\mathbf{k}}\,\frac{2\pi\tau_\alpha(\mathbf{k})N_\alpha(\mathbf{k})}{\hbar}\mathscr{A}_{i\gamma}^{\alpha}(\mathbf{k})\mathscr{A}_{j\gamma}^{*\alpha}(\mathbf{k})=\delta_{ij}. \tag{10}$$

Next, knowing the spectrum of eigenvalues and the eigenfunctions, the solution of the inhomogeneous equation (7) can be written in the form

$$C_{\gamma\delta}^{\alpha\beta}(\mathbf{k},\mathbf{k}',\mathbf{q})=\sum_{ij}a_{ij}(\mathbf{q})\mathscr{A}_{i\gamma}^{\alpha}(\mathbf{k})\mathscr{A}_{j\delta}^{*\beta}(\mathbf{k}'), \tag{11}$$

where the coefficients $a_{ij}(\mathbf{q})$ are unknown. To find them, we also expand the inhomogeneous term in Eq. (7) in the functions $\mathscr{A}_{i\gamma}^{\alpha}$:

$$\mathscr{N}V_{\alpha\beta}(-\mathbf{k},-\mathbf{k}')V_{\gamma\delta}(\mathbf{k},\mathbf{k}')=\sum_{ij}W_{ij}\mathscr{A}_{i\gamma}^{\alpha}(\mathbf{k})\mathscr{A}_{j\delta}^{*\beta}(\mathbf{k}'). \tag{12}$$

Next multiplying both sides of Eq. (7) by $\mathscr{A}_{m\gamma}^{*\alpha}(\mathbf{k})\mathscr{A}_{n\delta}^{\beta}(\mathbf{k}')$, integrating over $\mathbf{k}$ and $\mathbf{k}'$, and summing over the spin indices as in Eq. (10), we obtain a system of algebraic equations for the coefficients $a_{im}(\mathbf{q})$:

$$\sum_i[T_{ni}(\mathbf{q})+(1-\lambda_n)\delta_{ni}]a_{im}(\mathbf{q})=W_{nm}, \tag{13}$$

where the coefficients T_{ni}, disappearing in the limit $\mathbf{q}\to 0$ and $\tau_\varphi^{(\nu)}\to\infty$, are equal to

$$T_{ni}(\mathbf{q})=\mathscr{N}\sum_{\alpha\beta\gamma\delta}\zeta(\alpha,\gamma)\zeta(\beta,\delta)\int d\Omega_{\mathbf{k}}\,\frac{2\pi\tau_\alpha(\mathbf{k})N_\alpha(\mathbf{k})}{\hbar}\times\int d\Omega_{\mathbf{g}}\,\frac{2\pi\tau_\beta(\mathbf{g})N_\beta(\mathbf{g})}{\hbar}V_{\alpha\beta}(-\mathbf{k},-\mathbf{g})\times V_{\gamma\delta}(\mathbf{k},\mathbf{g})\mathscr{A}_{n\gamma}^{*\alpha}(\mathbf{k})\mathscr{A}_{i\delta}^{\beta}(\mathbf{g})\left\{i\mathbf{v}^{(\beta)}(\mathbf{g})\mathbf{q}\tau_\beta(\mathbf{g})+[\mathbf{v}^{(\beta)}(\mathbf{g})\mathbf{q}\tau_\beta(\mathbf{g})]^2+\frac{\tau_\beta(\mathbf{g})}{\tau_\varphi^{(\beta)}(\mathbf{g})}\right\}. \tag{14}$$

It can be seen from Eq. (13) that the contribution to the Cooperon that diverges as $\mathbf{q}\to 0$ comes only from solutions of Eq. (9) with $\lambda_n=1$. In this event, two cases are possible. For $\lambda_0=1$, a nondegenerate eigenvalue which corresponds to the solution $\mathscr{A}_{0\gamma}^{\alpha}(\mathbf{k})$,

$$C_{\gamma\delta}^{\alpha\beta}(\mathbf{k},\mathbf{k}',\mathbf{q})=\sum_m a_{0m}(\mathbf{q})\mathscr{A}_{0\gamma}^{\alpha}(\mathbf{k})\mathscr{A}_{m\delta}^{*\beta}(\mathbf{k}'), \tag{15}$$

$$a_{0m}(\mathbf{q})=\frac{W_{0m}}{T_{00}(\mathbf{q})-\sum_{i\neq 0}\dfrac{T_{0i}(\mathbf{q})T_{i0}(\mathbf{q})}{1-\lambda_i}}.$$

Summing over m, we finally obtain

$$C_{\gamma\delta}^{\alpha\beta}(\mathbf{k},\mathbf{k}',\mathbf{q})=\frac{\mathscr{A}_{0\gamma}^{\alpha}(\mathbf{k})\mathscr{N}\sum_{\mu\nu}\zeta(\mu,\nu)\int d\Omega_{\mathbf{g}}\,\dfrac{2\pi\tau_\nu(\mathbf{g})N_\nu(\mathbf{g})}{\hbar}V_{\mu\beta}(-\mathbf{g},-\mathbf{k}')V_{\nu\delta}(\mathbf{g},\mathbf{k}')\mathscr{A}_{0\nu}^{*\mu}(\mathbf{g})}{T_{00}(\mathbf{q})-\sum_{i\neq 0}\dfrac{T_{0i}(\mathbf{q})T_{i0}(\mathbf{q})}{1-\lambda_i}}. \tag{16}$$

If, on the other hand, the eigenvalue $\lambda=1$ is r-tuply degenerate, then the coefficients $a_{im}(\mathbf{q})$ are found by solving a system of equations of dimensionality $r\times r$:

$$\sum_{i=1}^{r}T_{ni}(\mathbf{q})a_{im}(\mathbf{q})=W_{nm},\quad n,\ m=1,\cdots,r. \tag{17}$$

Here

$$C_{\gamma\delta}^{\alpha\beta}(\mathbf{k},\mathbf{k}',\mathbf{q})=\sum_{i,j=1}^{r}a_{ij}(\mathbf{q})\mathscr{A}_{i\gamma}^{\alpha}(\mathbf{k})\mathscr{A}_{j\delta}^{*\beta}(\mathbf{k}'). \tag{18}$$

Let us now apply this general method to the particular problem of weak localization of holes with the spectrum considered in Sec. 2.

A. Bulk crystals

In this case, as a result of the factor $\zeta(\mu,\nu)$ in Eq. (9) only eight of the sixteen different spin components $\mathscr{A}_{i\gamma}^{\alpha}(\mathbf{k})$ are nonzero, namely those with α and γ such that $E_\alpha(\mathbf{k})=E_\gamma(\mathbf{k})$. What is more, since the functions $\hat{F}_\alpha$ follow one

from the other by time inversion, the remaining system of eight equations splits into two systems of respectively two and six equations.

Analysis shows that starting from zero deformation up to the point where $|b\varepsilon| \sim E_F$, the eigenvalue $\lambda = 1$ is nondegenerate and is contained in a system of two equations:

$$\lambda_i \mathscr{A}^{h1}_{i_{h2}}(\mathbf{k}) = \mathscr{N} \int d\Omega_{\mathbf{g}} \Bigg\{ \frac{2\pi\tau_h(\mathbf{g}) N_h(\mathbf{g})}{\hbar} \times [|V_{h1,h1}(\mathbf{k},\mathbf{g})|^2 + |V_{h1,h2}(\mathbf{k},\mathbf{g})|^2] \mathscr{A}^{h1}_{i_{h2}}(\mathbf{g}) + \frac{2\pi\tau_l(\mathbf{g}) N_l(\mathbf{g})}{\hbar} [|V_{h1,l1}(\mathbf{k},\mathbf{g})|^2 + |V_{h1,l2}(\mathbf{k},\mathbf{g})|^2] \mathscr{A}^{l1}_{i_{l2}}(\mathbf{g}) \Bigg\}, \tag{19}$$

$$\lambda_i \mathscr{A}^{l1}_{i_{l2}}(\mathbf{k}) = \mathscr{N} \int d\Omega_{\mathbf{g}} \Bigg\{ \frac{2\pi\tau_h(\mathbf{g}) N_h(\mathbf{g})}{\hbar} \times [|V_{l1,h1}(\mathbf{k},\mathbf{g})|^2 + |V_{l1,h2}(\mathbf{k},\mathbf{g})|^2] \mathscr{A}^{h1}_{i_{h2}}(\mathbf{g}) + \frac{2\pi\tau_l(\mathbf{g}) N_l(\mathbf{g})}{\hbar} [|V_{l1,l1}(\mathbf{k},\mathbf{g})|^2 + |V_{l1,l2}(\mathbf{k},\mathbf{g})|^2] \mathscr{A}^{l1}_{i_{l2}}(\mathbf{g}) \Bigg\}, \tag{20}$$

$$\mathscr{A}^{h2}_{i_{h1}} = -\mathscr{A}^{h1}_{i_{h2}}, \quad \mathscr{A}^{l2}_{i_{l1}} = -\mathscr{A}^{l1}_{i_{l2}},$$

where $\mathscr{A}^{\alpha}_{i_\alpha} = 0$. The eigenvalue $\lambda_0 = 1$ belongs to the solution

$$\mathscr{A}^{h1}_{0_{h2}}(\mathbf{k}) = \xi\tau_h^{-1}(\mathbf{k}), \quad \mathscr{A}^{l1}_{0_{l2}}(\mathbf{k}) = \xi\tau_l^{-1}(\mathbf{k}), \tag{21}$$

where ξ is a normalization factor. Note that in the absence of deformation the solution $\mathscr{A}^{\alpha}_{0_\gamma}(\mathbf{k})$ corresponds to zero total angular momentum, composed from the states α and γ. The expression for $C^{\alpha\beta}_{\gamma\delta}$ is obtained by substituting Eqs. (21) into expression (16).

For the particular case of scattering by a short-range potential, expression (16) simplifies since in this case $T_{0i}(\mathbf{q}) T_{i0}(\mathbf{q}) = 0$ as a result of the absence of a source term in the classical kinetic equation for the diffusion coefficient. Finally, the equation for the Cooperon takes the form

$$C^{\alpha\beta}_{\gamma\delta}(\mathbf{k},\mathbf{k}',\mathbf{q}) = \frac{\hbar}{4\pi(\overline{N}_h + \overline{N}_l)} \frac{\mathscr{A}^{\alpha}_{0_\gamma}(\mathbf{k}) \mathscr{A}^{\beta}_{0_\delta}(\mathbf{k}') \xi^{-2}}{D_\parallel q_\parallel^2 + D_\perp q_\perp^2 + \tau_\varphi^{-1}}. \tag{22}$$

Here

$$\overline{N}_\alpha = \int d\Omega_{\mathbf{k}} N_\alpha(\mathbf{k}) \tag{23}$$

is the total number of holes of species α on the Fermi surface. Expressions (22) have the standard form of the diffusion pole, which involves the average diffusion coefficient and the average dephasing time of the wave function:

$$D_{\parallel,\perp} = \frac{\overline{N}_h D^{(h)}_{\parallel,\perp} + \overline{N}_l D^{(l)}_{\parallel,\perp}}{\overline{N}_h + \overline{N}_l}, \tag{24}$$

where $D^{(\alpha)}_{\parallel,\perp}$ are the components of the diffusion tensor of holes of species α:

$$D^{(\alpha)}_{\parallel,\perp} = \frac{1}{\overline{N}_\alpha} \int d\Omega_{\mathbf{g}} N_\alpha(\mathbf{g}) [v^{(\alpha)}_{\parallel,\perp}(\mathbf{g})]^2 \tau_\alpha(\mathbf{g}), \tag{25}$$

$$\tau_\varphi^{-1} = \frac{\int d\Omega_{\mathbf{g}} [N_h(\mathbf{g})/\tau^{(h)}_\varphi(\mathbf{g}) + N_l(\mathbf{g})/\tau^{(l)}_\varphi(\mathbf{g})]}{\overline{N}_h + \overline{N}_l}. \tag{26}$$

In the limiting case of large deformations, when $|b\varepsilon| \gg E_F$, only one subband is filled, and the number of equations is reduced to four, each of which contains the eigenvalue $\lambda = 1$. Therefore the Cooperon is found from Eqs. (17) and (18) with $r = 4$.

In this limiting case it is convenient to classify the states $\hat{\Psi}_{\alpha\mathbf{k}}$ according to the projection of the total angular momentum on the deformation axis. In the case of uniaxial compression the states with $J_z = \pm 1/2$ are responsible for the upper valence subband, and in the case of dilation, the states with $J_z = \pm 3/2$. Such a pair of states, differing in the sign of J_z, we will label with the index $\alpha = 1, 2$.

For infinitely large deformation, only one component of the functions $\hat{F}_\alpha$ is different from zero. Consequently, transitions between two degenerate states are absent and we have $V_{\alpha\beta} \sim \delta_{\alpha\beta}$, and $\tau \equiv \tau_0$ does not depend on the direction of the quasimomentum. Therefore the four eigenfunctions $\mathscr{A}^{\alpha}_{i_\gamma}$ corresponding to the eigenvalue $\lambda = 1$ are also independent of $\mathbf{k}$ and can be chosen in the following way:

$$\mathscr{A}^1_{1_1} = \mathscr{A}^2_{2_2} = \mathscr{A}^1_{3_2} = \mathscr{A}^2_{4_1} = (2\pi\tau_0 \overline{N}/\hbar)^{-1/2}, \tag{27}$$

their remaining components are equal to zero.

In the basis (27), the coefficients T_{ni} and W_{mn} entering into Eqs. (17) are equal to

$$W_{mn} = \delta_{mn} \frac{\hbar}{2\pi\overline{N}\tau_0}, \tag{28}$$

$$T_{ni}(\mathbf{q}) = T_0(\mathbf{q}) \delta_{ni}, \quad T_0(\mathbf{q}) = (D^{(0)}_\parallel q_\parallel^2 + D^{(0)}_\perp q_\perp^2 + \tau_\varphi^{-1}) \tau_0. \tag{29}$$

Here the superscript (0) indicates that the corresponding quantities should be calculated for infinite deformation. The expression for the Cooperon has the standard form:

$$C^{11}_{11}(\mathbf{q}) = C^{22}_{22}(\mathbf{q}) = C^{11}_{22}(\mathbf{q}) = C^{22}_{11}(\mathbf{q}) = \frac{\hbar}{2\pi\overline{N}\tau_0} \frac{1}{T_0(\mathbf{q})}, \tag{30}$$

and the remaining components are equal to zero.

In order to obtain an expression for the Cooperon at finite deformation, we must again make use of Eqs. (17) and (18), where $\mathscr{A}^{\alpha}_{i_\gamma}$ are given in (27) and the T_{ni} are replaced by $\tilde{T}_{ni}$:

$$\tilde{T}_{ni}(\mathbf{q}) = T_{ni}(\mathbf{q}) + \mathscr{N} \sum_{\alpha\beta\gamma\delta} \int d\Omega_{\mathbf{k}} \frac{2\pi\tau_0 N_0(\mathbf{k})}{\hbar} \times \int d\Omega_{\mathbf{g}} \frac{2\pi}{\hbar} \mathscr{A}^{*\alpha}_{n_\gamma} \mathscr{A}^{\beta}_{i_\delta} [\tau_0 N_0(\mathbf{g}) V^{(0)}_{\alpha\beta} V^{(0)}_{\gamma\delta}$$

$$-\tau(\mathbf{g})N(\mathbf{g})V_{\alpha\beta}(-\mathbf{k},-\mathbf{g})V_{\gamma\delta}(\mathbf{k},\mathbf{g})], \qquad (31)$$

where $T_{ni}(\mathbf{q})$ are calculated according to formula (14) for infinitely large deformation and are given by formula (29). The difference between $\tilde{T}_{ni}(\mathbf{q})$ and $T_{ni}(\mathbf{q})$ is due both to transitions between states with opposite spins and to a change in the rate of transitions inside one branch and disappears as $|b\varepsilon|\to\infty$. The difference $\tilde{T}_{ni}(\mathbf{q})-T_{ni}(\mathbf{q})$ appears because in our foregoing perturbation-theory treatment of spin relaxation processes, besides the parameters $\mathbf{v}^{(\nu)}(\mathbf{g})\mathbf{q}\tau_\nu(\mathbf{g})$, $[\mathbf{v}^{(\nu)}(\mathbf{g})\mathbf{q}\tau_\nu(\mathbf{g})]^2$, and $\tau_\nu(\mathbf{g})/\tau_\varphi^{(\nu)}(\mathbf{g})$ taken into account in T_{ni}, quantities of the same order can also arise due to changes in $V_{\alpha\beta}(\mathbf{k},\mathbf{g})$ in Eq. (4) under the action of deformation.

Calculation shows that in the first nonvanishing order in $E_F/|b\varepsilon|$ the quantities $\tilde{T}_{ni}$ can be represented in the form

$$\begin{aligned}
&\tilde{T}_{11}=\tilde{T}_{22}=T_0+\tau_0/\tau_\parallel,\\
&\tilde{T}_{33}=\tilde{T}_{44}=T_0+\tau_0/(2\tau_\perp),\\
&\tilde{T}_{34}=\tilde{T}_{43}=\tau_0/(2\tau_\perp),
\end{aligned} \qquad (32)$$

the remaining components are equal to zero. Here

$$\frac{\tau_0}{\tau_\parallel}=\left(\frac{4}{45}\right)^2\left(\frac{E_F}{b\varepsilon}\right)^4\frac{(m_\parallel-m_\perp)^4}{m_\parallel^2m_\perp^2}\left(\frac{m_\parallel^2+m_\perp^2}{2m_\parallel m_\perp}\right)^2, \qquad (33)$$

$$\frac{\tau_0}{\tau_\perp}=\left(\frac{4}{45}\right)^2\left(\frac{E_F}{b\varepsilon}\right)^4\frac{(m_\parallel-m_\perp)^4}{m_\parallel^2m_\perp^2}, \qquad (34)$$

$m_{\parallel,\perp}$ are the effective hole masses for motion along the deformation axis and transverse to it. These expressions for the spin relaxation times, written in terms of $m_\parallel$ and $m_\perp$, do not depend on the sign of the deformation, but the values of the masses themselves are different for dilation and compression:[6]

$$m_\parallel=\frac{\hbar^2}{2(A\pm B)},\quad m_\perp=\frac{\hbar^2}{2(A\mp B/2)}. \qquad (35)$$

Here the upper sign corresponds to dilation, and the lower one, to compression. Solving the system of equations (17) with $\tilde{T}_{ni}$ (32), we obtain a set of expressions for the nonzero components of the Cooperon:

$$\begin{aligned}
C_{11}^{11}(\mathbf{q})&=C_{22}^{22}(\mathbf{q})=\frac{\hbar}{4\pi\overline{N}\tau_0}\frac{2}{T_0(\mathbf{q})+\tau_0/\tau_\parallel},\\
C_{21}^{12}(\mathbf{q})&=C_{12}^{21}(\mathbf{q})=\frac{\hbar}{2\pi\overline{N}\tau_0}\left[\frac{1}{T_0(\mathbf{q})+\tau_0/\tau_\perp}-\frac{1}{T_0(\mathbf{q})}\right],\\
C_{22}^{11}(\mathbf{q})&=C_{11}^{22}(\mathbf{q})=\frac{\hbar}{4\pi\overline{N}\tau_0}\left[\frac{1}{T_0(\mathbf{q})+\tau_0/\tau_\perp}-\frac{1}{T_0(\mathbf{q})}\right],
\end{aligned} \qquad (36)$$

its remaining components are equal to zero.

Comparing formulas (36) and (22) we can trace out the transition from the case of infinitely large deformation to zero deformation. Formulas (36) are valid as long as $\tau_{\parallel,\perp}\gg\tau_0$ holds. For $\tau_{\parallel,\perp}\sim\tau_0$ formulas (22) start to apply. Looking at formulas (36) we can say that as $\tau_{\parallel,\perp}$ decrease, the diffusion pole disappears from C_{11}^{11}, C_{22}^{22}, and from the first term of the remaining four Cooperons, this first term becomes $(D_\parallel^{(0)}q_\parallel^2+D_\perp^{(0)}q_\perp^2+\tau_\varphi^{-1})\tau_0$ times smaller than the second term in C_{21}^{12}, and formula (36) goes over to formula (22). It follows from formulas (33) and (34) that even for $|b\varepsilon|\sim E_F$ the inequality $\tau_{\parallel,\perp}\gg\tau_0$ can be preserved. This means that the phenomenon of weak localization is described by formula (36) for practically any deformation, and all changes are associated with $D_{\parallel,\perp}$, τ_φ, $\tau_{\parallel,\perp}$, and τ.

Note that for zero deformation the formula (22) does not coincide with the corresponding expression in Ref. 2. This is because Ref. 2 neglected the fact that the equation for the Cooperon should only treat transitions between states with the same energy for given $\mathbf{k}$. The symmetry arguments used in Ref. 2 are valid only for $m_h=m_l$, i.e., for $B=0$.

B. Quantum wells

Let us turn our attention now to the phenomenon of weak localization in p-type quantum wells based on compounds with a Γ_8 valence band. We consider a rectangular symmetric quantum well, employing as before the spherical approximation to describe the states in the valence band. For simplicity, we assume the barriers to be infinitely high. The spectrum and wave functions of the carriers under these assumptions have been used in many studies. We use the form of the wave functions suggested in Ref. 8:

$$\hat{\Psi}_{\alpha\mathbf{k}}^{(n)}=e^{i\mathbf{k}\cdot\boldsymbol{\rho}}\hat{F}_\alpha^{(n)}(\mathbf{k},z), \qquad (37)$$

$$\hat{F}_1=\begin{bmatrix}-v_0C(z)\\ iv_1S(z)e^{i\varphi_\mathbf{k}}\\ -v_2C(z)e^{2i\varphi_\mathbf{k}}\\ iv_3S(z)e^{3i\varphi_\mathbf{k}}\end{bmatrix},\quad \hat{F}_2=\begin{bmatrix}iv_3S(z)e^{-3i\varphi_\mathbf{k}}\\ v_2C(z)e^{-2i\varphi_\mathbf{k}}\\ iv_1S(z)e^{-i\varphi_\mathbf{k}}\\ v_0C(z)\end{bmatrix}.$$

Here $\mathbf{k}$ is the two-dimensional (in-plane) wave vector, $\varphi_\mathbf{k}$ is its azimuthal angle, $\boldsymbol{\rho}$ and z are the coordinates characterizing motion respectively in the plane of the quantum well and along its growth axis, n is the number of the size quantization level, α labels the two degenerate states (in a symmetric quantum well), and $C(z)$ and $S(z)$ are respectively a symmetric and an antisymmetric function of the coordinate z. The dispersion equation for finding the energy of such states $E_n(k)$, and also expressions for $C(z)$ and $S(z)$ and the $\varphi_\mathbf{k}$-independent real coefficients v_i $(i=1-3)$ are given in Ref. 8.

An equation for the Cooperon in the case when one size-quantization subband is filled can be derived from Eq. (7), where $\mathbf{k}$ and $\mathbf{g}$ are two-dimensional (in-plane) vectors and $\zeta(\mu,\nu)\equiv1$. In this case, when averaging over position of the impurities in a symmetric quantum well it is important to note that

$$\int dz\ C^3(z)S(z)=\int dz\ S^3(z)C(z)=0.$$

System of equations (9) takes the form

$$\lambda_i^{(00)}\mathscr{C}_i^{(00)}{}_2^1(\varphi_\mathbf{k})=\mathscr{N}\frac{2\pi\tau N}{\hbar}\int d\varphi_\mathbf{g}[|V_{11}(\varphi_\mathbf{k}-\varphi_\mathbf{g})|^2+|V_{12}(\varphi_\mathbf{k}-\varphi_\mathbf{g})|^2]\mathscr{C}_i^{(00)}{}_2^1(\varphi_\mathbf{g}), \qquad (38)$$

$$\mathscr{A}_i^{(00)2}{}_1=-\mathscr{A}_i^{(00)1}{}_2,\quad \mathscr{A}_i^{(00)1}{}_1=\mathscr{A}_i^{(00)2}{}_2=0,$$

$$\lambda_i^{(10)}\mathscr{A}_i^{(10)1}{}_2(\varphi_{\mathbf{k}})=\mathscr{N}\frac{2\pi\tau N}{\hbar}\int d\varphi_{\mathbf{g}}[|V_{11}(\varphi_{\mathbf{k}}-\varphi_{\mathbf{g}})|^2+|V_{12}(\varphi_{\mathbf{k}}-\varphi_{\mathbf{g}})|^2]\mathscr{A}_i^{(10)1}{}_2(\varphi_{\mathbf{g}}),\quad (39)$$

$$\mathscr{A}_i^{(10)2}{}_1=\mathscr{A}_i^{(10)1}{}_2,\quad \mathscr{A}_i^{(10)1}{}_1=\mathscr{A}_i^{(10)2}{}_2=0.$$

$$\lambda_i^{(11)}\mathscr{A}_i^{(11)1}{}_1(\varphi_{\mathbf{k}})=\mathscr{N}\frac{2\pi\tau N}{\hbar}\int d\varphi_{\mathbf{g}}[V_{11}^2\times(\varphi_{\mathbf{k}}-\varphi_{\mathbf{g}})\mathscr{A}_i^{(11)1}{}_1(\varphi_{\mathbf{g}})+V_{12}^2(\varphi_{\mathbf{k}},\varphi_{\mathbf{g}})\mathscr{A}_i^{(11)2}{}_2(\varphi_{\mathbf{g}})],\quad (40)$$

$$\lambda_i^{(11)}\mathscr{A}_i^{(11)2}{}_2(\varphi_{\mathbf{k}})=\mathscr{N}\frac{2\pi\tau N}{\hbar}\int d\varphi_{\mathbf{g}}[V_{11}^{*2}(\varphi_{\mathbf{k}},\varphi_{\mathbf{g}})\times\mathscr{A}_i^{(11)1}{}_1(\varphi_{\mathbf{g}})+V_{11}^{*2}\times(\varphi_{\mathbf{k}}-\varphi_{\mathbf{g}})\mathscr{A}_i^{(11)2}{}_2(\varphi_{\mathbf{g}})],$$

$$\mathscr{A}_i^{(11)1}{}_2=\mathscr{A}_i^{(11)2}{}_1=0.$$

Here the total relaxation time is given by

$$\tau^{-1}=\mathscr{N}\frac{2\pi N}{\hbar}\int d\varphi[|V_{11}(\varphi)|^2+|V_{12}(\varphi)|^2].\quad (41)$$

and the density of states at the Fermi level is expressed in terms of the particle velocity v_F and the quasimomentum k_F:

$$N=\frac{k_F}{(2\pi)^2\hbar v_F}.\quad (42)$$

Note that since the kernel of the integral equation (38) depends on a difference of angles, the equations for different Fourier harmonics separate.

If we have $E_F\sim\Delta$, where Δ is the minimum energy separating two lower subbands, the eigenvalue $\lambda=1$ is contained only in the first equation. The corresponding normalized solution is

$$\mathscr{A}_0^{(00)1}{}_2=(4\pi\tau\overline{N}/\hbar)^{-1/2},\quad (43)$$

where $\overline{N}=2\pi N$.

In contrast to the three-dimensional case for scattering by a short-range potential, in the classical kinetic equation the source term is different from zero. Consequently, the products satisfy $T_{0i}(\mathbf{q})T_{i0}(\mathbf{q})\neq 0$. However, as a consequence of the orthogonality of $\mathscr{A}_i^{00\alpha}{}_\gamma$ with the solutions $\mathscr{A}_i^{10\alpha}{}_\gamma$ and $\mathscr{A}_i^{01\alpha}{}_\gamma$, only the solutions $\mathscr{A}_i^{11\alpha}{}_\gamma\sim e^{\pm i\varphi_k}$ contribute to the sum in the denominator in Eq. (16). Carrying out the calculations indicated in formula (16), we obtain

$$C_{\gamma\delta}^{\alpha\beta}(\mathbf{k},\mathbf{k}',\mathbf{q})=\frac{\mathscr{A}_0^{(00)\alpha}{}_\gamma\mathscr{A}_0^{(00)\beta}{}_\delta}{Dq^2\tau+\tau/\tau_\varphi},\quad (44)$$

where the diffusion coefficient

$$D=\frac{1}{2}v_F^2\tau_{\mathrm{tr}},$$

and τ_{tr} is the so-called transport time arising in the solution of the kinetic equation in the case when the scattering probability depends on the difference of angles between the initial and final values of the quasimomentum:

$$\tau_{\mathrm{tr}}^{-1}=\mathscr{N}\frac{2\pi N}{\hbar}\int d\varphi[|V_{11}(\varphi)|^2+|V_{12}(\varphi)|^2](1-\cos\varphi).\quad (45)$$

In the limit $E_F\ll\Delta$ we have $|V_{12}|\ll|V_{11}|$ and the solutions corresponding to the eigenvalue $\lambda=1$ are contained in all four equations (38)–(40). Consequently, to calculate the Cooperon it is necessary to use formulas (17) and (18) with $r=4$, where T_{ni} have been replaced by $\tilde{T}_{ni}$:

$$\tilde{T}_{ni}(q)=D_0q^2\tau_0+\frac{\tau_0}{\tau_\varphi}+\mathscr{N}\sum_{\alpha\beta\gamma\delta}\int d\varphi_{\mathbf{k}}\frac{2\pi\tau_0N_0}{\hbar}\times\int d\varphi_{\mathbf{g}}\frac{2\pi}{\hbar}\mathscr{A}_{n\,\gamma}^{*\alpha}\mathscr{A}_{i\,\delta}^{\beta}[\tau_0N_0V_{\alpha\beta}^{(0)}V_{\gamma\delta}^{(0)}-\tau NV_{\alpha\beta}(\varphi_{\mathbf{k}}-\varphi_{\mathbf{g}})V_{\gamma\delta}(\varphi_{\mathbf{k}}-\varphi_{\mathbf{g}})],\quad (46)$$

The index 0 indicates quantities that have been calculated without allowing for the mixing of light and heavy holes arising for $k\neq 0$; the components

$$\mathscr{A}_{1_1}^1=\mathscr{A}_{2_2}^2=\mathscr{A}_{3_2}^1=\mathscr{A}_{4_1}^2=(2\pi\tau_0\overline{N}/\hbar)^{-1/2},\quad (47)$$

are found for $k=0$ with the remaining components equal to zero. Solving the system of equations (17) with $\tilde{T}_{ni}(q)$ given by formula (46), we obtain an expression for the Cooperon:

$$C_{11}^{11}(\mathbf{q})=C_{22}^{22}(\mathbf{q})=\frac{\hbar}{4\pi\overline{N}\tau_0}\frac{2}{D_0q^2\tau_0+\tau_0/\tau_\varphi+\tau_0/\tau_\parallel^{QW}},$$

$$C_{21}^{12}(\mathbf{q})=C_{12}^{21}(\mathbf{q})=\frac{\hbar}{4\pi\overline{N}\tau_0}\left[\frac{1}{D_0q^2\tau_0+\tau_0/\tau_\varphi+\tau_0/\tau_\perp^{QW}}-\frac{1}{D_0q^2\tau_0+\tau_0/\tau_\varphi}\right],\quad (48)$$

$$C_{22}^{11}(\mathbf{q})=C_{11}^{22}(\mathbf{q})=\frac{\hbar}{4\pi\overline{N}\tau_0}\left[\frac{1}{D_0q^2\tau_0+\tau_0/\tau_\varphi+\tau_0/\tau_\perp^{QW}}+\frac{1}{D_0q^2\tau_0+\tau_0/\tau_\varphi}\right],$$

where the remaining components are equal to zero. Here $\tau_{\parallel,\perp}^{QW}$ have the meaning of longitudinal and transverse spin relaxation times, where the role of the preferred axis is played by the normal to the plane of the quantum well:

$$\frac{\tau_0}{\tau_\parallel^{QW}}=\left(\frac{k_Fa}{\pi}\right)^4I_\parallel.\quad (49)$$

$$\frac{\tau_0}{\tau_\perp^{QW}}=\left(\frac{k_Fa}{\pi}\right)^6\left(1+\frac{m_h^2}{m_l^2}\right)I_\perp,\quad (50)$$

where a is the width of the quantum well and $I_{\parallel,\perp}$ depends only on the mass ratio m_l/m_h for infinitely high barriers:

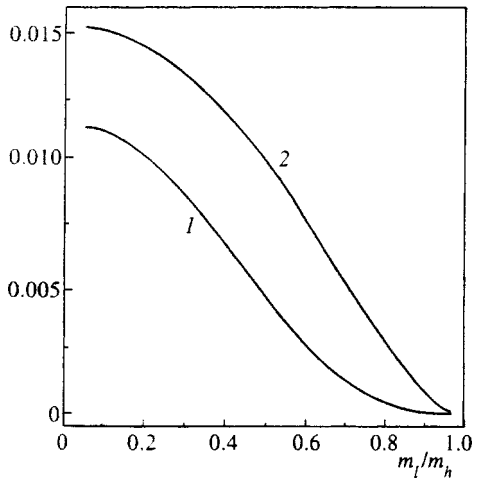


FIG. 1. Dependence of the longitudinal and transverse spin relaxation times in a quantum well on the ratio of the masses of the light and heavy holes: *1*—$I_\parallel$, *2*—$10I_\perp$.

$$I_\parallel=\frac{3}{2}\int_0^1 dx\left(\sin\frac{\pi x}{2}-\frac{\sin\kappa x}{\sin\kappa}\right)^4,$$

$$I_\perp=\frac{3}{2}\int_0^1 dx\,\cos^2\frac{\pi x}{2}\left(\sin\frac{\pi x}{2}-\frac{\sin\kappa x}{\sin\kappa}\right)^2,$$

$$\kappa=\sqrt{\frac{m_l}{m_h}}\frac{\pi}{2}.$$

$I_{\parallel,\perp}$ are plotted in Fig. 1 as functions of m_l/m_h. Formulas (49) and (50) are valid for

$$\frac{m_h}{m_l}\left(\frac{k_F a}{\pi}\right)^2\ll 1.$$

Note that the nonzero value of $\tau_0/\tau_\parallel^{QW}$ arises because the rate of the transitions changes with increasing k while conserving the projection of the angular momentum, and the nonzero value of $\tau_0/\tau_\perp^{QW}$ is due to transitions in which the projection of the angular momentum changes. The E_F/Δ dependence of the times $\tau_\parallel^{QW}$ and $\tau_\perp^{QW}$ is different:

$$\tau_0/\tau_\parallel^{QW}\sim(E_F/\Delta)^2,\quad \tau_0/\tau_\perp^{QW}\sim(E_F/\Delta)^3.$$

Formulas (48) apply for $\tau_0/\tau_{\parallel,\perp}^{QW}\ll 1$, i.e., for $E_F\ll\Delta$. With growth of E_F/Δ the spin relaxation rate grows, $\tau_{\parallel,\perp}^{QW}$ becomes $\sim\tau_0$, and the diffusion pole is preserved in formulas (48) only in the last terms in C_{21}^{12}, C_{12}^{21}, C_{22}^{11}, and C_{11}^{22}, whose form coincides with formula (44). Consequently, weak localization effects in quantum wells for arbitrary E_F/Δ (but for the case in which only one subband is filled) are described by formula (48), in which D, τ_φ, $\tau_{\parallel,\perp}^{QW}$, and τ depend on E_F/Δ.

Having obtained expressions for the Cooperons, we can examine the influence of weak localization on various kinetic phenomena in semiconductor structures. Below we calculate the magnetoresistance in classically weak fields such that $\omega_c\tau\ll 1$, where ω_c is the classical cyclotron frequency. The presence of a magnetic field has an effect on weak localization starting from

$$\omega_c\tau E_F\tau_\varphi/\hbar\sim 1.$$

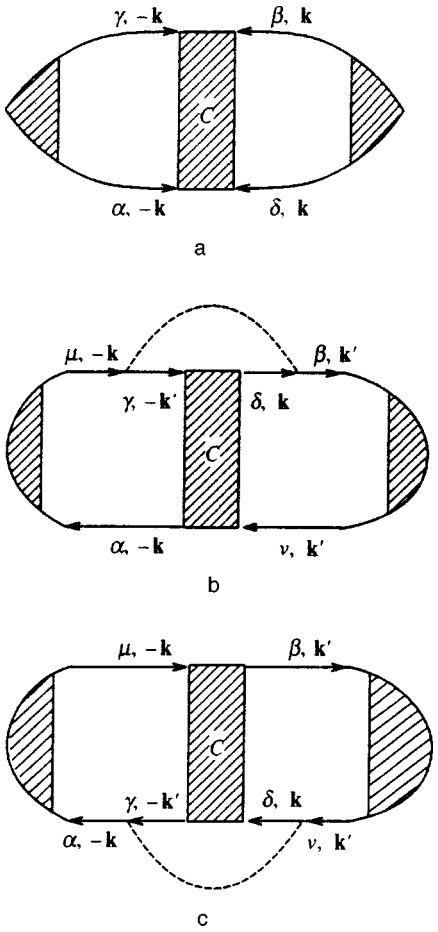


FIG. 2. Diagrammatic representation of the contributions to the conductivity: a—$\Delta\sigma^{(\mathrm{I})}$, b—$\Delta\sigma^{(\mathrm{II})}$, c—$\Delta\sigma^{(\mathrm{III})}$.

Here $E_F\tau_\varphi/\hbar\gg 1$.

4. CALCULATION OF THE DIFFUSION CONTRIBUTION TO THE CONDUCTIVITY

As is well known,[9] to calculate the contribution to the conductivity associated with weak localization it is necessary to sum the three diagrams depicted in Fig. 2. In the structures under study this contribution to the static conductivity can be represented in the form

$$\Delta\sigma_{ij}=\Delta\sigma_{ij}^{(\mathrm{I})}+\Delta\sigma_{ij}^{(\mathrm{II})}+\Delta\sigma_{ij}^{(\mathrm{III})},\tag{51}$$

$$\Delta\sigma_{ij}^{(\mathrm{I})}=\frac{e^2\hbar}{2\pi}\sum_{\alpha\beta\gamma\delta}\int\frac{d^zk}{(2\pi)^z}\int\frac{d^zq}{(2\pi)^z}\,\tilde v_i^{(\alpha\gamma)}(-\mathbf{k})\tilde v_j^{(\beta\delta)}(\mathbf{k})\times G_\gamma^A(-\mathbf{k})G_\alpha^R(-\mathbf{k})G_\beta^A(\mathbf{k})G_\delta^R(\mathbf{k})C_{\delta\alpha}^{\gamma\beta}(\mathbf{k},-\mathbf{k},\mathbf{q}),\tag{52}$$

$$\Delta\sigma_{ij}^{(\mathrm{II})}=\frac{e^2\hbar}{2\pi}\sum_{\alpha\beta\gamma\delta\mu\nu}\int\frac{d^zk}{(2\pi)^z}\int\frac{d^zk'}{(2\pi)^z}\int\frac{d^zq}{(2\pi)^z}\,\tilde v_i^{(\alpha\mu)}\times(-\mathbf{k})\tilde v_j^{(\beta\nu)}(\mathbf{k}')G_\mu^A(-\mathbf{k})G_\alpha^R(-\mathbf{k})\times G_\beta^A(\mathbf{k}')G_\nu^R(\mathbf{k}')V_{\mu\gamma}(-\mathbf{k},-\mathbf{k}')V_{\delta\beta}(\mathbf{k},\mathbf{k}')\times G_\gamma^A(-\mathbf{k}')G_\delta^A(\mathbf{k})C_{\nu\alpha}^{\gamma\delta}(\mathbf{k}',-\mathbf{k},\mathbf{q}),\tag{53}$$

$$\Delta\sigma_{ij}^{(\mathrm{III})}=\frac{e^2\hbar}{2\pi}\sum_{\alpha\beta\gamma\delta\mu\nu}\int\frac{d^zk}{(2\pi)^z}\int\frac{d^zk'}{(2\pi)^z}\times\int\frac{d^zq}{(2\pi)^z}\,\tilde{v}_i^{(\alpha\mu)}(-\mathbf{k})\tilde{v}_j^{(\beta\nu)}(\mathbf{k}')G_\mu^A(-\mathbf{k})\times G_\alpha^R(-\mathbf{k})G_\beta^A(\mathbf{k}')G_\nu^R(\mathbf{k}')V_{\gamma\alpha}(-\mathbf{k}',-\mathbf{k})\times V_{\nu\delta}(\mathbf{k}',\mathbf{k})G_\delta^R(\mathbf{k})G_\gamma^R(-\mathbf{k}')C_{\nu\alpha}^{\gamma\delta}(\mathbf{k},-\mathbf{k}',\mathbf{q}),\quad(54)$$

where the matrix element contrasted with the hatched vertex satisfies the equation

$$\tilde{\mathbf{v}}^{(\alpha\beta)}(\mathbf{k})=\mathbf{v}^{(\alpha)}(\mathbf{k})\,\delta_{\alpha\beta}+\mathscr{N}\sum_{\mu\nu}\int\frac{d^zg}{(2\pi)^z}\,V_{\alpha\mu}(\mathbf{k},\mathbf{g})V_{\nu\beta}(\mathbf{g},\mathbf{k})\tilde{\mathbf{v}}^{(\mu\nu)}\times(\mathbf{g})G_\nu^R(\mathbf{g})G_\mu^A(\mathbf{g}).\quad(55)$$

Equation (55) is similar to Eq. (4) for the Cooperon and can be solved by the method expounded above.

