Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 165–171.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 195–202.

GRAVITATION, ASTROPHYSICS

Nonself-Similar Regimes of Isothermal Collapse of Protostellar Clouds

A. E. Dudorov and A. G. Zhilkin*
Chelyabinsk State University, Chelyabinsk, 454021 Russia
**e-mail: zhag@csu.ru*
Received May 28, 2002

Abstract—We consider the development of inhomogeneity in the isothermal collapse of protostellar clouds. The initial and boundary conditions correspond to the classical statement of the problem on the contraction of a homogeneous cloud from a given volume. A centered rarefaction wave is shown to propagate from the outer boundary of the cloud toward its center at the first collapse stage. Analysis reveals two possible regimes of isothermal collapse, depending on the relationship between the rarefaction wave focusing time t_* and the cloud free-fall collapse time t_{ff}. For cold clouds, $t_* = t_{ff}$ and the rarefaction wave is not reflected. In this case, as time elapses, the cloud collapse becomes self-similar with the characteristic density profile $\rho \sim r^{-2}$. In hot clouds, $t_* < t_{ff}$ and the focusing can take place before the formation of an opaque core. Since the velocities of the rarefaction wave along and across magnetic field lines in a magnetized cloud are different, its front assumes a shape elongated along magnetic field lines. Depending on the initial conditions, based on analytical estimates, we investigate various possible scenarios for the collapse of magnetic protostellar clouds. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

At the initial stages of collapse (supersonic gasdynamic contraction under self-gravity), a protostellar cloud is transparent to its intrinsic infrared radiation. Therefore, the cloud temperature is constant with a high accuracy over a wide density range. Studying the isothermal collapse of protostellar clouds is an important astrophysical problem, because the characteristic profiles of the velocity, density, and other quantities that determine the basic parameters of protostars with accretion disks are formed precisely at this stage.

The emerging flow pattern is determined by two dimensional parameters—the cloud temperature T and the gravitational constant G. If the statement of the problem contains no other dimensional constants, then the flow is self-similar [1]. Similarity solutions for the problem of isothermal collapse were considered by many authors [2–6]. These similarity solutions were used to take into account the effects of slow rotation [7] and weak magnetic field [8] based on perturbation theory.

Consider the problem of the collapse of an isothermal protostellar cloud in the classical statement on the contraction of an initially homogeneous cloud from a given volume (see [2]). Under these initial conditions, the subsequent evolution of the cloud can be treated in terms of the piston problem (or the more general Riemann problem of the decay of an arbitrary discontinuity), where the gas self-gravity acts as the piston. The discontinuity decays to form a centered rarefaction wave at the boundary of the cloud that propagates toward its center [9]. Thus, the rarefaction wave front separates up the entire mass of the collapsing gas into two parts. In the inner region, the density is uniform and, hence, the matter collapses freely (there is no pressure gradient). In the outer region, a nonuniform density profile is formed behind the rarefaction wave front.

For spherically symmetric isothermal clouds, the law of motion of the rarefaction wave front and the criterion that separates the two types of flow were found in [10]. This solution was generalized to rotating isothermal clouds by Tsuribe and Inutsuka [11]. They showed that the rarefaction wave front surface in rotating clouds is flattened along the rotation axis. This is because the gas velocity in the equatorial plane is lower than its velocity along the rotation axis due to the centrifugal force. In this case, the speed of propagation of the weak discontinuity through the gas is equal to the isothermal speed of sound c_T, as in a nonrotating cloud.

Here, we consider the evolution of the rarefaction wave in collapsing magnetic nonrotating protostellar clouds. In the next section, this problem is considered for a spherically symmetric isothermal cloud with no magnetic field. In the third section, we analyze the evolution of the rarefaction wave and the various types of flow for a collapsing isothermal magnetized protostellar cloud. In the Conclusions, we present our main results.

1063-7761/03/9602-0165$24.00

2. SPHERICALLY SYMMETRIC COLLAPSE

Consider the gravitational contraction of a spherically symmetric isothermal cloud from a given volume. Assume that at the initial time, the cloud matter is uniformly distributed over a sphere of radius R_0. The pressure in the cloud is related to the density by the equation of state for an ideal gas at constant temperature:

$$P = c_T^2\rho, \quad c_T^2 = \frac{\mathcal{R}T}{\mu} = \text{const}, \tag{1}$$

where c_T is the isothermal speed of sound, $\mathcal{R}$ is the universal gas constant, and μ is the molecular weight. The initial statement of the problem can be treated in terms of the standard gasdynamic piston problem (see, e.g., [12]), in which the gas self-gravity acts as a kind of a piston.

At the first contraction stage, a centered rarefaction wave emerges and propagates from the boundary of the cloud toward its center. The boundary R between the inner region and the rarefaction wave region moves through the gas with the speed of sound c_T. Given the gas motion, the equation for the coordinate R of the weak discontinuity can be written as

$$\frac{dR}{dt} = v - c_T, \tag{2}$$

where $v = v(R, t)$ is the gas velocity at the rarefaction wave front.

Since the rarefaction wave is immediately adjacent to the region of freely collapsing gas, the velocity v can be determined from the solution of the problem of free-fall collapse (cloud contraction under self-gravity without a pressure gradient and other forces). Let us change to dimensionless variables:

$$\rho(r, t) = \rho_0\sigma(\tau), \quad v(r, t) = \frac{r}{t_0}h(\tau), \quad t = t_0\tau, \tag{3}$$

where ρ_0 is the initial density in the cloud and

$$t_0 = \frac{1}{\sqrt{4\pi G\rho_0}} \tag{4}$$

is the characteristic time scale. The flow of gas in the inner region is described by the relations (see, e.g., [1])

$$\sigma = \frac{1}{(1-\eta)^3}, \quad h = -\sqrt{\frac{2}{3}}\frac{\sqrt{\eta}}{(1-\eta)^{3/2}}, \tag{5}$$

where the parameter η varies between 0 and 1. The value of $\eta = 1$ corresponds to the free-fall collapse time

$$t_{ff} = \sqrt{\frac{3\pi}{32G\rho_0}}. \tag{6}$$

The dimensionless time τ is related to η by

$$\sqrt{