To start with, let us consider corrections to the conductivity in a bulk sample. For scattering by a short-range potential for any deformation $V_{\alpha\mu}(\mathbf{k},\mathbf{g})V_{\nu\beta}(\mathbf{g},\mathbf{k})$ is an even function of $\mathbf{g}$. Therefore the integral in Eq. (55) is equal to zero, and

$$\tilde{\mathbf{v}}^{(\alpha\beta)}(\mathbf{k})=\mathbf{v}^{(\alpha)}(\mathbf{k})\,\delta_{\alpha\beta}.\quad(56)$$

For the same reason

$$\Delta\sigma_{ij}^{(\mathrm{II})}=\Delta\sigma_{ij}^{(\mathrm{III})}=0.$$

It follows from Eq. (56) that only $C_{\beta\alpha}^{\alpha\beta}(\mathbf{k},-\mathbf{k},\mathbf{q})$ Cooperons contribute to $\Delta\sigma_{ij}^{(1)}$. For $|b\varepsilon|<E_F$ according to formula (22)

$$C_{\beta\alpha}^{\alpha\beta}(\mathbf{k},\mathbf{k}',\mathbf{q})=-\frac{\hbar}{4\pi(\overline{N}_h+\overline{N}_l)}\,\frac{\tau_\alpha^{-1}(\mathbf{k})\tau_\beta^{-1}(\mathbf{k}')}{D_\parallel q_\parallel^2+D_\perp q_\perp^2+\tau_\varphi^{-1}},\quad(57)$$

and after summation over α and β and integration with respect to $\mathbf{k}$, the expression for the conductivity becomes

$$\Delta\sigma_{ij}=\frac{e^2}{\pi\hbar}\,D_{ij}\int\frac{d^zq}{(2\pi)^z}\,\frac{1}{D_\parallel q_\parallel^2+D_\perp q_\perp^2+\tau_\varphi^{-1}}.\quad(58)$$

In the strong deformation limit the Cooperons are defined by Eqs. (36) and

$$\Delta\sigma_{ij}=-\frac{e^2}{\pi\hbar}\,D_{ij}^{(0)}\int\frac{d^zq}{(2\pi)^z}\left[\frac{2}{D_\parallel^{(0)}q_\parallel^2+D_\perp^{(0)}q_\perp^2+\tau_\varphi^{-1}+1/\tau_\parallel}+\frac{1}{D_\parallel^{(0)}q_\parallel^2+D_\perp^{(0)}q_\perp^2+\tau_\varphi^{-1}+1/\tau_\perp}-\frac{1}{D_\parallel^{(0)}q_\parallel^2+D_\perp^{(0)}q_\perp^2+\tau_\varphi^{-1}}\right].\quad(59)$$

The changeover from formula (59) to formula (58) as the deformation is decreased takes place when $\tau_{\parallel,\perp}$ becomes $\sim\tau_0$ and the first two terms in formula (59) disappear. Note that if the inequality $\tau_{\parallel,\perp}\gg\tau_0$ is still satisfied for $|b\varepsilon|\geqslant E_F$, then the first two terms have the form of diffusion poles at the same time that the coefficients of $q_{\parallel,\perp}^2$ cannot coincide with the corresponding diffusion coefficients.

For two-dimensional carriers in a symmetric quantum well the scattering probability depends on the difference of angles between the initial and final directions of the quasimomentum. Therefore the components $\tilde{\mathbf{v}}^{(\alpha\beta)}(\mathbf{k})$ can be expressed in terms of the total relaxation and transport times:

$$\tilde{\mathbf{v}}^{(\alpha\beta)}(\mathbf{k})=\mathbf{v}^{(\alpha)}(\mathbf{k})\,\frac{\tau_{\mathrm{tr}}}{\tau}\,\delta_{\alpha\beta}.\quad(60)$$

Substituting this expression in Eqs. (52)–(54), we obtain

$$\Delta\sigma^{(\mathrm{I})}=-\frac{e^2}{\pi\hbar}\,D\,\frac{2\pi\overline{N}\tau\tau_{\mathrm{tr}}}{\hbar}\int\frac{d^zq}{(2\pi)^z}\sum_{\alpha\beta}C_{\beta\alpha}^{\alpha\beta}(\mathbf{q}),$$

$$\Delta\sigma^{(\mathrm{II})}=\Delta\sigma^{(\mathrm{III})}=\frac{1}{2}\,\frac{\tau-\tau_{\mathrm{tr}}}{\tau_{\mathrm{tr}}}\,\Delta\sigma^{(\mathrm{I})}.\quad(61)$$

Using expressions (44) and (48) for the Cooperon, it is possible to obtain a formula for the contribution to the conductivity. For $E_F\ll\Delta$

$$\Delta\sigma=-\frac{e^2}{\pi\hbar}\,D_0\int\frac{d^2q}{(2\pi)^2}\left[\frac{2}{D_0q^2+\tau_\varphi^{-1}+1/\tau_\parallel^{QW}}+\frac{1}{D_0q^2+\tau_\varphi^{-1}+1/\tau_\perp^{QW}}-\frac{1}{D_0q^2+\tau_\varphi^{-1}}\right],\quad(62)$$

and for $E_F\sim\Delta$

$$\Delta\sigma=\frac{e^2}{\pi\hbar}\,D\int\frac{d^2q}{(2\pi)^2}\,\frac{1}{Dq^2+\tau_\varphi^{-1}}.\quad(63)$$

The changeover from formula (62) to formula (63) with growth of E_F/Δ is completely analogous to the changeover from formula (59) to formula (58) with decrease of the deformation.

As was already mentioned, in a magnetic field an additional phase-breaking of the wave function takes place, destroying the weak localization and decreasing the value of $|\Delta\sigma_{ij}|$. To calculate the magnetoresistance according to Refs. 1 and 2 the integral in q in formulas (58) and (59) should be replaced according to the following rule:

$$\int\frac{d^zq}{(2\pi)^z}\to\frac{\omega_c}{4\pi D_a}\sum_n\int\frac{d^{z-2}q}{(2\pi)^{z-2}},\quad(64)$$

$$D_\parallel q_\parallel^2+D_\perp q_\perp^2\to(D_aq^2)^{z-2}+\omega_c(n+1/2),$$

where

$$D_a=(D_\parallel^{z-2}D_\perp^2)^{1/z},$$

and ω_c is the cyclotron frequency of a particle with charge $2e$ and inverse effective mass tensor $m_{ij}^{-1}=2D_{ij}/\hbar$,

$$\omega_c=\frac{4eH}{\hbar c}\,D_c,\quad D_c=\sqrt{D_\perp(D_\perp\cos^2\theta+D_\parallel\sin^2\theta)}.\quad(65)$$

Here H is the magnitude of the magnetic field and θ is the angle between the field and the deformation axis.

In formulas (64) $z=3$ holds if the dimensions of the sample exceed $\sqrt{D_a\tau_\varphi}$. If the length of the sample in the

direction of the magnetic field is less than this length, then from the point of view of diffusive motion the sample is two-dimensional and we have $z=2$.

For the case of a quantum well

$$\int \frac{d^2q}{(2\pi)^2} \to \frac{\omega_c}{4\pi D_a} \sum_n , \tag{66}$$

$$Dq^2 \to \omega_c(n+1/2);$$

and in this case ω_c is also given by formula (65), where by θ we mean the angle between the magnetic field and the normal to the quantum well, $D_\parallel = 0$ and $D_\perp = D$.

It is convenient to represent the final expressions for the conductivity in the form of a difference

$$\delta\sigma_{ij}(H) = \Delta\sigma_{ij}(H) - \Delta\sigma_{ij}(0).$$

For a bulk sample for deformations $|b\varepsilon| < E_F$ we have

$$\delta\sigma_{ij}(H) = -\frac{D_{ij}}{D_a} \frac{e^2}{4\pi^2\hbar} \sqrt{\frac{eH}{\hbar c} \frac{D_c}{D_a}} f_3\left(\frac{4D_c eH}{\hbar c} \tau_\varphi\right), \tag{67}$$

and for $|b\varepsilon| \gg E_F$

$$\delta\sigma_{ij}(H) = \frac{D_{ij}^{(0)}}{D_a^{(0)}} \frac{e^2}{4\pi^2\hbar} \sqrt{\frac{eH}{\hbar c} \frac{D_c^{(0)}}{D_a^{(0)}}} \times \left[2 f_3\left(\frac{4D_c^{(0)} eH}{\hbar c} \frac{\tau_\varphi \tau_\parallel}{\tau_\varphi + \tau_\parallel}\right) + f_3\left(\frac{4D_c^{(0)} eH}{\hbar c} \frac{\tau_\varphi \tau_\perp}{\tau_\varphi + \tau_\perp}\right) - f_3\left(\frac{4D_c^{(0)} eH}{\hbar c} \tau_\varphi\right) \right]. \tag{68}$$

Here f_3 is a function introduced in Ref. 10:

$$f_3(x) = \sum_{n=0}^{\infty} \left\{ 2[\sqrt{n+1+x} - \sqrt{n+x}] - \frac{1}{\sqrt{n+1/2+x}} \right\}. \tag{69}$$

In the case of two-dimensional diffusion in a bulk sample for $|b\varepsilon| < E_F$ we have

$$\delta\sigma_{ij}(H) = -\frac{D_{ij}}{D_a} \frac{e^2}{4\pi^2\hbar} f_2\left(\frac{4D_c eH}{\hbar c} \tau_\varphi\right), \tag{70}$$

and for $|b\varepsilon| \gg E_F$

$$\delta\sigma_{ij}(H) = \frac{D_{ij}^{(0)}}{D_a^{(0)}} \frac{e^2}{4\pi^2\hbar} \left[2 f_2\left(\frac{4D_c^{(0)} eH}{\hbar c} \frac{\tau_\varphi \tau_\parallel}{\tau_\varphi + \tau_\parallel}\right) + f_2\left(\frac{4D_c^{(0)} eH}{\hbar c} \frac{\tau_\varphi \tau_\perp}{\tau_\varphi + \tau_\perp}\right) - f_2\left(\frac{4D_c^{(0)} eH}{\hbar c} \tau_\varphi\right) \right]. \tag{71}$$

Here

$$f_2(x) = \ln x + \psi(1/2 + 1/x), \tag{72}$$

where $\psi(y)$ is the digamma function.

The variation of the conductivity in a magnetic field in a quantum well within the framework of the model considered above is isotropic and is described by formula (70) for $E_F \sim \Delta$ and by formula (71) for $E_F \ll \Delta$, where in the latter case $\tau_{\parallel,\perp}$ have been replaced by $\tau_{\parallel,\perp}^{QW}$.

As the deformation is decreased, formulas (68) and (71) give way to formulas (67) and (70) since $\tau_{\parallel,\perp}$ decrease and the first two terms in expressions (68) and (71) disappear since we have $f_3(0) = f_2(0) = 0$. The same thing happens in a quantum well as E_F/Δ is increased. From formulas (67), (68), (70), and (71) it is clear that the magnetoresistance changes sign when $|b\varepsilon|/E_F$ or E_F/Δ varies in the quantum well.

Formulas (67) and (70) differ from the result obtained in Ref. 2 by a factor of two. This difference is due to the above-mentioned inaccuracy in the calculation of the Cooperon.

5. CONCLUSION

In the present paper we have constructed a theory of weak localization for the case of strong spin–orbit coupling. We have obtained equations for the Cooperons in semiconductors with a complex valence band with allowance for strong transitions between subbands. We have examined the dependence of the magnetoresistance on the external parameters: on deformation in bulk samples and on the doping level in quantum wells. Expressions have been obtained for the variation of the anomalous contribution to the conductivity in a magnetic field. We have shown that the magnetoresistance in a nondeformed bulk sample is positive and changes sign with growth of the deformation. Such behavior also takes place in a quantum well when the doping level is reduced.

Note that since the interparticle interaction was not taken into account in the construction of the given theory, it is capable of describing the experimental data only at temperatures $T \ll \hbar/\tau_\varphi$ (Ref. 2).

ACKNOWLEDGMENTS

This work was partially supported by the Russian Fund for Fundamental Research (Projects No. 96-02-17849, 96-02-16959a, and 96-15-96955), the program ''Physics of Solid-State Nanostructures,'' and by the Volkswagen Foundation.

*)E-mail: golub@coherent.ioffe.rssi.ru

[1] S. Hikami, A. Larkin, and Y. Nagaoka, Prog. Theor. Phys. **63**, 707 (1980).
[2] B. L. Al'tshuler, A. G. Aronov, A. I. Larkin, and D. E. Khmel'nitskiĭ, Zh. Éksp. Teor. Fiz. **81**, 768 (1981) [Sov. Phys. JETP **54**, 411 (1981)].
[3] S. V. Ioranskiĭ, Yu. B. Lyanda-Geller, G. E. Pikus, JETP Lett. **60**, 206 (1994).
[4] F. G. Pikus and G. E. Pikus, Phys. Rev. B **51**, 16 928 (1995).
[5] W. Knap, C. Skierbiszewski, A. Zduniak, E. Litwin-Staszewska, D. Bertho, F. Kobbi, J. L. Robert, G. E. Pikus, F. G. Pikus, S. V. Iordanskii, V. Moser, K. Zekentes, and Yu. B. Lyanda-Geller, Phys. Rev. B **53**, 3912 (1996).

[6] G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors* (Wiley, New York, 1974; Nauka, Moscow, 1972).
[7] L. P. Gor'kov, A. I. Larkin, and D. E. Khmel'nitskiĭ, JETP Lett. **30**, 228 (1979).
[8] I. A. Merkulov, V. I. Perel', and M. E. Portnoĭ, Zh. Éksp. Teor. Fiz. **99**, 1202 (1991) [Sov. Phys. JETP **72**, 669 (1991)].
[9] D. Rainer and G. Bergmann, Phys. Rev. B **32**, 3522 (1985).
[10] A. Kawabata, Solid State Commun. **34**, 431 (1980).

Translated by Paul F. Schippnick

Intraband transitions in magnetoexcitons in coupled double quantum wells

A. B. Dzyubenko*)

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia;
Walter Schottky Institute, Technische Universität München, D-85748 Garching, Germany
(Submitted 8 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 1446–1459 (April 1998)

A theory of far-infrared (FIR) magneto-optical intraband $s \rightarrow p^{\pm}$ transitions of direct and indirect excitons in semiconductor coupled double quantum wells has been developed. The case of symmetric strained $In_xGa_{1-x}As/GaAs$ quantum wells with nondegenerate valence band in the regime of both narrow and wide barriers has been analyzed. The energies and dipole matrix elements of transitions between the ground s and excited $p^{\pm}$ states in a quantizing magnetic field $B>2$ T and electric field $\mathscr{E}$ perpendicular to the quantum well plane have been studied. The regimes of direct (in a weak electric field) and indirect (in a strong electric field) transitions, and the transition between the direct and indirect regimes, have been investigated.
 [S1063-7761(98)02304-X]

1. INTRODUCTION

Two-dimensional (2D) spatially separated electron–hole ($e-h$) systems in a strong magnetic field have been studied theoretically for a number of years.[1] Depending on the separation d between e- and h-layers and the population number of excitons at the lowest Landau level, $\nu_X = 2\pi l_B^2 n_X$ (where n_X is the exciton density and $l_B = (\hbar c/eB)^{1/2}$), such systems demonstrate an abundance of possible low-temperature phases. In particular, at small d Bose–Einstein condensation of magnetoexcitons in the state with momentum $K=0$ is possible (see also Ref. 2, where exact many-body results in the limit $d=0$ were obtained).

In order to check theoretical predictions, real quasi-two-dimensional systems with sufficiently long exciton lifetimes are necessary. Recently experimenters' attention has been focused on systems of this kind.[3–7] Some evidence in favor of condensation of indirect excitons in a strong magnetic field was provided by interband magneto-optical spectroscopy (with a temporal and spatial resolution) of type II GaAs/AlAs quantum wells.[3] In addition, anomalies were detected in low-temperature transport properties of excitons in a magnetic field.[8] Other semiconductor structures that have been intensely studied in recent times are InGaAs/GaAs[4] and GaAs/GaAlAs[5] coupled double quantum wells (DQW). When an electric field $\mathscr{E}$ is applied normally to the quantum well plane, the exciton ground state is modified (direct-to-indirect crossover). In a strong electric field $\mathscr{E}$ the ground state is an indirect exciton (Fig. 1), whose radiative lifetime is considerably longer. This makes it possible to investigate many-body effects in neutral $e-h$ systems in a strong magnetic field B at low exciton temperatures.

Identification of many-body effects in optical spectra demands detailed knowledge of optical properties of excitons in DQW in a strong magnetic field. The theory of magneto-optical transitions of excitons in InGaAs/GaAs DQW in a low-density regime was presented in our previous publication[9] and is in good agreement with experimental data.[4] The intraband FIR magnetospectroscopy proved to be an efficient tool in studies of the ground and excited states of excitons in bulk indirect semiconductors (see Ref. 10 and references therein). Experimental results concerning quasi-two-dimensional excitons obtained by this technique began to emerge relatively recently. FIR magnetospectroscopy was used[6] in measurements of $e-h$ interaction as a function of the population number in type II InAs/AlGaSb quantum wells. Evidence in favor of the existence of a stable exciton state in a strong magnetic field (in the presence of excess free electrons) was obtained.[7] FIR spectra of type II GaAs/AlAs quantum wells in a strong magnetic field were also measured in the regime of low exciton density.[11] Another highly sensitive technique, namely the optically detected cyclotron resonance, was used in studies of direct excitons in GaAs quantum wells.[12–15]

No detailed theoretical study of intraband magneto-optical properties of quasi-two-dimensional excitons has been published as yet. On the contrary, one can even find in the literature erroneous claims[16] about the energy of $s \rightarrow p$ intraexciton transitions as a function of the magnetic field (a drop in the transition energy with B), which contradict experimental data.[15] Previously we analyzed changes in the $1s \rightarrow np^{\pm}$ transitions in DQW due to the magnetic field in the regime of a wide barrier between wells at $\mathscr{E}=0$.[17]

This paper reports on a theoretical investigation of the energies and dipole matrix elements of FIR transitions in symmetric InGaAs/GaAs DQW as functions of the barrier width in a strong magnetic field $B=10$ T (Sec. 3.1), changes in the transitions caused by an applied electric field $\mathscr{E}$ in a fixed magnetic field (Sec. 3.2), and changes in these parameters with a magnetic field under strong and intermediate electric fields $\mathscr{E}$ (Sec. 2.1). Sections 2.1–2.3 describe the calculation techniques, and Sec. 2.4 gives a qualitative description of magnetoexciton spectra in DQW. Some results of this work were briefly reported in our previous publications.[17,18]

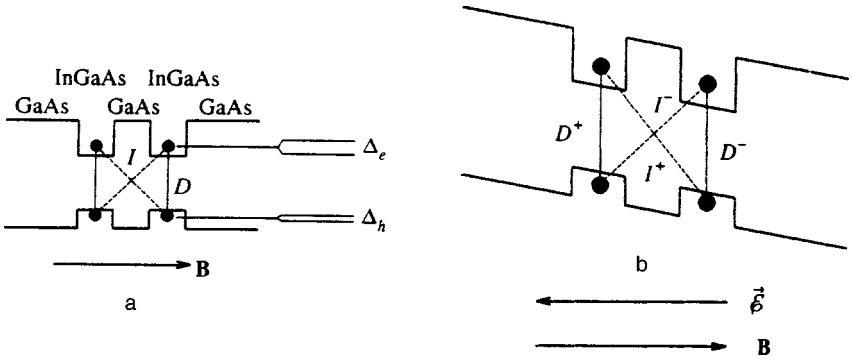


FIG. 1. a) Direct D and indirect I excitons in DQW. Splittings between symmetric and antisymmetric states of electrons and holes, Δ_e and Δ_h, are shown. b) Excitons in DQW in an electric field

2. THEORETICAL MODEL

2.1. System Hamiltonian

Consider a heavy-hole exciton in a symmetric strained $In_{x_1}Ga_{1-x_1}As/GaAs/In_{x_2}Ga_{1-x_2}As$ DQW with $x_1=x_2=0.2$, well widths $L_1=L_2$, and barrier width L_b (Fig. 1). Light holes, whose branch is separated from that of heavy holes by several tens of meV, can be neglected.[19] The exciton Hamiltonian in DQW in perpendicular electric $\vec{\mathscr{E}}=(0,0,\mathscr{E})$ and magnetic $\mathbf{B}=(0,0,B)$ fields can be expressed as

$$H=H_{ez}+H_{hz}+H_{2D}+U_{eh}\equiv H_0+U_{eh}, \tag{1}$$

where the Hamiltonian components

$$H_{ez}=-\frac{\hbar^2}{2m_e}\frac{\partial^2}{\partial z_e^2}+V_e(z_e)+e\mathscr{E}z_e, \tag{2}$$

$$H_{hz}=-\frac{\hbar^2}{2m_{hz}}\frac{\partial^2}{\partial z_h^2}+V_h(z_h)-e\mathscr{E}z_h \tag{3}$$

describe the motion of free electrons and holes along the z-axis. The well depths for electrons and holes are assumed to be $V_{ei}=0.8\Delta E_g(x_i)$ and $V_{hi}=0.2\Delta E_g(x_i)$, where $\Delta E_g(x_i)=E_g(0)-E_g(x_i)$ is the band-gap offset, $E_g(x)=1.519-1.47x+0.375x^2$ eV is the gap width in $In_xGa_{1-x}As$, and the effective masses are $m_e=0.067$ and $m_h=0.35$.[9,19] The exciton energy is measured with respect to $E_g(0)$.

The solutions of the one-dimensional Schrödinger equations

$$H_{ez}\zeta_i(z_e)=E_i^{(e)}\zeta_i(z_e),\quad H_{hz}\xi_j(z_h)=E_j^{(h)}\xi_j(z_h), \tag{4}$$

corresponding to the lowest discrete levels are calculated numerically. In order to avoid difficulties with the continuum in an electric field $\mathscr{E}\neq 0$, boundary conditions corresponding to infinite energy barriers at sufficient distances (200–500 Å) from the DQW are invoked. When $\mathscr{E}=0$ and the DQW is symmetric (i.e., the two wells are identical at $x_1=x_2$ and $L_1=L_2$), the subscripts $i,j=s,a$ correspond to the symmetric ground state (s) and antisymmetric first excited state (a) of electrons and holes, respectively:

$$\zeta_{s(a)}(z_e)=\pm\zeta_{s(a)}(-z_e),\quad \xi_{s(a)}(z_h)=\pm\xi_{s(a)}(-z_h). \tag{5}$$

The splittings between symmetric and antisymmetric states, $\Delta_e=E_a^{(e)}-E_s^{(e)}$ and $\Delta_h=E_a^{(h)}-E_s^{(h)}$, are determined by penetration of the wave functions under the barrier (see Sec. 3.1).

The Hamiltonian H_{2D} of relative motion of a noninteracting electron–hole pair with magnetic momentum of the center of mass $\mathbf{K}$ in a perpendicular magnetic field B has the form[20,21]

$$H_{2D}=-\frac{\hbar^2}{2\mu}\nabla_{\boldsymbol{\rho}}^2+\frac{1}{2}\hbar(\omega_{ch}-\omega_{ce})\hat{l}_z+\frac{e^2B^2}{8\mu c^2}\rho^2+\frac{e}{Mc}\mathbf{B}\cdot(\boldsymbol{\rho}\times\mathbf{K})+\frac{K^2}{2M}, \tag{6}$$

where $\boldsymbol{\rho}=\boldsymbol{\rho}_e-\boldsymbol{\rho}_h$ is the relative separation, $\mu^{-1}=m_e^{-1}+m_{h\|}^{-1}$ is the reduced mass, $\omega_{ce(h)}=eB/m_{e(h\|)}c$ is the electron (hole) cyclotron frequency, and $\hat{l}_z=-i(\boldsymbol{\rho}\times\nabla_{\boldsymbol{\rho}})_z$ is the z-projection of the orbital angular momentum of relative motion. In this formula we have taken advantage of the existence of an exact integral of the motion, namely the magnetic center-of-mass momentum,[20] whose operator is

$$\hat{\mathbf{K}}=-i\hbar\nabla_{\mathbf{R}}-\frac{e}{c}\mathbf{A}(\boldsymbol{\rho}),$$

where $\mathbf{R}=(m_e\boldsymbol{\rho}_e+m_{h\|}\boldsymbol{\rho}_h)/M$ is the center-of-mass location, and $M=m_e+m_{h\|}$. The vector potential is expressed in the symmetric gauge $\mathbf{A}=\frac{1}{2}\mathbf{B}\times\boldsymbol{\rho}$, and $\mathbf{r}=(\boldsymbol{\rho},z)$. Note that in Eqs. (2), (3), and (6), an isotropic electron spectrum is assumed, while the masses of holes moving in the quantum well plane and in the perpendicular direction are different, $m_{h\|}\neq n_{hz}$ (see Appendix to Ref. 9, where the nonparabolicity of $m_{h\|}$ is discussed). In what follows, we will neglect the difference between effective masses in the InGaAs wells and GaAs barriers. The energy of the Coulomb interaction between electrons and holes can be written in the form

$$U_{eh}=U_{eh}(|\mathbf{r}_e-\mathbf{r}_h|)=-\frac{e^2}{\epsilon|\mathbf{r}_e-\mathbf{r}_h|}, \tag{7}$$

where $\epsilon=12.5$. In a InGaAs/GaAs DQW, the effect of electrostatic image forces is very weak, owing to the small difference between the dielectric constants of GaAs ($\epsilon=12.5$) and $In_{0.2}Ga_{0.8}As$ ($\epsilon=13$),[9] so this effect is neglected.

2.2. Wave functions of magnetoexcitons with $K=0$

In order to calculate the eigenfunctions of Hamiltonian (1), we diagonalize the term U_{eh} of electron–hole interaction in the basis of the wave functions of noninteracting electron–hole pairs in a DQW in a magnetic B and electric $\mathscr{E}$ fields. The wave function of an exciton with center-of-mass momentum $\mathbf{K}=0$ (see Sec. 2.3) can be expressed in the form of the expansion[9]

$$\Psi_{\mathbf{K}=0,l_z}(\mathbf{r}_e,\mathbf{r}_h)=\exp\left(\frac{i[\boldsymbol{\rho}\times\mathbf{R}]_z}{2l_B^2}\right)\Phi_{l_z}(\boldsymbol{\rho},z_e,z_h), \tag{8}$$

$$\Phi_{l_z}(\boldsymbol{\rho},z_e,z_h)=\sum_{i,j=1,2}\sum_{n-m=l_z}A_{ijnm}\zeta_i(z_e)\xi_j(z_h)\phi_{nm}(\boldsymbol{\rho}), \tag{9}$$

where $l_B=(\hbar c/eB)^{1/2}$, $\zeta_i(z_e)$ and $\xi_j(z_h)$ are the electron and hole wave functions determined by Eq. (4), $\phi_{nm}(\boldsymbol{\rho})=(a^\dagger)^n(b^\dagger)^m|00\rangle/\sqrt{n!m!}$ are the factored wave functions in a magnetic field B,[22,23] and $\boldsymbol{\rho}=\boldsymbol{\rho}_e-\boldsymbol{\rho}_h$. For magnetoexcitons the quantum numbers n and m label the Landau levels of electrons and holes, respectively, and the angular momentum projection $l_z=n-m$. Note that the wave functions $\phi_{nm}(\boldsymbol{\rho})$ of the e–h-pairs correspond to bound states in a field B (the characteristic length scale $\langle nm|\boldsymbol{\rho}^2|nm\rangle=2(n+m+1)l_B^2$). Therefore Eq. (9) can be considered an expansion in exciton wave functions. Note also that Eq. (9) takes into account mixing of different subbands i and j, which is important for accuracy of calculations (compare to the discussion in Ref. 24).

The energy eigenvalues E and eigenfunctions of Hamiltonian (1) for a magnetoexciton with angular momentum projection $l_z=n-m$ are calculated by numerically solving the secular equation

$$\mathrm{Det}\left(\left[E_i^{(e)}+E_j^{(h)}+\hbar\omega_{ce}\left(n+\frac{1}{2}\right)+\hbar\omega_{ch}\left(m+\frac{1}{2}\right)-E\right]\times\delta_{ii'}\delta_{jj'}\delta_{nn'}\delta_{mm'}+U_{ijnm}^{i'j'n'm'}\right)=0, \tag{10}$$

where the matrix elements of the e–h interaction have the form

$$U_{ijnm}^{i'j'n'm'}=\langle i'j'n'm'|U_{eh}|ijnm\rangle=\delta_{n'-m',n-m}\times\int\frac{d^2q}{(2\pi)^2}\left(-\frac{2\pi e^2}{\varepsilon q}\right)F_{ij}^{i'j'}(q)\mathscr{D}_{nm}^{n'm'}(\mathbf{q}), \tag{11}$$

$$\delta_{n'-m',n-m}\mathscr{D}_{nm}^{n'm'}(\mathbf{q})=\left(\frac{\min(n,n')!}{\max(n,n')!}\frac{\min(m,m')!}{\max(m,m')!}\right)^{1/2}\times\left(\frac{q^2l_B^2}{2}\right)^{|n-n'|}L_{\min(n,n')}^{|n-n'|}\left(\frac{q^2l_B^2}{2}\right)\times L_{\min(m,m')}^{|m-m'|}\left(\frac{q^2l_B^2}{2}\right)\exp\left(-\frac{q^2l_B^2}{2}\right), \tag{12}$$

where L_n^m are the generalized Laguerre polynomials and

$$F_{ij}^{i'j'}(q)=\int_{-\infty}^{\infty}dz_e\int_{-\infty}^{\infty}dz_h\exp(-q|z_e-z_h|)\times\zeta_i(z_e)\zeta_{i'}(z_e)\xi_j(z_h)\xi_{j'}(z_h) \tag{13}$$

are the form factors related to the wave functions of one-dimensional motion. The integrals in Eq. (13) and then in Eq. (11) are calculated numerically; the calculation is based on an expansion that includes the two lowest electron and hole levels (i and j), at least ten Landau levels at $B=12$ T, and up to 36 Landau levels at $B=2$ T. An approximate technique of taking into account the nonparabolicity of heavy holes is described in Appendix to Ref. 9.

2.3. Interaction between excitons and FIR radiation

In the Faraday geometry (the wave vector of light is aligned with the magnetic field **B**), the Hamiltonian describing light absorption due to interaction between the excitons and FIR electric field with amplitude $\mathscr{F}_0$ and frequency ω has the form

$$\delta\hat{V}^{\pm}=\frac{e\mathscr{F}_0}{\omega}\left(\frac{\pi_e^{\pm}}{m_e}-\frac{\pi_h^{\pm}}{m_{h\|}}\right)\exp(-i\omega t). \tag{14}$$

Here the plus and minus signs denote left-handed (right-handed) circular polarization $\sigma^{\pm}$; $\pi_\alpha^{\pm}=\pi_{\alpha x}\pm i\pi_{\alpha y}$ ($\alpha=e,h$), and

$$\boldsymbol{\pi}_e=-i\hbar\nabla_e+\frac{e}{c}\mathbf{A}_e,\quad \boldsymbol{\pi}_h=-i\hbar\nabla_h-\frac{e}{c}\mathbf{A}_h$$

are the kinematic momentum operators. One can show that

$$[\delta\hat{V}^{\pm},\hat{\mathbf{K}}]=0. \tag{15}$$

This means that the magnetic momentum does not change during an FIR transition. All populated exciton states contribute to intraband FIR transitions, including those with finite **K**. This is the difference between intraband and interband transitions, since in the latter only excitons with $\mathbf{K}=0$ are optically active. In this paper we consider only FIR transitions of excitons with center-of-mass momentum $\mathbf{K}=0$, which can be characterized by a constant angular momentum projection l_z (see Eq. (6)). Therefore, the selection rules for excitons with $\mathbf{K}=0$ in a magnetic field B have the usual form

$$\langle\Psi'_{\mathbf{K}=0,l'_z}|\delta\hat{V}^{\pm}|\Psi_{\mathbf{K}=0,l_z}\rangle\sim\delta_{l'_z,l_z\pm1}. \tag{16}$$

Effects related to FIR absorption by two-dimensional magnetoexcitons with $\mathbf{K}\neq0$ were discussed in Ref. 18b. By using expansion (9) and the formula

$$\delta\hat{V}^{+}=\frac{i\sqrt{2}e\hbar\mathscr{F}_0}{\omega l_B}\left(\frac{a^\dagger}{m_e}-\frac{b}{m_{h\|}}\right)e^{-i\omega t}, \tag{17}$$

where $a^\dagger$ ($b^\dagger$) is the ladder operator corresponding to electron (hole) Landau levels (see Eq. (9)), we can express the matrix elements of intraband transitions between s and (for example) p^+ exciton states by the formula

$$|f|^2\sim|\langle\Psi_{\mathbf{K}=0,p^+}|\delta\hat{V}^{+}|\Psi_{\mathbf{K}=0,s}\rangle|^2$$

$$\sim \left| \sum_{ij=1,2} \sum_{n} \sqrt{n+1} A^{*}_{ijn+1\,n} \left(\frac{A_{ijnn}}{m_e} - \frac{A_{ijn+1\,n+1}}{m_{h\|}} \right) \right|^2 l_B^{-2}. \tag{18}$$

For symmetric DQW (subscripts $i,j=s,a$), FIR transitions in the Faraday geometry are allowed only between exciton states with the same spatial parity $i \otimes j$ under inversion ($z_e \to -z_e$, $z_h \to -z_h$): $S \to S$ and $A \to A$ (see also Sec. 2.4.1).

2.4. Magnetoexcitons in DQW: qualitative description

2.4.1. Classification of states. The DQW have four exciton terms (instead of one in an isolated quantum well when the lowest size-quantized level is taken into account).[4,9,24,25] The classification of the states depends on the R, which is the ratio between one-particle symmetric–antisymmetric e–h splitting Δ_e, Δ_h and the difference between the binding energies of direct (D) and indirect (I) excitons: $\delta E_{ID} = E_D - E_I$, $R = \max(\Delta_e, \Delta_h)/\delta E_{DI}$.

When $R \ll 1$, the wide-barrier regime of DQW is realized, and the exciton states in DQW are predominantly either direct or indirect.[1)] In addition, there is splitting due to tunneling through the barrier: for example, in symmetric DQW at $\mathscr{E}=0$ each direct and indirect state is split into states symmetric (S) and antisymmetric (A) under inversion ($z_e \to -z_e$, $z_h \to -z_h$). In the case of a wide barrier, the symmetric–antisymmetric splitting is governed by two-particle e–h tunneling through the barrier, $\Delta_X \simeq \Delta_e \Delta_h / \delta E_{DI}$.[9] The splitting Δ_X is suppressed by a rise in excitonic effects ($\sim \delta E_{DI}^{-1}$); in particular, it decreases with increasing magnetic field B. In the wide-barrier regime, we will label exciton states by quantum numbers of the high magnetic field limit (D_{nm}, I_{nm}) by indicating the numbers of e and h Landau levels that are dominant in expansion (9), and by the spatial character of the states. When necessary, we will indicate the state inversion symmetry (S or A) at $\mathscr{E}=0$, and under strong $\mathscr{E}$ the lower (D^-_{nm} and I^-_{nm}) and upper (D^+_{nm} and I^+_{nm}) branches of the exciton spectrum (one example is shown in Fig. 2).

For a sufficiently thin barrier, the opposite limit is encountered, $\Delta_e, \Delta_h \gg \delta E_{DI}$ and $R \gg 1$. In this regime, excitons cannot be classified as direct or indirect, since these states are mixed. Many of the characteristic features of the narrow-barrier regime can be understood in the one-particle approximation, neglecting excitonic effects.[26,27] Exciton states in symmetric DQW at $\mathscr{E}=0$ can be classified as ij_{nm}, where $i, j = s, a$, in accordance with the quantum numbers of electron and hole wave functions $\zeta_i \xi_j \phi_{nm}$, which dominate expansion (9). The states ss_{nm} and aa_{nm} (sa_{nm} and as_{nm}) correspond to exciton states that are spatially symmetric S (antisymmetric A).

2.4.2. FIR transitions. In a strong magnetic field, the exciton $1s$ states are formed predominantly by the wave function ϕ_{00} of the lowest e and h Landau levels. Owing to the Coulomb e–h interaction, there is a small admixture of states ϕ_{nn} of higher Landau levels proportional to $\sim l_B / a_{Be(h)} \ll 1$, where $a_{Be(h)} = \epsilon \hbar^2 / m_{e(h)} e^2$. Similarly, the $2p^+(2p^-)$ exciton states are formed predominantly by the wave function $\phi_{10}(\phi_{01})$ with a small admixture of $\phi_{n+1n}(\phi_{n\,n+1})$ states. Therefore the $1s \to 2p^+$ ($1s \to 2p^-$) excition transition can be considered an electron (hole) cyclotron resonance, $\phi_{00} \to \phi_{10}$ ($\phi_{00} \to \phi_{01}$), which is modified by excitonic effects. The evolution of the energy and matrix elements of transitions from the symmetric $1s$ ground states or $D_{00}S$ to various $p^{\pm}$ states in a magnetic field B at $\mathscr{E}=0$ in $In_{0.2}Ga_{0.8}As/GaAs$ DQW were discussed in a previous publication.[17] For example, the strongest $1s \to p^+$ transitions are $D_{00}S \to D_{10}S$ and $D_{00}S \to I_{10}S$, i.e., the transition to the first electron Landau level. The transition energy is higher than the free-electron cyclotron energy, since the original $1s$ state is more tightly bound than the final $2p^{\pm}$ state. In DQW, the dipole matrix element $|f^2|$ of the $D_{00}S \to D_{10}S$ transition only increases with B. An explanation of such behavior was given in Ref. 17.

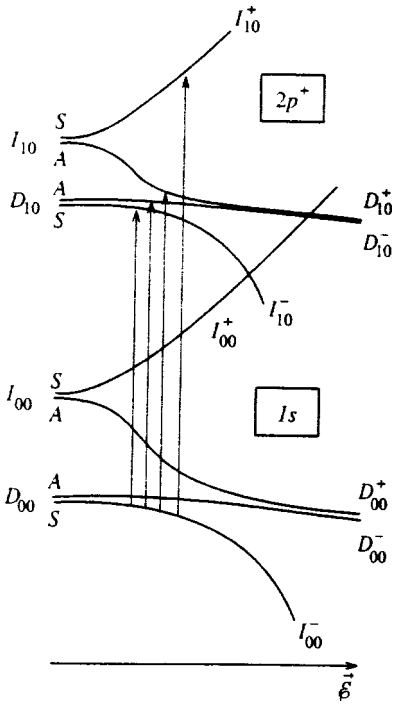


FIG. 2. Energies of $1s$ and $2p^+$ exciton states in DQW in the wide-barrier regime as functions of the electric field $\mathscr{E}$. Vertical arrows show the four lowest transitions from the $1s$ ground state to $2p^+$ excited states, depicted in detail in Fig. 4a.

3. NUMERICAL CALCULATIONS AND DISCUSSION

In this part of the paper, we discuss results for symmetric $In_{0.2}Ga_{0.8}As/GaAs$ DQW with $L_1 = L_2 = 60$ Å. Section 3.1 is dedicated to the dependence of energies and oscillator strengths of intraexciton FIR transitions on the barrier thickness L_b in a magnetic field $B=10$ T at $\mathscr{E}=0$, and Secs. 3.2 and 3.3 to their dependence on the magnetic and electric fields at fixed $L_b = 60$ Å.

3.1. Dependence on the barrier thickness

An important parameter that determines many of the features of excitons in DQW (in particular, the character of a crossover from the direct to indirect regime in an applied electric field $\mathscr{E}$) is the tunnel barrier thickness L_b, which

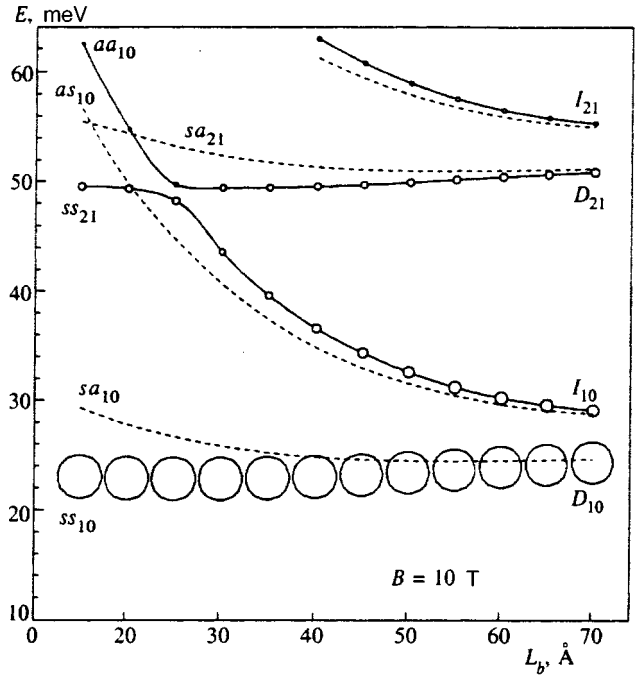


FIG. 3. Energies and dipole matrix elements of exciton transitions from the symmetric $1s$ ground state to p^- excited states as functions of the barrier width L_b in an InGaAs/GaAs DQW with $L_1=L_2=60$ Å and $x=0.2$ at $\mathscr{E}=0$. The areas of open circles are proportional to $|f^2|$ in Eq. (18). Dashed lines correspond to forbidden ($S\to A$) transitions to antisymmetric final states. Characteristics of final states are labeled in the graph.

determines the coupling between the quantum wells. Energies and matrix elements of transitions from the symmetric $1s$ ground state to p^+ states are plotted versus L_b in Fig. 3. In the narrow-barrier regime, the initial state is formed mostly of the $\zeta_s\xi_s\phi_{00}$ wave function denoted by ss_{00} (see Sec. 2.4.1), and two possible symmetric $2p^+$ final states are ss_{10} and aa_{10} with the wave function $\zeta_s\xi_s\phi_{10}$ and $\zeta_a\xi_a\phi_{10}$, respectively. The energies of these two transitions are $\hbar\omega_{ce}+\delta E_1$ and $\hbar\omega_{ce}+\Delta_e+\Delta_h+\delta E_2$, respectively. Here δE_i are the energy corrections due to the differences in the Coulomb binding energies of the $1s$ and $2p^+$ states, and $\delta E_2 \ll \Delta_e+\Delta_h$ for small L_b. The matrix element of the $ss_{00}\to ss_{10}$ transition is large, and that of the $ss_{00}\to aa_{10}$ transition is very small (and is due to the admixture of the aa_{00} state to ss_{00} and ss_{10} to aa_{10}). As $\Delta_e+\Delta_h>\hbar\omega_{ce}$ in the narrow-barrier regime, an anticrossing between the $3p^+$ and ss_{21} states takes place at $L_b\simeq 25$ Å, which leads to a redistribution of oscillator strengths between the transitions.

After the crossover to the wide-barrier regime, excitons become predominantly either direct (D) or indirect (I). For example, the ground $1s$ state is the $D_{00}S$ exciton with the wave function $(\zeta_s\xi_s+\zeta_a\xi_a)\phi_{00}/\sqrt{2}$, and the two $2p^+$ final states are the direct $D_{10}S$ and indirect $I_{10}S$ excitons with the wave functions $(\zeta_s\xi_s\pm\zeta_a\xi_s)\phi_{10}/\sqrt{2}$. Figure 3 shows that the energy of the transition to the first $2p^+$ excited state increases very slowly with L_b because the changes in the binding energies of the initial and final states cancel each other. The transition energy to the next $2p^+$ state (with the wave function $\zeta_a\xi_a\phi_{10}$ at small L_b and $(\zeta_s\xi_s-\zeta_a\xi_a)\phi_{10}/\sqrt{2}$ at large L_b) rapidly drops with L_b, since the symmetric–antisymmetric splitting Δ_α exponentially drops with the barrier width L_b. In the wide-barrier regime the parameter Δ_α is determined by one-particle tunneling across the barrier: $\Delta_\alpha \simeq |E_\alpha|\exp(-\mathscr{S}_\alpha)/\pi$, where $\mathscr{S}_a=\sqrt{2m_{\alpha z}|E_\alpha|}L_\mathrm{b}/\hbar$ and E_α is the energy of the level in a single quantum well.[28]

In the considered case of $In_{0.2}Ga_{0.8}As/GaAs$ DQW with $L_1=L_2=60$ Å and, for example, $L_\mathrm{b}\simeq 60$ Å, the numerical calculation yields[9] $\Delta_e\simeq 4.9$ meV and $\Delta_h\simeq 0.6$ meV. For large L_b the two lowest transitions to the $2p^+$ states are the transitions to the direct ($D_{00}S\to D_{10}S$) and indirect ($D_{00}S\to I_{10}S$) excitons with the energy difference between them $\simeq \delta E_{DI}=E_D-E_I$. The separation from the optically forbidden transitions to A-states $\approx\Delta_X\simeq\Delta_e\Delta_h/\delta e_{DI}$. The transitions to the next Landau levels, $D_{00}S\to D_{21}S$ and $D_{00}S\to I_{21}S$, correspond to the final $3p^+$ states, and their oscillator strengths are considerably smaller.

3.2. Evolution of FIR transitions in an electric field at $B\neq 0$

A perpendicular electric field $\mathscr{E}$ breaks the symmetry under inversion $z\to -z$ and allows all $s\to p^\pm$ transitions in DQW. Exciton $1s$ and $2p^\pm$ levels in an electric field $\mathscr{E}$ in the wide-barrier regime are shown in Fig. 2. In a weak electric field, all levels shift quadratically due to the Stark effect. In intermediate fields, depending on Landau level numbers nm, the crossover between direct and indirect exciton states occurs. Owing to the lower Coulomb energy, this crossover happens in weaker electric fields for the $2p^+$ state than for the $1s$ state. This effect can be seen in the FIR absorption spectra.

Let us consider evolution of transitions from the ground $1s$ state to the excited $p^\pm$ states in the electric field $\mathscr{E}$ and fixed magnetic field (Fig. 4). The transition to the first excited $2p^+$ state experiences a red shift, which saturates in strong electric fields $\mathscr{E}$. This shift is also a function of B: the higher the magnetic field, the larger the red shift. This shift is controlled by excitonic effects. Indeed, in a weak field $\mathscr{E}$ both the initial D_{00} and final D_{10} states are direct excitons. In a strong electric field, they become the indirect magnetoexcitons I^-_{00} and I^-_{10} with lower binding energies. As a result, the exciton transition energy drops and approaches that of the cyclotron resonance of free carriers (these energies are marked by arrows in Fig. 4).

Note also the nonmonotonic dependence of the energy of transition to the third $2p^+$ excited state (at $\mathscr{E}=0$ this is the $D_{00}S\to I_{10}A$ transition, which is strictly forbidden by symmetry selection rules). This nonmonotonic behavior is due to successive crossovers from the direct state to the indirect state, first for the initial state and then for the final state in the FIR transition. The first crossover (when the third excited $2p^+$ state transforms from the indirect $I_{10}A$ to direct D^+_{10} magnetoexciton) occurs in a lower field $\mathscr{E}$, when the initial state is predominantly a spatially direct exciton D_{00}. This explains both the growth in the oscillator strength and red shift due to the larger Stark effect for the $2p^+$ state. Then the initial $1s$ state undergoes a crossover from $D_{00}S$ to I^-_{00}. As a result, we have the $I^-_{00}\to D^+_{10}$ transition, which has an oscillator strength decreasing with the field strength and a shift

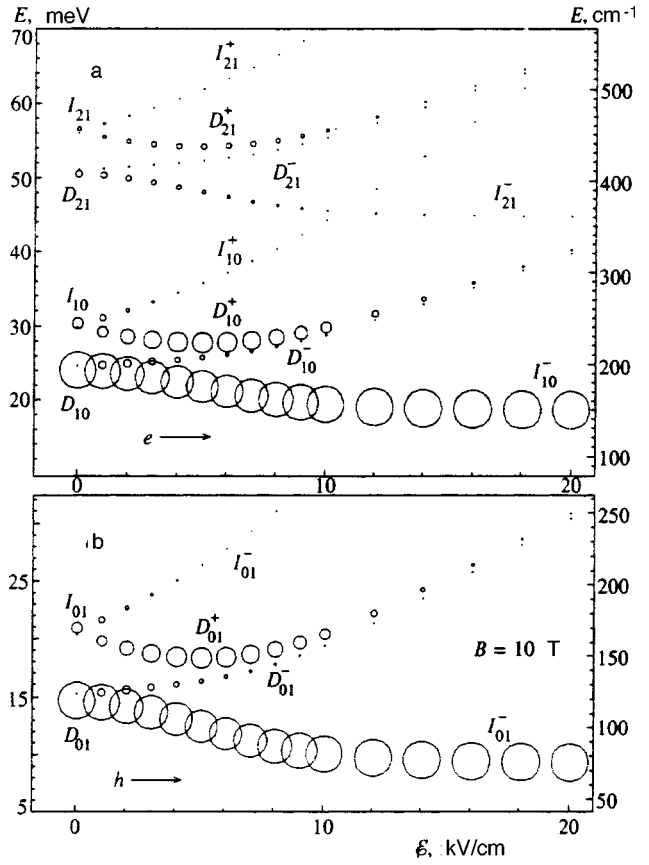


FIG. 4. Evolution in an electric field $\mathscr{E}$ of energies and dipole matrix elements of transitions from the $1s$ ground state to (a) excited p^+ states and (b) p^- states in a magnetic field $B=10$ T for symmetric InGaAs/GaAs DQW with $L_1=L_2=60$ Å and $x=0.2$. The areas of open circles are proportional to the transition matrix element squared, $|f^2|$. The horizontal arrows indicate energies of cyclotron resonances for free electrons and holes. Characteristics of final states in the transitions are labeled in the graph.

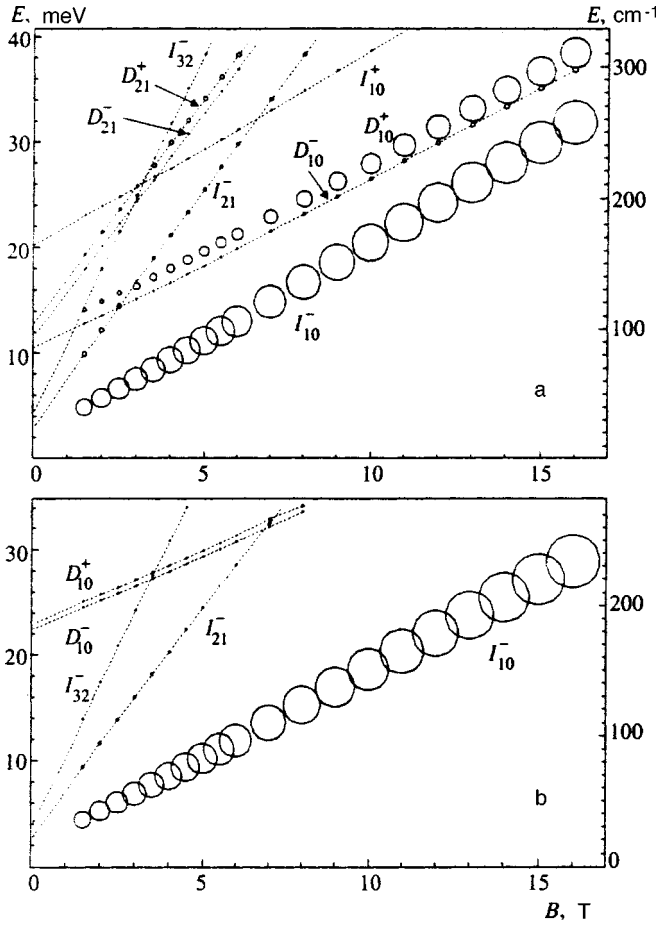


FIG. 5. Evolution in a magnetic field B of energies and dipole matrix elements of $1s\rightarrow p^+$ transitions as functions of the magnetic field B in electric fields (a) $\mathscr{E}=7$ kV/cm and (b) $\mathscr{E}=17.2$ kV/cm. The dotted lines show positions of several weak transitions. The areas of open circles are proportional to dipole matrix elements squared, $|f|^2$ [Eq. 18)].

almost linear in $\mathscr{E}$ due to the Stark effect in the initial state I_{00}^- of the indirect exciton.

3.3. Evolution of FIR transitions in a magnetic field at fixed $\mathscr{E}\neq 0$

The binding energy of indirect excitons increases with B more slowly than that of direct excitons. Therefore a magnetic field B induces a crossover from an indirect to direct state in a strong fixed electric field $\mathscr{E}$, which depends on the Landau level numbers of the exciton states.[4,9,25] This effect can be seen in exciton FIR absorption spectra. The evolution of both the energies and dipole matrix elements of the $1s\rightarrow p^+$ transition with the magnetic field B in the electric field $\mathscr{E}=7$ kV/cm is illustrated by Fig. 5a, and in the field $\mathscr{E}=17.2$ kV/cm by Fig. 5b.

In the stronger electric field $\mathscr{E}$ (Fig. 5a), the initial $1s$ state is the indirect exciton I_{00}^-. In the magnetic field range under consideration, $B<16$ T, no crossover between the direct and indirect states occurs, so only the $I_{00}^-\rightarrow I_{10}^-$ transition has a large matrix element, which rapidly (essentially linearly) increases with B. Transitions to all remaining higher levels have much lower intensities.

In the weaker electric field (Fig. 5b), the gap between the I_{10}^- and $D_{10}^{\pm}$ states is considerably smaller. Furthermore, the mixing between direct and indirect exciton states is notable.[9] Therefore the $I_{00}^-\rightarrow D_{10}^+$ transition has a notable matrix element even at intermediate magnetic fields B. At $B<4$ T, the behavior of spectral lines is complicated owing to numerous anticrossings between levels of direct np^+ and indirect $n'p^+$ excitons, where $n'>n$. This results in small splittings of lines and redistribution of their intensities, which is similar to the behavior of interband transitions discussed in Refs. 4 and 9. At $B>10$ T, the $I_{00}^-\rightarrow D_{10}^+$ transition amplitude increases rapidly because of the indirect-to-direct crossover in the initial state: the ground state gradually evolves[4,9] and transforms from the indirect I_{00}^- to direct D_{00}^- exciton. Since the excitonic effects in $2p^{\pm}$ states are considerably weaker, such a crossover occurs in much stronger magnetic fields.

Note that the transition to the final D_{10}^- state remains very weak because of the large difference between the shapes of wave functions of (almost degenerate) D_{10}^- and D_{10}^+ direct excitons. Indeed, it follows from the probability distribution for the excitons (Fig. 6), i.e.,

$$P_{\mathbf{K}=0,l_z}(z_e,z_h)=\int d^2\rho\,|\Psi_{\mathbf{K}=0,l_z}(\mathbf{r}_e,\mathbf{r}_h)|^2, \qquad (19)$$

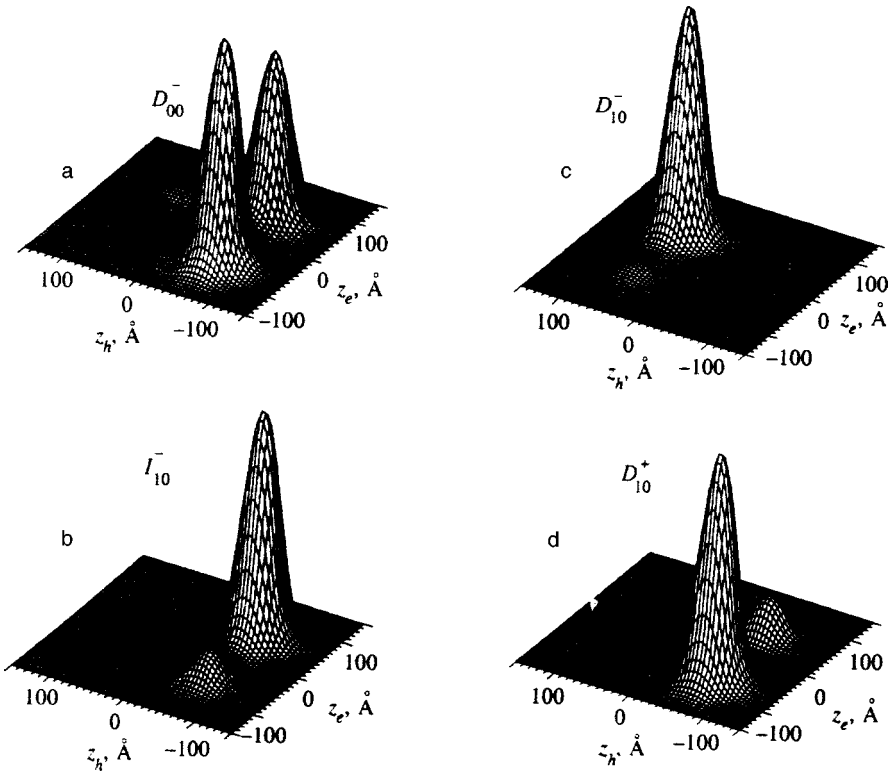


FIG. 6. Probability distributions $P(z_e, z_h)$ (Eq. (19)) for exciton states involved in transitions shown in Fig. 5a ($B=10$ T): a) initial $1s$ state D_{00}^-; three of the various low-lying $2p^+$ final states: b) I_{10}^-, c) D_{10}^-, and d) D_{10}^+.

that in fields $\mathscr{E}=7$ kV/cm and $B=10$ T, the ground state at the lowest Landau levels (i.e., the initial state in the transitions in question) is predominantly direct (we denote it by D_{00}^-). This state is predominantly a direct exciton in the left quantum well (z_e, $z_h<0$) with a large admixture of an indirect component ($z_e<0$, $z_h>0$) and an extremely small component corresponding to a direct exciton in the right quantum well (z_e, $z_h>0$). In the same fields, the $2p^+$ ground state is polarized because the Coulomb excitonic effects are not as strong, i.e., this is predominantly an indirect exciton I_{10}^- ($z_e>0$, $z_h<0$) with a small admixture of the direct exciton (z_e, $z_h<0$). The dipole matrix element of the $D_{00}^-\rightarrow I_{10}^-$ transition is large due to the large spatial overlap between the wave functions of these states. The next two excited $2p^\pm$ states (D_{10}^- and D_{10}^+) are predominantly direct excitons in the right and left quantum wells, respectively. As a result, only the $D_{00}^-\rightarrow D_{10}^+$ transition is strong, whereas the $D_{00}^-\rightarrow D_{10}^-$ transition is very weak because of the small spatial overlap between the wave functions of these two states.

4. CONCLUSIONS

We have analyzed theoretically intraband magneto-optical transitions of direct and indirect excitons in InGaAs/GaAs coupled double quantum wells. Features in the behavior of transition energies and matrix elements due to the crossover from the narrow-barrier regime to the wide-barrier regime in an electric field $\mathscr{E}$ and magnetic field B have been described. In particular, a red shift of the transition from the $1s$ ground state to the first $2p^\pm$ excited state in DQW due to the direct–indirect crossover in an electric field $\mathscr{E}$ has been predicted. This effect is due to Coulomb excitonic effects and increases with B. These theoretical results may be useful in planning experiments and interpreting their results.

The author is indebted to A. L. Yablonskii for designing computer codes used in numerical calculations, and to G. E. W. Bauer, L. V. Butov, A. A. Dremin, J. Kono, B. D. McCombe, and V. B. Timofeev for useful discussions. This work was supported by grants from Volkswagen, the Russian Fund for Fundamental Research, and INTAS.

*)E-mail: dzyub@gpi.ac.ru

1)This classification is approximately applicable to the case of a barrier with an intermediate width, when $R\sim 1$. Note also that at a fixed L_b in sufficiently strong field B one can satisfy the condition $R<1$.

[1] Y. Kuramoto and C. Horie, Solid State Commun. **25**, 713 (1978); I. V. Lerner, Yu. E. Lozovik, and D. R. Musin, J. Phys. C **14**, L311 (1980); D. Yoshioka and A. H. MacDonald, J. Phys. Soc. Jpn. **59**, 4211 (1990); X. M. Chen and J. J. Queen, Phys. Rev. Lett. **67**, 895 (1991).

[2] I. V. Lerner and Yu. E. Lozovik, Zh. Eksp. Teor. Fiz. **80**, 1488 (1981) [Sov. Phys. JETP **53**, 763 (1981)]; A. B. Dzyubenko and Yu. E. Lozovik, Fiz. Tverd. Tela (Leningrad) **25**, 1519 (1983); **26**, 1540 (1984) [Sov. Phys. Solid State **25**, 874 (1983); **26**, 938 (1984)].

[3] L. V. Butov, A. Zrenner, G. Abstreiter, G. Böhm, and G. Weimann, Phys. Rev. Lett. **73**, 304 (1994).

[4] L. V. Butov, A. Zrenner, G. Abstreiter, A. V. Petinova, and K. Eberl, Phys. Rev. B **52**, 12153 (1995).

[5] M. Bayer, V. B. Timofeev, F. Faller, T. Gutbrod, and A. Forchel, Phys. Rev. B **54**, 8799 (1996).

[6] J. Kono, B. D. McCombe, J.-P. Cheng, I. Lo, W. C. Mitchel, and C. E. Stutz, Phys. Rev. B **50**, 12492 (1994).

[7] J.-P. Cheng, J. Kono, B. D. McCombe, I. Lo, W. C. Mitchel, and C. E. Stutz, Phys. Rev. Lett. **74**, 450 (1995).

[8] L. V. Butov, in *Proceedings of the 23rd International Conference on Physics of Semiconductors (ICPS-23)*, M. Scheffler and R. Zimmermann (eds.), World Scientific, Singapore (1996), p. 1927.

[9] A. B. Dzyubenko and A. L. Yablonskii, Phys. Rev. B **53**, 16355 (1996).

[10] D. Labrie, M. L. W. Thewalt, I. J. Booth, and G. Kirczenov, Phys. Rev. Lett. **61**, 1882 (1988).
[11] C. C. Hodge, C. C. Phillips, M. S. Skolnick, G. W. Smith, C. R. Whitehouse, P. Dawson, and C. T. Foxon, Phys. Rev. B **41**, 12319 (1990).
[12] M. Salib, H. A. Nickel, G. S. Herold, A. Petrou, B. D. McCombe, R. Chen, K. K. Bajaj, and W. Schaff, Phys. Rev. Lett. **77**, 1135 (1996).
[13] J. Cerne, J. Kono, M. S. Sherwin, M. Sundaram, A. C. Gossard, and G. E. W. Bauer, Phys. Rev. Lett. **77**, 1131 (1996).
[14] A. A. Dremin, V. B. Timofeev, D. Birkedal, and J. Hvam, Phys. Status Solidi A **164**, 557 (1997).
[15] H. A. Nickel, G. H. Herold, M. S. Salib, G. Kioseoglou, A. Petrou, B. D. McCombe, and D. Broido, in *Proceedings of EP2DS-12*, Tokyo (1997), to be published in Physica B; J. Kono, M. Y. Su, J. Cerne, M. S. Sherwin, S. J. Allen Jr., T. Inoshita, T. Noda, and H. Sakaki, *ibid.*
[16] J. Cen and K. K. Bajaj, Phys. Rev. B **47**, 1392 (1993).
[17] A. B. Dzyubenko and A. L. Yablonskii, JETP Lett. **64**, 213 (1996).
[18] a) A. B. Dzyubenko and A. L. Yablonskii, in *Proceedings of the 12th International Conference on High Magnetic Fields in the Physics of Semiconductors II*, G. Landwehr and W. Ossau (eds.), World Scientific, Singapore (1997), p. 693; b) A. B. Dzyubenko, JETP Lett. **66**, 617 (1997).
[19] T. G. Andersson, Z. G. Chen, V. D. Kulakovskii, A. Uddin, J. T. Vallin, Phys. Rev. B **37**, 4032 (1988).
[20] L. P. Gor'kov and I. E. Dzyaloshinskiĭ, Zh. Éksp. Teor. Fiz. **53**, 717 (1967) [Sov. Phys. JETP **26**, 449 (1968)].
[21] I. V. Lerner and Yu. E. Lozovik, Zh. Éksp. Teor. Fiz. **78**, 1167 (1980) [Sov. Phys. JETP **51**, 588 (1980)].
[22] I. A. Malkin and V. I. Man'ko, Zh. Éksp. Teor. Fiz. **55**, 1014 (1968) [Sov. Phys. JETP **28**, 527 (1969)].
[23] A. H. MacDonald and D. S. Ritchie, Phys. Rev. B **33**, 8336 (1986).
[24] M. M. Dignam and J. E. Sipe, Phys. Rev. B **43**, 4084 (1991).
[25] G. W. Bryant, Phys. Rev. B **46**, 1893 (1992).
[26] Y. J. Chen, E. S. Koteles, B. S. Elman, and C. A. Armiento, Phys. Rev. B **36**, 4562 (1987).
[27] S. Charbonneau, M. L. Thewalt, E. S. Koteles, and B. Elman, Phys. Rev. B **38**, 6287 (1988).
[28] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics. Nonrelativistic Theory*, Pergamon Press, New York (1976).

Translation was provided by the Russian Editorial office.

Scaling near the upper critical dimensionality in the localization theory

I. M. Suslov*)

P. L. Kapitza Institute for Physical Problems, Russian Academy of Sciences, 117334 Moscow, Russia
(Submitted 21 October 1997)
Zh. Éksp. Teor. Fiz. **113**, 1460–1473 (April 1998)

The phenomenon of upper critical dimensionality d_{c2} has been studied from the viewpoint of the scaling concepts. The Thouless number $g(L)$ is not the only essential variable in scale transformations, because there is the second essential parameter connected with the off-diagonal disorder. The investigation of the resulting two-parameter scaling has revealed two scenarios, and switching from one to another scenario determines the upper critical dimensionality. The first scenario corresponds to the conventional one-parameter scaling and is characterized by the parameter $g(L)$ invariant under scale transformations when the system is at the critical point. In the second scenario, the Thouless number $g(L)$ grows at the critical point as $L^{d-d_{c2}}$, which leads to a violation of the Wegner relation $s=\nu(d-2)$ between the critical exponents for conductivity s and localization radius ν, which takes the form $s=\nu(d_{c2}-2)$. The resulting formulas for $g(L)$ are in agreement with the symmetry theory suggested in a previous publication, I. M. Suslov, Zh. Éksp. Teor. Fiz. **108**, 1686 (1995) [JETP **81**, 925 (1995)]. A more rigorous version of Mott's argument concerning localization due to topological disorder has been proposed. © *1998 American Institute of Physics.* [S1063-7761(98)02404-4]

1. INTRODUCTION

The one-parameter scaling hypothesis[1] has played an important role in development of the contemporary localization theory[2–8] and stimulated creation of the theory of quantum corrections[9] unambiguously supported by an experiment. The criticism of the one-parameter scaling[10–13] in fact refers not to underlying physical ideas, but rather to its justification in the formalism of σ-models.[14–16] The justification problem remains a pressing one, and may require more accurate definitions of the basic notions as well as lead to a restriction of the range of applicability. Here we discuss modifications of scaling concepts that we believe are inevitable in high-dimensional spaces.

Experience with phase-transition theory[17,18] indicates that scaling is applicable only to spaces with dimensionalities d within an interval between the upper and lower critical dimensionalities, d_{c1} and d_{c2}. For $d<d_{c1}$, there is no phase transition, and for $d>d_{c2}$, the mean-field theory is valid. There is no doubt that $d_{c1}=2$ in the localization theory,[1] whereas the issue of the upper critical dimensionality has remained a subject for discussions for many years.[19–25] As concerns the problem of the density of states (determined by the averaged Green's function $\langle G\rangle$), a comprehensive solution was recently found[26–29] by the author of this paper. It was demonstrated that $d_{c2}=4$ and how the condition $d>4$ simplifies the problem. The singularity at $d=4$ was also investigated, and the $(4-\epsilon)$-dimensional theory was developed. As concerns conductivity, which is determined by correlator $\langle G^R G^A\rangle$, the upper critical dimensionality could be, in principle, different for this quantity. The latter statement was made in Ref. 21, but there are some serious errors.[26] In fact, this conjecture is not true: the special role of dimensionality $d=4$ is a fundamental fact manifesting itself in the renormalizability of the theory,[26–29] and the renormalization properties of both density-of-states and conductivity problems are similar. This clearly follows from the fact that the same diagrammatic technique is used in both problems. Nonrenormalizability of the theory at $d>4$ indicates the importance of the Hamiltonian structure on the atomic scale, which is the reason why the scaling invariance is broken. This reasoning is supported by the previously developed ''symmetry theory,''[30] which yields the results that are in agreement with those of a one-parameter scaling only for $d<4$.

The present paper was motivated by two factors. On one hand, the opinion that $d_{c2}=\infty$ has recently become quite popular.[10,24,25] This opinion is not absolutely groundless since the one-parameter scaling theory *gives no indication* of the existence of an upper critical dimensionality. So there are certain drawbacks in the existing physical picture of localization, although it remains unchanged after many years of discussions.

On the other hand, the Wegner relation

$$s=(d-2)\nu \tag{1}$$

between the critical exponents for conductivity (s) to those of localization radius (ν), which derives from the scaling theory, can be obtained under less demanding conditions.[31] Namely, it suffices to postulate the symmetry of correlation length on both sides of the transition and independence of the Thouless number at the critical point of the length scale. These two assumptions are taken for granted, so the mechanism responsible for a violation of the Wegner relation at $d>4$[23,30,32] deserves a consideration on the physical level.

The aim of the reported work was to fill these gaps and investigate the phenomenon of the upper critical dimensionality from the standpoint of the scaling concepts.

2. PROBLEM OF THE SECOND PARAMETER

The scaling theory[1] is based on the Thouless scaling consideration,[4,33] which is similar to the well-known Kadanov scheme in the theory of critical phenomena.[17,18] The disordered system in question, which is described by the Anderson model on a d-dimensional cubic lattice with the coupling integrals J between nearest neighbors and the spread of the energy levels W, is divided into blocks of size L. In the absence of interaction between the blocks, the system has random energy levels with a characteristic spacing $\Delta(L)\sim J(a_0/L)^d$, where a_0 is the lattice constant. If the interaction is ''switched on,'' the matrix elements between the states of the neighboring blocks appear and result in hybridization of ''block'' functions. The hybridization is the strongest between the states with close energies, and on a qualitative level we can consider only such states. By selecting in each block a level closest to a given energy E, we obtain the effective Anderson model with the spread of levels $W(L)\sim\Delta(L)$ and coupling integrals $J(L)$ determined by the corresponding matrix elements. The effective Anderson model provides a reduced description of the system on scales larger than L, and its properties are controlled by the Thouless number

$$g(L)=\frac{J(L)}{W(L)}, \tag{2}$$

related to the conductance $G(L)$ of a block with dimension L:

$$g(L)\sim\frac{\hbar}{e^2}G(L),\quad G(L)=\sigma(L)L^{d-2}. \tag{3}$$

Repeating the Thouless consideration for the effective Anderson model, we obtain an algorithm for calculating $g(bL)$ with integer b, given $g(L)$:

$$g(bL)=F(b,g(L)). \tag{4}$$

Abrahams *et al.*[1] considered the limit $b\to1$ for this equation, when it can be rewritten in the form suggested by Gell-Mann and Low:

$$\frac{d\ln g}{d\ln L}=\beta(g). \tag{5}$$

The transition point g_c is determined by a condition $\beta(g_c)=0$, and the conductivity $\sigma=\lim_{L\to\infty}\sigma(L)$ and localization radius ξ behave in the vicinity of the transition as

$$\sigma\propto(g_0-g_c)^s,\quad \xi\propto(g_c-g_0)^{-\nu}, \tag{6}$$

where g_0 is the value of $g(L)$ at $L\sim a_0$, $1/\nu=g_c\beta'(g_c)$, and the critical exponent s is determined by Eq. (1).

The theory developed by Abrahams *et al.*[1] corresponds to the simplest scenario of one-parameter scaling. In principle, one can imagine alternative situations. For example, if *two parameters*, $g(L)$ and $h(L)$, are important, we have, by analogy with Eq. (4),

$$g(bL)=F(b,g(L),h(L)),\quad h(bL)=G(b,g(L),h(L)), \tag{7}$$

which in the limit $b\to1$ yields

$$\frac{d\ln g}{d\ln L}=\beta(g,h), \tag{8a}$$

$$\frac{d\ln h}{d\ln L}=\gamma(g,h), \tag{8b}$$

and the results are determined by the properties of two functions, $\beta(g,h)$ and $\gamma(g,h)$.

The arguments presented in Ref. 1 in favor of *one parameter* $g(L)$ scaling in spite of their peculiarity[1)] were well grounded. If the basic physical concept proposed in that paper is correct, the parameter $g(L)$ changes over a distance of the order of the correlation radius ξ, which can be arbitrarily large near the transition point, and Eq. (8b) can be analyzed at a constant g. If parameter $h(L)$ varies between the finite limits and is a monotonic function, it should on a certain scale $L_0\ll\xi$ tend to a limiting value $h_\infty(g)$, and after substituting this value into Eq. (8a) we return to a one-parameter scaling. An oscillating behavior of parameter $h(L)$ would only indicate its inadequate definition, since averaging out the oscillations[34] would lead to an equation system like (8) with a smoothed parameter $\overline{h}(L)$, which varies monotonically. The parameter $h(L)$ can only be important if it tends to zero or infinity, but then can be detected on the level of order-of-magnitude estimates, and it would have had a clear physical sense. The entire scientific community has failed to suggest such a parameter throughout the period starting with the year 1979.

There are two candidates to the role of the second parameter which appear as a matter of course, but are rejected after a closer scrutiny.

a) While the Thouless scheme is constructed without approximations, the effective Anderson model contains a large number $n(L)$ of levels at each lattice site, which increases with L and can be considered as the second parameter. But hybridization of states in neighboring blocks with energies E and E' is determined by the parameter $J(L)/|E-E'|$ and is inessential for $|E-E'|\gg J(L)$. Therefore one can take into account only $n(L)\sim J(L)/\Delta(L)$ levels around energy E, and the parameter $n(L)$ does not generate a new scale since it is of the same order as the Thouless number $g(L)$. Nonetheless, this modification of the Thouless scheme reveals new opportunities and will be considered in future work.

b) The overlap integrals in the Thouless construction are random values, and the ratio $\varphi(L)=\delta J(L)/J(L)$ between their fluctuation $\delta J(L)$ and their typical value $J(L)$ can be treated as the second parameter. But fluctuations can be neglected if $\delta J(L)\ll J(L)$, and the opposite case $\delta J(L)\gg J(L)$ is impossible since the extreme limit of off-diagonal disorder corresponds to a symmetric distribution of coupling integrals around zero when $\delta J(L)\sim J(L)$. Hence, the parameter $\varphi(L)$ can only play some role when it is of the order of unity and does not generate a new scale. Nonetheless, the off-diagonal disorder is significant, although a more appropriate definition of the corresponding parameter is required.

Estimates based on the optimal fluctuation technique[35,36] show that a typical wave function of localized states has a behavior

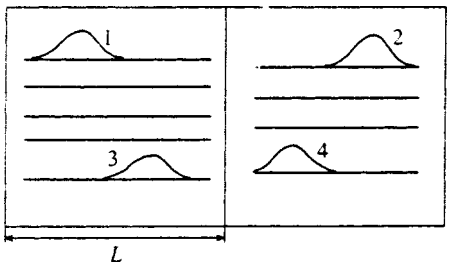


FIG. 1. At large ζ in Eq. (10), the block eigenfunctions are highly localized on scales $L<\xi$, which leads to strong off-diagonal disorder. For example, the overlap between the states 1 and 2 is substantially smaller than between 3 and 4.

$$|\Psi(r)|\propto\begin{cases} r^{-\zeta}, & r\ll\xi \\ \exp(-r/\xi), & r\gg\xi, \end{cases} \tag{9}$$

where $\zeta=d-2$ increases with the space dimensionality. This result is valid outside the close neighborhood of the transition point, i.e., in the region similar to that where the Landau theory[7] can be used, but such results have a tendency to become rigorous in spaces with a high dimensionality. In the critical region, a similar result is associated with investigations of multifractal properties of the wave functions[37]

$$\langle|\Psi(r)|^2|\Psi(r')|^2\rangle\propto|r-r'|^{-\eta}, \quad |r-r'|\ll\xi, \tag{10}$$

where $\eta\sim\epsilon$ for $d=2+\epsilon$ and $\eta\sim1$ for $d=3$, i.e., it also increases with the space dimensionality. Therefore, let us assume that Eq. (10) holds in the critical region and ζ increases without bound as d increases, and let us consider whether this property can lead to a catastrophe. A large value of ζ means that the block wave functions in the Thouless scheme are strongly localized on a scale smaller than ξ (Fig. 1), which leads to strong off-diagonal disorder. For example, the overlap integral coupling states 1 and 2 is much smaller than that coupling states 3 and 4. The anticipated catastrophe is a localization due to the pure off-diagonal disorder, which can occur even if $W(L)=0$, i.e., when the spread of energy levels is neglected. The Thouless number $g(L)$ in this case is infinite and cannot play any role, and the hybridization of block states is controlled by a different parameter related to off-diagonal disorder.

3. LOCALIZATION IN THE CASE OF OFF-DIAGONAL DISORDER

A possibility of localization due to off-diagonal disorder was discussed in connection with the problem of formation of an impurity band in a semiconductor, which in fact stimulated the creation of the localization theory.[38] An isolated impurity in a semiconductor can generate a state with energy E_0 within the band gap. When the concentration of such impurities is finite, they form an impurity band, which is described in the site representation by the Anderson model with off-diagonal disorder (sometimes this is termed the Lifshitz model[5]):

$$\sum_{n'} J_{nn'}\Psi_{n'}+E_0\Psi_n=E\Psi_n. \tag{11}$$

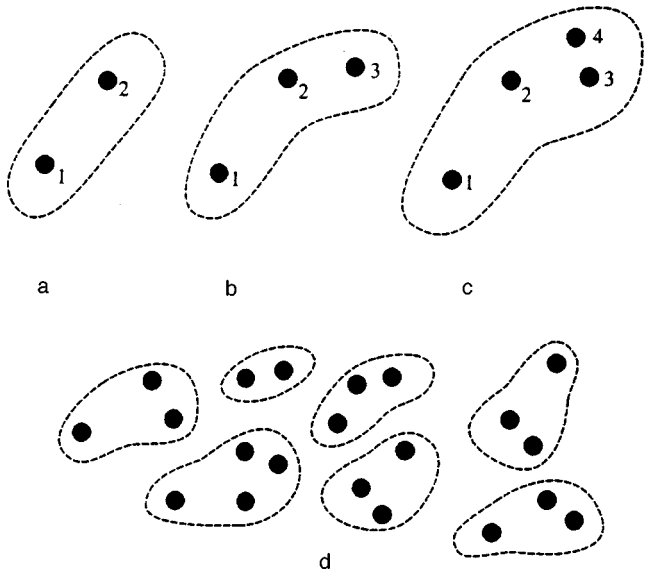


FIG. 2. Decomposition of an arbitrary configuration of impurities into clusters.

If the coupling integral drops exponentially,

$$J_{nn'}\propto\exp\{-\kappa|\mathbf{r}_n-\mathbf{r}_{n'}|\}, \tag{12}$$

where $\mathbf{r}_n$ is the coordinate of the nth impurity, the impurity band is completely localized in the limit of low concentration. Intuitive arguments in favor of this conjecture were suggested by Mott[3] on the basis of Lifshitz's classification of states.[35,36] Here we present a refined version of Mott's argument with the aim of attracting attention to physically significant aspects ignored by both Mott and Lifshitz.

The density of states $\nu(E)$ of the impurity band is a continuous function of energy and is formed by levels of which the overwhelming majority have energies different from that of an isolated impurity E_0. In order to obtain such levels, one should take into account the interaction between an arbitrary impurity atom 1 and its environment, no matter how weak it is. According to Lifshitz, the main factor is ''collisions'' between impurity atoms, i.e., random encounters among the latter. If the unit distance is the average distance between impurities, the limit of zero concentration corresponds to $\kappa\to\infty$ in Eq. (12). Since the overlap integral decays exponentially with the distance, only interaction between the nearest neighbors should be taken into account. Nevertheless, the analysis cannot be limited to pairwise ''collisions.''

Indeed, suppose that the nearest neighbor of atom 1 is atom 2. If the nearest neighbor of atom 2 is atom 1, the 1–2 pair can be treated in isolation from its environment (Fig. 2a). If the nearest neighbor of atom 2 is atom 3, we must consider the 1–2–3 cluster (Fig. 2b): first the hybridization of states of atoms 2 and 3 should be taken into account, then their interaction with atom 1. If the nearest neighbor of atom 3 is not atom 2 but atom 4, we must consider the 1–2–3–4 cluster (Fig. 2c), etc. If this construction process starts with atom 1 and ends with atom i, we consider by definition that atom 1 belongs to ith cluster. It is evident that atoms 2, 3, ... specified in this process belong to the same ith cluster.

Taking each impurity atom in turn as atom 1, we obtain a decomposition of an arbitrary configuration into clusters (Fig. 2d). The decomposition is unique since each atom in this scheme belongs to a certain cluster, and no atom can belong to two clusters at once (neglecting an infinitesimal probability of detecting an exact equality between two interatomic distances). Formally we should take into account arbitrarily large clusters, but in fact all clusters contain with an overwhelming probability a number of atoms on the order of unity (the existence of an infinite cluster would mean concentration of an infinite number of impurities in a finite volume).

Let us introduce parameter R_1, which is the characteristic interatomic distance inside a cluster, and parameter R_2, which is the characteristic separation between clusters. A thorough investigation is needed to give rigorous definitions of these parameters, but for any reasonable definition we have

$$R_1 < R_2, \tag{13}$$

since clusters are formed from the nearest atoms.

By neglecting interaction between clusters and diagonalizing Hamiltonians of isolated clusters, we obtain the zero approximation for the density of states $\nu(E)$ of the impurity band, whose width is determined by the parameter $\exp(-\kappa R_1)$. This approximation is asymptotically exact in the limit of zero concentration, since the nearest neighbor of each atom is in the same cluster, and the shift of its level with respect to E_0 is calculated correctly in the lowest approximation.

Regarding each cluster as a site of a new lattice and taking into account interaction between clusters, we obtain the effective Anderson model with the spread of levels $W \propto \exp(-\kappa R_1)$ and overlap integrals $J \propto \exp(-\kappa R_2)$. By virtue of Eq. (13), we have $J/W \to 0$ as $\kappa \to \infty$, and in the zero-concentration limit, all states are localized inside the clusters. The latter clarifies the physical sense of these clusters.

Thus, we have proved the basic feasibility of localization of all states due to the pure off-diagonal disorder. Note that the pattern of hybridization between the eigenstates of separate blocks (Fig. 1), neglecting the spread of energy levels and in the limit $\zeta \to \infty$, is similar to the case of topological disorder in a system of impurities with exponential overlap.

4. TWO-PARAMETER SCALING

In the presence of off-diagonal disorder, a disordered system can be characterized by two parameters:

$$g(L) = \frac{J(L)}{W(L)}, \quad \varphi(L) = \frac{\delta J(L)}{J(L)}, \tag{14}$$

the latter having as an upper bound a certain value $\varphi_{\max}$ (Sec. 2). A phase diagram in coordinates (g,φ) is shown in Fig. 3. At $\varphi = 0$, the boundary between localized and delocalized states is located at $g \sim 1$. An increase in φ leads to greater disorder in the system, and the boundary AB between the two phases displaces to higher g and tends to infinity at some φ_c (a curve like AB' precludes localization due to the pure off-diagonal disorder, when the Thouless number is infinite).

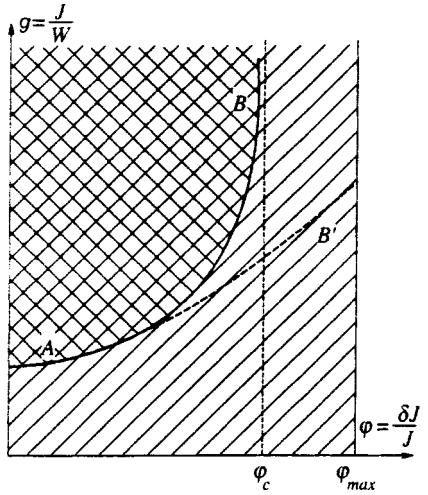


FIG. 3. Phase diagram in coordinates (g,φ). The hatched area corresponds to localized states, the cross-hatched area to delocalized states.

The existence of the critical point φ_c solves the problem of the second parameter in the renormalization group: the new nontrivial scale is associated not with φ, but with $\varphi - \varphi_c$.

If the parameters g and φ uniquely determine the state of a disordered system, then in the course of the Thouless scale transformation one point of plane (g,φ) turns into another point of this plane. If the system is at a critical point, it can move only along the critical AB surface, which is the locus of such points.

In order to return to the conventional scheme of one-parameter scaling, we should postulate, in accordance with the conventional concepts of the theory of critical phenomena (Ref. 17, Ch. 6), the existence of a fixed point F (Fig. 4a), which is stable for states on the critical surface but unstable for states off the critical surface. In the theory of differential equations,[39] such a property is associated with a saddle point characterized by two asymptotes, AB and CD, and hyperbolic trajectories in the vicinity of this point (Fig. 4a). Changes in the Thouless number $g(L)$ with scale L for this case are shown in Fig. 5a. It has a constant value g_c at point F (curve *1*), relaxes to g_c at a finite scale L_0 for the points on the critical surface different from F (curves *2* and *3*), approaches g_c at the scale L_0 and departs from this value at the scale ξ for the points close to the critical surface (curves *4* and *5*). Roughly speaking, evolution in the (g,φ) plane consists of two stages, namely the fast relaxation to the curve CD and slow motion along this curve. At scales $L \gg L_0$ the (g,φ) plane is in fact compressed to the line CD, and positions on this line are determined by the Thouless number.[2)] Thus, we have returned to the conventional scheme, and we assume it to be valid for low dimensions.

Suppose that there is no stationary point on the critical surface at large d. Then a system at a critical point moves upward along curve AB as L increases (Fig. 4b). The downward motion is impossible because this means that off-diagonal disorder disappears asymptotically at large L and contradicts the physical arguments of Sec. 2. The Thouless number $g(L)$ increases with L at the transition point (curve *1* in Fig. 5b), in the metallic phase it increases faster,[1] $g(L)$

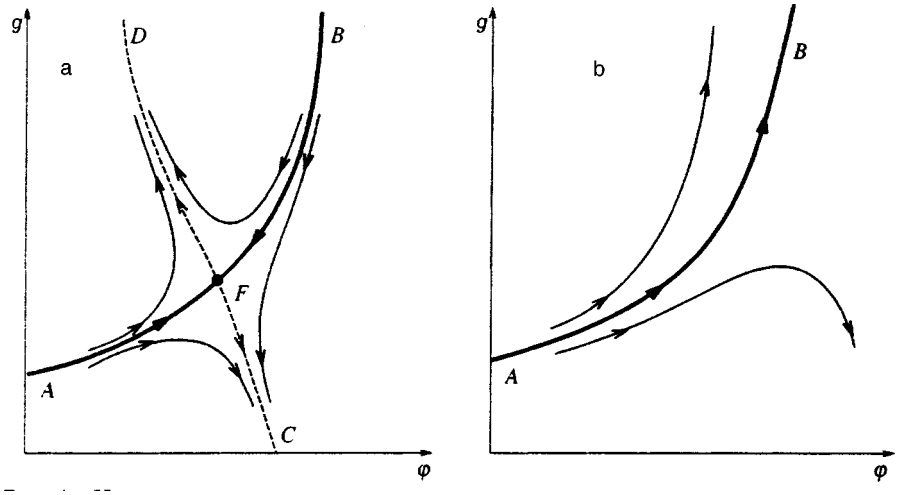


FIG. 4. Flow diagram for Thouless' scale transformations (a) in the presence of a stationary point F on the critical surface AB and (b) in the absence of such a point.

$\sim\sigma L^{d-2}$ (curve *2*), and in the localized phase the curve exhibits reentrant behavior (curve *3*).

At first sight, such reentrant behavior is absurd from the physical standpoint. This means[1] that the degree of hybridization between block states increases at smaller L, but then drops for an unclear reason. In reality, this is not so, since the hybridization is not determined entirely by the Thouless number, but is also a function of $\varphi(L)$. At the transition point, the effective disorder (hence the hybridization degree) remains at the same level but is transferred from the diagonal type to off-diagonal one. In the localized phase, the effective disorder increases monotonically, but in the first stage the Thouless number grows, and the diagonal disorder characterized by this parameter decreases owing to transformation to the off-diagonal disorder. Only when $L>\xi$ and the total disorder has increased considerably does diagonal disorder also begin to grow.

As the space dimensionality increases, the first scenario (Fig. 4a) should gradually transfer to the second one (Fig. 4b), so the stationary point should move upwards along the curve *AB* and go to infinity at a certain dimensionality d_{c2}. We identify this value with the upper critical dimensionality. The aim of subsequent analysis is to develop a phenomenological theory of this bifurcation.

The phenomenological description is possible because the functions $\beta(g,h)$ and $\gamma(g,h)$ in the two-parameter scaling equations (8) admit regular expansions. By virtue of Eq. (7), they describe a relation between two finite systems, whereas all singularities emerge in the thermodynamic limit.[17] This argument assumes, however, an adequate choice of scaling variables, which do not have their built-in singularities. In this sense, the variable φ is not appropriate because it has a singular point φ_c. Therefore we introduce a new variable $h=F(g,\varphi)$ such that in the (g,h) plane the curves of Fig. 4a take the form shown in Fig. 6, i.e., curve *AB* has an asymptote $g\sim h$ as $g,h\to\infty$ and curve *CD* becomes a vertical line. The first condition is adopted so that the critical surface, which is associated with no singularities, should have regular projections on both coordinate axes, and the second is assumed to simplify the equations (see below).

In investigating the bifurcation, it is sufficient to analyze Eq. (8) in the region of large g and h, where it can be transformed to

$$\frac{d\ln g}{d\ln L}=(d-2)+\frac{Ah}{g}+\frac{Bh^2}{g^2}+\frac{Ch^3}{g^3}+\ldots\equiv(d-2)+\tilde{\beta}\left(\frac{g}{h}\right), \tag{15a}$$

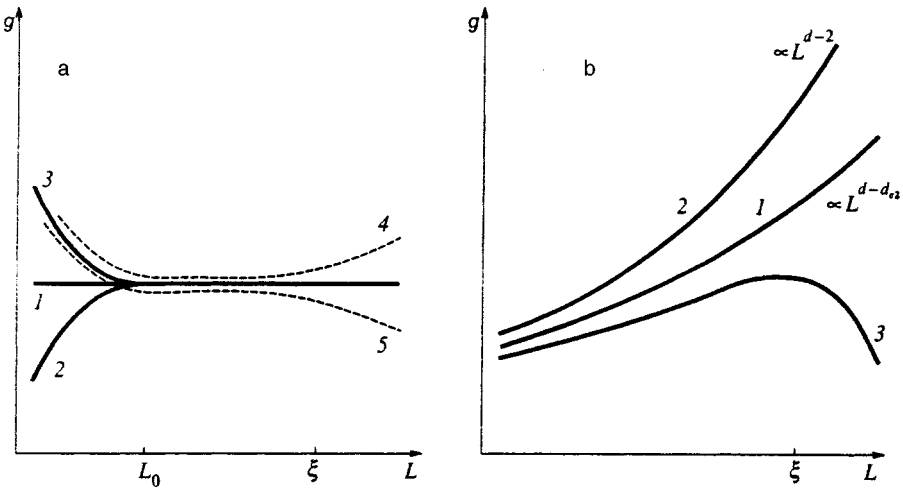


FIG. 5. Evolution of the Thouless parameter in scenarios illustrated by Figs. 4a and 4b.

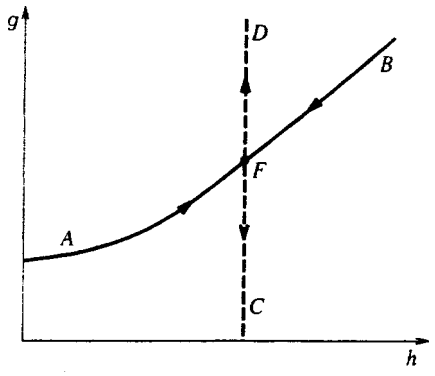


FIG. 6. Diagram of Fig. 4a after the variable change $h=F(g,\varphi)$.

$$\frac{d\ \ln h}{d\ \ln L}=\mu+\frac{b}{h}, \tag{15b}$$

where parameter μ changes sign at $d=d_{c2}$,

$$\mu=\alpha(d-d_{c2}), \quad d\to d_{c2}, \tag{16}$$

and $\alpha>0$, $b>0$, and $A<0$. Indeed, at $h=$const all conclusions from Ref. 1 apply to function $\beta(g,h)$, i.e., it has asymptotes $d-2$ and $\ln g$ for large and small g, and at $d>2$ has a root g_c, which is a function of h in this specific case. By expanding $\beta(g,h)$ in powers of $1/g$,

$$\beta(g,h)=(d-2)+\frac{A_1(h)}{g}+\frac{A_2(h)}{g^2}+\ \ldots, \tag{17}$$

we find that the expansion $A_n(h)$ in powers of $1/h$ should begin with h^n in order to yield a root $g_c\sim h$ (Fig. 6). By retaining the leading terms of the expansion in h, we obtain Eq. (15a).

As follows from the foregoing, at $d>d_{c2}$ the function $\gamma(g,h)$ should lead to unbounded growth in h, which, however, should not be faster than that in g, so that the root $g_c\sim h$ should retain its physical sense. Given that $g(L)$ increases no faster than L^{d-2},[1] we have at large h the condition $0<\gamma(g,h)<d-2$, which indicates that the expansion of $\gamma(g,h)$ in powers of $1/g$ and $1/h$ begins with a zero-order term:

$$\gamma(g,h)=\mu+\frac{a}{g}+\frac{b}{h}+\ \ldots \tag{18}$$

If the variables are defined so that curve CD is a vertical line, the coordinate h_c of the stationary point is independent of g and the coefficient a in Eq. (18) is zero. The stationary point should be stable for $d<d_{c2}$, and absent for $d>d_{c2}$, which means that b is positive and μ changes sign at $d=d_{c2}$, as can be seen in Eqs. (15) and (16).

Equation system (15) is easy to analyze. For $d<d_{c2}$, Eq. (15b) has a stationary point $h_c=b/|\mu|$, and the variable change $g\to gh_c$ in Eq. (15a) returns us to the one-parameter scaling with the critical exponents given by equations

$$1/\nu=g_c\tilde{\beta}'(g_c), \quad s=\nu(d-2), \quad (d-2)+\tilde{\beta}(g_c)=0. \tag{19}$$

For $d>d_{c2}$ and large h, we have $h(L)\propto L^{\mu}$, and after the change $g\to gL^{\mu}$, Eq. (15b) is reduced to a one-parameter form, but with $d-2-\mu$ instead of $d-2$. For $L\lesssim\xi$, the Thouless number follows the law

$$g(L)=g_c\left(\frac{L}{a_0}\right)^{\mu}+(g_0-g_c)\left(\frac{L}{a_0}\right)^{\mu+1/\nu}, \tag{20}$$

and the critical exponents are determined by the equations

$$1/\nu=g_c\tilde{\beta}'(g_c), \tag{21a}$$

$$s=\nu(d-2-\mu), \tag{21b}$$

$$(d-2-\mu)+\tilde{\beta}(g_c)=0. \tag{21c}$$

The localization radius is defined as the distance at which the parameter $g(L)$ begins to drop for $g_0<g_c$ (i.e., in the localized phase) and the exponent s is determined by matching the function defined by Eq. (20) and $g(L)\sim\sigma L^{d-2}$ at $L\sim\xi$. At the transition point, the Thouless number increases according to the law

$$g(L)\propto L^{\mu}, \tag{22}$$

which is the reason why the Wegner relation fails (see Eq. (21b)). The comparison between Eqs. (19) and (21) demonstrates that critical exponents as functions of d have cusps at $d=d_{c_2}$.

Usually, one feature of the upper critical dimensionality is that the critical exponents are independent of d above d_{c2}. As follows from Eq. (21b), this is possible if $\mu=d+$const, which yields in combination with Eq. (16)

$$\mu=d-d_{c2}. \tag{23}$$

Given this relation, we obtain the Thouless number as a function of the length scale for $L\lesssim\xi$:

$$g(L)=g_c+(g_0-g_c)(L/a_0)^{1/\nu}, \quad d<d_{c2}, \tag{24a}$$

$$g(L)=g_c(L/a_0)^{d-d_{c2}}+(g_0-g_c)(L/a_0)^{d-d_{c2}+1/\nu}, \quad d>d_{c2}. \tag{24b}$$

Equation (24b) is the main result of our phenomenological approach. Equation (24a) is a well-known consequence of one-parameter scaling, but its range of applicability is limited.

5. COMPARISON TO THE SYMMETRY THEORY

The symmetry theory[30] yields the same values of critical exponents as the Vollhardt and Wölfle self-consistent theory[32]:

$$\nu=1/(d-2), \quad s=1 \ \text{ for } \ 2<d<4, \qquad \nu=1/2, \quad s=1 \ \text{ for } \ d>4. \tag{25}$$

For $d<4$ they are compatible with the one-parameter scaling because the Wegner relation $s=\nu(d-2)$ holds. Its failure at $d>4$ means that $d_{c2}=4$.

In order to compare the results given by Eq. (24) to the symmetry theory, let us derive from the latter[30] the diffusion coefficient D_L for a finite block of size L. It is calculated using the diffusion coefficient $D(\omega,q)$ for an infinite system using the formula[3)]

$$D_L \sim D\left(i\frac{D_L}{L^2}, L^{-1} \right). \tag{26}$$

It was shown in Ref. 30 that

$$D(\omega,q) = D_0(\omega) q^0, \quad q \ll a_0^{-1}, \tag{27}$$

and $D_0(\omega)$ is given by the equation

$$D_0(\omega) = A\tau + B\left(-\frac{i\omega}{D_0(\omega)} \right)^{1/2\nu}, \tag{28}$$

where τ is the distance to the transition point. Given that $g(L) \propto D_L L^{d-2}$ and parameter τ is proportional to $g_0 - g_c$, we can easily derive from Eqs. (26)–(28)

$$g(L) = g_c (L/a_0)^{d-2-1/\nu} + (g_0 - g_c)(L/a_0)^{d-2}. \tag{29}$$

This result is similar to Eq. (24) but not identical in the general case. The results expressed by Eqs. (24) and (29) are identical only for specific values of critical exponents given by Eq. (25):

$$\begin{aligned} &g(L) = g_c + (g_0 - g_c)(L/a_0)^{d-2}, \quad d<4, \\ &g(L) = g_c (L/a_0)^{d-4} + (g_0 - g_c)(L/a_0)^{d-2}, \quad d>4. \end{aligned} \tag{30}$$

Thus, the phenomenological model developed in the reported work is in full agreement with the symmetry theory.[30] This correspondence between the two theories is far from trivial because the symmetry theory is based on different principles and does not use in any way the scaling concepts.

This work was stimulated by discussions with V. E. Kravtsov, A. D. Mirlin, and M. V. Feigel'man, who are gratefully acknowledged. I am also grateful to participants of seminars at the Kapitza Institute for Physical Problems and Lebedev Institute of Physics for interesting discussions on the results of this research.

This work was supported by INTAS (application No. 580) and Russian Fund for Fundamental Research (Project 96-02-19527).

*)E-mail: suslov@kapitza.ras.ru

1)"We cannot see how any statistical feature of the energy levels other than this ... ratio can be relevant" (Ref. 1).

2)The assumption that only two parameters, g and φ, are essential means in reality that all other parameters relax rapidly to a surface which can be mapped one-to-one onto the (g,φ) plane.

3)Equation (26) fails in the localized phase for $L \gtrsim \xi$ due to a nonlocal response.[32]

[1] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishman, Phys. Rev. Lett. **42**, 673 (1979).
[2] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).
[3] N. F. Mott and E. A. Davis, *Electron Processes in Non-Crystalline Materials*, Clarendon Press, Oxford (1979), p. 31.
[4] D. J. Thouless, Phys. Rep. **13**, 92 (1974).
[5] A. L. Éfros, Usp. Fiz. Nauk **126**, 41 (1978) [Sov. Phys. Usp. **21**, 746 (1978)].
[6] M. V. Sadovskiĭ, Usp. Fiz. Nauk **133**, 223 (1981) [Sov. Phys. Usp. **24**, 96 (1981)].
[7] M. V. Sadovskii, Sov. Sci. Rev. A. Phys. **7**, 1 (1986).
[8] D. Belitz and T. R. Kirkpatrick, Rev. Mod. Phys. **66**, 261 (1994).
[9] B. L. Al'tshuler, A. G. Aronov, D. E. Khmel'nitskii, and A. I. Larkin, in *Quantum Theory of Solids*, I. M. Lifshitz (ed.), Mir Publishers, Moscow (1982).
[10] K. B. Efetov, Zh. Éksp. Teor. Fiz. **88**, 1032 (1985) [Sov. Phys. JETP **61**, 606 (1985)].
[11] K. B. Efetov, Zh. Éksp. Teor. Fiz. **92**, 638 (1987) [Sov. Phys. JETP **65**, 360 (1987)].
[12] B. L. Al'tshuler, V. E. Kravtsov, and I. V. Lerner, Zh. Éksp. Teor. Fiz. **91**, 2276 (1986) [Sov. Phys. JETP **64**, 1352 (1986)].
[13] V. E. Kravtsov, I. V. Lerner, and V. I. Yudson, Zh. Éksp. Teor. Fiz. **94**, 255 (1988) [Sov. Phys. JETP **67**, 147 (1988)].
[14] L. Schäfer and F. Wegner, Z. Phys. B **38**, 113 (1980).
[15] K. B. Efetov, A. I. Larkin, and D. E. Khmel'nitskiĭ, Zh. Éksp. Teor. Fiz. **79**, 1120 (1980) [Sov. Phys. JETP **52**, 568 (1980)].
[16] K. B. Efetov, Adv. Phys. **32**, 53 (1983).
[17] S. Ma, *Modern Theory of Critical Phenomena*, Benjamin/Cummings, Reading (USA) (1976).
[18] A. Z. Patashinskiĭ and V. L. Pokrovskiĭ, *Fluctuation Theory of Phase Transitions* [in Russian], Nauka, Moscow (1982).
[19] D. J. Thouless, J. Phys. C **9**, L603 (1976).
[20] T. Lukes, J. Phys. C **12**, L797 (1979).
[21] A. B. Harris and T. C. Lubensky, Phys. Rev. B **23**, 2640 (1981).
[22] J. P. Straley, Phys. Rev. B **28**, 5393 (1983).
[23] H. Kunz and R. Souillard, J. de Phys. Lett. **44**, L503 (1983).
[24] A. D. Mirlin and I. V. Fyodorov, Phys. Rev. Lett. **72**, 526 (1994).
[25] M. Schreiber and H. Grussbach, Phys. Rev. Lett. **76**, 1687 (1996).
[26] I. M. Suslov, Zh. Éksp. Teor. Fiz. **102**, 1951 (1992) [Sov. Phys. JETP **75**, 1049 (1992)].
[27] I. M. Suslov, Zh. Éksp. Teor. Fiz. **106**, 560 (1994) [JETP **79**, 307 (1994)].
[28] I. M. Suslov, Zh. Éksp. Teor. Fiz. **111**, 220 (1997) [JETP **84**, 120 (1997)].
[29] I. M. Suslov, Zh. Éksp. Teor. Fiz. **111**, 1896 (1997) [JETP **84**, 1036 (1997)].
[30] I. M. Suslov, Zh. Éksp. Teor. Fiz. **108**, 1686 (1995) [JETP **81**, 925 (1995)].
[31] I. M. Suslov, JETP Lett. **43**, 704 (1986).
[32] D. Vollhardt and P. Wölfle, Phys. Rev. B **22**, 4666 (1980); Phys. Rev. Lett. **48**, 699 (1982).
[33] D. C. Licciardello and D. J. Thouless, Phys. Rev. Lett. **35**, 1475 (1975).
[34] N. N. Moiseev, *Asymptotic Methods in Nonlinear Mechanics* [in Russian], Nauka, Moscow (1969).
[35] I. M. Lifshitz, Usp. Fiz. Nauk **83**, 617 (1964) [Sov. Phys. Usp. **7**, 549 (1965)].
[36] I. M. Lifshitz, S. A. Gredeskul, and L. A. Pastur, *Introduction to Theory of Disordered Systems* [in Russian], Nauka, Moscow (1982).
[37] T. Brandes, B. Huckestein, and L. Schweitzer, Ann. Phys. **5**, 633 (1996).
[38] P. W. Anderson, Usp. Fiz. Nauk **127**, 19 (1979).
[39] L. É. Él'sgol'ts, *Differential Equations and Calculus of Variations* [in Russian], Nauka, Moscow (1969).

Translation was provided by the Russian Editorial office.

Superconducting transition temperature in mercury HTSC-cuprates under hydrostatic pressure

E. S. Itskevich

Institute of High Pressure Physics, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia
(Submitted 20 March 1997)
Zh. Éksp. Teor. Fiz. **113**, 1474–1483 (April 1998)

A study is made of the properties of the homologous series of mercury HTSC-cuprates $HgBa_2Ca_{n-1}Cu_nO_{2n+2+\delta}$ with $n=1-8$. Experiments are conducted under pressure for samples with $n=1-5$. The Hg-1223 and Hg-1234 phases were synthesized using a controlled high pressure chamber. The oxygen content of an initial mixture corresponding to the Hg-1234 phase was varied by changing the composition of the initial BaO/BaO_2 oxides. The dependence of the superconducting transition temperature T_c on the lattice constant a (and, therefore, on the oxygen content) and of $T_c^{\max}$ and $dT_c^{\max}/dp$ on n are convex upward up to $n=4, 5$. The maximum values always correspond to the Hg-1223 phase. Experimental $T_c^{\max}(n)$ curves for the phases with $n=1-6$ and $dT_c^{\max}/dp$ curves for $n=1-5$ are compared with Anderson's theory (the so-called RVB model). A general analysis of these results indicates that the mercury cuprates have an ideal structure for HTSC. The Hg-1223 phase is the ''champion'' in this ideal structure and the critical temperature corresponding to this phase ($T_c=135$ K) is the highest at atmospheric pressure. © *1998 American Institute of Physics.* [S1063-7761(98)02504-9]

1. INTRODUCTION

The homologous series of layered mercury HTSC-cuprates is described by the general formula $HgBa_2Ca_{n-1}Cu_nO_{2n+2+\delta}$. The layers of CuO_2 determine the superconductivity.

In connection with the appearance of this series of tetragonal cuprates with a high superconducting transition temperature ($T_c=135$ K, $n=3$)[2] and the available theories, we have decided to examine our experimental data on mercury HTSC compounds from a unified standpoint. The superconducting phases with $n=1-5$ have been studied at pressures of up to 2.5 GPa. The 1245 phase was produced with an 85% content of the pure phase, while the 1256 and 1267 phases are mixtures and, because of the strong overlap in the conditions for synthesizing them at high pressures, it was impossible to work with them.

We attempt to understand our experimental results in terms of the resonant valence bond model proposed by Anderson.[1] All the CuO_2 layers in a unit cell are assumed to be equivalent with a uniform charge distribution and only the bonds between neighboring layers of CuO_2 are considered.[1,3]

2. EXPERIMENT

We used a well worked out method for measuring the $T_c(p)$ curves from the jumps in the two magnetic susceptibilities. Chambers were used in which the pressure was created in a liquid medium and recorded at room temperature. Then the chamber was cooled to the required low temperatures. The pressure was measured at room temperature and a suitable calibration was used to determine its value at the given low temperature. Our experiments over many years with these chambers on quantum effects in single crystals have shown that hydrostatic behavior and almost the full pressure are maintained down to liquid-helium temperatures. The temperature was measured using a (Cu+0.15%Fe)/Cu thermocouple.

The measurement samples with $n=1, 2$ were prepared in closed vials in the Department of Inorganic Chemistry at the Chemistry Faculty of Moscow State University and those with $n=3-5$ were prepared in high quasihydrostatic pressure chambers at the Institute of High Pressure Physics of the Russian Academy of Sciences. We used our earlier[4] data on $T_c(p)$, except for the data for $n=1$. The data on $T_c(p)$ for the 1201 phase were taken from corresponding combined data in the literature. New experimental data, including samples synthesized in a controlled quasihydrostatic pressure chamber,[5] were also used. All the $T_c(p)$ curves in the new experiments were obtained for the maximum T_c of the given phase for monophase samples of 1223, 1234, and 1245, whose superconducting characteristics are listed in Table I. We did not use our values for the pressure derivative dT_c/dp of the 1201 phase, which differed from the data published by others. The discrepancy can be explained by the fact that in our chambers, the pressure created at room temperature redistributes the excess oxygen. Similar arguments have been advanced and proven elsewhere[6] for Tl-2201 samples which yielded low pressure derivatives, as we have also found.[4] To a certain extent, this is consistent with the results of Ref. 6, where a reduction in T_c was observed during peroxidation in the Hg-1223 phase and, perhaps, in the Hg-1234 phase. The oxygen content was varied by changing the BaO/BaO_2 ratio in the initial mixture.

The cations of doubly valent mercury in the structure of the layered HTSC-oxides are characterized by a dumbell coordination with oxygen atoms necessarily present only in neighboring layers and only optionally in the mercury layers.

TABLE I. Parameters of the HTSC-samples (experimental data).

	n				
	1	2	3	4	5
$T_c^{\max}$, K	98	127	135	126	111
$dT_c^{\max}/dp$, K/GPa	2.4	3.2	4.6	2.2	4.3
$d\ \ln a/dp$, 1/GPa	–	–	−0.003*	–	–
$d\ \ln c/dp$, 1/GPa	–	–	−0.0067*	–	–
d_{out}^{**}, nm	–	–	0.94	–	–
d_{in}^{**}, nm	–	–	0.32	–	–
$d\ \ln d_{\text{out}}/dp$, 1/GPa	–	–	−0.0098	–	–
$d\ \ln d_{\text{in}}/dp$, 1/GPa	–	–	−0.0054	–	–
$d\ \ln A_{\text{out}}/dp$, 1/GPa	0.0245	–	–	–	–
$d\ \ln A_{\text{in}}/dp$, 1/GPa	–	0.025	–	–	–

Notes: *Data from Ref. 15: $d\ \ln a/dp=-0.003$, $d\ \ln c/dp=-0.006$; **d_{out} is the separation between CuO_2 layers of neighboring unit cells, d_{in} is that within a unit cell.

The weak coupling with neighboring cations makes it possible to change the population δ of the anion position in the mercury layer with oxygen atoms and to vary the concentration of holes in the conduction band. There is no incompatibility of the layers because of this weak Hg–O bond in the layer. The amount of oxygen in the layer (δ), because it changes the lattice parameter a, determines T_c (a convex-upward or dome-shaped $T_c(a)$ curve[7]); see Fig. 1. Because of the distinctive features of the crystalline structure, T_c also depends on the pressure (Fig. 2). The layers of Ba–O and Hg–O form an NaCl-type ionic lattice which alternates with perovskite units. The perovskite Cu–O layers are responsible for the superconductivity. In order for superconductivity to appear, the degree of oxidation must be within the range 2.05–2.25. Carrier delocalization requires that the Cu–O separations in the layer be in the range 0.190–0.197 nm. The Cu–O(3) bond length is 0.27 nm, i.e., the bond is much weaker. For high T_c, all these structural relationships must be satisfied ideally. For example, if the cation dimensions on the two sides of the Cu–O layer are not the same, then the structure is distorted and T_c is reduced.

The NaCl units ensure stability of the structure and serve as reservoirs of holes for the CuO_2 layer. The cation–oxygen distance in any layer is equal to the distance in the neighboring layer. All the parameters a for the tetragonal structure in the Cu–O, Hg–O, and Ba–O layers are the same.

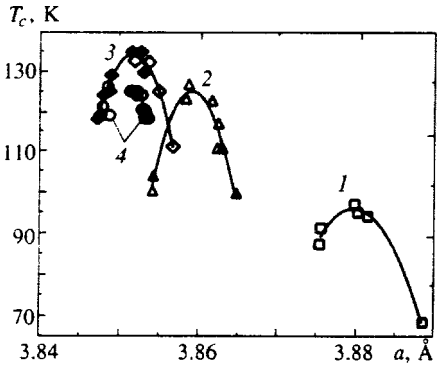


FIG. 1. T_c as a function of the lattice parameter a for mercury HTSC-cuprates: *1*—Hg-1201; *2*—Hg-1212; *3*—Hg-1223; *4*—Hg-1234.

The crystalline structure (tetragonal for all phases) was determined for each sample during synthesis and is given in Refs. 2 and 5. The dependence of its parameters on pressure is given in Table I. The phase analysis was based on x-ray patterns. The 1223 and 1234 phases were synthesized with different peroxidation. Here the hole concentration in the CuO_2 layers was varied by changing the amount x of oxygen in the Hg–O layer, which makes it possible to control T_c;[7] see Fig. 1. The derivatives with respect to pressure were determined from T_c measurements at a pressure of 1 GPa. The results are given in Table 1 and in Figs. 3 and 4.

Note that our data for the pressure derivatives of the various phases (up to the 1234 phase) form the same sort of dome-shaped curves as for $T_c^{\max}(n)$, only much more distinctly, with a maximum in the 1223 phase. The difference in $T_c^{\max}$ for the various n is about 40%, while the pressure derivatives vary by about a factor of two (Figs. 3 and 4). The 1245 phase does not have a dome-shaped curve (see below).

3. DISCUSSION OF RESULTS

We have examined our results in terms of the RVB model.[1] Wheatley *et al.*[1] have stated the basic assumptions of the model and given formulas for T_c of the various phases of layered HTSC-cuprates. This makes a comparison with experiment possible. It is assumed without question that the CuO_2 layers are responsible for the superconductivity.

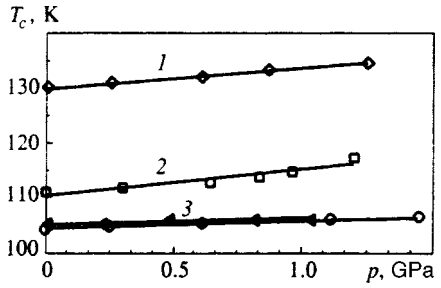


FIG. 2. $T_c(p)$ curves for different phases: *1*—Hg-1223 ($\delta_{\text{opt}}=0.4$); *2*—Hg-1245; *3*—data for underoxidized samples of Hg-1223 ($\delta=0.10$ and $\delta=0.15$).

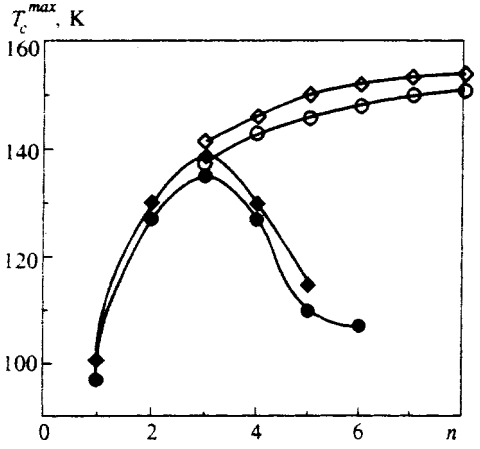


FIG. 3. Experimental and calculated values of $T_c^{\max}$ as a function of the number n of CuO_2 sublayers. The solid symbols are from experiments at the initial pressure (circles) and at a pressure of 1 GPa (diamonds); the hollow symbols are calculations for, respectively, the initial pressure and 1 GPa.

We have calculated the pressure dependence of the main model parameters, expressing them in terms of the interlayer transport integrals $t_{\perp\,\mathrm{in}}$ and $t_{\perp\,\mathrm{out}}$ between CuO_2 layers. The exchange integral within the layers is $J=4t_{\parallel}^2/U$, where U is the interatomic Hubbard energy,[3] which is assumed to be independent of pressure.[8]

Using our RVB model calculations for the first two terms of the homologous series, we have obtained the pressure derivatives of the coupling parameters A_{in} and A_{out} from the experimental data; here A_{in} is the coupling parameter between the CuO_2 layers in a unit cell and A_{out} is that between the layers of neighboring cells. The two parameters are related to the interlayer transport layers $t_{\perp\,\mathrm{in}}$ and $t_{\perp\,\mathrm{out}}$: $A_{\mathrm{in,out}}=t_{\perp\,\mathrm{in,out}}^2/J$.

For a layered structure with one CuO_2 layer, we obtain the following expression:

$$T_c(1)=2A_{\mathrm{out}}R\delta, \tag{1}$$

where R is of order unity and δ is the amount of hole doping, relative to half filling and defined as the number of holes per Cu atom. Superconductivity is observed experimentally for $0.05<\delta<0.2$.[9,10] Starting with the data for the 1201 phase when optimally doped with oxygen ($T_c^{\max}=98$ K) and the data for the pressure derivatives (2.4 K/GPa) at these critical temperatures, we obtain the pressure derivatives for A_{out}. We assume that δ corresponds to optimum doping of the Hg–O layer. The results are shown in Table I.

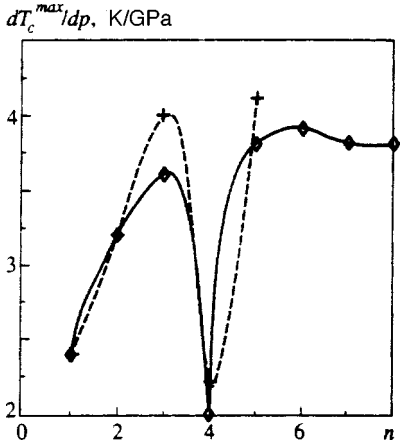


FIG. 4. The experimental (crosses) and calculated (diamonds) pressure derivatives $dT_c^{\max}/dp$ as functions of the number n of sublayers.

For a layered structure with two CuO_2 layers in a unit cell,[1] we have

$$T_c(2)=(A_{\mathrm{out}}+A_{\mathrm{in}})R\delta. \tag{2}$$

Using the data for $A_{\mathrm{out}}(p)$, with Eq. (2) we calculated the pressure derivatives for A_{in}, as well. In the calculation we used a critical temperature $T_c^{\max}=127$ K, which corresponds to optimum doping of the 1212 phase. For the 1223 and 1234 phases we calculated $T_c^{\max}$ and the pressure derivatives as functions of $A_{\mathrm{out}}(p)$ and $A_{\mathrm{in}}(p)$ in accordance with the formulas from Ref. 1 reduced to the form

$$T_c(3)=\frac{1}{2}\,(A_{\mathrm{out}}+\sqrt{A_{\mathrm{out}}^2+8A_{\mathrm{in}}^2})R\delta, \tag{3}$$

and

$$T_c(4)=\frac{1}{2}\,(A_{\mathrm{out}}+A_{\mathrm{in}}+\sqrt{A_{\mathrm{out}}^2+5A_{\mathrm{in}}^2-2A_{\mathrm{out}}A_{\mathrm{in}}})R\delta. \tag{4}$$

The results of the calculations on the effect of pressure on T_c are given in Table I and Figs. 3 and 4. Analogously to the expressions for $n=3$ and 4,[1] we have obtained formulas for the phases with $n=5-8$ and calculated T_c:[11]

$$T_c(5)=146\ \mathrm{K}, \quad T_c(6)=148\ \mathrm{K},$$

$$T_c(7)=150\ \mathrm{K}, \quad T_c(8)=151\ \mathrm{K}.$$

For $n=5-8$ the calculated values of T_c are almost invariant. Although they differ from the much higher calculated values (the deviation is as high as 40 K for large n), the experimental values[12] shown in Fig. 3 generally tend to become independent of n for $n>4$. A comparison of the pressure derivatives with experiment showed good agreement for the 1223, 1234, and 1245 phases (Fig. 4).

For $n\leqslant 5$ the pressure derivatives in Fig. 4 are essentially the same, although there is a sharp transition from a dome-shaped curve to a jump at $n=5$ followed by a plateau (calculations). This kind of behavior in the pressure derivatives is encountered in high pressure studies. The pressure is a good parameter for testing theoretical models.[13]

We have neglected $\delta(p)$. The contribution of charge transport in the mercury cuprates is much smaller than in HTSC materials such as Y-123 and it is the pressure, as such, which mainly affects the magnitude of T_c.[14] Our data confirm this. (See below.) It has been shown[14] that the mercury cuprates are the best candidates for studying the effect of pressure on the properties of HTSC materials. It appears to us that the increase in T_c to 135 K in the 1223 phase and the reduction compared to theory in the 1234 phase and for higher n are related to possible disruptions of the ideal ordering. For example, nonstoichiometric oxygen in the HgO_δ layer increases the oxygen surroundings of part of the Ba atoms (increases the coordination number) and the resulting additional interaction displaces the Ba atoms and disrupts the

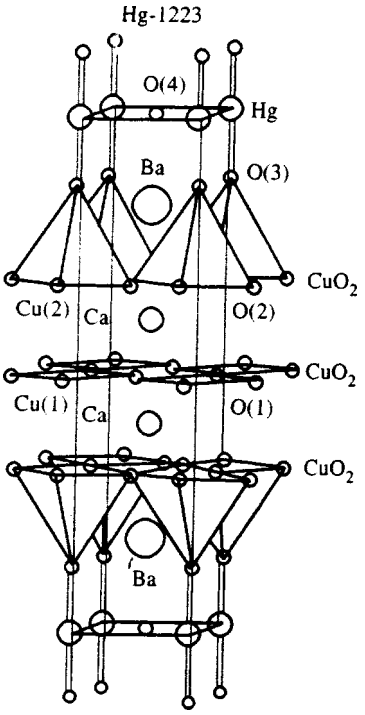


FIG. 5. The $HgBa_2Ca_2Cu_3O_{8+\delta}$ ($n=3$) phase. The atoms in the structure are labelled. O(4) is superstoichiometric oxygen in an amount δ.

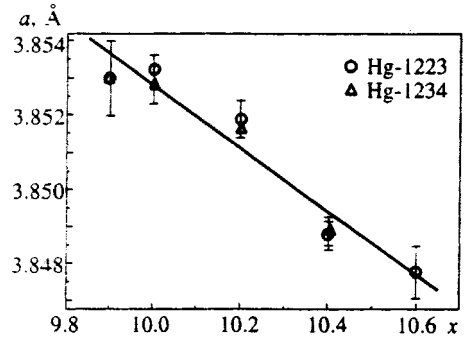


FIG. 6. The lattice parameter a for the 1223 and 1234 phases as a function of the amount x of initial oxygen.

ideal structure, even in the 1223 phase. It is possible that $T_c(p=0)=135$ K is an upper bound for layered HTSC materials, since for $n=3$ the ''idealness'' is the maximum possible. In the following we shall discuss in more detail what we mean by this term.

Measurements of the compressibility at pressures up to 4.5 GPa revealed a roughly twofold anisotropy $(dc/c)\times(da/a)$, which is not a very large amount. Our compressibility data are confirmed elsewhere[15] up to 6.0 GPa.

The sharp difference in the effect of pressure on the spacing in a unit cell of the 1223 phase (Fig. 5) along the c axis observed in Ref. 15 should be mentioned: the distance $d(\mathrm{Cu(2)–O(3)})=0.2727$ nm is greatly reduced at a pressure of 8 GPa (to 0.244 nm, i.e., by 10%). At the same time, the Hg–O(3) spacing varies little: from 0.198 to 0.196 nm at the same pressure. Along the a axis, similarly small and equal compressibilities were observed for the Cu(1)–O(1), Cu(2)–O(2), and Hg–O(4) spacings.

Armstrong *et al.*[15] point out that the relationship of the Cu(2)–O(3) structural parameter to the role of the electronic level of the O(3) ion, which controls the electronic structure of the CuO_2 plane and therefore T_c, has been discussed extensively in the literature. On the other hand, the singularity we have observed in the pressure derivatives at $n=4$ may be related to a change in the compressibility of the structural parameter $d(\mathrm{Cu(2)–O(3)})$ on going from one internal CuO_2 layer to another. A further rise in the number of internal layers causes a monotonic but tiny growth in T_c and makes the pressure derivatives approach saturation.

Note that both underoxygenation and peroxygenation cause a drop in T_c. The increase in δ in an underoxygenated sample is related to a reduction in the lattice constant a[6] owing to oxidation of copper and an enhancement of the copper-oxygen bond in the layer and takes place in parallel with the pressure effect which raises $T_c^{\max}$. On the other hand, increased peroxidation leads to a reduction in T_c and in the parameter a,[7] i.e., the opposite of the effect of pressure on T_c. In the dome-shaped plots of T_c as a function of the amount of oxidation (the parameter a), the position of δ_{opt} (the value of δ corresponding to $T_c^{\max}$) depends only on the amount of CuO_2 layers in the structure for $n\leqslant 3$. Our studies of peroxygenated samples of the 1223 and 1234 phases (see Figs. 1 and 6) showed that as the number of CuO_2 layers increases, the shift in the curves along the a axis becomes smaller.

Based on these data, the $T_c(n)$ curve can be divided into two regions with respect to n. In the first (up to and including $n=3$), the energy level of the apical oxygen O(3) plays a fundamental role. This has been proven many times, e.g., in Ref. 16, where it is shown that the Hall effect confirms the invariance in the hole concentration when pressure is applied. The hydrostatic pressure changes the volume, but does not redistribute the charges. At the same time, the pressure changes the length of the Cu–O(3) bond greatly.[15] Applying pressure to HTSC materials containing O(3) produces higher pressure derivatives than when O(3) is absent. For $n=1,2$ the parameter δ is independent of the pressure, so that the character of the distribution of holes among the CuO_2 layers cannot change. Equations (1) and (2) are valid for $\delta=\mathrm{const}$. The critical temperature $T_c^{\max}$ is entirely determined by the parameters A_{out} and A_{in}. At the same time, the pressure affects $T_c^{\max}$, especially in the 1223 phase. A quantitative analysis (comparison) of the compressibilities at the initial pressure along the a axis and of the $T_c^{\max}(p)$ and $T_c(a)$ curves (Figs. 1 and 6) shows that there are different mechanisms by which T_c and δ are affected. Our experiments (see Fig. 1) imply that T_c decreases as a is reduced. At $\Delta a=0.0007$ nm there is a transition into another phase with a change $\Delta c=0.3$ nm, a change in the number of CuO_2 layers, and another value of $T_c^{\max}$ (Figs. 1, 6, and 7). This follows from the method of oxidation proposed in Ref. 7, in which the amount of initial oxygen during synthesis of the 1234 phase was changed and new phases were added. For small n, the same thing happens, but the parameter a changes more (to 0.002 nm). We may assume that δ retains a certain constancy as the relationship of the phases changes. The main

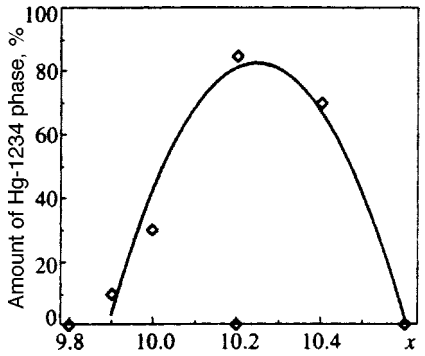


FIG. 7. Percent content of the 1234 phase in synthesized samples as a function of the amount x of initial oxygen (synthesis conditions: $T_s = 900$ °C, $t_s = 2$ h, $p_s = 2$ GPa).

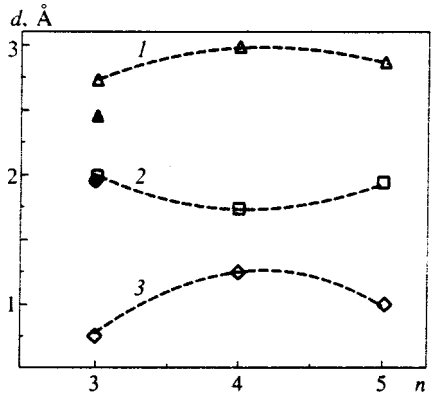


FIG. 8. The separations Cu–O(3) (*1*) and Hg–O(3) (*2*) and their difference Δd (*3*) as functions of the number n of layers. Data obtained at a pressure of 8.5 GPa are shown for the 1223 phase:[15,17] (solid circle) d(Hg–O(3)) and (solid triangle) d(Cu–O(3)).

factor when pressure is applied is the variation in the lattice parameter c, i.e., in the bond lengths along it, as well as the variation in A_{out} and A_{in}. This is related to the results of Refs. 15 and 16 for the 1223 phase and the assumption that the pressure has little effect on δ_{opt}.

The influence of the number of layers on T_c decreases as their number increases. This correlates with our results: neither the critical temperature nor the pressure derivatives increase for large n in either the RVB theory or in the experiments. It is possible that this is related to the fact that the presence of superstoichiometric oxygen does not cause peroxidation of the layers in phases with large n. Most likely, another pattern sets in at $n = 3, 4$: the dome-shaped curves vanish and T_c and the pressure derivatives become constant owing to the small role played by the interior layers of CuO_2. The change in the influence of the pressure, i.e., the constancy of the pressure derivatives, may be related to a large increase in the separation between O(3) and CuO_2 layers with large n. Furthermore, this may still be caused by the fact that the presence of superstoichiometric oxygen does not cause peroxidation of the layers in phases with large n and the overall hole concentration does not change in all the superconducting layers. In the phases with $n < 3$, there is a rapid reduction in the Cu–O separation in the layers as the degree of oxidation increases, as is indicated by the magnitude of the parameter a. This effect becomes smaller as the number of layers increases from $n = 1$ and vanishes for $n = 3$. (See Fig. 1 and Ref. 12.)

The structural balance fails for $n > 3$, since more nonstoichiometric oxygen is required with increasing n.[6] As a result, the Ba–O(Hg) separation decreases, while the Ba–O(Cu) separation increases and the interaction between Ba and O(Cu) atoms decreases. This also leads to a slight reduction in the Cu–O separation in the CuO_2 layer. Thus, increasing the thickness of the perovskite unit and this unbalancing reduce $T_c^{\max}$.

The main reason for the reduction in $T_c^{\max}$ and the change in the pressure derivatives in the phases with $n = 4$ and 5 must be assumed to be a reduction in the average oxidation of the copper owing to the weakened interaction, i.e., to a reduction in the parameters A_{out} and A_{in}, which determine $T_c^{\max}$ for the ideal 1223 phase.

We have already noted that the distribution of holes in the layers does not change. This is even more true of the phases with $n > 4$. Preliminary data indicate that the range of variation of a, i.e., the Cu–O separation in the layer, does not change during peroxidation in phases with $n > 3$. This is all confirmed by an examination of the correlations between the pressure derivatives and separations in the lattice along the c axis (a simplification of the ideas of Ref. 16). At a pressure of 1 GPa, in the 1223 phase the separations d(Hg–O(3)) and d(Cu(2)–O(3)) and the separation between the CuO_2 layers decrease in different ways. According to Ref. 16, the energy level of the O(3) atom is determined by the electronic states in the CuO_2 layer. This yields a maximum pressure derivative for the 1223 phase. The values of T_c are also set by a high oxidation (large δ), which depends weakly on the pressure but strongly on the amount of O(4). However, the effect of the separations d(Hg–O(3)) and d(Cu(2)–O(3)) should not be neglected. See Fig. 8, on which the values of d for the Hg-1223, 1234, and 1245 phases,[15,17] as well as their difference Δd, are indicated. A quantity inversely proportional to Δd controls the interaction. It is very similar to the pressure derivatives of $T_c^{\max}$ for different n. Pressure strongly reduces d(Cu–O(3)) in the 1223 phase, i.e., leads to a larger pressure derivative. Increasing the number of layers reduces the effect of the Hg–O(3) and Cu–O(3) bonds on the interaction. This correlation leads to a reduction in $T_c^{\max}$ and to a rise in the pressure derivatives to values corresponding to the 1223 phase. All the changes in $T_c^{\max}$ are determined by the outer layers of CuO_2. This is also indicated by calculations[11] based on the model of Wheatley *et al.*[1]

Experiments should be conducted regarding the effect of pressure on the dome-shaped $T_c(a)$ curves for phases with $n > 3$. This is impossible for the phases with $n > 5$ owing to difficulties in obtaining the required samples.

4. CONCLUSION

We have measured the $T_c^{\max}(p)$ curves at pressures up to 2.5 GPa for the Hg-1223, 1234, and 1245 phases. For the first four phases (1201, 1212, 1223, and 1234), the plots of the pressure derivatives and $T_c^{\max}$ as functions of the phase

number have a dome shape. All of our measurements yield pressure derivatives of 4 K/GPa for the 1223 phase when T_c is near $T_c^{\max}$ and 2 K/GPa for the 1234 phase. When we go over to the 1245 phase, the pressure derivative again increases to 4 K/GPa.

Using the Anderson model[1] to describe the data has made it possible to confirm the experimental data on $T_c^{\max}$ for the 1223 and 1234 phases and the observed pressure derivatives for the 1223, 1234, and 1245 phases, as well as to demonstrate that $T_c^{\max}$ is independent of the number of layers for $n>4$, in qualitative agreement with experiment.

An examination of the bond lengths along the c axis in the 1223, 1234, and 1245 phases reveals a nonmonotonicity that correlates with the observed variation in the pressure derivative of $T_c^{\max}$ with the number of layers.

The pressure dominates the other interactions. The major changes in $T_c^{\max}$ are determined by the external layers of CuO_2. For layered HTSC materials $T_c^{\max}=135$ K is an upper bound.

The author thanks K. A. Loshkin and E. V. Antipov for help and discussions, as well as the following colleagues at the Institute of High Pressure Physics: T. I. Dyuzheva and L. M. Lityagina for measuring the compressibility, I. G. Kuzemskaya and L. F. Kulikova for synthesizing the samples, A. E. Luppov and I. V. Morozov for the measurements under pressure, and E. V. Zhasinas for help with the calculations.

This work was supported by the State Program on HTSC (Grant No. 96045).

[1] J. W. Wheatley, T. C. Hsu, and P. W. Anderson, Nature (London) **333**, 121 (1988).
[2] S. M. Kazakov, E. S. Itskevich, and L. N. Bogacheva, JETP Lett. **58**, 343 (1993).
[3] N. F. Mott, *Metal-insulator Transition*, Taylor and Francis, London (1974).
[4] A. E. Luppov, E. S. Itskevich, A. G. Gapotchenko, and S. M. Kazakov, Zh. Éksp. Teor. Fiz. **105**, 1785 (1994) [JETP **78**, 966 (1994)].
[5] E. S. Itskevich, T. I. Dyuzheva, I. G. Kuzemskaya, and K. A. Lokshin, JETP Lett. **62**, 512 (1995).
[6] J. S. Schilling and A.-K. Klehe, *Proc. High Press. Conf.*, Warsaw (1995), pp. 673, 679.
[7] K. A. Lokshin, I. G. Kuzemskaya, L. F. Kulikova, E. V. Antipov, and E. S. Itskevich, Physica C **279**, 11 (1997).
[8] D. T. Jover, Academisch Proefschrift (Ph.D. Thesis), Vrije Universiteit, Amsterdam (1996).
[9] M. S. Shafeer and T. Penney, Eur. J. Solid State Inorg. Chem. **27**, 191 (1990).
[10] J. L. Tallon and J. R. Cooper, in *Advances in Superconductivity*, Vol. 5, Springer Verlag, Tokyo (1993), p. 339.
[11] E. S. Itskevich and E. V. Zhasinas, submitted to Physica C.
[12] E. V. Antipov, *Dissertation for Candidate's Degree in Phys.-Math. Sciences*, Moscow State University, Moscow (1997).
[13] N. V. Zavaritskiĭ, E. S. Itskevich, and A. N. Voronovskiĭ, Zh. Éksp. Teor. Fiz. **60**, 1408 (1971) [Sov. Phys. JETP **33**, 762 (1971)].
[14] W. Struzkin, Yu. A. Timofeev, R. T. Downs, R. J. Hemley, and H. K. Mao, *Papers from Geophys. Lab.*, No. 2568, Carnegie Inst. of Washington, USA.
[15] A. R. Armstrong, W. I. F. David, I. Gameson *et al.*, Phys. Rev. B **52**, 2968 (1991).
[16] Y. Ohta, Y. Tohyama, and T. Maekawa, Phys. Rev. B **43**, 2968 (1991).
[17] S. M. Loreiro, E. V. Antipov, E. M. Kopnin *et al.*, Physica C **257**, 117 (1996).

Translated by D. H. McNeill

Reciprocity relations for the effective electrical conductivity of randomly inhomogeneous media in the fractal regime

A. A. Snarskiĭ*) and K. V. Slipchenko

Ukrainian National Technical University (Kiev Polytechnical Institute), 252056 Kiev, Ukraine

I. V. Bezsudnov

Nauka-Service, 103473 Moscow, Russia

(Submitted 10 July 1997)
Zh. Éksp. Teor. Fiz. **113**, 1484–1490 (April 1998)

An exact relation for the realization-averaged effective conductivities in the fractal region is found for two-dimensional randomly inhomogeneous media. It has the form $\{\sigma_e(\tau,L)\} \times\{1/\sigma_e(-\tau,L)\}^{-1}=\sigma_e^2(\tau=0,L\gg\xi)$, where ξ is the correlation length (the self-averaging scale), L is the size of the system, $\tau=(p-p_c)/p_c$, and p_c is the percolation threshold. For $L\gg\xi$, the system is self-averaged, and the relation transforms into the Dykhne reciprocity relation, A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1971)] $\sigma_e(\tau)\sigma_e(-\tau)=\sigma_e^2(\tau=0)=\sigma_1\sigma_2$. A similar relation is obtained for media with an exponentially broad distribution of local conductivities, as well as for individual realizations of some deterministic structures. © *1998 American Institute of Physics.* [S1063-7761(98)02604-3]

1. INTRODUCTION

As is widely known, a macroscopically inhomogeneous medium is described by its own effective characteristics, for example, the effective electrical conductivity σ_e, which, by definition, relates the volume-averaged electric fields $\langle\mathbf{E}\rangle$ and current densities $\langle\mathbf{j}\rangle$:

$$\langle\mathbf{j}\rangle=\sigma_e\langle\mathbf{E}\rangle, \tag{1}$$

where $\mathbf{E}=\mathbf{E}(\mathbf{r})$ and $\mathbf{j}=\mathbf{j}(\mathbf{r})$ are the local electric field intensity and current density. The averaging is performed over a volume V with the characteristic dimension $L\sim V^{1/3}$, which is much greater than the correlation length ξ. Self-averaging of the system occurs for $L>\xi$ (in this case L is customarily called the representative dimension).

Two-phase media, in which all the elements can be clearly separated into two types, such as ''black'' and ''white,'' ''highly conductive'' and ''poorly conductive,'' ''metallic'' and ''insulating,'' etc., represent a thoroughly studied type of randomly inhomogeneous media.

In the two-dimensional case ($d=2$), a series of exact relations[1] for σ_e can be obtained for two-phase media. At the percolation threshold, i.e., for $p=p_c=1/2$, where p is the concentration of the highly conductive phase with a specific conductance $\sigma_1>\sigma_2$, we obtain Dykhne's exact expression,[1] which is suitable for any value of the ratio $h=\sigma_2/\sigma_1$,

$$\sigma_e(p=p_c)=\sqrt{\sigma_1\sigma_2}, \quad L\gg\xi. \tag{2}$$

In the case of $p\neq p_c$ the exact analytical expression for σ_e is unknown and is probably impossible to obtain, but an exact relation can be written for so-called reciprocal media.

If a reciprocal (tilde-labeled) medium, which differs from the original medium by interchange of the conductivity values of the phases $\sigma_1\leftrightarrow\sigma_2$, is introduced, then, according to Ref. 1, σ_e and the effective conductivity of the reciprocal medium $\tilde{\sigma}_e$ obey an exact relation $\sigma_e(p)\tilde{\sigma}_e(p)=\sigma_1\sigma_2$, i.e., a so-called reciprocity relation. The local fields and currents in these media are related by the symmetry transformations[1]

$$\mathbf{j}=\Lambda\mathbf{n}\times\tilde{\mathbf{E}}\equiv\Lambda P_{\pi/2}\tilde{\mathbf{E}}, \quad \mathbf{E}=\Lambda^{-1}\mathbf{n}\times\tilde{\mathbf{j}}\equiv\Lambda^{-1}P_{\pi/2}\tilde{\mathbf{j}},$$
$$\Lambda=\sqrt{\sigma_1\sigma_2},$$

where $\mathbf{j}$ and $\mathbf{E}$ are the local current and field in the original medium, $\tilde{\mathbf{j}}$ and $\tilde{\mathbf{E}}$ are the analogous quantities in the reciprocal medium, $\mathbf{n}$ is a unit vector that is normal to the plane of the medium, and $P_{\pi/2}$ is the operator for rotation by $\pi/2$. In randomly inhomogeneous media $\tilde{\sigma}_e(p)=\sigma_e(1-p)$, and the reciprocity relation takes the form

$$\sigma_e(p)\sigma_e(1-p)=\sigma_1\sigma_2 \quad [=\sigma_e^2(p=p_c)],$$

whence follows, in particular, the expression (2) for $p=p_c=1/2$.

For media of dimensions $L<\xi$ (which are shorter than self-averaging scale), the system is mesoscopic, and the measured characteristics fluctuate from realization to realization. In this case the realization-averaged quantities are well defined.

2. RECIPROCITY RELATION FOR A TWO-PHASE TWO-DIMENSIONAL MEDIUM IN THE FRACTAL REGION

For percolation systems the averages of the specific conductance and resistance over realizations exhibit power-law dependences on the size of the system;[2] such systems are referred to as self-similar and fractal systems.[2,3] For the extreme case of strong inhomogeneity ($h=0$) the realization-averaged conductivity $\{\sigma_e\}$ and resistivity $\{\rho_e\}$ have the form[2]

$$\{\sigma_e\}=\sigma_1(L/a_0)^{-t/\nu}, \quad \tau>0, \quad L<\xi,$$

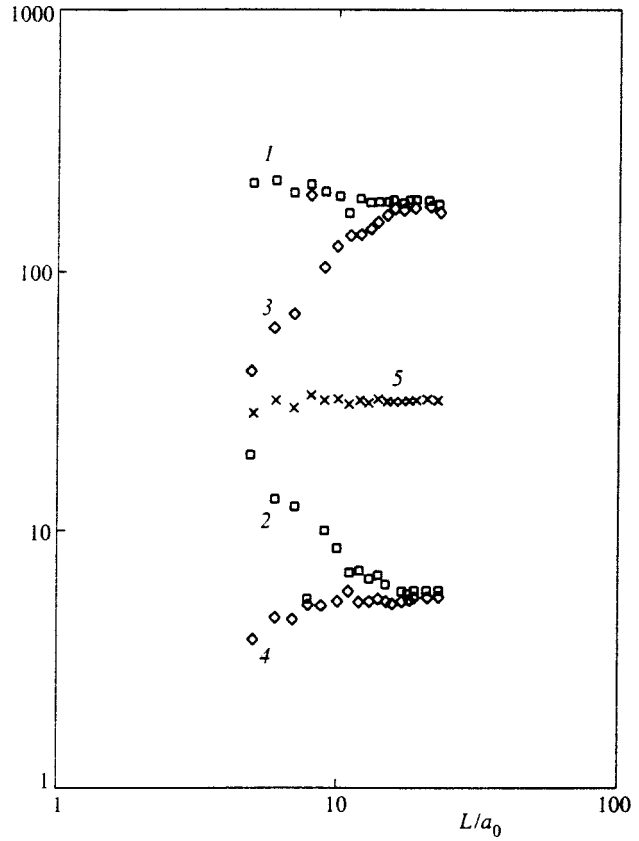


FIG. 1. Dependence of the realization-averaged conductances and resistances on the size of the sample for $\sigma_2/\sigma_1 \approx 10^{-2}$. *1* — $\{G(\tau=0.1,L)\}$, *2* — $\{G(\tau=-0.1,L)\}$, 3 — $1/\{R(\tau=0.1,L)\}$, *4* — $1/\{R(\tau=-0.1),L\}$, *5* — $\sqrt{\{G(\tau=0.1,L)\}/\{R(\tau=-0.1,L)\}}$.

$$\{\rho_e\}=\rho_2(L/a_0)^{g/\nu}, \quad \tau<0, \quad L<\xi,$$

where a_0 is the minimum size in the medium, and in the case of a network problem it is the length of a link. Although $\{\sigma_e\}$ and $\{\rho_e\}$ depend strongly (according to a power law) on L, the combination of these two parameters, which generalizes the reciprocity relation to the fractal region, scarcely depends on L in the two-dimensional case, as follows from the expressions presented above:

$$\{\sigma_e(\tau,L)\}\{\rho_e(-\tau,L)\}^{-1}=\sigma_e^2(\tau=0,L\gg\xi)\approx\sigma_1\sigma_2.$$

Figure 1 presents the results of numerical simulation of the conductivity of a two-dimensional network.

The reciprocity relation can be rigorously proved in the fractal region. For this purpose we consider a sample of size $L_\parallel\times L_\perp$ ($L_{\parallel,\perp}<\xi$) and apply current contacts to it, first along the vertical sides (in this case $\langle\mathbf{j}\rangle\| x$) and then along the horizontal sides. Ohm's law can then be written in the form

$$I_{\parallel,\perp}=G_{\parallel,\perp}U_{\parallel,\perp}, \tag{3}$$

where $G_{\parallel,\perp}$ is the conductance of the sample in the directions parallel and perpendicular to x.

The symmetry transformations in terms of the total current and the voltage drop ($I_{\parallel,\perp}$ and $U_{\parallel,\perp}$) have the form

$$I_{\parallel,\perp}=-\Lambda\tilde{U}_{\perp,\parallel}, \quad U_{\parallel,\perp}=-\Lambda^{-1}\tilde{I}_{\perp,\parallel}. \tag{4}$$

We note that in the general case the determination of the conductance or the resistance of a finite sample implies the assignment of specific boundary conditions (for example, ideal current contacts on some faces and insulation on others). It is easily seen that the symmetry transformations alter the boundary conditions so that the ideal current contacts become insulators, and, conversely, boundaries with an insulator become ideal current contacts. Thus, if we consider a rectangular sample, whose vertical faces are covered by ideal current contacts with a potential difference and whose horizontal faces are ideal insulators, in the reciprocal sample the current contacts with a potential difference are the horizontal faces, and the insulators are vertical faces. Integration of the transformations along contours joining the opposite faces gives the relations between the corresponding potential differences and currents in the dual samples. Substituting (4) into (3), for Ohm's law in the reciprocal medium we obtain

$$\tilde{I}_{\parallel,\perp}=\tilde{G}_{\parallel,\perp}\tilde{U}_{\parallel,\perp}, \quad \tilde{G}_{\perp,\parallel}=\Lambda^2/G_{\parallel,\perp}. \tag{5}$$

This relation in a finite $N\times N$ network was indicated in Ref. 4. Averaging the latter expression over realizations, we obtain

$$\{\tilde{G}_{\perp,\parallel}\}\{1/G_{\parallel,\perp}\}=\sigma_1\sigma_2. \tag{6}$$

In the case of $L_\parallel=L_\perp$, after moving over to specific characteristics, we obtain

$$\{\sigma_e(p,L)\}/\{\tilde{\rho}_e(p,L)\}=\sigma_1\sigma_2, \tag{7}$$

where $\rho_e=1/\sigma_e$, and since a randomly inhomogeneous medium satisfies $\{\tilde{\rho}_e(p,L)\}=\{\rho_e(1-p,L)\}$, the reciprocity relation (7) takes the form

$$\{\sigma_e(p,L)\}/\{\rho_e(1-p,L)\}=\sigma_1\sigma_2. \tag{8}$$

When $L>\xi$, it is no longer necessary to average over realizations, and (8) becomes the familiar reciprocity relation,[1] and when $p=p_c=1/2$, it transforms into Dykhne's formula for the effective conductivity (2).

Since the operations of division and averaging over realizations do not commute, it is impossible to obtain an expression analogous to Dykhne's formula (2) from (8) even for $p=p_c$. It is possible, however, to write such a reciprocity relation containing only one effective conductivity at the percolation threshold. To derive this relation, before averaging the second expression in (5), we take its logarithm and use the fact that $\{\ln G_\parallel\}=\{\ln G_\perp\}=\{\ln G\}$ when $L_\parallel=L_\perp=L$. Then, after the transition to the effective conductivity, for $p=p_c$ we obtain

$$\{\ln\,\sigma_e(p_c,L)\}=\ln\sqrt{\sigma_1\sigma_2} \quad (=\ln\,\sigma_e(p_c,L>\xi)). \tag{9}$$

3. RECIPROCITY RELATION FOR A STRIP

Let us consider a sample of a randomly inhomogeneous two-phase medium in the form of a strip, such that its dimension along the x axis satisfies $L_\parallel\gg\xi$, and its dimension along the y axis satisfies $L_\perp<\xi$. Although one of the dimensions of the sample is less than the correlation radius, both the conductance of the medium along the x axis (when the current contacts assigning the boundary conditions are such that $\langle\mathbf{j}\rangle$

is parallel to the x axis) $G_\parallel$ and the conductance in the direction perpendicular to it $G_\perp$ are well defined quantities and do not require additional averaging over realizations of the random structure. In fact, in one extreme case we have $L_\perp = a_0$, and self-averaging occurs for $L_\parallel \gg \xi_1 \propto |\tau|^{-\nu_1}$ and $\nu_1 = 1$.[5] In the other, two-dimensional, case, it occurs for $L_\perp = L_\parallel \gg \xi_2 \propto |\tau|^{-\nu_2}$ and $\nu_2 = 4/3$.[2] As $L_\perp$ increases from a_0 to $L_\parallel \gg \xi_2$, there is a transition from one-dimensional to two-dimensional behavior. The characteristic self-averaging length then goes from ξ_1 to ξ_2, and there are no conceivable reasons why this transition would not be monotonic. Therefore, as in the case of randomly inhomogeneous media (where $L_\parallel = L_\perp \gg \xi$), in our case we have

$$\tilde{G}_{\parallel,\perp}(p) = G_{\parallel,\perp}(1-p),$$

and we can thus write the following reciprocity relation for a strip:

$$G_\parallel(p) G_\perp(1-p) = \sigma_1 \sigma_2 . \tag{10}$$

At the percolation threshold, or, more precisely, in the blurred region, the reciprocity relation (10) relates $G_\parallel$ and $G_\perp$ for the same sample to one another as follows:

$$G_\parallel(p_c) G_\perp(p_c) = \sigma_1 \sigma_2 .$$

We note that in the special case in which $L_\perp$ is equal to the minimum size in an inhomogeneous system and the sample becomes homogeneous, the resistance along the sample is simply the sum of the elementary resistances, and the conductance across the sample is the sum of the elementary conductances. The reciprocity relation (10) then holds identically, as is easily demonstrated by a direct calculation.

4. RECIPROCITY RELATIONSHIP FOR A MEDIUM WITH AN EXPONENTIALLY BROAD DISTRIBUTION OF LOCAL CONDUCTIVITIES

One of the major problems in percolation theory is the determination of σ_e for a random network with an exponentially broad distribution of resistivities. The problem of hopping conduction reduces to this problem in a special case.[2,6] In the network variant the resistance of the ith link is given in the form $r_i = r_0 e^{\lambda x}$, where x is a random variable in a unit interval having a smooth distribution.[6] In the continuum variant we have $\sigma(x) = \sigma_0 e^{-\lambda x}$. Such a medium represents a special case of the two-dimensional media considered in Ref. 1 with the Ohm's law

$$\mathbf{j}(\mathbf{r}) = \exp(\langle \ln \sigma \rangle + \chi) \mathbf{E}(\mathbf{r}), \tag{11}$$

where $\chi(x,y) = \ln \sigma - \langle \ln \sigma \rangle$, and is treated, according to Ref. 1, as an ensemble of systems such that the multiple-point distribution function of the conductivity is an even function of χ. For $L > \xi$ the expression for σ_e is derived in analogy to the two-phase case, and, as was shown in Ref. 1,

$$\sigma_e(L > \xi) = \exp(\langle \ln \sigma \rangle).$$

Applying the symmetry transformations to a sample of size $L < \xi$, we can easily obtain

$$\{\sigma_e\}/\{\rho_e\} = \exp\langle \ln \sigma \rangle. \tag{12}$$

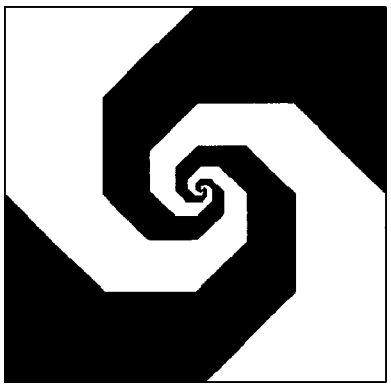


FIG. 2. Two-phase structure in the form of a ''broken spiral.'' The resistance of the sample remains unchanged upon interchange of the two phases and rotation by $\pi/2$.

This is the reciprocity relation for a system with an exponentially broad spread of resistivities.

As in the case of a two-phase medium, the relation for $\{\ln \sigma_e\}$ can be obtained for a medium with an exponentially broad distribution of local conductivities. Applying the arguments presented at the end of Sec. 1, we obtain an expression analogous to (9), which can be written in two forms:

$$\{\ln \sigma_e\} = \langle \ln \sigma \rangle, \quad \{\ln \sigma_e(L < \xi)\} = \ln \sigma_e(L > \xi), \tag{13}$$

i.e., in a medium with an exponentially broad distribution of local conductivities, the realization-averaged effective conductivity is equal to the volume-averaged local conductivity.

The reciprocity relation in a strip can be obtained in a similar manner. Without presenting the derivation, we at once write

$$G_\parallel G_\perp = \exp(2\langle \ln \sigma \rangle). \tag{14}$$

If the local conductivity is distributed according to the law $\sigma(x) = \sigma_0 e^{-\lambda x}$, where x is a random quantity with a uniform distribution function that takes values from zero to unity, then

$$G_\parallel G_\perp = \sigma_0^2 e^{\lambda}.$$

5. RESISTANCE OF DETERMINISTIC STRUCTURES

The reciprocity relations for realization-averaged conductances were considered above. There are, of course, no such relations for an individual realization. However, for a certain class of inhomogeneous structures the symmetry transformations makes it possible to obtain not only relations like the reciprocity relations, but also the exact value of the resistances and conductances for arbitrary conductivities of the phases.

As an example, let us consider the spiral structure depicted in Fig. 2. A structure consisting of two phases with the specific conductances σ_1 (black) and σ_2 (white) is constructed so that it remains unchanged under the interchange $\sigma_1 \leftrightarrow \sigma_2$ and rotation by $\pi/2$ about the axis which is perpendicular to the surface and passes through the center of the square. In structures which satisfy such symmetry, the concentrations of the phases are clearly equal. The expression for the conductance (and the resistance) of the sample G

$=\sqrt{\sigma_1\sigma_2}$ then follows at once from (5) (we recall that the thickness of the sample is assumed to be equal to unity). The result obtained means that $G_\parallel$ (with vertical contacts) and $G_\perp$ (with horizontal contacts) are identical. We note that this does not follow directly from the drawing of the structure in Fig. 2. Rotation by $\pi/2$ without the interchange $\sigma_1 \leftrightarrow \sigma_2$ does not give an identical result, since the structures are geometrically different. Similar structures of conducting paths (if, for example, it is assumed that the ''black'' phase conducts much better than the ''white'' phase) were considered in describing the flow of current in polycrystalline media in the case of strong conductivity anisotropy in Ref. 7. The concept of a trap for lines of good conductivity was introduced in that work. When a trap is approached, a line of good conductivity winds around its center, the current density increases, and current flows through lines of poor conductivity into the neighboring set of lines of good conductivity and ''grows'' along them. The conductance of such a trap was estimated using the variational principle for the case of strong anisotropy and had the value $G \sim \sqrt{\sigma_\parallel \sigma_\perp}$, where $\sigma_\parallel$ and $\sigma_\perp$ are the principal values of the conductivity tensor of the polycrystalline medium. The conductance of a similar two-phase trap, which was calculated exactly above, is also expressed in terms of the geometric mean, but now, of course, of the isotropic conductivities of the phases.

6. CONCLUSIONS

Several examples of two-dimensional media, for which a reciprocity relation for the realization-averaged conductance of the sample can be rigorously derived using Dykhne's symmetry transformations, have been considered here. These examples do not, of course, exhaust the class of media in which similar relations are possible. Symmetry transformations were used for $L > \xi$ in Ref. 1 to calculate the effective electrical conductivity in polycrystalline films and in Ref. 8 to calculate the effective galvanomagnetic characteristics. When $L > \xi$ holds, reciprocity relations are also possible for kinetic phenomena which are more complicated than conduction, for example, for thermoelectric phenomena.[9–12] Similar relations can be obtained for $L < \xi$ in all these cases.

The question of the existence of reciprocity relations in three-dimensional randomly inhomogeneous media merits a separate study. Such relations can clearly be only approximate and can hold in the region of the universal behavior of the effective characteristics, for example, near the percolation threshold.

We thank A. N. Lagar'kov, A. K. Sarychev, and P. M. Tomchuk for discussing some relevant questions. This work was carried out with partial financial support from the Russian Fund for Fundamental Research (Grants Nos. 95-02-04432-a and 97-02-16923a).

*)E-mail: asnar@phys.carrier.kiev.ua

[1] A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59** 110 (1970) [Sov. Phys. JETP **32**, 63 (1971)].
[2] D. Stauffer and A. Aharony, *Introduction to Percolation Theory*, Taylor and Francis, London–Washington, DC (1992).
[3] J. Feder, *Fractals*, Plenum Press, New York (1988) [Russ. transl., Mir, Moscow (1991)].
[4] J. Helsing and G. Grimvall, Phys. Rev. B **41**, 11 364 (1990).
[5] T. Ohtsuki and T. Keyes, J. Phys. A: Math. Gen. **17**, L559 (1984).
[6] B. I. Shklovskii and A. L. Efros, *Electronic Properties of Doped Semiconductors*, Springer-Verlag, Berlin (1984).
[7] Yu. A. Dreĭzin and A. M. Dykhne, Zh. Éksp. Teor. Fiz. **84**, 1756 (1983) [Sov. Phys. JETP **57**, 1024 (1983)].
[8] A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 641 (1970) [Sov. Phys. JETP **32**, 348 (1971)].
[9] B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **81**, 665 (1981) [Sov. Phys. JETP **54**, 355 (1981)].
[10] J. P. Straley, J. Phys. D **14**, 2101 (1981).
[11] V. Halpern, J. Phys. C: Solid State Phys. **16**, L217 (1983).
[12] B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **85**, 568 (1983) [Sov. Phys. JETP **58**, 331 (1983)].

Translated by P. Shelnitz

Mechanisms of Auger recombination in quantum wells

G. G. Zegrya[1)] and A. S. Polkovnikov[2)]

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted 24 July 1997)
Zh. Éksp. Teor. Fiz. **113**, 1491–1521 (April 1998)

The main mechanisms for the Auger recombination of nonequilibrium carriers in semiconductor quantum-well heterostructures are investigated. It is shown for the first time that there are three fundamentally different Auger recombination mechanisms in quantum wells: 1) a threshold-free mechanism, 2) a quasithreshold mechanism, and 3) a threshold mechanism. The rate of the threshold-free process has a weak temperature dependence. The rate of the quasithreshold Auger process exhibits an exponential temperature dependence. However, the threshold energy depends significantly on the quantum-well width and is close to zero for narrow quantum wells. It is shown that the threshold-free and quasithreshold processes are dominant in fairly narrow quantum wells, while the quasithreshold and threshold Auger processes are dominant in wide quantum wells. The limiting transition to a three-dimensional Auger process is accomplished for a quantum-well width tending to infinity. The value of the critical quantum-well width, at which the quasithreshold and threshold Auger processes combine to form a single three-dimensional Auger recombination process, is found. [S1063-7761(98)02704-8]

1. INTRODUCTION

Two recombination processes predominate in semiconductors at high nonequilibrium-carrier excitation levels: 1) a radiative process and 2) a nonradiative Auger process associated with an electron–electron interaction. The mechanisms of Auger recombination in homogeneous semiconductors have been studied by many investigators.[1–4] In narrow-gap semiconductors the Auger recombination processes involving two electrons and a heavy hole (the CHCC Auger process) or an electron and two heavy holes with a transition of one of them into a spin–orbit split-off band (the CHHS Auger process) are most probable.[2,4,5] Both these processes have thresholds, and the Auger recombination rate varies with the temperature according to an exponential law.[1,2] Semiconductors in which the spin–orbit splitting constant is close to the gap width (GaSb and InAs) are exceptions. In these semiconductors the rate of the CHHS process depends weakly on the temperature under certain conditions.[6] It is generally assumed that the phonon-assisted Auger recombination processes are dominant in weakly doped semiconductors at low temperatures and high nonequilibrium-carrier excitation levels.[3,5] Because of the large momentum transfer to the phonon, the threshold for the heavy holes is removed, and the rate of such an Auger process is a power function of the temperature. However, removal of the threshold as a result of interactions with phonons is not the only possibility. At high carrier concentrations hole–hole or hole–electron scattering can be a more effective mechanism for removing the threshold to the Auger process. Therefore, the question of the mechanism of Auger recombination in homogeneous semiconductors at low temperatures remains open. We have discussed this problem in greater detail in a separate paper.[7] Auger recombination processes involving impurities are not considered in the present work.

Solitary semiconductor heterostructures, quantum wells, quantum wires, and quantum dots are spatially inhomogeneous due to the existence of heterobarriers. The presence of an interface influences not only the energies and wave functions of the carriers, but also the macroscopic properties of the heterostructures, the influence on the latter being of fundamental importance.[8] It is generally assumed that the mechanism of Auger recombination in quantum wells is the same as the mechanism in a homogeneous semiconductor.[5,9–12] Nevertheless, the presence of an interface significantly influences the electron–electron interaction in quantum wells, and this influence has a fundamental character. An interface removes the constraints imposed by the energy and momentum conservation laws on electron–electron interaction processes. More specifically, the conservation law for the quasimomentum component perpendicular to the interface is removed. This, in turn, leads to the appearance of new threshold-free channels for Auger recombination in heterostructures.[8] The rate of the threshold-free Auger recombination process varies with the temperature according to a power law. At low temperatures the threshold-free process, unlike the threshold process, is a fairly effective channel of nonradiative recombination, and for narrow quantum wells it is dominant in comparison with the phonon-assisted Auger recombination process at high carrier concentrations.[7] The presence of a threshold-free matrix element for the electron–electron interaction also has a significant influence on the phonon-assisted Auger recombination process. The latter becomes a resonant process and is intensified significantly in comparison with the three-dimensional phonon-assisted Auger process.[7] However, at high concentrations of

the nonequilibrium carriers the phonon-assisted Auger process is considerably weaker than the phononless threshold-free Auger process down to extremely low temperatures. The first direct experiment devised to observe the threshold Auger recombination channel at $T=77$ K was described in Ref. 13.

A detailed analysis of the threshold and threshold-free Auger recombination mechanisms for a single heterobarrier was performed in Ref. 8. The conditions under which the threshold-free channel is more dominant than the threshold channel were analyzed. There has been no such detailed analysis of the Auger recombination mechanisms for quantum wells. The possibility of removing the threshold for the Auger recombination process for quantum wells associated with the passage of excited carriers into the continuous part of the spectrum was analyzed in Ref. 14. However, that paper does not include a microscopic theory for the threshold-free process, and there is no theoretical analysis of the competition between the threshold, quasithreshold, and threshold-free Auger recombination mechanisms at different temperatures for different quantum-well widths. In Refs. 15 and 16 only the threshold-free Auger recombination channel corresponding to small momentum transfers during the Coulomb interaction between the particles was considered (for the CHCC process) with neglect of the spin–orbit interaction.

The purpose of the present work is to theoretically investigate the principal mechanisms of the Auger recombination of nonequilibrium carriers in semiconductor quantum wells. It will be shown that there are three fundamentally different Auger recombination mechanisms in quantum wells: 1) a threshold mechanism, which is similar to an Auger process in a homogeneous semiconductor, 2) a quasithreshold mechanism, whose threshold energy depends significantly on the quantum-well width, and 3) a threshold-free mechanism, which is not observed in a homogeneous semiconductor. The threshold energy for a threshold Auger recombination process in a quantum well is close to the threshold energy for a homogeneous semiconductor. Conversely, because of the small value of the threshold energy, the rate of the quasithreshold process in narrow quantum wells depends weakly on the temperature. For this reason, in fairly narrow quantum wells there is no clear distinction between the threshold-free and quasithreshold Auger recombination mechanisms, and they combine to form a single threshold-free Auger process. As the quantum-well width increases, the threshold energy of the quasithreshold process increases and tends to the bulk value. The behavior of the threshold-free Auger recombination mechanism is totally different. As the quantum-well width increases, its rate decreases sharply, and this mechanism vanishes when we move over to a homogeneous semiconductor. The conditions under which the threshold-free Auger recombination mechanism is dominant in comparison with the threshold mechanism are obtained below. The critical value of the quantum-well width, at which the quasithreshold and threshold Auger recombination mechanisms combine to form one three-dimensional Auger process, are also found.

2. BASIC EQUATIONS

The wave functions of the charge carriers are needed to analyze the Auger recombination mechanisms and to find the rate of the Auger process. As has already been established for bulk Auger processes, the wave functions of the carriers must be calculated in a multiband approximation.[2] We shall use the four-band Kane model, which faithfully describes the wave functions and spectrum of the carriers in narrow-gap III–V semiconductors.[17]

For most III–IV semiconductors the wave functions of the conduction band at the Brillouin-zone center are described by the Γ_6^+ representation, and the wave functions of the valence band are described by the Γ_7^+ and Γ_8^+ representations. The first two of these are doubly degenerate, and the last representation is quadruply degenerate. The equations for the corresponding wave functions can be written in differential form. The basis wave functions of the conduction band and the valence band are usually taken in the form of eigenfunctions of the angular momentum.[17,18] However, for our purposes there is another more suitable representation of the basis functions:

$$|s\uparrow\rangle,\ |s\downarrow\rangle,\ |x\uparrow\rangle,\ |x\downarrow\rangle,\ |y\uparrow\rangle,\ |y\downarrow\rangle,\ |z\uparrow\rangle,\ |z\downarrow\rangle, \tag{1}$$

where $|s\rangle$ and the functions $|x\rangle$, $|y\rangle$, and $|z\rangle$ are s- and **p**-type Bloch functions with angular momenta equal to 0 and 1, respectively. The former describes the state of the conduction band, and the latter describe the state of the valence band at the Γ point. The arrows denote the direction of the spin. The wave function ψ of the carriers can be represented in the form

$$\psi=\Psi_s|s\rangle+\mathbf{\Psi}|\mathbf{p}\rangle,$$

where Ψ_s and $\mathbf{\Psi}$ are spinors. Near the Γ point the equations for the envelope functions of Ψ_s and $\mathbf{\Psi}$ in the spherical approximation have the following form:

$$\begin{cases} (E_c-E)\Psi_s-i\hbar\gamma\nabla\cdot\mathbf{\Psi}=0, \\ (E_v-\delta-E)\mathbf{\Psi}-i\hbar\gamma\nabla\Psi_s+\dfrac{\hbar^2}{2m}(\tilde{\gamma}_1+4\tilde{\gamma}_2)\nabla(\nabla\cdot\mathbf{\Psi}) \\ -\dfrac{\hbar^2}{2m}(\tilde{\gamma}_1+2\tilde{\gamma}_2)(\nabla\times(\nabla\times\mathbf{\Psi}))+i\delta(\boldsymbol{\sigma}\times\mathbf{\Psi})=0. \end{cases} \tag{2}$$

Here γ is the Kane matrix element, which has the dimensions of velocity, $\tilde{\gamma}_1$ and $\tilde{\gamma}_2=\tilde{\gamma}_3$ are generalized Luttinger parameters,[18] $\delta=\Delta_{so}/3$, Δ_{so} is the spin–orbit splitting constant, E_c and E_v are the energies of the lower edge of the conduction band and the upper edge of the valence band, m is the mass of a free electron, and $\boldsymbol{\sigma}=(\sigma_x,\sigma_y,\sigma_z)$ denotes the Pauli matrices. If the heavy-hole mass, which describes the interaction with the upper bands, is introduced phenomenologically instead of the Luttinger parameters, Eqs. (2) transform into the equations obtained in Ref. 19. It can be seen that Eqs. (2) do not differ from the equations usually used in the literature.[18,20–22] In the first equation of system (2) for the electrons we neglect the term with a heavy mass.

The Fourier transformation of Eqs. (2) makes it possible to obtain the spectra for the electrons and holes:

a) for heavy holes,

$$E_h=\delta-\frac{\hbar^2k^2}{2m_h};$$

b) for light holes and spin-split-off holes,

$$E_{l,\mathrm{so}}=-\frac{\delta}{2}-\frac{\hbar^2k^2}{4}(m_l^{-1}+m_h^{-1}) \pm\sqrt{2\delta^2+\left(\frac{\delta}{2}-\frac{\hbar^2k^2}{4}(m_l^{-1}+m_h^{-1})\right)^2}; \tag{3}$$

c) for electrons,

$$k^2=\frac{\mathscr{E}}{\hbar^2\gamma^2}\,\frac{\mathscr{E}^2+\mathscr{E}(2E_g+3\delta)+(E_g+3\delta)E_g}{E_g+\mathscr{E}+2\delta}.$$

Here k is the quasimomentum of the carriers,

$$m_l^{-1}=\frac{2\gamma^2}{E_g+\delta-E}+m^{-1}(\tilde{\gamma}_1+4\tilde{\gamma}_2),$$

$$m_h^{-1}=m^{-1}(\tilde{\gamma}_1-2\tilde{\gamma}_2),$$

m_h coincides with the heavy-hole mass, and m_l coincides with the light-hole mass in the case of a spin–orbit coupling constant equal to zero. For convenience, we set the energy of the upper edge of the valence band E_v equal to δ. We calculate the electron energy $\mathscr{E}$ from the lower edge of the conduction band.

The expression for the probability flux density can be derived from Eq. (2) by making the substitution $E\to -i\hbar\partial/\partial t$ and then using a procedure similar to the procedure employed in quantum mechanics.[23] It can also be derived using the $\mathbf{k}\cdot\mathbf{p}$ approximation in second-order perturbation theory. As a result, for the holes we obtain

$$\mathbf{j}_h=\frac{E_g+\delta-E}{2m_l\gamma}[\Psi_s\boldsymbol{\Psi}^*+\Psi_s^*\boldsymbol{\Psi}]-\frac{i\hbar}{2m_h}[(\boldsymbol{\Psi}\times\mathrm{curl}\boldsymbol{\Psi}^*) -(\boldsymbol{\Psi}^*\times\mathrm{curl}\boldsymbol{\Psi})]. \tag{4}$$

For electrons in the conduction band the expression for the probability flux density takes on a simpler form:

$$\mathbf{j}_e=\gamma[\Psi_s\boldsymbol{\Psi}^*+\Psi_s^*\boldsymbol{\Psi}]. \tag{5}$$

The exact procedure for deriving the boundary conditions for the wave functions at an interface are still not entirely clear. However, several approximate methods for solving this problem have been developed in the last few years. Following the method developed by Burt,[20] from the system of equations (2) we obtain the Kane equations, which can be integrated through an interface

$$\begin{cases}(E_g+\delta-E)\Psi_s-i\hbar\gamma\nabla\cdot\boldsymbol{\Psi}=0,\\ -E\boldsymbol{\Psi}-i\hbar\gamma\nabla\Psi_s+\dfrac{\hbar^2}{2m}\nabla(6\tilde{\gamma}_2\nabla\cdot\boldsymbol{\Psi})\\ +\dfrac{\hbar^2}{2m}\dfrac{\partial}{\partial x_k}(\tilde{\gamma}_1-2\tilde{\gamma}_2)\dfrac{\partial}{\partial x_k}\boldsymbol{\Psi}+i\delta(\boldsymbol{\sigma}\times\boldsymbol{\Psi})=0.\end{cases} \tag{6}$$

Using these equations and the probability flux density conservation law, we can obtain the boundary conditions for the wave-function envelopes [see Eqs. (24) and (25)].

Carrier states in a quantum well

The wave functions of carriers in a quantum well can be found by utilizing the symmetry properties of the Hamiltonian in the quantum well. The spinless Hamiltonian $\mathscr{H}_0$ is invariant under the replacement $x\to -x$. Let us consider the operator $\mathscr{R}$, which is such that

$$\mathscr{R}\colon (x,y,z)\to(-x,y,z),\quad \mathscr{R}=\mathscr{T}\mathscr{C}_{\pi x}, \tag{7}$$

$$\mathscr{H}_0\mathscr{R}=\mathscr{R}\mathscr{H}_0,$$

where $\mathscr{T}$ is the inversion operator, $\mathscr{C}_{\pi x}$ is the operator for rotation by π about the x axis, and the x axis is perpendicular to the plane of the quantum well.

With consideration of the spin–orbit coupling the Hamiltonian can be written in the form

$$\mathscr{H}=\mathscr{H}_0+\frac{\hbar}{4m^2c^2}(\nabla V\times\mathbf{p})\cdot\boldsymbol{\sigma}, \tag{8}$$

where $\mathbf{p}$ is the momentum operator and V is the potential energy of an electron in the crystal. The latter term does not commute with $\mathscr{R}$. Therefore, the symmetry operator $\mathscr{D}$ can be sought in the form of the product of $\mathscr{R}$ and a certain spin matrix S, which must be found: $\mathscr{D}=\mathscr{R}\otimes S$. Since inversion does not alter the sign of a vector product, the matrix S should satisfy the relations:

$$\begin{cases}S\sigma_x=\sigma_xS\\ S\sigma_y=-\sigma_yS,\\ S\sigma_z=-\sigma_zS\end{cases}\quad \sigma_x=\begin{bmatrix}0&1\\1&0\end{bmatrix},$$

$$\sigma_y=\begin{bmatrix}0&-i\\i&0\end{bmatrix},\quad \sigma_z=\begin{bmatrix}1&0\\0&-1\end{bmatrix}. \tag{9}$$

It is clear that the Pauli matrix σ_x can be taken as S: $S=\sigma_x$. In the basis (1) the matrix $\mathscr{D}$ takes the form

$$\mathscr{D}=\begin{bmatrix}0&1&0&0&0&0&0&0\\1&0&0&0&0&0&0&0\\0&0&0&-1&0&0&0&0\\0&0&-1&0&0&0&0&0\\0&0&0&0&0&1&0&0\\0&0&0&0&1&0&0&0\\0&0&0&0&0&0&0&1\\0&0&0&0&0&0&1&0\end{bmatrix},\quad \mathscr{D}=\mathscr{D}^{-1}. \tag{10}$$

The matrix of the Hamiltonian $\mathscr{H}$ must satisfy the commutation relations[22]

$$\mathscr{D}\mathscr{H}(-x,y,z)\mathscr{D}^{-1}=\mathscr{H}(x,y,z)\leftrightarrow\mathscr{H}(-x,y,z) =\mathscr{D}^{-1}\mathscr{H}(x,y,z)\mathscr{D}. \tag{11}$$

Consequently, $\Psi(x,y,z)$ and $\mathscr{D}\Psi(-x,y,z)$ satisfy the same equation. Therefore, the eigenfunctions of the Hamiltonian can be sought in the form of the eigenfunctions of the operator $\mathscr{D}$:

$$\Psi(x,y,z)+\nu\mathscr{D}\Psi(-x,y,z),\ \text{where } \nu=\pm1. \tag{12}$$

The values $\nu=\pm 1$ correspond to carrier states with different symmetry. When the wave functions are so chosen, the boundary conditions need to be satisfied on only one interface, since they will hold automatically on the second interface. Expanding Eq. (12), we find the necessary conditions for different components of the symmetrized wave functions:

$$\begin{aligned}
&\Psi_{s\uparrow}(x,y,z)=\pm\Psi_{s\downarrow}(-x,y,z),\\
&\Psi_{x\uparrow}(x,y,z)=\mp\Psi_{x\downarrow}(-x,y,z),\\
&\Psi_{y\uparrow}(x,y,z)=\pm\Psi_{y\downarrow}(-x,y,z),\\
&\Psi_{z\uparrow}(x,y,z)=\pm\Psi_{z\downarrow}(-x,y,z),
\end{aligned} \tag{13}$$

where the plus sign corresponds to $\nu=1$, while the minus sign corresponds to $\nu=-1$ for the s, y, and z components, and the opposite is true for the x component.

A. Holes in a rectangular quantum well

Selecting the coordinate system so that the longitudinal component of the wave vector corresponds to the y axis and performing Fourier transformation in that plane, we obtain the following expressions for the wave functions of the carriers.

For the heavy holes we have

$$\begin{aligned}
\mathbf{\Psi}_h(q,x)=H_1&\begin{pmatrix} q\,\cos k_h x\xi \\ -ik_h\,\sin k_h x\xi \\ -k_h\,\sin k_h x\xi+q\,\cos k_h x\eta \end{pmatrix}\\
&+H_2\begin{pmatrix} q\,\sin k_h x\eta \\ ik_h\,\cos k_h x\eta \\ -q\,\sin k_h x\xi-k_h\,\cos k_h x\eta \end{pmatrix}.
\end{aligned} \tag{14}$$

Here q and k_h are the y and x components of the quasimomentum of the heavy holes,

$$\xi=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix},\quad \eta=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix},$$

and H_1 and H_2 are normalization constants.

For the light holes we have

$$\begin{aligned}
\mathbf{\Psi}_l(q,x)=L_1&\begin{pmatrix} k_l\,\sin k_l x\eta-\lambda_l q\,\cos k_l x\xi \\ -iq\,\cos k_l x\eta+i\lambda_l k_l\,\sin k_l x\xi \\ -\lambda_l k_l\,\sin k_l x\xi+\lambda_l q\,\cos k_l x\eta \end{pmatrix}\\
&+L_2\begin{pmatrix} -k_l\,\cos k_l x\xi-\lambda_l q\,\sin k_l x\eta \\ -i\lambda_l k_l\,\cos k_l x\eta-iq\,\sin k_l x\xi \\ -\lambda_l q\,\sin k_l x\xi-\lambda_l k_l\,\cos k_l x\eta \end{pmatrix},
\end{aligned} \tag{15}$$

$$\Psi_{sl}=\frac{i\hbar\gamma(k_l^2+q^2)}{E_g+\delta-E}[L_1\,\cos k_l x\eta+L_2\,\sin k_l x\xi], \tag{16}$$

where $\lambda_l=\delta/(E+2\delta-\hbar^2k_l^2/2m_h)$, and q and k_l are the y and x components of the light-hole quasimomentum. The wave functions of the spin-split-off holes are similar to the wave functions of the light holes.

The transition to functions of another symmetry in the expressions presented above can be accomplished using the formal replacement $\xi\leftrightarrow\eta$ for the s, x, and y components and the replacement $\xi\leftrightarrow-\eta$ for the z component. In the barrier region ($x>a/2$) the wave functions of the bound holes decay exponentially as the distance from the interface increases:

$$\begin{aligned}
\mathbf{\Psi}_h=\tilde{H}_1&\begin{pmatrix} q\xi \\ -i\kappa_h\xi \\ -\kappa_h\xi+q\eta \end{pmatrix}\exp(-\kappa_h(x-a))\\
&+\tilde{H}_2\begin{pmatrix} q\eta \\ -i\kappa_h\eta \\ -q\xi+\kappa_h\eta \end{pmatrix}\exp(-\kappa_h(x-a)),
\end{aligned} \tag{17}$$

$$\begin{aligned}
\mathbf{\Psi}_l=\tilde{L}_1&\begin{pmatrix} \kappa_l\eta-\tilde{\lambda}_l q\xi \\ -iq\eta+i\tilde{\lambda}_l\kappa_l\xi \\ -\lambda_l\kappa_l\xi+\tilde{\lambda}_l q\eta \end{pmatrix}\exp(-\kappa_l(x-a))\\
&+\tilde{L}_2\begin{pmatrix} \kappa_l\xi-\tilde{\lambda}_l q\eta \\ i\tilde{\lambda}_l\kappa_l\eta-iq\xi \\ -\lambda_l q\xi+\tilde{\lambda}_l\kappa_l\eta \end{pmatrix}\exp(-\kappa_l(x-a)),
\end{aligned} \tag{18}$$

$$\Psi_{sl}=\frac{i\hbar\gamma(-\kappa_l^2+q^2)}{E_g+\tilde{\delta}+U_c-E}[\tilde{L}_1\eta+\tilde{L}_2\xi]\exp(-\kappa_l(x-a)), \tag{19}$$

$$\tilde{\lambda}_l=\frac{\tilde{\delta}}{U_v+E+2\tilde{\delta}-\hbar^2k_l^2/2m_h},\quad \tilde{\delta}=\frac{\tilde{\Delta}_{so}}{3}.$$

Here U_c and U_v are the heights of the barriers for the electrons and holes, $\tilde{\Delta}_{so}$ is the spin–orbit coupling constant in the broad-band region, and κ_l and κ_h are the magnitudes of the x components of the quasimomenta of the light and heavy holes under the barrier, which are related to the corresponding energy values according to Eqs. (3). For $x<-a/2$ the wave functions can be obtained from (17)–(19) using (13). If we simultaneously consider the wave functions of two or more particles, it is impossible to cause the z components of their quasimomentum to vanish simultaneously by selecting the coordinate system in a general form. The transition to a function with arbitrary quasimomentum directions is accomplished using the rotation matrix

$$D_\varphi=R_\varphi\otimes S_\varphi, \tag{20}$$

where R_φ acts on the coordinate components of the wave function, and S_φ acts on the spinor components. The Euler angles for rotation in the yz plane by an angle φ are

$$\Phi=-\pi/2,\quad \Theta=\varphi,\quad \Psi=\pi/2.$$

Thus,

$$R_\varphi=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\varphi & \sin\varphi \\ 0 & 0 & -\sin\varphi & \cos\varphi \end{bmatrix},$$

$$S_\varphi = \begin{bmatrix} \cos\varphi/2 & -i\sin\varphi/2 \\ i\sin\varphi/2 & \cos\varphi/2 \end{bmatrix}. \tag{21}$$

If the vector $\mathbf{q}$ has the components $q(0,\cos\varphi,\sin\varphi)$ in the x,y,z coordinate system, the wave function can be written in the form

$$\Psi_{\mathbf{q}} \equiv \Psi_\varphi = D_{-\varphi}\Psi_0. \tag{22}$$

The wave function previously found is labeled by a subscript zero. Here we present the wave function of the heavy holes obtained using (22), since we shall need it below:

$$\Psi_h(q,x,\phi) = H_1 \begin{bmatrix} q\cos k_h x e^{-i\phi}\xi \\ -ik_h \sin k_h x\xi - q\cos k_h x \sin\phi\eta \\ -k_h \sin k_h x\xi + q\cos k_h x\cos\phi\eta \end{bmatrix} + H_2 \begin{bmatrix} q\sin k_h x e^{i\phi}\eta \\ ik_h\cos k_h x\eta + q\sin k_h x\sin\phi\xi \\ -k_h\cos k_h x\eta - q\sin k_h x\cos\phi\xi \end{bmatrix}. \tag{23}$$

The boundary conditions for the hole wave functions can be derived by integrating the Kane equations (6) through the interface. We shall use the approximation $\gamma = \text{const}$, which is a good approximation for semiconductor heterostructures based on III–V compounds. Then we obtain the continuity conditions for the following quantities at an interface:

1) Ψ_x,

2) $\dfrac{E_g+\delta-E}{2m_l\gamma}\Psi_s + \dfrac{\hbar^2}{2m_h}\mathbf{q}\cdot\boldsymbol{\Psi}_\perp \Leftrightarrow (m_h^{-1}-m_l^{-1})\mathbf{q}\cdot\boldsymbol{\Psi}_\perp + im_l^{-1}\dfrac{\partial\Psi_x}{\partial x}$,

3) $m_h^{-1}\dfrac{\partial}{\partial x}\boldsymbol{\Psi}_\perp$,

4) $m_h^{-1}\dfrac{\partial}{\partial x}|\boldsymbol{\Psi}_\perp|^2$,

5) $m_h^{-1}|\boldsymbol{\Psi}|^2$. (24)

The boundary conditions are further simplified, if the value of the modified Luttinger parameters on the two sides of the interface $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$, which characterize the interactions with the higher bands, are assumed to be identical. In this case we have $m_h = \text{const}$. Taking into account $m_l^{-1} \approx 2\gamma^2/(E_g+\delta-E) \gg m_h^{-1}$, instead of (24) we obtain the continuity condition for the following quantities:

1) $\boldsymbol{\Psi}$,

2) $\dfrac{\partial}{\partial x}\boldsymbol{\Psi}_\perp$,

3) $\dfrac{1}{E_g+\delta-E}\,\text{div}\boldsymbol{\Psi}$. (25)

Generally speaking, the wave functions of the holes and in a quantum well are results of the superposition of three branches of the valence band: heavy, light, and spin-split-off holes. However, the last of the branches decays strongly, i.e., exponentially, with the exponent $\kappa_{so} \approx \sqrt{4m_h\Delta_{so}/3\hbar^2}$ as the distance from the interface increases. As a result, this branch mainly influences the value of the derivative of the wave function near the interface, and its influence on the value of the wave function itself is negligibly small. We stress that such an approximation is not equivalent to the use of a 4×4 Hamiltonian from the onset. We shall seek the wave function as the result of the superposition of the branches for the heavy and light holes. Near the upper edge of the valence band we have $|\lambda_{so}| \approx m_h/m_l \gg 1$. This means that only the first and third of the boundary conditions (25) can be employed. In this approximation the light and heavy holes do not mix and have different spectra, which are obtained when the wave functions are substituted into the boundary conditions (25).

The spectrum of the heavy holes coincides with the quantum-mechanical spectrum of a particle in a rectangular quantum well. The dispersion equation for states with even and odd x components of the wave function for the heavy holes takes the form

$$\tan\frac{k_h a}{2} = \frac{\kappa_h}{k_h}, \quad \text{for even states;}$$

$$\cot\frac{k_h a}{2} = \frac{k_h}{\kappa_h}, \quad \text{for odd states.} \tag{26}$$

For the light holes the states with different parity are no longer separated, and the dispersion equation becomes somewhat more complicated:

$$\left[\frac{E_g+\delta+V_c-E}{E_g+\delta-E}\frac{k_l^2+q^2}{\kappa_l^2-q^2}\kappa_l\cot\frac{k_l a}{2} + k_l\frac{2\lambda_l-1}{2\tilde{\lambda}_l-1}\right] \times \left[\frac{E_g+\tilde{\delta}+V_c-E}{E_g+\delta-E}\frac{k_l^2+q^2}{\kappa_l^2-q^2}\kappa_l\tan\frac{k_l a}{2} - k_l\frac{2\lambda_l-1}{2\tilde{\lambda}_l-1}\right] = q^2\left[\frac{2\lambda_l-1}{2\tilde{\lambda}_l-1} + \frac{E_g+\tilde{\delta}+V_c-E}{E_g+\delta-E}\frac{k_l^2+q^2}{\kappa_l^2-q^2}\right]^2. \tag{27}$$

Here a is the width of the quantum well. We note that the states of the light holes also split into states with different parity when $q=0$.

The opposite situation is observed for the spin-split-off holes. The components of the wave functions of the light and heavy holes oscillate rapidly, and their contribution to the overlap integrals is negligibly small. In just the same way, it can be seen that Ψ_x and $\text{div}\boldsymbol{\Psi}/(E_g+\delta-E)$ should be considered continuous for the split-off spin–orbit component. The form of the wave functions of the spin-split-off holes is similar to the form of the wave functions (15) and (16) of the light holes. Strictly speaking, when the condition $E_g-\Delta > U_v$ is satisfied, the spectrum of the spin-split-off holes is continuous. However, when the rapidly oscillating contributions of the light- and heavy-hole subbands are neglected, it can be either continuous or discrete. In the general case there is a density-of-states peak with small values for the hole momentum components that are perpendicular to the inter-

face near such a quasidiscrete level. The spectrum of these holes has a form similar to the spectrum of the light holes.[7]

B. Electrons in a rectangular quantum well.

Electrons obey the same symmetry rules as holes. Their wave functions have a form similar to that of the wave functions of the light holes and can be written as follows:

for $|x|<a/2$,

$$\Psi_{sc}=A_1\cos k_c x\,\eta+A_2\sin k_c x\,\xi,$$

$$\mathbf{\Psi}_c=\frac{i\hbar\gamma}{Z}A_1\begin{pmatrix} k_c\sin k_c x\,\eta-\lambda_c q\cos k_c x\,\xi \\ -iq\cos k_c x\,\eta+i\lambda_c k_c\sin k_c x\,\xi \\ -\lambda_c k_c\sin k_c x\,\xi+\lambda_c q\cos k_c x\,\eta\end{pmatrix}+\frac{i\hbar\gamma}{Z}A_2\begin{pmatrix} -k_c\cos k_c x\,\xi-\lambda_c q\sin k_c x\,\eta \\ -i\lambda_c k_c\cos k_c x\,\eta-iq\sin k_c x\,\xi \\ -\lambda_c q\sin k_c x\,\xi-\lambda_c k_c\cos k_c x\,\eta\end{pmatrix}, \tag{28}$$

for $x>a/2$,

$$\Psi_{sc}^{(1)}=[\tilde{A}_1\eta+\tilde{A}_2\xi]\exp\left(-\kappa_c\left(x-\frac{a}{2}\right)\right),$$

$$\mathbf{\Psi}_c^{(1)}=\frac{i\hbar\gamma}{\tilde{Z}}\tilde{A}_1\begin{pmatrix}\kappa_c\eta-\tilde{\lambda}_c q\xi \\ -iq\eta+i\tilde{\lambda}_c\kappa_c\xi \\ -\tilde{\lambda}_c\kappa_c\xi+\tilde{\lambda}_c q\eta\end{pmatrix}\exp\left(-\kappa_c\left(x-\frac{a}{2}\right)\right)+\frac{i\hbar\gamma}{\tilde{Z}}\tilde{A}_2\begin{pmatrix}\kappa_c\xi-\tilde{\lambda}_c q\eta \\ i\tilde{\lambda}_c\kappa_c\eta-iq\xi \\ -\tilde{\lambda}_c q\xi+\tilde{\lambda}_c\kappa_c\eta\end{pmatrix}\exp\left(-\kappa_c\left(x-\frac{a}{2}\right)\right), \tag{29}$$

where

$$Z=\frac{\mathscr{E}^2+\mathscr{E}(2E_g+2\delta)+(E_g+3\delta)E_g}{\mathscr{E}+E_g+2\delta},$$

$$\tilde{Z}=\frac{\mathscr{E}^2+\mathscr{E}(2E_g+2U_v+2\tilde{\delta})+(E_g+U_v+3\tilde{\delta})(E_g+U_v)}{\mathscr{E}+E_g+U_v+2\tilde{\delta}},$$

$$\tilde{\lambda}_c=\frac{\tilde{\delta}}{\mathscr{E}+E_g+U_v+2\tilde{\delta}}.$$

Here k_c is the x component of the electron quasimomentum in the quantum well, q is the electron longitudinal momentum, and κ_c is the magnitude of the x component of the electron quasimomentum under the barrier. Functions with different symmetries can be derived using the same procedure as in the case of holes. The continuity of Ψ_s and Ψ_x follows from the boundary conditions. This leads to the dispersion equation

$$\left(k_c\tan\frac{k_c a}{2}-\frac{Z}{\tilde{Z}}\kappa_c\right)\left(k_c\cot\frac{k_c a}{2}+\frac{Z}{\tilde{Z}}\kappa_c\right)=-q^2\left(\lambda_c-\tilde{\lambda}_c\frac{Z}{\tilde{Z}}\right)^2. \tag{30}$$

The spectrum splits into even and odd states if the longitudinal wave vector q is small or if the expression in parentheses on the right-hand side of the equation is close to zero. The latter condition usually holds, since as a rule we have $U_v\ll E_g$, which corresponds to semiconductors with a similar band structure.

3. PROBABILITY OF AUGER RECOMBINATION

The probability of Auger recombination per unit time is calculated within first-order perturbation theory with respect to the electron–electron interaction:

$$W_{i\to f}=\frac{2\pi}{\hbar}|M_{fi}|^2\delta(\varepsilon_f-\varepsilon_i), \tag{31}$$

where

$$M_{fi}=\langle\Psi_f(\mathbf{r}_1,\mathbf{r}_2,\nu_1,\nu_2)\left|\frac{e^2}{\kappa_0|\mathbf{r}_1-\mathbf{r}_2|}\right|\Psi_i(\mathbf{r}_1,\mathbf{r}_2,\nu_1,\nu_2)\rangle \tag{32}$$

is the matrix element of the electron–electron interaction, $\mathbf{r}_1$ and $\mathbf{r}_2$ are the coordinates of the carriers, ν_1 and ν_2 are the spin variables [see (12)], e is the charge of an electron, and κ_0 is the static dielectric constant of the semiconductor.

When the antisymmetrization of the wave functions is taken into account, the matrix element of the Auger transition can be written in the form

$$M_{fi}=M_{\rm I}-M_{\rm II}, \tag{33}$$

where

$$M_{\rm I}=\langle\Psi_3(\mathbf{r}_1,\nu_1)\Psi_4(\mathbf{r}_2,\nu_2)\times\left|\frac{e^2}{\kappa_0|\mathbf{r}_1-\mathbf{r}_2|}\right|\Psi_1(\mathbf{r}_1,\nu_1)\Psi_2(\mathbf{r}_2,\nu_2)\rangle,$$

$$M_{\rm II}=\langle\Psi_3(\mathbf{r}_1,\nu_1)\Psi_4(\mathbf{r}_2,\nu_2)\times\left|\frac{e^2}{\kappa_0|\mathbf{r}_1-\mathbf{r}_2|}\right|\Psi_1(\mathbf{r}_2,\nu_2)\Psi_2(\mathbf{r}_1,\nu_1)\rangle. \tag{34}$$

In quantum wells, as in homogeneous semiconductors, the main contribution to the Auger recombination rate is made by the CHCC and CHHS processes. Strictly speaking, such terminology is not applicable to the carriers in a quantum well, since there is mixing of the states of the light, heavy, and spin-split-off holes. However, as was indicated above, in the limit $m_c\ll m_h$ the mixing of the heavy and light holes is small, and the admixture of spin-split-off holes to the heavy and light holes can be neglected when the condition $\Delta_{so}\gg T$ is satisfied. The latter condition practically always holds for III–V semiconductors. Therefore, the terminology indicated above can be used. In the present work we confine ourselves to consideration of the CHCC Auger recombination process. All the dependences for the Auger recombina-

tion coefficients in the CHHS process with $\Delta_{so}<E_g$ are qualitatively identical to those for the CHCC process. Detailed expressions for the Auger recombination coefficients in the CHHS process were presented in Ref. 7. In the CHCC process there are two electrons and one heavy hole in the initial state, which we shall assume to be localized. The wave functions of one of the bound electrons and the hole will be denoted by φ_c and φ_h, respectively. The symbols Ψ_1 and Ψ_4 will be retained for the wave functions of the two departing carriers, and we shall interpret Ψ_4 as the final state of the high-energy electron.

Matrix element of an Auger transition

The matrix element of the electron–electron Coulomb interaction is most conveniently calculated using the Fourier representation:

$$M_1=\frac{4\pi e^2}{\kappa_0}\int\frac{I_{23}(p)I_{14}(-p)}{p^2+q^2}\frac{dp}{2\pi}\delta_{\mathbf{q}_1+\mathbf{q}_2-\mathbf{q}_3-\mathbf{q}_4},$$

where

$$I_{ij}(p)=\int\Psi_i^*(x)\Psi_j(x)e^{ipx}\,dx, \tag{35}$$

$$\delta_q=\begin{cases}1, & q=0,\\ 0, & q\neq 0,\end{cases}$$

$q=|\mathbf{q}_1-\mathbf{q}_4|=|\mathbf{q}_3-\mathbf{q}_2|$ is the momentum transfer in the plane of the quantum well during the Coulomb interaction. In narrow-gap III–V semiconductors the mass of an electron in the conduction band is generally much less than the mass of a heavy hole in the valence band. On the basis of these arguments we shall neglect the electron momenta in comparison to the hole momenta wherever they appear in the form of simple algebraic expressions. Such an approximation is not permissible for determining the threshold energy [see, for example, (60)].

It is convenient to represent the matrix element of the electron–electron interaction in the form

$$M_1=\frac{8\pi e^2}{\kappa_0}\int_0^\infty\Psi_1^*(x)\Psi_4(x)[\mathscr{I}(x)-\tilde{\mathscr{I}}(x)]\,dx, \tag{36}$$

where

$$\mathscr{I}(x)=\int_{-\infty}^{+\infty}\frac{dp}{2\pi}\frac{1}{p^2+q^2}\exp(-ipx)\int_{-a/2}^{a/2}\varphi_c(y)\varphi_h(y)\exp(ipy)\,dy$$

$$\approx\frac{1}{2q(q^2+k_h^2)}\begin{cases}2q\varphi_c(x)\varphi_h(x)\\ -\exp\left\{q\left(x-\frac{a}{2}\right)\right\}\left[q\varphi_c^<\left(\frac{a}{2}\right)\varphi_h^<\left(\frac{a}{2}\right)+(\varphi_c^<\varphi_h^<)'_{a/2}\right], & 0<x<\frac{a}{2},\\ \exp\left\{q\left(\frac{a}{2}-x\right)\right\}\left[q\varphi_c^<\left(\frac{a}{2}\right)\varphi_h^<\left(\frac{a}{2}\right)-(\varphi_c^<\varphi_h^<)'_{a/2}\right], & x>\frac{a}{2},\end{cases}$$

$$\tilde{\mathscr{I}}(x)=\int_{-\infty}^{+\infty}\frac{dp}{2\pi}\frac{1}{p^2+q^2}\exp(-ipx)\int_{|x|>a/2}\varphi_c(y)\varphi_h(y)\exp(ipy)\,dy$$

$$\approx-\frac{1}{2q(k_h^2-q^2)}\begin{cases}-\exp\left\{q\left(x-\frac{a}{2}\right)\right\}\left[q\varphi_c^>\left(\frac{a}{2}\right)\varphi_h^>\left(\frac{a}{2}\right)+(\varphi_c^>\varphi_h^>)'_{a/2}\right], & 0<x<a/2,\\ -2q\varphi_c(x)\varphi_h(x)+\exp\left\{q\left(\frac{a}{2}-x\right)\right\}\left[q\varphi_c^>\left(\frac{a}{2}\right)\varphi_h^>\left(\frac{a}{2}\right)-(\varphi_c^>\varphi_h^>)'_{a/2}\right], & x>a/2,\end{cases} \tag{37}$$

The $<$ and $>$ signs on the wave functions and their derivatives indicate that the values of the latter must be taken within the quantum well and in a barrier, respectively. Formulas (37) were derived using the approximation

$$e^{-qa}\ll 1,\quad e^{-\kappa_h a/2}\ll 1,$$

$$k_c^2\ll q^2+k_h^2. \tag{38}$$

Although the last inequality does not hold for fairly narrow quantum wells, the expressions obtained above, nevertheless, give a value of the Auger coefficient with a sufficiently high accuracy over the entire range of quantum-well widths.

Depending on the parity of the product $\varphi_c^*(x)\varphi_h(x)$, the integrals $\mathscr{I}(x)$ and $\tilde{\mathscr{I}}(x)$ satisfy the following relations[3)]

$$\mathscr{I}(-x),\tilde{\mathscr{I}}(-x)=\pm\mathscr{I}(x),\tilde{\mathscr{I}}(x),$$

where the plus and minus signs correspond to even and odd values of $\varphi_c^*(x)\varphi_h(x)$. At small values of the momentum transfer q the expressions (37) for the integrals $\mathscr{I}$ and $\tilde{\mathscr{I}}$ cease to be valid. This is because relations (38) do not hold at small values of q, since the interference processes due to scattering of the carriers on the two interfaces become significant. In the case of fairly high temperatures (greater than

the size-quantization energy of a heavy hole) the incorrect dependence of the matrix element on the momentum transfer at small values of q does not play a significant role, since the main contribution to the Auger recombination rate is made by momenta that are greater than or equal to the thermal value. In the case of low temperatures (below the size-quantization energy of a heavy hole) the bound carriers are located in the ground size-quantization level, and expressions (37) are easily modified. For this purpose it is sufficient to write $2\ \sinh(qx)e^{-qa}$ instead of the multipliers $e^{q(x-a)}$ and $2\ \sinh(qa)e^{-qx}$ instead of $e^{q(a-x)}$.

It is noteworthy that the matrix element $M_{\rm I}$ is nonzero only if the product $\Psi_1^*(x)\Psi_4(x)\varphi_c^*(x)\varphi_h(x)$ is an even function [see Eqs. (36) and (37)]. The expression for $M_{\rm II}$ is obtained from (36) by replacing the subscript 1 in $\Psi_1^*(x)$ by a subscript 2. The labels I and II will be omitted in the matrix elements up to the end of this section. In the approximation of an infinitely deep quantum well for the localized heavy holes, the term containing $\tilde{\mathscr{T}}$ vanishes and can, therefore, be neglected.

The state of an excited electron can lie in either the continuous or the discrete spectrum,[4)] the latter situation arising when the longitudinal momentum of the electron (in the plane of the quantum well) significantly exceeds the transverse momentum.

In determining the Auger recombination rate, both the localized and delocalized states must be taken into account as final states in the integral (36). The possibility of the transition of the excited electron into a bound or free state leads to significantly different Auger recombination mechanisms in quantum wells.

a) *Calculation of the Auger recombination matrix element for a transition into the continuous spectrum*. The wave function of the excited carriers $\Psi_4(x)$ satisfies the relations

$$\Psi_4''=\begin{cases} -k_4^2\Psi_4, & \text{for } |x|<a/2, \\ -\tilde{k}_4^2\Psi_4, & \text{for } |x|>a/2. \end{cases} \tag{39}$$

With consideration of relation (39), within the quantum well we can write

$$F_4^n=(-1)^n\frac{\Psi_4^{(n)}}{k_4^{2n}},$$

where F_4^n is the nth original form of Ψ_4. Similarly,

$$\Phi_4^n=(-1)^n\frac{(e^{qx}\Psi_4(x))^{(n)}}{(k_4^2+q^2)^n}e^{-2qx},$$

is the nth original form of $\Psi_4(x)e^{-qx}$. The expressions for the original F_4^n and Φ_4^n are needed below to calculate the matrix element M according to (36).

It is seen from (37) that the matrix element is a sum of two parts. The first of these is related to the presence of an interface and contains a multiplier which decays exponentially on the two sides of an interface. The second part originates from integration of the wave functions of the carriers within the quantum well. Therefore, it is natural to divide the matrix element into two parts:

$$M=\frac{4\pi e^2}{\kappa_0}2(M_1+M_2),$$

where

$$\begin{aligned} M_1=&\frac{1}{2q(q^2+k_h^2)}\Bigg\{(\varphi_c\varphi_h)'_{a-0}\sum_{n\geqslant 1}\frac{1}{(q^2+k_4^2)^n} \\ &\times\Bigg[\left(\Psi_4^*\exp\left\{q\left(\frac{a}{2}-x\right)\right\}\right)^{(n)}_{a/2-0}\Psi_1^{(n-1)}\Bigg|_{a/2-0} \\ &-\left(\Psi_4^*\exp\left\{q\left(x-\frac{a}{2}\right)\right\}\right)^{(n)}_{a/2+0}\Psi_1^{(n-1)}\Bigg|_{a/2+0}\frac{(q^2+k_4^2)^n}{(q^2+\tilde{k}_4^2)^n} \\ &+2qn\Psi_4^{(n-1)}\Big|_{a/2-0}\Psi_1^{(n-1)}\Big|_{a/2-0}\Bigg] \\ &+q(\varphi_c\varphi_h)_{a/2-0}\sum_{n\geqslant 1}\frac{1}{(q^2+k_4^2)^n}\Bigg[\left(\Psi_4\exp\left\{q\left(\frac{a}{2}\right.\right.\right. \\ &\left.\left.\left.-x\right)\right\}\right)^{(n)}_{a/2-0}\Psi_1^{(n-1)}\Bigg|_{a/2-0}\Bigg] \\ &+\frac{(q^2+k_4^2)^n}{(q^2+\tilde{k}_4^2)^n}(\Psi_4e^{q(x-a/2)})^{(n)}_{a/2+0}\Psi_1^{(n-1)}\Bigg|_{a/2+0} \\ &-2\Psi_4^{(n)}\Big|_{a/2-0}\Psi_1^{(n-1)}\Big|_{a/2-0}\Bigg] \\ &+2q\Bigg[\sum_{n\geqslant 1}C_{n+1}^2\frac{1}{(q^2+k_4^2)^{n+1}} \\ &\times\Psi_4^{(n)}\Psi_1^{(n-1)}\Big|_{a/2-0}(\varphi_c\varphi_h)''_{a/2-0} \\ &-\sum_{n\geqslant 1}C_{n+2}^3\frac{1}{(q^2+k_4^2)^{n+1}}\Psi_4^{(n-1)}\Psi_1^{(n-1)}\Big|_{a/2-0} \\ &\times(\varphi_c\varphi_h)'''_{a/2-0}\Bigg]\Bigg\}, \end{aligned} \tag{40}$$

$$\begin{aligned} M_2=&\frac{1}{q^2+k_h^2}\int_0^{a/2}(\Psi_4\Psi_1)(\varphi_c\varphi_h)\,dx \\ &-\sum_{n\geqslant 1}\frac{1}{(q^2+k_4^2)^n}\Bigg(-\Psi_4^{(n)}\Psi_1^{(n-1)}\varphi_c\varphi_h|_{a/2-0} \\ &+C_n^1\Psi_4^{(n-1)}\Psi_1^{(n-1)}(\varphi_c\varphi_h)'|_{a/2-0} \\ &+C_{n+1}^2\Psi_4^{(n)}\Psi_1^{(n-1)}(\varphi_c\varphi_h)''|_{a/2-0}\frac{1}{q^2+k_h^2} \end{aligned}$$

$$-C_{n+2}^{3}\Psi_4^{(n-1)}\Psi_1^{(n-1)}(\varphi_c\varphi_h)'''|_{a/2-0}\frac{1}{q^2+k_4^2}\Bigg).$$

In expanding the matrix elements M_1 and M_2 into series, we assume that the condition $q^2+k_4^2>(k_1^2,k_h^2)$ holds. In other words, the total momentum of the excited particle exceeds the momentum of the localized particle.

The value of M_1 is determined by the values of the wave functions and their derivatives at the interface, and $M_1\to 0$ as $a\to\infty$. Conversely, a δ-function contribution appears in M_2 as $a\to\infty$, and it transforms into the ordinary Auger recombination matrix element for a homogeneous semiconductor [see (62) and (64)]. We note that M_1, M_2, and, therefore, M are essentially threshold-free matrix elements. In fact, none of the constraints imposed on the initial momenta of the carriers k_1, k_c, and k_h apply to them. However, the mechanisms which lead to violation of the momentum conservation law ($k_1+k_2\neq k_3+k_4$) in these terms are different. In M_1 the violation of the conservation law is associated with the scattering of carriers on the interfaces. The same mechanism leads to the appearance of a threshold-free Auger process in the case of scattering on a single heterobarrier.[8] In M_2 the reason for the violation of the conservation law is the confinement of the integration volume with respect to x to the region of the quantum well, which leads to the appearance of a function of the form $k^{-1}\sin(ka/2)$ instead of $\delta(k)$. The differences between M_1 and M_2 just indicated physically mean that the matrix element M_1 corresponds to a true threshold-free process, whose appearance is attributed to the scattering of momentum on the heterobarriers. The matrix element M_2 corresponds to the quasithreshold process, and when the quantum-well width a tends to infinity, it transforms into the ordinary threshold matrix element. The additional sum appearing in the curly brackets in the expression for M_2 and the corresponding terms in M_1 were introduced so that at small q the expansion in the small parameter k/k_4 would begin at a higher power for M_2 than for M_1. Here k is the characteristic momentum of the localized carriers.

When M_1 is calculated, it can be assumed that $q\ll k_4$. In fact, at large values of q there is an exponentially small number of carriers.[5)] When M_1 and M_2 are calculated, it is convenient to utilize the following relations:

$$\sum_{k\geqslant 0}x^k=\frac{1}{1-x},\quad \sum_{k\geqslant 0}kx^k=\frac{x}{(1-x)^2},$$

$$\sum_{k\geqslant 0}k^2x^k=\frac{x(1+x)}{(1-x)^3},\quad \sum_{k\geqslant 0}k^3x^k=\frac{x(1+4x+x^2)}{(1-x)^4}.$$

Then

$$M_1\approx\frac{(\varphi_c\varphi_h)'_{a/2-0}}{2(q^2+k_h^2)}\Bigg[\Psi_4\Psi_1|_{a/2-0}\frac{1}{k_4^2}\frac{1+k_1^2/k_4^2}{(1-k_1^2/k_4^2)^2}$$

$$-\Psi_4\Psi_1|_{a/2+0}\frac{1}{\tilde{k}_4^2}\frac{1-\kappa_1^2/\tilde{k}_4^2}{(1+\kappa_1^2/\tilde{k}_4^2)^2}$$

$$+\frac{2}{k_4^4}\frac{1}{1-(k_1^2/k_4^2)^2}\Psi_4'\Psi_1'|_{a/2-0}$$

$$-\frac{2}{\tilde{k}_4^4(1+\kappa_1^2/\tilde{k}_4^2)^2}\Psi_4'\Psi_1'|_{a/2+0}\Bigg]$$

$$+\frac{(\varphi_c\varphi_h)_{a/2-0}}{2(q^2+k_h^2)}\Bigg[\frac{1}{k_4^2(1-k_1^2/k_4^2)}\Psi_4'\Psi_1|_{a/2-0}$$

$$-\frac{1}{\tilde{k}_4^2(1+\kappa_1^2/\tilde{k}_4^2)}\Psi_4'\Psi_1|_{a/2+0}$$

$$-\frac{1}{k_4^2(1-k_1^2/k_4^2)}\Psi_4\Psi_1'|_{a/2-0}$$

$$+\frac{1}{\tilde{k}_4^2(1+\kappa_1^2/\tilde{k}_4^2)}\Psi_4\Psi_1'|_{a/2+0}\Bigg]$$

$$-\frac{(\varphi_c\varphi_h)''_{a/2-0}}{q^2+k_h^2}\Bigg[\frac{1}{k_4^4}\frac{3+k_1^2/k_4^2}{(1-k_1^2/k_4^2)^3}\Psi_4\Psi_1'|_{a/2-0}$$

$$-\frac{1}{k_4^4}\frac{1+3k_1^2/k_4^2}{(1-k_1^2/k_4^2)^3}\Psi_4'\Psi_1|_{a/2-0}\Bigg]$$

$$-\frac{(\varphi_c\varphi_h)^{(3)}_{a/2-0}}{q^2+k_h^2}\Bigg[\frac{4}{3k_4^6}\frac{3+k_1^2/k_4^2}{(1-k_1^2/k_4^2)^4}\Psi_4'\Psi_1'|_{a/2-0}$$

$$+\frac{1}{3k_4^4}\frac{(3+k_1^2/k_4^2)(1+3k_1^2/k_4^2)}{(1-k_1^2/k_4^2)^4}\Psi_4\Psi_1|_{a/2-0}\Bigg]. \qquad (41)$$

This expression can be simplified appreciably, if we leave only the $O(k_1^2/k_4^2)$ terms in it. The expression in the second set of brackets is identically equal to zero according to the orthogonality condition of Ψ_1 and Ψ_4. As a result,

$$M_1\approx\frac{(\varphi_c\varphi_h)'_{a/2-0}}{2(q^2+k_h^2)}\Bigg[\Psi_4\Psi_1|_{a/2-0}\frac{1}{k_4^2}\left(1+\frac{3k_1^2}{k_4^2}\right)$$

$$-\Psi_4\Psi_1|_{a/2+0}\frac{1}{\tilde{k}_4^2}\left(1-\frac{3\kappa_1^2}{\tilde{k}_4^2}\right)\Bigg]-\frac{(\varphi_c\varphi_h)''_{a/2-0}}{(q^2+k_h^2)}$$

$$\times\Bigg[-\frac{1}{k_4^4}\Psi_4'\Psi_1|_{a/2-0}+\frac{3}{k_4^4}\Psi_4\Psi_1'|_{a/2-0}\Bigg]$$

$$-\frac{(\varphi_c\varphi_h)^{(3)}_{a/2-0}}{(q^2+k_h^2)}\frac{1}{k_4^4}\Psi_4\Psi_1|_{a/2-0}. \qquad (42)$$

The final expressions for M_1 can be obtained after substituting the expressions obtained in the preceding section for the wave functions of the carriers at an interface into (42) [see (14) and (29)]. As we have already noted above, the approximation of an infinitely deep quantum well was employed during this process for the heavy holes in the bound state. This approximation is justified in view of the large mass of the heavy holes. The condition $E_g>U_v$ remains in force. As a result we have

1) for $\nu_c = \pm \nu_h$,

$$M_1 \approx \frac{1}{2(q_h^2+k_h^2)} \frac{\hbar^3\gamma^3}{E_g^3}\left(\frac{1+\alpha/3}{1+\alpha/2}\right)^2 \frac{1+2\alpha/3}{1+\alpha} \frac{q_h k_c \kappa_c k_h}{\sqrt{k_c^2+\kappa_c^2}} \times\left(\frac{3V_c+V_v}{4E_g}+\frac{9}{2}\frac{E_{0c}}{E_g}\right) A_c A_h \Psi_c \Psi_f|_{a/2-0} - \frac{1}{2(q_h^2+\kappa_h^2)} \frac{\hbar^3\gamma^3}{E_g^3}\left(\frac{1+\alpha/3}{1+\alpha/2}\right)^2 \frac{1+2\alpha/3}{1+\alpha} \frac{q_h k_c k_h}{\sqrt{k_c^2+\kappa_c^2}} \times \frac{E_{0c}}{E_g} A_c A_h \Psi_c \Psi_f'|_{a/2-0}; \tag{43}$$

2) for $\nu_c = \mp \nu_h$,

$$M_1 \approx \frac{i}{2(q_h^2+k_h^2)} \frac{\hbar^3\gamma^3}{E_g^3} \frac{(1+\alpha/3)^2}{(1+\alpha/2)^2} \frac{1+2\alpha/3}{1+\alpha} \frac{q_c q_h k_c k_h}{\sqrt{k_c^2+\kappa_c^2}} \times \sin\phi\left(\frac{3V_c+V_v}{4E_g}+\frac{9E_g-6V_c}{2E_g}\right) A_c A_h \Psi_c \Psi_f|_{a/2-0} + \frac{i}{2(q_h^2+k_h^2)} \frac{\hbar^3\gamma^3}{E_g^3} \frac{(1+\alpha/3)^2}{(1+\alpha/2)^2} \frac{1+2\alpha/3}{1+\alpha} \times \sin\phi \frac{q_c q_h \kappa_c k_h}{k_c\sqrt{k_c^2+\kappa_c^2}} \frac{E_{0c}}{E_g} A_c A_h \Psi_c \Psi_f'|_{a/2-0}. \tag{44}$$

Here we have written $\alpha = \Delta_{so}/E_g$, and ϕ is the angle between the directions of the electron and heavy-hole longitudinal momenta q_c and q_h.

Let us proceed to an analysis of M_2. The integral appearing in M_2 is proportional to the sum

$$\int_0^a (\Psi_4\Psi_1)\varphi_c\varphi_h \, dx \propto \sum \pm \frac{\sin(k_4-k)a/2}{k_4-k}, \tag{45}$$

where k takes eight different values: $k = \pm k_1 \pm k_c \pm k_h$. The remaining terms in (40) are obtained by subtracting from it the expansion of this integral in the small parameter k/k_4 to fourth order inclusively. Here $\sqrt{k_4^2+q_4^2}$ must be used instead of k_4. For this reason, the terms of the form $\sin(k_4-k)a/(k_4-k)$ must be replaced by

$$\frac{\sin(k_4-k)a}{k_4-k}\chi(k,q),$$

where

$$\chi(k,q) = \frac{q^8+q^6(3k_4^2+kk_4)+q^4(3k_4^4+2kk_4^3+k^2k_4^2)+q^2(k_4^6+kk_4^5+k^3k_4^3)+k^4k_4^4}{(q^2+k_4^2)^4}. \tag{46}$$

In the limit $k \ll k_4$ we have $M_2 \propto k_4^{-5}$, and hence, $M_2 \sim M_1(k/k_4) \ll M_1$. When $k = k_4$, the multiplier accompanying M_2 becomes equal to unity, signifying that the expansion in the small parameter virtually vanishes. In the limit $k_4 \ll q$ the function χ is also approximately equal to unity. In addition, for $a \to \infty$

$$\sin\frac{(k_4-k)a/2}{k_4-k} \to \frac{\pi}{2}\delta(k-k_4),$$

whence it follows that M_2 transforms into the bulk Auger coefficient in the limit of an infinitely wide quantum well.

Of all the terms appearing in the sum (45), the largest is the one for which $k = k_1 + k_c + k_h$, since this term has the smallest threshold energy. (We interpret the threshold energy as the mean energy of a heavy hole participating in an Auger transition.) The contribution of the remaining terms to the sum is significantly smaller, and we neglect it for simplicity. Then the expression for the matrix element of the quasithreshold Auger process is written in the following form:

$$M_2 \approx \frac{1}{8(q_h^2+k_h^2)} e^{i\delta}\frac{\hbar\gamma}{E_g}\frac{1+2/3\alpha}{1+\alpha} A_c A_f A_c A_h \times \frac{\sin(k_f-k_{c1}-k_{c2}-k_h)a/2}{k_f-k_{c1}-k_{c2}-k_h} \times \chi(k_{c1}+k_{c2}+k_h, q_h)\begin{cases} q_h k_c e^{i\phi}+q_c k_h, & \nu_c = \pm\nu_h, \\ q_c q_h \sin\phi, & \nu_c = \mp\nu_h. \end{cases} \tag{47}$$

Here δ is an insignificant phase multiplier.

b) *Calculation of the Auger recombination matrix element for a transition into the discrete spectrum*. Let us now move on to an analysis of the matrix element of an Auger transition under which a particle with a large energy remains in the bound state. This case corresponds to the approximation $q_f \gg k_f$, where q_f and k_f are interpreted as the values of the momentum of the highly excited particle in the plane of the quantum well (q_f) and in directions perpendicular to it (k_f). Accordingly, no expansions need be performed in the matrix element, and we can write

$$M_3 \approx \frac{1}{q^2+k_h^2}\int_0^{a/2} (\Psi_4\Psi_1)(\varphi_c\varphi_h) \, dx. \tag{48}$$

This integral can easily be calculated, but the general formula obtained is quite involved, and we shall not present it here. We present only an estimate of M_3 here, which is valid when the bound carriers are in the ground quantum state:

$$M_3 \approx \frac{1}{q_h^2+k_h^2} e^{i\delta} \frac{\hbar \gamma a}{2Z} A_c A_f A_c A_h \alpha q_c q_h \sin\phi \quad (\nu_c = -\nu_h). \tag{49}$$

Here α is a multiplier of the order of unity, which arises from the integration of the product of the envelopes of the carrier wave functions over the region of the quantum well:

$$\int_0^{a/2} f_1 f_2 f_3 f_4 \, dx \approx \frac{a}{2\alpha}, \quad \text{where } f_i = \cos k_i x, \tag{50}$$

and i labels the initial and final states of the particles participating in the Auger recombination process. We note that α is nonzero only when the parities of the pairwise products $\Psi_1\Psi_4$ and $\Psi_c\Psi_h$ coincide; therefore, there is only one term in (49), unlike (47). For wide quantum wells, where the particles can be in different quantum bound states, α is given by the expression [compare (45)]

$$\alpha = \frac{1}{16} \sum_{\nu_1,\nu_2,\nu_3,\nu_4=0,1} (-1)^{\nu_i \sigma_i} \frac{\sin((-1)^{\nu_i} k_i) a/2}{(-1)^{\nu_i} k_i a/2}. \tag{51}$$

Here the label i refers to summation from 1 to 4, and σ_i characterizes the parity of the function f_i ($\sigma_i = 1$ for an odd function, and $\sigma_i = 0$ for an even function).

4. AUGER RECOMBINATION COEFFICIENT

The calculation of the Auger recombination rate requires summation of the probability of the Auger transition per unit time (31) over all the initial and final carrier states with the corresponding weights, i.e., occupation numbers. We present the expression for M_{I} and M_{II} (33) in the form of the sums

$$\begin{aligned} M_{\mathrm{I}} &= M_{\mathrm{I}}^{(1)} \delta_{\nu_3\nu_1} \delta_{\nu_4\nu_2} + M_{\mathrm{I}}^{(2)} \delta_{\nu_3 - \nu_1} \delta_{\nu_4\nu_2} \\ &\quad + M_{\mathrm{I}}^{(3)} \delta_{\nu_3\nu_1} \delta_{\nu_4 - \nu_2} + M_{\mathrm{I}}^{(4)} \delta_{\nu_3 - \nu_1} \delta_{\nu_4 - \nu_2}, \\ M_{\mathrm{II}} &= M_{\mathrm{II}}^{(1)} \delta_{\nu_3\nu_2} \delta_{\nu_4\nu_1} + M_{\mathrm{II}}^{(2)} \delta_{\nu_3 - \nu_2} \delta_{\nu_4\nu_1} \\ &\quad + M_{\mathrm{II}}^{(3)} \delta_{\nu_3\nu_2} \delta_{\nu_4 - \nu_1} + M_{\mathrm{II}}^{(4)} \delta_{\nu_3 - \nu_2} \delta_{\nu_4 - \nu_1}, \end{aligned} \tag{52}$$

where $M_{\mathrm{I}}^{(i)}$ and $M_{\mathrm{II}}^{(i)}$ are the matrix elements of transitions with fixed carrier spin states.

After averaging over the initial spin states and summing over the final spin states, we obtain

$$\left\langle \sum_{\nu_3,\nu_4} |M_{fi}|^2 \right\rangle = \sum_i \left(|M_{\mathrm{I}}^{(i)}|^2 + |M_{\mathrm{II}}^{(i)}|^2 - \mathrm{Re}[M_{\mathrm{I}}^{(i)*} M_{\mathrm{II}}^{(i)}] \right). \tag{53}$$

In the derivation of (53) we took into account that $M_{\mathrm{I}}^{(3)} \approx M_{\mathrm{II}}^{(3)} \approx M_{\mathrm{I}}^{(4)} \approx M_{\mathrm{II}}^{(4)} \approx 0$. It is convenient to use $\langle M^2 \rangle$ to denote the matrix element summed over the spin variables:

$$\langle M^2 \rangle = \sum_{\nu_3,\nu_4,\nu_1,\nu_2} |M_{fi}|^2. \tag{54}$$

Then the Auger recombination rate takes the form

$$G = \frac{2\pi}{\hbar} \sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_4} \langle M^2 \rangle f_1 f_2 (1-f_3)(1-f_4) \delta \times (E_3 + E_4 - E_1 - E_2). \tag{55}$$

Here f_1 and f_2 are the Fermi distribution functions of the carriers in the initial state, and f_3 and f_4 are the analogous functions in the final state. For highly excited states, f_4 can be set equal to zero. It is noteworthy that we can write $\tilde{f}_3$ instead of $1-f_3$, where $\tilde{f}_3$ is the hole distribution function.

The contributions of the matrix elements M_1, M_2, and M_3 to the Auger recombination rate differ, since the excited particles for such processes lie in different quantum states. In the case of M_1 and M_2 the excited particles lie in states in the continuous spectrum, and in the case of M_3 the excited particle lies in a state in the discrete spectrum. It is more complicated with separation of the contributions between M_1 and M_2. Although the physical difference between them is maintained, there is an interference term between them. At small values of the quantum-well width the interference is significant, since both processes are virtually threshold-free, but a calculation neglecting it continues to lead to a result of the correct order of magnitude, which reflects all the main features of the Auger recombination coefficient as a function of the temperature and the parameters of the quantum-well structure. If the quantum well is sufficiently wide, the interference between M_1 and M_2 can be neglected. In fact, while M_1 as a function of the quasimomenta does not have any peaks, the magnitude of M_2 has a maximum at the point $k_4(q) + k_3 = k_1 + k_2$. When the quantum-well width tends to infinity, the maximum at that point has the character of a δ function. Therefore, for wide quantum wells the Auger recombination probability corresponding to M_2 as a function of the heavy-hole longitudinal momentum q_h has a maximum at larger values of q_h than does the probability corresponding to M_1. As the quantum-well width decreases, the maxima of these probabilities approach one another, and the region of overlap between these matrix elements grows.

The Auger transition probabilities corresponding to M_1 and M_2 as functions of the heavy-hole longitudinal momentum for various quantum-well widths are presented in Fig. 1. It is seen from the figure that interference between the threshold-free process corresponding to M_1 and the quasithreshold process corresponding M_2 takes place only for narrow quantum wells, in accordance with the foregoing statements. It should be noted that the Auger recombination probabilities are fairly smooth functions of the heavy-hole longitudinal momentum, since summation over discrete quantum states of the carriers was performed during their calculation. When q_h is close to the maximum value specified by the longitudinal momentum and energy conservation laws, the Auger recombination probability has a square-root divergence, which can be eliminated by integrating over q_h, i.e., by calculating the Auger recombination rate.

Thus, we represent the Auger recombination rate in the form

$$G = G_1 + G_2 + G_3, \tag{56}$$

where the rate G_1 corresponds to the threshold-free Auger process with the matrix element M_1, G_2 corresponds to the

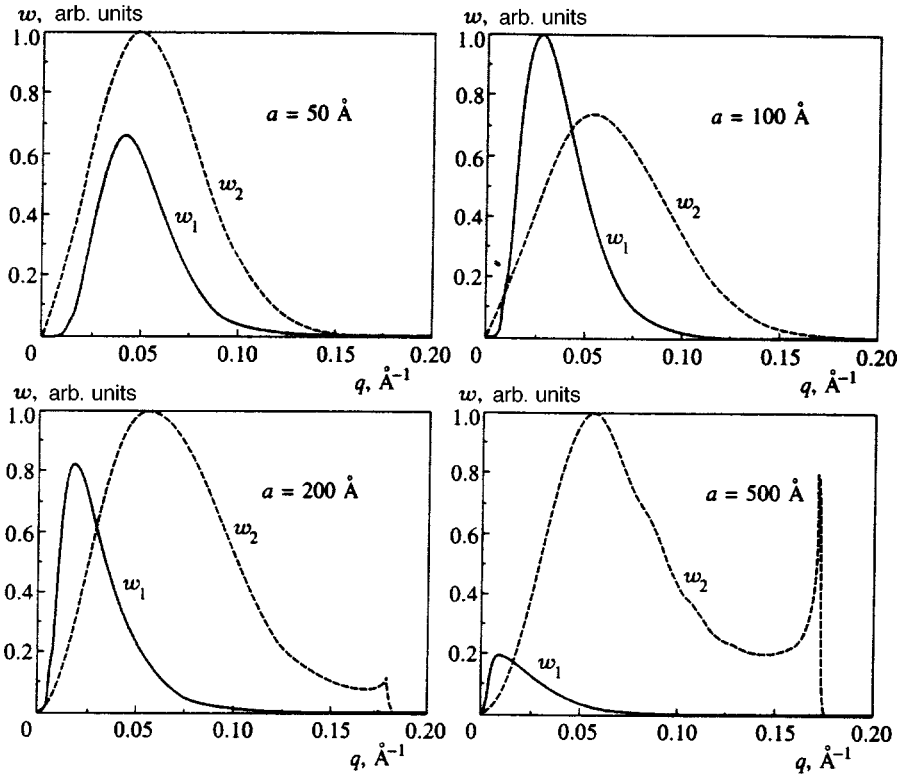


FIG. 1. Auger transition probabilities w_1 and w_2 corresponding to the threshold-free and quasithreshold matrix elements M_1 and M_2 vs. the heavy-hole longitudinal momentum at $T=300$ K for various values of the quantum-well width a. The parameters characteristic of InGaAsP/InP quantum wells with $E_g \approx 1$ eV were used in the calculation.

quasithreshold Auger process with the matrix element M_2, and G_3 corresponds to the threshold Auger process with the matrix element M_3.

Expressions for G_1 and G_2 can be obtained using Eq. (55) by replacing the summation over k_4 in it by integration and going over from a δ function with respect to the energy to a δ function with respect to the momentum. Below we shall investigate the Auger recombination coefficient C, which is related to the rate G by the expression

$$G = Cn^2 p.$$

Here n and p denote the two-dimensional concentrations of the electrons and holes, respectively. For the Auger recombination coefficient C_1 we have

$$C_1 \approx \frac{256\pi^2}{3}\frac{E_B}{\hbar}\lambda_g^4 \frac{1}{a(a+2/\kappa_c)^2} F\left(\frac{\Delta_{so}}{E_g}\right)\frac{k_c^2\kappa_c^2}{(k_c^2+\kappa_c^2)^2} \times\left\langle \frac{q_h^2 k_h^2}{(q_h^2+k_h^2)^3}\frac{1}{k_f(q_h)}\right\rangle \left\{ \frac{V_c}{E_g}\left[\left(\frac{3V_c+V_v}{4E_g}+\frac{9}{2}\frac{E_{0c}}{E_g}\right)^2 + \frac{k_f^2}{\kappa_c^2}\frac{E_{0c}^2}{E_g^2}\right] + \frac{T}{2E_g}\frac{k_c^2+\kappa_c^2}{\kappa_c^2}\left[\left(\frac{V_v-9V_c}{4E_g}+\frac{9}{2}\frac{E_{0c}}{E_g}\right)^2 + \frac{k_f^2\kappa_c^2}{k_c^4}\frac{E_{0c}^2}{E_g^2}\right]\right\}, \tag{57}$$

where

$$F(x) = \frac{1+x/2}{1+4x/9}\frac{1+7x/9+2x^2/9}{(1+x/2)^2}, \quad E_B = \frac{m_c e^4}{2\hbar^2\kappa_0^2},$$

$$\lambda_g = \frac{\hbar}{\sqrt{2m_c E_g}}.$$

For simplicity, here we have replaced the electron longitudinal momentum by its thermal value. For degenerate carriers the electron Fermi energy E_F must be used instead of the temperature. The angle brackets denote averaging over the distribution function of the heavy holes. In the case of a Boltzmann distribution, which is usually observed for holes, this averaging has the form

$$\langle f(q)\rangle = \frac{2}{q_T^2}\int_0^\infty q f(q)\exp\left(-\frac{q^2}{q_T^2}\right) dq,$$

where q_T is the heavy-hole thermal momentum:

$$q_T = \sqrt{\frac{2m_h T}{\hbar^2}}.$$

Similarly, for C_2 we obtain

$$C_2 \approx \frac{32\pi^2}{3}\frac{E_B}{\hbar}\lambda_g^2 \left\langle \frac{\chi^2(q_h, k_h+2k_c)}{a(a+2/\kappa_c)^2}\frac{q_c^2 k_h^2 + q_h^2(k_c^2+q_c^2/2)}{(q_h^2+k_h^2)^3 k_f} \times\frac{1-\cos(k_f-k_h-2k_c)a}{2(k_f-k_h-2k_c)^2}\right\rangle, \tag{58}$$

where $\chi(q,k)$ is the function defined above [see (46)]. Finally, for C_3 we have

$$C_3 \approx \frac{2048\pi^2}{3}\frac{E_B}{\hbar}\lambda_g^2 \frac{a}{(a+2/\kappa_c)^3}$$

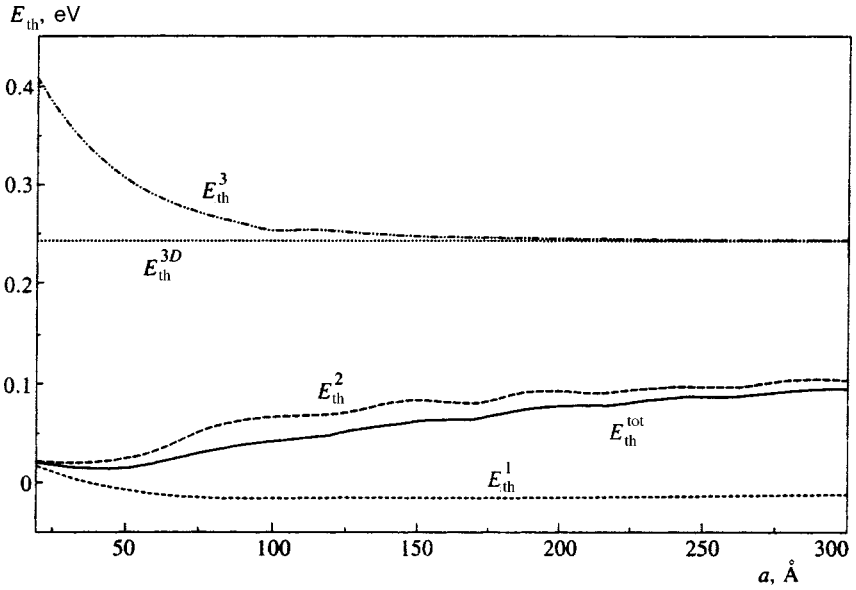


FIG. 2. Dependence of the threshold energy for the CHCC process on the quantum-well width for three Auger recombination mechanisms: the threshold-free mechanism ($E_{\rm th}^1$), the quasithreshold mechanism ($E_{\rm th}^2$), and the threshold mechanism ($E_{\rm th}^3$) at $T=300$ K. The solid curve corresponds to the threshold energy $E_{\rm th}^{\rm tot}$ for the total Auger recombination coefficient ($C=C_1+C_2+C_3$). The horizontal dotted line corresponds to the threshold energy $E_{\rm th}^{\rm 3D}$ for the three-dimensional Auger process.

$$\times\left\langle \frac{q_{\rm th}^2}{q_T^2}\frac{q_c^2}{(q_{\rm th}^2+k_h^2)^3}\exp\left(-\frac{q_{\rm th}^2}{q_T^2}\right)\alpha^2\right\rangle_n . \tag{59}$$

In the latter case the averaging is performed only over the discrete heavy-hole quantum states and the electron momenta. The threshold value of $q_{\rm th}$ is found from the energy conservation law and the longitudinal component of the momentum:

$$E_f(\sqrt{k_f^2+q_{\rm th}^2})=E_g+\frac{\hbar^2(q_{\rm th}^2+k_h^2)}{2m_h}+\frac{\hbar^2(k_{c1}^2+k_{c2}^2)}{2m_c}.$$

For simplicity, we neglected the electron longitudinal momenta in determining the threshold energy, since they are small; however, we took into account the size-quantization energy of the electrons, which alters the effective value of the gap width in the quantum well. If the energy of the excited electron E_f is expanded into a series in the momenta near $q_{\rm th}=Q$, where Q is the value of the electron momentum corresponding to an energy equal to E_g ($Q\approx\sqrt{4m_cE_g/\hbar^2}$), the following estimate can be obtained for the threshold momentum:

$$q_{\rm th}\approx\sqrt{\frac{4m_cE_g}{\hbar^2}+\frac{3}{2}k_c^2}. \tag{60}$$

If the quantum-well width tends to infinity, the threshold momentum tends to its bulk value.[2] In addition, it must be taken into account that for wide quantum wells with a large number of levels the multiplier α [see (49)] tends to a δ function. This is a manifestation of the conservation law for the transverse component of the quasimomentum:

$$\alpha^2\to\frac{\pi}{128}a\sum\delta(k_h\pm k_{c1}\pm k_{c2}\pm k_{c4}).$$

When the quantum-well width is large and $V_c\ll E_g$ holds, the inequality $C_3\ll C_2$ is satisfied, since $C_3/C_2\approx\sqrt{V_c/E_g}$. Therefore, for wide wells C_3 can be neglected in comparison to C_2. If $V_c\lesssim E_g$, the relation $C_3/C_2\approx\sqrt{V_c/(E_g-V_c)}\geqslant 1$ holds. For narrow quantum wells the threshold energy for C_3 increases [see (60)], and the Auger recombination coefficient (59) decreases in comparison to the bulk value by the factor

$$\exp\left(\frac{k_c^2}{q_T^2}\right)\approx\exp\left(\frac{m_c}{m_h}\frac{E_{0c}}{T}\right).$$

By equating the exponent to unity it is easy to estimate the characteristic quantum-well width at which this effect becomes significant:

$$E_{0c}\approx T\cdot\frac{2}{3}\frac{m_h}{m_c}\Leftrightarrow a\approx\frac{\pi}{q_T}. \tag{61}$$

Thus, at values of the quantum-well width a that are smaller than several reciprocal thermal momenta, $a\lesssim\pi/q_T$, the threshold energy $E_{\rm th}(a)$ becomes significantly greater than the bulk value $E_{\rm th}^{\rm 3D}$ (see Fig. 2). For III–V semiconductor compounds at room temperature the equality (61) is achieved when the quantum-well width is of the order of a hundred angstroms.

Let us consider the Auger recombination coefficient C_2 in greater detail for the quasithreshold process. In the limit $a\to\infty$ the following replacement can be made in the averaged function in (58):

$$\frac{1-\cos(k_f-k_h-2k_c)a}{2(k_f-k_h-2k_c)^2}\to\frac{\pi a}{2}\delta(k_f-k_h-2k_c). \tag{62}$$

This formula clearly reveals the presence of a threshold in this limit, and after multiplication by a^2, C_2 transforms into the three-dimensional expression. For comparison, we present the result from Ref. 2 for $C^{\rm 3D}$ and our limiting expression:

$$C^{\rm 3D}=6\sqrt{2\pi^5}\frac{e^4m_c\hbar^3}{\kappa_0^2}\frac{1}{E_g^{5/2}T^{1/2}m_c^{1/2}m_h^{3/2}}$$

$$\times\exp\left(-\frac{2m_c}{m_h}\frac{E_g}{T}\right), \tag{63}$$

$$C_2a^2=6\frac{16\sqrt{2\pi^5}}{27}\frac{e^4m_c\hbar^3}{\kappa_0^2}\frac{1}{E_g^{5/2}T^{1/2}m_c^{1/2}m_h^{3/2}}$$

$$\times\exp\left(-\frac{2m_c}{m_h}\frac{E_g}{T}\right). \tag{64}$$

The multiplier 4 in (64) appears, because the calculation of M_2 according to (47) requires considering not only the term with $k=k_{c1}+k_{c2}+k_h$, but also the terms with $k=k_{c1}-k_{c2}+k_h$, $k=-k_{c1}+k_{c2}+k_h$, and $k=-k_{c1}-k_{c2}+k_h$. When the quantum-well width tends to infinity, all four terms make identical contributions to C_2. As we see, there is a difference between expressions (63) and (64) only in the numerical multiplier. The small discrepancy by a factor of 2/3 appears, because the size-quantization momenta of the electrons must be distinguished from one another when there is a large number of electronic levels: $k_{c1}\neq k_{c2}$. In addition, expression (63) was obtained in a simplified model, in which the magnitude of the spin–orbit splitting is assumed to be infinitely large. The condition $\Delta<E_g$, which holds for most narrow-gap III–V semiconductors, was actually used in deriving (64). When we calculated (64), we neglected V_c in comparison to E_g. In the general case C_2a^2 must be replaced by $(C_2+C_3)a^2$, and (64) remains valid. However, the limiting transition from the quasithreshold process to the threshold process [see (62)] can be made only for very wide quantum wells. A qualitative criterion for this transition can be obtained from an analysis of the Auger transition probability as a function of the heavy-hole momentum. As we have already noted above, the quasithreshold process with C_2 dominates for wide quantum wells. The probability of this process has two characteristic peaks [see (58)]. The first of them corresponds to a maximum of the square of the matrix element for the transition near the threshold value of the heavy-hole momentum. The width of this maximum is of the order of the reciprocal width of the quantum well. The second peak is near the value of the hole thermal momentum. The Auger recombination coefficient C_2 can be estimated by multiplying the values of the Auger transition probability near these maxima by the corresponding widths. Then

$$C_2\approx C_2^{\rm th}(Q_h\approx q_{\rm th})+C_2^T(Q_h\approx q_T), \tag{65}$$

where Q_h is the value of the heavy-hole momentum: $Q_h^2=k_h^2+q_h^2$.

$$C_2^T\approx C_2^{\rm th}\frac{\lambda_{E_g}}{a}\left(\frac{T}{E_{\rm th}}\right)^{3/2}\exp\frac{E_{\rm th}}{T}.$$

Here $\lambda_{E_g}\approx 2\pi/q_{\rm th}$ is the wavelength of the highly excited electron. A comparison of the contributions from $C_2^{\rm th}$ and C_2^T provides a natural criterion for transformation of the quasithreshold process into the three-dimensional threshold Auger process. It can be written as follows:

$$a\gg a_c,$$

where

$$a_c=\lambda_{E_g}\left(\frac{T}{E_{\rm th}}\right)^{3/2}\exp\left(\frac{E_{\rm th}}{T}\right). \tag{66}$$

For semiconductors with a wide gap of order 1 eV at room temperature the critical width a_c can reach several thousand angstroms. However, the criterion $a\gg a_c$ for passage from the two-dimensional Auger process to the three-dimensional process was obtained without consideration of the momentum relaxation processes of the particles due to interactions with phonons and electron–electron scattering. In reality, the mean free path of the particles is generally smaller than a_c, and the transition to the three-dimensional Auger process actually occurs when the quantum-well width takes a value of the order of the mean free path of the carriers. It follows directly from these statements that the Auger coefficient in the three-dimensional case can be correctly calculated only when the processes involving scattering of the electron and hole momenta are taken into account.[7]

Quantum wells with a width $a<a_c$ exhibit displacement of the probability maximum w_2 for the quasithreshold process as a function of the heavy-hole momentum toward longer wavelengths (see Fig. 1) as a decreases. This leads to a decrease in the threshold energy of this process and, as a consequence, in weakening of the temperature dependence of the Auger recombination coefficient.

Figure 2 presents plots of the dependence of the threshold energy on the quantum-well width for all three Auger recombination mechanisms with C_1, C_2, and C_3 taken individually and for the overall Auger recombination process with $C=C_1+C_2+C_3$, which were determined from the formula

$$E_{\rm th}^{(i)}(T)=T^2\frac{d\ \ln C_i}{dT},\quad i=1,2,3. \tag{67}$$

The threshold energy for the quasithreshold process is smaller than the three-dimensional value, since the value of the critical width $a_c\approx 1000$ Å is greater than the maximum quantum-well width presented in the figure. For the threshold-free Auger process the threshold energy decreases as the quantum-well width increases, and it becomes negative at a certain width. Such behavior of the threshold energy is associated with the fact that the Auger recombination coefficient C_1 becomes a decreasing function of the temperature for sufficiently wide quantum wells. As the quantum-well width increases, the threshold energy for the overall Auger process tends to the limiting value $E_{\rm th}^{\rm 3D}$, which is marked in the figure.

Let us move on to an examination of the threshold-free Auger process. As we have already noted above, the probability of the threshold-free Auger transition does not have any features as a function of the heavy-hole momentum. Therefore, the coefficient C_1 has a weak nonexponential temperature dependence. It was investigated in detail for the first time in Ref. 8. In addition, the function $C_1(T)$ is nonmonotonic and has a maximum. The presence of this maximum is easily explained. At low temperatures and, accordingly, small longitudinal momenta of the carriers, their wave functions are almost orthogonal, and the value of C_1 is small. As the temperature rises, the characteristic momentum trans-

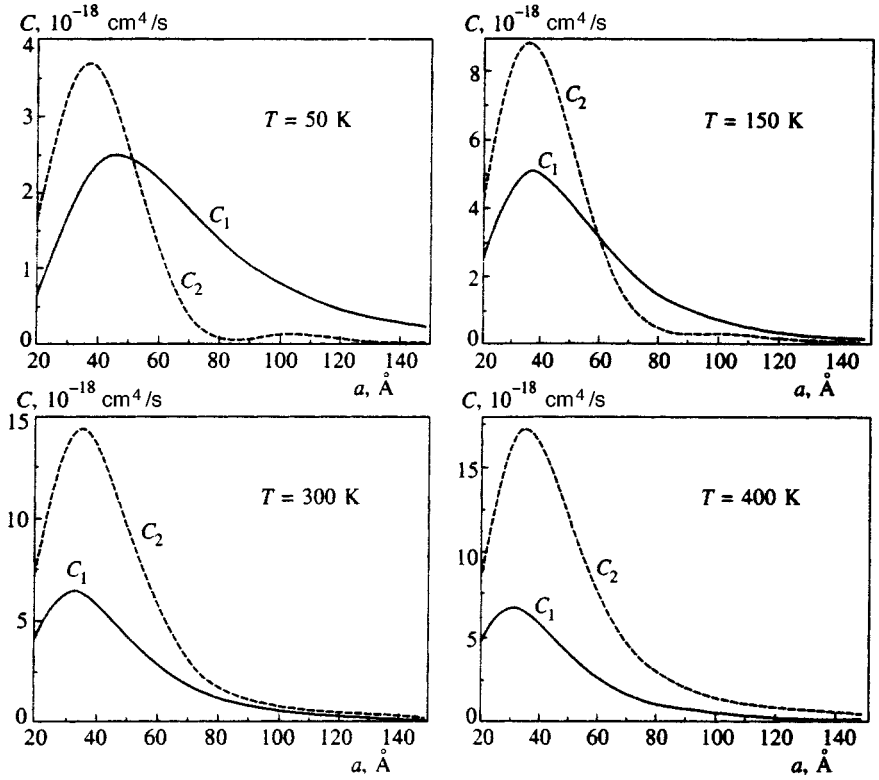


FIG. 3. Dependence of the Auger recombination coefficients C_1 and C_2 for the threshold-free and quasithreshold processes on the quantum-well width a at various temperatures.

fer in the Coulomb interaction increases (it is approximately equal to the heavy-hole thermal momentum). Therefore, at low temperatures the Auger coefficient is an increasing function of the temperature. As the temperature rises further, the Auger recombination coefficient $C_1(T)$ reaches a maximum and begins to decrease, since the long-range Coulomb interaction responsible for the Auger process is weak when the momentum transfer is large. The temperature at which the maximum occurs is easily estimated on the basis of the argument that the size-quantization energy of the holes is equal to the temperature $T \approx \hbar^2 \pi^2 / 2 m_h a^2$. We note that this maximum would not be present, if we assumed that the overlap integral I_{ch} is proportional to the momentum transfer. Such an assumption, which, in our opinion, has no justification for most of the structures investigated, has often been used in the literature (see, for example, Ref. 24) and leads to incorrect expressions for the Auger recombination rate and to incorrect dependences of this rate on the temperature and the quantum-well parameters.

The Auger coefficient C_1 as a function of the quantum-well width a has an abruptly descending form. Depending on which term is predominant in (57), C_1 decreases with increasing a either as $1/a^3$, as $1/a^5$, or as $1/a^7$. In any case, even after multiplication by a^2, C_1 remains a decreasing function of the quantum-well width. Therefore, such a process can be dominant only for fairly narrow quantum wells. For $a \approx 1/\kappa_c$, the coefficient C_1 has a maximum, which is associated with the weak overlap of the carrier wave functions. As the quantum-well width is diminished further, the rate of the threshold-free Auger process decreases smoothly. The simplified expression for C_1 in the CHCC process obtained in Refs. 15 and 16 gives a $1/a^7$ dependence on the quantum-well width.

Figure 3 presents the dependence of the Auger recombination coefficients C_1 and C_2 on the quantum-well width at various temperatures for a model structure based on the compound InGaAsP. First, the dependences of C_1 and C_2 on the quantum-well width have a pronounced maximum. It is noteworthy that the positions of the maxima scarcely depend on the temperature. Second, as is seen from the figure, the relative role of the quasithreshold Auger recombination processes increases with increasing temperature. The threshold process is not indicated in the figure, since the value of C_3 at the quantum-well widths considered is considerably smaller (by several orders of magnitude) than C_1 and C_2. Therefore, the dependence of C_3 on the quantum-well width is presented separately in Fig. 4. We note that the maximum is achieved at a far greater quantum-well width for this process than for the quasithreshold and threshold-free processes. This, in turn, is attributed to the decrease in the value of the threshold energy for the threshold process as the quantum-well width increases (see Fig. 2), rather than to the overlap of the wave functions.

Figure 5 presents the temperature dependence of the total Auger recombination coefficient and the partial contributions of the threshold-free and quasithreshold mechanisms for various quantum-well widths. It is seen from this figure for sufficiently wide quantum wells that the threshold-free Auger process dominates ($C_1 > C_2$)at low temperatures, while the quasithreshold process dominates ($C_2 > C_1$) at high temperatures. Therefore, the temperature dependence of the total Auger recombination coefficient has a characteristic form with a maximum and a minimum. As the quantum-well width is increased, both the maximum and the minimum of the Auger recombination coefficient shift toward lower temperatures, and they vanish in the limit of an infinitely wide

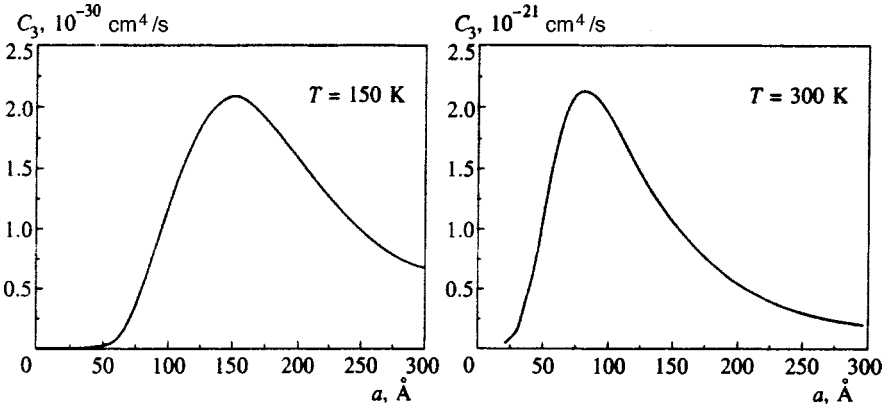


FIG. 4. Dependence of the coefficient C_3 for the threshold Auger recombination process on the quantum-well width at various temperatures.

quantum well. Thus, in the case of a homogeneous semiconductor the Auger recombination coefficient is a monotonic function of the temperature. We note that a Boltzmann distribution of the carriers was used to calculate the temperature dependence of the Auger recombination coefficients. At low temperatures electrons and holes generally obey Fermi–Dirac statistics. Therefore, the mean momenta of the electrons and holes participating in the Auger process depend weakly on the temperature and do not vanish at $T=0$. Figure 6 presents the temperature dependence of the threshold-free Auger recombination coefficient C_1 for various values of the hole Fermi energy for quantum wells with various widths. There are significant differences between the Auger recombination coefficients for the Fermi–Dirac and Boltzmann distributions in the limit $T \ll E_F$, where E_F is the hole Fermi energy. This condition is usually achieved only in the case of very low temperatures, at which the Auger recombination process is not important.

5. DISCUSSION

The analysis that we performed showed that in semiconductor quantum-well structures there are three different Auger recombination mechanisms, viz., threshold, quasithreshold, and threshold-free mechanisms. The Auger recombination coefficients C_1, C_2, C_3 of these processes depend differently both on the temperature and on the parameters of the quantum well: the heights of the heterobarriers for electrons and holes (V_c and V_v) and its width (see Figs. 3–5). In the limit $a \to \infty$ the sum of the quasithreshold and threshold Auger recombination coefficients multiplied by the square of the quantum-well width, $C_2 a^2 + C_3 a^2$, tends to the bulk value of the Auger recombination coefficient C^{3D}, while $C_1 a^2$ tends to zero. For sufficiently narrow quantum wells the value of the two-dimensional Auger recombination coefficient multiplied by a^2 is greater than the three-dimensional value due to the predominance of the threshold-

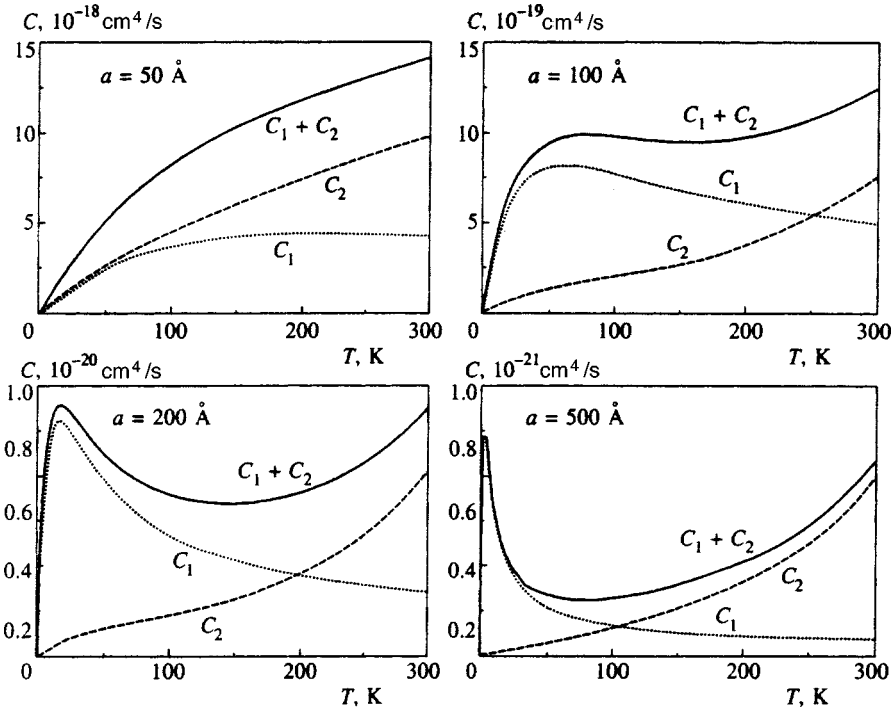


FIG. 5. Temperature dependence of the total Auger recombination coefficient and partial contributions of the threshold-free and quasithreshold mechanisms for various quantum-well widths.

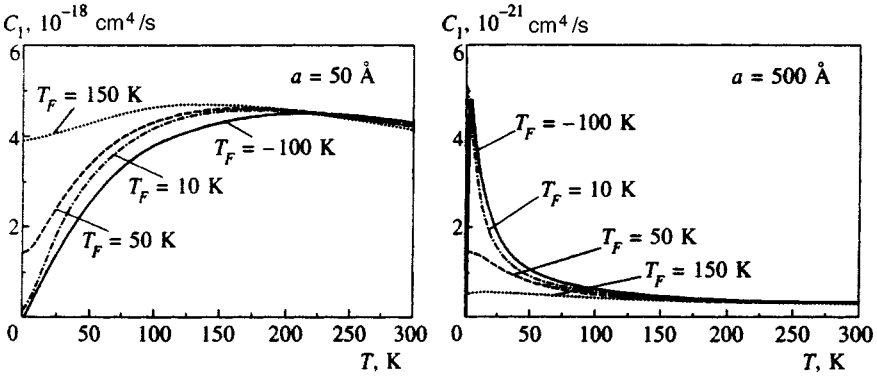


FIG. 6. Comparison of the temperature dependences of the threshold-free Auger recombination coefficient C_1 for various values of the Fermi energy for two different quantum-well widths. In the figure T_F denotes the Fermi energy expressed in degrees. The curve with $T_F=-100$ K corresponds approximately to Boltzmann statistics.

free and quasithreshold Auger recombination processes. Thus, the Auger recombination process is enhanced in quantum wells in comparison to a homogeneous semiconductor. This enhancement is even more significant at low temperatures. Under these conditions the three-dimensional Auger recombination coefficient C^{3D} is small due to the presence of a small exponential multiplier [see (63)]. We note that the entire analysis of the dependence of the Auger recombination coefficients (C_1, C_2, and C_3) on the temperature and the parameters of the quantum well is qualitatively applicable to both the CHCC and CHHS Auger processes to the same extent. However, since we did not specify model quantum-well structures, we illustrated these dependences in the case of the CHCC process. The Auger recombination coefficients C_1, C_2, and C_3 for the CHHS Auger recombination process were analyzed in detail in Ref. 7.

We note that significant suppression of the Auger recombination processes in quantum wells is possible when the conditions $V_c, V_v > E_g$ and $E_2 - E_1 > E_g$ are satisfied (E_1 and E_2 are the energies of the first and second size-quantization levels of the carriers),[25] i.e., if the energy of the excited particle is insufficient for a transition into the continuous spectrum or to the next size-quantization level. To satisfy these equations, structures with deep and narrow quantum wells for both electrons and holes must be fabricated. The presently existing technologies permit the fabrication of such structures on the basis of InAs/AlSb (Ref. 26) and InAs/GaSb/AlSb (Ref. 27). Only the threshold Auger recombination process corresponding to C_3 takes place in such deep quantum wells. This coefficient can be several orders of magnitude smaller than the Auger coefficients for the threshold-free and quasithreshold processes (C_1 and C_2) in shallow quantum wells ($V_c, V_v < E_g$).

Note that the phonon-assisted Auger recombination process also undergoes significant changes for quantum wells.[7] In analogy to the phononless Auger recombination process there are three different mechanisms for the phonon-assisted process ($C^3_{\rm ph}$, $C^2_{\rm ph}$, and $C^1_{\rm ph}$), which correspond to the threshold, quasithreshold, and threshold-free matrix elements for the electron–electron interaction. The first process is fully analogous to the three-dimensional process. However, in narrow quantum wells this process is appreciably weaker than the threshold-free and threshold Auger recombination processes. In the literature just this phonon-assisted process is considered the main Auger recombination process in quantum wells.[28,29] The phonon-assisted Auger recombination processes with the quasithreshold and threshold-free matrix elements for the electron–electron interaction can be resonant processes.[7] At low temperatures they can compete with the phononless Auger recombination processes. However, because there is no strong dependence of the latter on the temperature, such competition is possible at far lower temperatures than in the three-dimensional case. Thus, in narrow quantum wells the phononless (threshold-free) Auger process is dominant in comparison to the phonon-assisted process over a broader temperature range than in the three-dimensional case. As the quantum-well width increases the resonant scattering on the phonons weakens, and we go over to the usual three-dimensional conditions.[7]

It is important to note that at high concentrations of nonequilibrium carriers in a homogeneous semiconductor the phonon-assisted Auger recombination process can be weaker than the Auger process followed by hole–hole scattering, which eliminates the threshold.[7]

6. CONCLUSIONS

The following are the main results of this work.

1. It has been shown that three different Auger recombination mechanisms operate simultaneously in quantum wells: a) a threshold mechanism, b) a quasithreshold mechanism, and c) a threshold-free mechanism. In the limit of an infinitely wide quantum well the first two processes transform into the three-dimensional Auger process, and the threshold-free Auger process tends to vanish.

2. It has been demonstrated for narrow quantum wells that the Auger coefficients of the quasithreshold and threshold-free processes have a weak power-law dependence on the temperature. In addition, their values significantly exceed the three-dimensional coefficient divided by the square of the quantum-well width. At the same time, the coefficient of the two-dimensional threshold Auger process has a larger threshold energy than in the three-dimensional case ($E^{2D}_{\rm th} > E^{3D}_{\rm th}$). The corresponding Auger recombination coefficient is smaller than the three-dimensional expression divided by the square of the quantum-well width.

3. The limiting transition from the two-dimensional Auger recombination process to the three-dimensional process has been accomplished in this work. It has been shown that

$(C_2+C_3)a^2$ transforms as $a\to\infty$ into the three-dimensional Auger coefficient C^{3D} obtained in Ref. 2.

4. It has been shown for sufficient wide quantum wells that the dominant process is the quasithreshold Auger recombination process, whose threshold energy is an increasing function of the quantum-well width (see Fig. 2). The value of the critical quantum-well width, at which the threshold energy for the quasithreshold process becomes comparable to the threshold energy for the three-dimensional Auger process has been shown. The critical quantum-well width a_c is a strong (exponential) function of the temperature. For structures with $E_g\approx 1$ eV at room temperature, a_c can be as large as several thousand angstroms.

In conclusion we thank R. A. Suris and V. I. Perel' for stimulating the writing of this paper and for partially discussing the results. This work was partially support by the Russian Fund for Fundamental Research (Projects 96-02-17952 and 97-02-18151), INTAS (Grant 94-1172), and the ''Physics of Solid-State Nanostructures'' Russian State Program (Projects 97-0003 and 97-1035).

[1)]E-mail: zegrya@theory.ioffe.rssi.ru
[2)]E-mail: tol@theory.ioffe.rssi.ru

[3)]Although the functions $\varphi_c(x)$ and $\varphi_h(x)$ do not have a definite parity, their scalar product has a fixed parity.

[4)]We assume that, as is often the case, $V_c, V_v < E_g$.

[5)]We assume that the carriers have a Boltzmann distribution function. Similar arguments are also valid for Fermi–Dirac statistics in the case of $q_F \ll q_{\rm th}$, where q_F is the two-dimensional heavy-hole Fermi momentum.

[1] A. R. Beattie and P. T. Landsberg, Proc. R. Soc. London, Ser. A **249**, 16 (1959).
[2] B. L. Gel'mont, Zh. Éksp. Teor. Fiz. **75**, 536 (1978) [Sov. Phys. JETP **48**, 268 (1978)].
[3] A. Haug, J. Phys. C: Solid State Phys. **16**, 4159 (1983).
[4] M. Takeshima, Phys. Rev. B **26**, 917 (1982); **28**, 2039 (1983).
[5] G. P. Agrawal and N. K. Dutta, *Long-Wavelength Semiconductor Lasers*, Van Nostrand Reinhold, New York (1993).
[6] M. W. Aĭdaraliev, G. G. Zegrya, N. V. Zotova, B. A. Matveev, S. A. Karandashev, N. N. Stus', and G. N. Talalakin, Fiz. Tekh. Poluprovodn. **26**, 249 (1992) [Sov. Phys. Semicond. **26**, 138 (1992)].
[7] G. G. Zegrya and A. S. Polkovnikov, submitted to Phys. Rev. B.
[8] G. G. Zegrya and V. A. Kharchenko, Zh. Éksp. Teor. Fiz. **101**, 327 (1992) [Sov. Phys. JETP **74**, 173 (1992)].
[9] *Quantum Well Lasers*, P. S. Zory, Jr. (ed.), Academic Press, San Diego, California (1993).
[10] R. I. Taylor, R. A. Abram, M. G. Burt, and C. Smith, IEE Proc. **132**, 364 (1985).
[11] W. W. Lui, T. Yamanaka, Y. Yoshikuni, S. Seki, and K. Yoko, Phys. Rev. B **48**, 8814 (1993).
[12] N. K. Dutta, J. Appl. Phys. **54**, 1236 (1983).
[13] G. G. Zegrya, V. P. Evtikhiev, I. B. Kudryashov, and V. E. Tokranov, in *23rd International Symposium on Compound Semiconductors (ISCS-23), St. Petersburg, Russia, 23–27 September (Institute of Physics Conference Series No. 155)*, Institute of Physics Publishing, London (1996), p. 795.
[14] R. I. Taylor, R. A. Abram, M. G. Burt, and C. Smith, Semicond. Sci. Technol. **5**, 90 (1990).
[15] M. I. Dyakonov and V. Yu. Kachorovskii, Phys. Rev. B **49**, 17130 (1994).
[16] G. G. Zegrya, A. D. Andreev, N. A. Gun'ko, and E. V. Frolushkina, Proc. SPIE **2399**, 307 (1995).
[17] E. O. Kane, J. Phys. Chem. Solids **1**, 249 (1957).
[18] P. C. Sercel and K. J. Vahala, Phys. Rev. B **42**, 3690 (1990).
[19] R. A. Suris, Fiz. Tekh. Poluprovodn. **20**, 2008 (1986) [Sov. Phys. Semicond. **20**, 1258 (1986)].
[20] M. G. Burt, J. Phys.: Condens. Matter **4**, 6651 (1992).
[21] B. A. Foreman, Phys. Rev. B **49**, 1757 (1994).
[22] G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors*, Wiley, New York 1975; Nauka, Moscow, 1972.
[23] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, Nauka, Moscow, 1989; Pergamon Press, Oxford, 1977.
[24] J. Wang, P. von Allmen, J.-P. Leburton, and K. J. Linden, IEEE J. Quantum Electron. **QE-31**, 864 (1995).
[25] G. G. Zegrya, in *Antimonide Related Strained Layer Heterostructures*, M. O. Manasreh (ed.), Gordon and Breach, Amsterdam (1997).
[26] S. Ideshita, A. Furukawa, Y. Mochiziki, and M. Mizuta, Appl. Phys. Lett. **60**, 2594 (1992).
[27] M. Sweeny and J. Xu, Appl. Phys. Lett. **54**, 546 (1989).
[28] E. P. O'Reilly and M. Silver, Appl. Phys. Lett. **63**, 3318 (1993).
[29] E. P. O'Reilly and A. R. Adams, IEEE J. Quantum Electron. **QE-30**, 366 (1994).

Translated by P. Shelnitz

Influence of structural disorder on the current–voltage characteristic of a quasi-one-dimensional tunnel junction

V. Ya. Kirpichenkov

Novocherkassk State Technical University, 346400 Novocherkassk, Russia
(Submitted 14 November 1996; resubmitted 21 November 1997)
Zh. Éksp. Teor. Fiz. **113**, 1522–1530 (April 1998)

The influence of the subbarrier impurity scattering of tunneling electrons on the current–voltage characteristic of a quasi-one-dimensional insulator layer with weak structural disorder (a small impurity concentration) is considered in the one-electron approximation at $T=0$. An expansion in powers of the impurity concentration gives the form of the current–voltage characteristic and the conditions for small mesoscopic fluctuations of the static tunneling conductance of such a layer in the cases of resonant and nonresonant tunneling. © *1998 American Institute of Physics.* [S1063-7761(98)02804-2]

1. INTRODUCTION

Real tunnel junctions must contain some defects or impurities, at least in small concentrations. This naturally raises the question of how weak disorder affect the parameters of such junctions. The single-particle problem of the tunneling transmission of particles through a homogeneous potential barrier with identical scattering centers, i.e., impurities, randomly arranged within it was examined in the quasi-one-dimensional and three-dimensional cases at low impurity concentrations in Refs. 1 and 2.

This paper, which is closely related to Refs. 1 and 2, considers how the subbarrier impurity scattering of tunneling electrons affect the current–voltage characteristic of a quasi-one-dimensional barrier layer and the magnitude of the mesoscopic fluctuations of its static tunneling conductance in the cases of both nonresonant and resonant tunneling. Lower bounds on the transverse dimensions of the barrier layer are obtained from the requirement that these fluctuations be small.

2. MODEL: BASIC RELATIONS

Let us consider an N–I–N sandwich consisting of two identical layers of a normal metal (N) separated by a layer of an insulator (I) of thickness L and area S that is impregnated with impurities within the following model. For the conduction electrons in the N layers we adopt the three-dimensional isotropic quadratic dispersion law $\varepsilon=k^2$ $(\hbar^2/2m=1)$ with the Fermi energy ε_F, and for the barrier potential $U(\mathbf{r})$ in the region $0\leqslant x\leqslant L$, which is occupied by the insulator, in the absence of a potential difference on the barrier $(V=0)$ we take the model[1]

$$U(\mathbf{r})=U_0+v(\mathbf{r}), \quad \mathbf{r}=(x,\boldsymbol{\rho}), \quad 0\leqslant x\leqslant L. \tag{1}$$

A homogeneous insulator without impurities would have the potential $U_0=\text{const}>\varepsilon_F$, and disorder is generated in the model of structural disorder in the simplest case by the random potential $v(\mathbf{r})$ of identical impurities randomly distributed throughout the insulator layer:

$$v(\mathbf{r})=\sum_{0\leqslant x_j\leqslant L} u(\mathbf{r}-\mathbf{r}_j), \tag{2}$$

where the points $\mathbf{r}_j$ are macroscopically uniformly distributed with the density n in the layer at $0\leqslant x\leqslant L$, and $u(\mathbf{r})$ is the local single-impurity potential with a radius of action r_0. In the case under consideration of small impurity concentrations, the following relations are assumed to hold for the characteristic parameters with the dimensions of length:

$$r_0\ll\alpha_F^{-1}\ll n^{-1/3}<L, \tag{3}$$

where $\alpha_F^{-1}=(U_0-\varepsilon_F)^{-1/2}$ is the characteristic damping length of an electronic state with the energy ε_F in a homogeneous barrier.

In the quasi-one-dimensional case considered here it is assumed that the insulator layer is composed of $M=S/b^2$ insulator filaments (with impurities) with the transverse dimension b. In addition, the mean free path with respect to electron jumps from filament to filament is assumed to be large in comparison to the filament length L, so that the tunneling dynamics in each filament is independent of the dynamics in the other filaments. We shall first consider just this case as the simplest case for theoretical analysis and, in addition, as a case of interest in itself, since such a quasi-one-dimensional situation is possible, for example, in the case of strong anisotropy of the effective mass of the tunneling electrons in three-dimensional barriers or in the case of the elastic tunneling of electrons through a bundle of homopolymer molecules with impurities.

At a small bias $V\ll\varepsilon_F, U_0-\varepsilon_F$ (the electron charge $e=1$) and the temperature $T=0$ and under the assumption that the tunneling dynamics in different filaments is independent, the formulas for the tunneling current $\langle i(V)\rangle$, the tunnel conductance of the layer $\langle g(V)\rangle$ (per filament), and their relative root-mean-square fluctuations $\langle\delta^2\rangle^{1/2}$ can be represented in the form

$$\langle i(V)\rangle=\frac{1}{M}\sum_{k=1}^{M}\int_{\varepsilon_F}^{\varepsilon_F+V}\langle g_k(\varepsilon)\rangle d\varepsilon=\int_{\varepsilon_F}^{\varepsilon_F+V}\langle g(\varepsilon)\rangle d\varepsilon,$$

$$\langle g(V)\rangle=\frac{1}{V}\langle i(V)\rangle, \tag{4}$$

$$\langle\delta^2\rangle^{1/2}=\left[\frac{1}{M}\frac{\int_{\varepsilon_F}^{\varepsilon_F+V}[\langle g(\varepsilon)g(\varepsilon')\rangle-\langle g(\varepsilon)\rangle\langle g(\varepsilon')\rangle]d\varepsilon\,d\varepsilon'}{\int_{\varepsilon_F}^{\varepsilon_F+V}\langle g(\varepsilon)\rangle\langle g(\varepsilon')\rangle d\varepsilon\,d\varepsilon'}\right]^{1/2}. \tag{5}$$

For filaments of arbitrary thickness

$$g(\varepsilon)\equiv g(\varepsilon,G_m)=\int D(\varepsilon,\mathbf{q},\boldsymbol{\rho};G_m)\frac{d^2q}{(2\pi)^2}d^2\rho, \tag{6}$$

where $D(\varepsilon,\mathbf{q},\boldsymbol{\rho},G_m)$ is the tunneling transmission coefficient of a filament with a random impurity configuration $G_m=\{\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_m\}$ for electrons with the energy ε having a fixed transverse component of the momentum $\mathbf{q}$ at the ''entrance'' to the barrier and a fixed transverse coordinate $\boldsymbol{\rho}$ at the ''exit.'' The integration over $\mathbf{q}$ is performed over the range $0\leqslant q^2\leqslant\varepsilon$ $(q>0)$, and the integration over $\boldsymbol{\rho}$ is performed over the cross section of the filament. The averaging in (4) and (5) is carried out over the impurity configurations G_m:

$$\langle g(\varepsilon)\rangle=\sum_{m=0}^{N}C_N^m\frac{(V-V_1)^{N-m}}{V^N}\int_{(V_1)}g(\varepsilon,G_m)dG_m,$$

$$\langle g(\varepsilon)g(\varepsilon')\rangle=\sum_{m=0}^{N}C_N^m\frac{(V-V_1)^{N-m}}{V^N}\times\int_{(V_1)}g(\varepsilon,G_m)g(\varepsilon',G_m)dG_m, \tag{7}$$

where $dG_m=d\mathbf{r}_1d\mathbf{r}_2\ldots d\mathbf{r}_m$, $V_1=b^2L$, $V=SL=MV_1$, $N=nV$ is the total number of impurity atoms in the barrier layer, n is the three-dimensional impurity density, and $C_N^m=N(N-1)\times\ldots\times(N-m+1)/m!$.

For sufficiently thin filaments, such that $b\ll\alpha_F^{-1}$ and $M\gg1$, Eqs. (6) and (7) lead to the one-dimensional relations

$$\langle g(\varepsilon)\rangle\sim b^2\varepsilon\langle D(\varepsilon)\rangle,$$

$$\langle g(\varepsilon)g(\varepsilon')\rangle\sim b^4\varepsilon\varepsilon'\langle D(\varepsilon)D(\varepsilon')\rangle, \tag{8}$$

where $D(\varepsilon)=D(\varepsilon,\mathbf{q}=0,\boldsymbol{\rho}=0;\Gamma_m)$ is the tunneling transmission coefficient of a strictly homogeneous filament with the impurity configuration $\Gamma_m=\{x_1,x_2,\ldots,x_m\}$, which corresponds to the one-dimensional barrier potential

$$U(x)=U_0+\sum_{0\leqslant x_i\leqslant L}u(x-x_i). \tag{9}$$

In the one-dimensional case the impurities can be regarded as point scatterers[1] and the local single-impurity potential can be taken in the δ-function form:

$$u(x)=k_0\delta(x). \tag{10}$$

The averages in (8) are taken over the one-dimensional configurations Γ_m:

$$\langle D(\varepsilon)\rangle=\sum_{m=0}^{\infty}p_m\frac{1}{\Delta\Gamma_m}\int D(\varepsilon,\Gamma_m)d\Gamma_m,$$

$$\langle D(\varepsilon)D(\varepsilon')\rangle=\sum_{m=0}^{\infty}p_m\frac{1}{\Delta\Gamma_m}\times\int D(\varepsilon,\Gamma_m)D(\varepsilon',\Gamma_m)d\Gamma_m, \tag{11}$$

where

$$p_m=\frac{(n_1L)^m}{m!}e^{-n_1L}$$

is the probability that m impurities are found in a system with a linear impurity concentration $n_1=nb^2$ in a filament of length L, $d\Gamma_m=dx_1\ldots dx_m$ is a volume element of the configuration phase space, and

$$\Delta\Gamma_m=\int d\Gamma_m=\int_{(0<x_1<x_2<\ldots<x_m<L)}dx_1dx_2\ldots dx_m=\frac{L^m}{m!}$$

is the total volume of the configuration phase space of a system of m impurities in a filament of length L.

The tunneling transmission coefficients $D(\varepsilon,\Gamma_m)$ appearing in (11) are found by solving the tunneling problem for the one-dimensional Schrödinger equation with potential (10).[1]

The calculations of the current–voltage characteristic $\langle i(V)\rangle$ (4) and the root-mean-square fluctuation $\langle\delta^2(V)\rangle^{1/2}$ (5) reduce to calculations of the mean $\langle g(\varepsilon)\rangle$ and the correlator $\langle g(\varepsilon)g(\varepsilon')\rangle$, which can be expressed in terms of $\langle D(\varepsilon)\rangle$ and $\langle D(\varepsilon)D(\varepsilon')\rangle$ in accordance with (8) in the quasi-one-dimensional case, followed by their integration according to (4) and (5).

3. CURRENT–VOLTAGE CHARACTERISTIC OF A QUASI-ONE-DIMENSIONAL BARRIER IN THE CASE OF NONRESONANT TUNNELING

Let us first consider the case of nonresonant tunneling, which is observed for all $\varepsilon_F<U_0$, if the impurities are repulsive [$k_0>0$ in (10)], and for $\varepsilon_F<\varepsilon_g$, where ε_g is the true lower boundary of the single-particle electron energy spectrum in an infinite filament with impurities, if they are attractive ($k_0<0$).[1,2]

In the case of nonresonant tunneling $\langle D(\varepsilon)\rangle$ and $\langle D(\varepsilon)D(\varepsilon')\rangle$, as well as $\langle g(\varepsilon)\rangle$ and $\langle g(\varepsilon)g(\varepsilon')\rangle$, are smooth functions of the arguments ε and ε'. Therefore, for $V\ll\varepsilon_F,U_0-\varepsilon_F$ it follows from (4) and (5) that

$$\langle i(V)\rangle=\langle g(\varepsilon_F)\rangle V, \tag{12}$$

$$\langle\delta^2\rangle^{1/2}=\left[\frac{1}{M}\frac{\langle g^2(\varepsilon_F)\rangle-\langle g(\varepsilon_F)\rangle^2}{\langle g(\varepsilon_F)\rangle^2}\right]^{1/2}, \tag{13}$$

where, in accordance with (8),

$$\langle g(\varepsilon_F)\rangle \sim b^2\varepsilon_F\langle D(\varepsilon_F)\rangle, \quad \langle g^2(\varepsilon_F)\rangle \sim b^4\varepsilon_F^2\langle D^2(\varepsilon_F)\rangle. \tag{14}$$

To find the values of $\langle g(\varepsilon_F)\rangle$ and $\langle g^2(\varepsilon_F)\rangle$ appearing in (12) and (13) we must calculate $\langle D(\varepsilon_F)\rangle$ and $\langle D^2(\varepsilon_F)\rangle$ in accordance with (14), which we do below to exponential accuracy. For this purpose, in the limit of a small impurity density under discussion here ($n_1\alpha_F^{-1}\ll 1$) we write virial expansions[2] for the corresponding dimensionless decrements

$$\eta_1 = -\frac{1}{\mathscr{L}}\ln\langle D(\varepsilon_F)\rangle, \quad \eta_2 = -\frac{1}{\mathscr{L}}\ln\langle D^2(\varepsilon_F)\rangle,$$

$$\mathscr{L} \equiv \alpha_F L \gg 1, \tag{15}$$

in powers of $c = n_1\alpha_F^{-1} \ll 1$, confining ourselves, for simplicity, to the first-order terms:

$$\begin{aligned}
\eta_1 &= -\frac{1}{\mathscr{L}}\ln D^{(0)}(\varepsilon_F) - \frac{c}{\mathscr{L}} \\
&\quad \times \int_0^{\mathscr{L}} \left[\frac{D(\varepsilon_F,t_1)}{D^{(0)}(\varepsilon_F)} - 1\right] dt_1 + \ldots, \\
\eta_2 &= -\frac{2}{\mathscr{L}}\ln D^{(0)}(\varepsilon_F) - \frac{c}{\mathscr{L}} \\
&\quad \times \int_0^{\mathscr{L}} \left[\frac{D^2(\varepsilon_F,t_1)}{D^{(0)2}(\varepsilon_F)} - 1\right] dt_1 + \ldots,
\end{aligned} \tag{16}$$

where $\mathscr{L} = \alpha_F L$ is the dimensionless length of the filament, $t_1 = \alpha_F x_1$ is the dimensionless coordinate of an impurity at the point x_1, and $D^{(0)}(\varepsilon_F)$ and $D(\varepsilon_F, t_1)$ are the transmission coefficients of the one-dimensional barrier without impurities and with one impurity at the point x_1, respectively.

For nonresonant values of ε_F and $\mathscr{L} \gg 1$ we have[1]

$$D^{(0)}(\varepsilon_F) = e^{-2\mathscr{L}}, \quad \frac{D(\varepsilon_F,t_1)}{D^{(0)}(\varepsilon_F)} = (1-\mu_F)^2, \tag{17}$$

where $\mu_F = k_0(2\alpha_F + k_0)^{-1}$ is the subbarrier scattering amplitude (for ε_F) in the single-impurity potential (10). In the nonresonant case presently under consideration $|\mu_F| \sim 1$.

Substituting (17) into (16), we find

$$\begin{aligned}
\eta_1 &= 2 + c\mu_F(2-\mu_F), \\
\eta_2 &= 4 + 2c\mu_F(2-\mu_F) - c\mu_F^2(2-\mu_F)^2.
\end{aligned} \tag{18}$$

Taking into account (18), from (12)–(15)

$$\langle g(\varepsilon_F)\rangle \sim b^2\varepsilon_F \exp(-\eta_1\mathscr{L}), \tag{19}$$

$$\langle \delta^2\rangle^{1/2} \sim M^{-1/2}\exp(\eta\mathscr{L}/2), \tag{20}$$

where $\eta = \eta_2 - 2\eta_1 = c\mu_F^2(2-\mu_F)^2$ and $\eta\mathscr{L} \sim c\mathscr{L} > 1$.

A bound on the parameters of a quasi-one-dimensional insulator layer follows from (20). It ensures (in an ensemble of identical samples) that the relative fluctuations of the tunnel conductance will remain small, $\langle \delta^2\rangle^{1/2} \ll 1$, at nonresonant values of ε_F upon passage from sample to sample:

$$\sqrt{M} \gg \exp(\eta\mathscr{L}/2), \quad M = S/b^2. \tag{21}$$

When this condition is satisfied, real self-averaging of the static tunnel conductance (4) of a quasi-one-dimensional, structurally disordered insulator layer takes place in the nonresonant case.

4. CURRENT–VOLTAGE CHARACTERISTIC OF A QUASI-ONE-DIMENSIONAL BARRIER IN THE CASE OF RESONANT TUNNELING

Resonant tunneling takes place when ε_F falls within the region of the discrete spectrum in the associated problem for an infinite ($-\infty < x < \infty$) filament with attractive ($k_0 < 0$) impurities located in the former layer ($0 < x_j < L$).[1,3]

The resonant effects reach their greatest values in situations in which the energy ε_F is close to the subbarrier scattering band of amplitude $\mu(\varepsilon)$ at ε_0 ($\varepsilon_0 = U_0 - k_0^2/4$), which is the energy of a bound electron state in one impurity center (the energy of a local single-impurity level) in the associated spectral problem mentioned above. Therefore, below we shall consider the case where ε_F coincides with ε_0 when $V = 0$.

To find $\langle i(V)\rangle$ and $\langle \delta^2(V)\rangle$ we must first calculate the means $\langle D(\varepsilon)\rangle$ and $\langle D(\varepsilon)D(\varepsilon')\rangle$ (11) at resonant energies.

Let us consider the calculation of $\langle D(\varepsilon)\rangle$. For further analysis, it is convenient to represent $\langle D(\varepsilon)\rangle$ in the following form:

$$\langle D(\varepsilon)\rangle = \sum_{m=0}^{\infty} p_m D_m(\varepsilon), \tag{22}$$

$$D_m(\varepsilon) = \frac{1}{\Delta\Gamma_m}\Omega_m(\varepsilon),$$

$$\Omega_m(\varepsilon) = \int_{\{\Delta\Gamma_m\}} D(\varepsilon,\Gamma_m)\, d\Gamma_m, \tag{23}$$

where $D_m(\varepsilon)$ is the transmission of the filament averaged over the configurations with a fixed number of impurities m.

When the energy ε is close to ε_0, there are always impurity configurations $\Gamma_m^{\mathrm{res}}(\varepsilon)$, called resonant configurations at the energy ε, in which $D(\varepsilon, \Gamma_m^{\mathrm{res}}(\varepsilon)) \sim 1$ holds, while the transmission is exponentially small for all the remaining (nonresonant) configurations.[1] Accordingly, at each energy ε the phase space of a system of m impurities in a filament of length L is factorized into a set of resonant and nonresonant regions. When $\mathscr{L} \gg 1$ holds and ε is close to ε_0, the main contribution to the integral $\Omega_m(\varepsilon)$ is made by the resonant regions of the phase space [see (27) and (28) below]:

$$\begin{aligned}
\Omega_m(\varepsilon) &= \int_{\{\Delta\Gamma_m\}} D(\varepsilon,\Gamma_m)\, d\Gamma_m \\
&\sim \int_{\{\Delta\Gamma_m^{\mathrm{res}}(\varepsilon)\}} d\Gamma_m = \Delta\Gamma_m^{\mathrm{res}}(\varepsilon),
\end{aligned} \tag{24}$$

where $\Delta\Gamma_m^{\mathrm{res}}(\varepsilon)$ is the resonant (at the energy ε) phase volume of an m-impurity system in a filament of length L.

To find $\langle i(V)\rangle$ we must calculate the integrals for $m \geqslant 1$ in accordance with (4), (8), and (22)–(24) ($V \ll \varepsilon_F$)

$$\mathscr{J}_m(V)=\int_{\varepsilon_F}^{\varepsilon_F+V}\varepsilon\Omega_m(\varepsilon)d\varepsilon\sim\int_{\varepsilon_F}^{\varepsilon_F+V}\varepsilon\Delta\Gamma_m^{\rm res}(\varepsilon)d\varepsilon$$
$$\sim\varepsilon_F\int_{\varepsilon_F}^{\varepsilon_F+V}\Delta\Gamma_m^{\rm res}(\varepsilon)d\varepsilon. \tag{25}$$

Only the intermediate asymptotes found in Ref. 1 are known for the function $\Delta\Gamma_m^{\rm res}(\varepsilon)$, which specifies the amplitude and shape of the resonant transmission maximum in the vicinity of the impurity level at ε_0, for $m>1$. Therefore, to calculate the overall structure of the current–voltage characteristic and obtain numerical estimates, we confine ourselves to an approximation for $\Delta\Gamma_m^{\rm res}(\varepsilon)$, which correctly takes into account the two main parameters of the real function, viz., its characteristic width and height:

$$\Omega_m(\varepsilon)\sim\Delta\Gamma_m^{\rm res}(\varepsilon)\sim\begin{cases}\alpha_F^{-m}, & |\varepsilon-\varepsilon_0|\leqslant\gamma_m/2,\\ 0, & |\varepsilon-\varepsilon_0|>\gamma_m/2,\end{cases} \tag{26}$$

where $\gamma_m\sim k_0^2\exp(-\mathscr{S}/m)$ is the characteristic energy width of the resonant transmission band of a barrier of length $L=\mathscr{S}/\alpha_F$ with m impurities.

Returning to the estimate (24), we note that it is valid provided the contribution of the resonant regions of the phase space to the integral $\Omega_m(\varepsilon)$ is much greater than the contribution of the nonresonant regions:

$$\Delta\Gamma_m^{\rm res}(\varepsilon)\gg e^{-2\mathscr{S}}\Delta\Gamma_m, \quad \Delta\Gamma_m=L^m/m!, \quad \mathscr{S}=\alpha_F L. \tag{27}$$

In the resonant transmission band we have $|\varepsilon-\varepsilon_0|\leqslant\gamma_m/2$. With consideration of (26), this inequality takes the form

$$e^{-2\mathscr{S}}\mathscr{S}^m/m!\ll 1. \tag{28}$$

For $\mathscr{S}\gg 1$ (typical values are $\mathscr{S}\sim 10$), this condition holds for all $m\geqslant 1$.

Using (26) for the integral $\mathscr{J}_m(V)$ (25), we obtain the estimate

$$\mathscr{J}_m(V)\sim\varepsilon_F\alpha_F^{-m}\begin{cases}V, & |V|\leqslant\gamma_m/2,\\ (\gamma_m/2)\,\mathrm{sign}\ V, & |V|>\gamma_m/2.\end{cases} \tag{29}$$

Finally, taking into account (4), (8), (11), and (22)–(26), we obtain the current–voltage characteristic in the following form:

$$\langle i(V)\rangle=i_0(V)+\sum_{m=1}^{\infty}c^m e^{-c\mathscr{S}}\varphi_m(V), \tag{30}$$

where $i_0(V)\sim b^2\varepsilon_F\exp[-(2+c)\mathscr{S}]V$ is the contribution from the configurations with $m=0$ (the ''empty'' barrier) to the mean current,

$$\varphi_m(V)\sim b^2\alpha_F^m\mathscr{J}_m(V)$$
$$\sim b^2\varepsilon_F\begin{cases}V, & |V|\leqslant\gamma_m/2,\\ (\gamma_m/2)\,\mathrm{sign}\ V, & |V|>\gamma_m/2,\end{cases} \tag{31}$$

$c=n_1\alpha_F^{-1}$, and $\gamma_m\sim k_0^2\exp(-\mathscr{S}/m)$.

Thus, at small impurity concentrations ($c\ll 1$) the current–voltage characteristic $\langle i(V)\rangle$ (30) is represented in the form of a series in powers of the concentration, whose mth term gives the contribution of the m-impurity resonances to the mean tunneling current.

In the case considered here we are interested in the range of parameters in which the total contribution of the resonances with $m\geqslant 1$ significantly exceeds the contribution of the empty barrier $i_0(V)$:

$$\sum_{m=1}^{\infty}c^m e^{-c\mathscr{S}}\varphi_m(V)\gg i_0(V). \tag{32}$$

This inequality demarcates an extensive region in the (V,c) plane (for fixed $\mathscr{S}$, k_0) in which the resonant impurity tunneling current is decisive. For example, if we confine ourselves to only the term with $m=1$ on the left-hand side of (32), we have (taking into account $c\ll 1$)

$$1\gg c\gg\frac{i_0(V)}{\varphi_1(V)}e^{c\mathscr{S}}\sim\begin{cases}e^{-2\mathscr{S}}, & |V|\leqslant\gamma_1/2,\\ (2|V|/\gamma_1)e^{-2\mathscr{S}}, & |V|>\gamma_1/2,\end{cases} \tag{33}$$

where $\gamma_1\sim k_0^2\exp(-\mathscr{S})$ and $|V|\ll\varepsilon_F$. For example, for $\mathscr{S}=10$ and the characteristic values $k_0^2\sim\varepsilon_F\sim 10$ eV the range of concentrations (33) covers several orders of magnitude at $|V|\leqslant 1$ V ($10^{-5}\ll c\ll 1$). On the other hand, for any fixed concentration in this range, the contribution of the first resonance is formally comparable to $i_0(V)$ at such large values of $|V|\sim c\gamma_1\exp(2\mathscr{S})\sim c\varepsilon_F\exp(\mathscr{S})>\varepsilon_F$ that they are already far outside the range of applicability of the approximation used ($|V|\ll\varepsilon_F$).

Thus, in the range of parameters of interest to us, the quantity $i_0(V)$ in (30) can be omitted, and the current–voltage characteristic can be written in the form

$$\langle i(V)\rangle=\sum_{m=1}^{\infty}c^m e^{-c\mathscr{S}}\varphi_m(V). \tag{34}$$

Differentiating $\langle i(V)\rangle$ (34), we find

$$\frac{d\langle i\rangle}{dV}=\sum_{m=1}^{\infty}c^m e^{-c\mathscr{S}}\frac{d\varphi_m}{dV}, \tag{35}$$

$$\frac{d^2\langle i\rangle}{dV^2}=\sum_{m=1}^{\infty}c^m e^{-c\mathscr{S}}\frac{d^2\varphi_m}{dV^2}, \tag{36}$$

where in the approximation (31)

$$\frac{d\varphi_m}{dV}\sim b^2\varepsilon_F\begin{cases}1, & |V|\leqslant\gamma_m/2,\\ 0, & |V|>\gamma_m/2,\end{cases} \tag{37}$$

$$\frac{d^2\varphi_m}{dV^2}\sim -b^2\varepsilon_F\,\mathrm{sign}(V)\,\delta\left(|V|-\frac{\gamma_m}{2}\right). \tag{38}$$

Of course, the functions $\varphi_m(V)$ found in this approximation and its derivatives specify only the ''skeletons'' of the functions $\langle i(V)\rangle$, $d\langle i\rangle/dV$, and $d^2\langle i\rangle/dV^2$, which reflect only the most features of the real behavior. These functions, are, of course, smoothed, and the δ functions in (38) spread out to form peaks of finite width and height. The qualitative character of these functions is shown in Fig. 1.

The function $\langle i(V)\rangle$ is significantly nonlinear [because of the significant nonlinearity of the $\varphi_m(V)$ given by (31)]

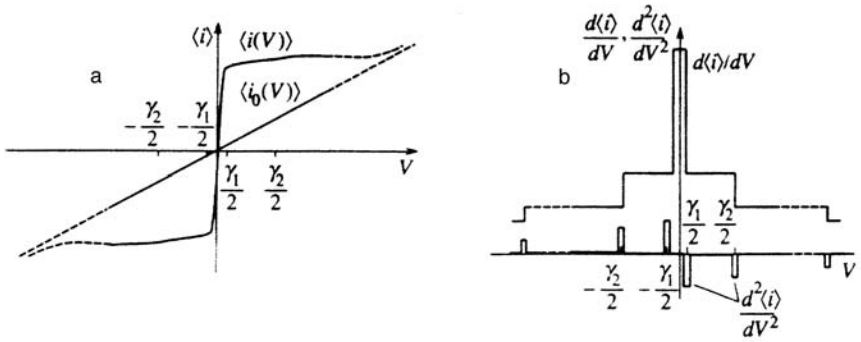


FIG. 1. Qualitative forms of $\langle i(V)\rangle$ (a) and of $d\langle i\rangle/dV$ and $d^2\langle i\rangle/dV^2$ (b).

and thus differs radically from the dependence in a barrier without impurities (or in the case of nonresonant tunneling in a barrier with impurities). The resonant current in the range of parameters considered (33) is several orders of magnitude greater than the current in an empty barrier. The m-impurity resonances are displayed in the plot of $d\langle i\rangle/dV$ in the form of a plateau with the characteristic width $\gamma_m \sim k_0^2 \times \exp(-\mathscr{L}/m)$, and peaks appear on the plot of $d^2\langle i\rangle/dV^2$ at $|V_m| \sim \gamma_m/2$.

If, as above, $\mathscr{L}=10$ and $k_0^2 \sim \varepsilon_F \sim 10$ eV are taken for estimates, then for the characteristic energy widths of the first three resonances we have

$$\gamma_1 \sim 10^{-3}\ \text{eV}, \quad \gamma_2 \sim 10^{-1}\ \text{eV}, \quad \gamma_3 \sim 1\,\text{eV}. \tag{39}$$

This means that these resonances can be observed experimentally only at the temperatures

$$T_1 \ll 10\ \text{K}, \quad T_2 \ll 10^3\ \text{K}, \quad T_3 \ll 10^4\ \text{K}, \tag{40}$$

respectively, and that they should be manifested on the voltage scales

$$V_1 \sim 10^{-3}\ \text{V}, \quad V_2 \sim 10^{-1}\ \text{V}, \quad V_3 \sim 1\ \text{V}. \tag{41}$$

In accordance with (5), (8), and (11), the calculation of $\langle \delta^2 \rangle$ requires calculations not only of $\mathscr{J}_m(V)$ for $m \geqslant 1$, but also of integrals of the form

$$K_m(V) = \int d\Gamma_m \left(\int_{\varepsilon_F}^{\varepsilon_F + V} \varepsilon D(\varepsilon, \Gamma_m) d\varepsilon \right)^2. \tag{42}$$

Within the approximations that were used to calculate the $\mathscr{J}_m(V)$, we obtain

$$K_m(V) \sim \varepsilon_F^2 \alpha_F^{-m} \begin{cases} V^2, & |V| \leqslant \gamma_m/2, \\ \gamma_m^2/4, & |V| > \gamma_m/2. \end{cases} \tag{43}$$

Then, it follows from (5) that

$$\langle \delta^2 \rangle^{1/2} = \frac{1}{\sqrt{M}} \left[\frac{\sum_{m=1}^{\infty} c^m e^{-c\mathscr{L}} \chi_m(V)}{\left(\sum_{m=1}^{\infty} c^m e^{-c\mathscr{L}} \varphi_m(V) \right)^2} - 1 \right]^{1/2}, \tag{44}$$

where

$$\chi_m(V) \sim b^4 \alpha_F^m K_m(V) \sim b^4 \varepsilon_F^2 \begin{cases} V^2, & |V| \leqslant \gamma_m/2, \\ \gamma_m^2/4, & |V| > \gamma_m/2. \end{cases} \tag{45}$$

Leaving only the terms with $m=1$ in the sums (44) for an estimate, from the condition $\langle \delta^2 \rangle^{1/2} \ll 1$ we obtain the following bounds on the tunnel-junction parameters:

$$\sqrt{M} \gg \frac{1}{\sqrt{c}} \exp\left(\frac{c\mathscr{L}}{2} \right), \quad M = \frac{S}{b^2}, \tag{46}$$

which ensure real self-averaging of the static tunnel conductance of a quasi-one-dimensional insulator layer with structural disorder in the resonant case.

We thank A. M. Kosevich, L. A. Pastur, G. A. Gogadze, A. N. Omel'yanchuk, and A. S. Rozhavskiĭ for discussing the questions addressed here.

[1] I. M. Lifshits and V. Ya. Kirpichenkov, Zh. Éksp. Teor. Fiz. **77**, 989 (1979) [Sov. Phys. JETP **50**, 499 (1979)].
[2] I. M. Lifshits, S. A. Gredeskul, and L. A. Pastur, Zh. Éksp. Teor. Fiz. **83**, 2362 (1982) [Sov. Phys. JETP **56**, 1370 (1982)].
[3] I. M. Lifshits, S. A. Gredeskul, and L. A. Pastur, *Introduction to the Theory of Disordered Systems*, Wiley, New York (1988); Nauka, Moscow (1982).

Translated by P. Shelnitz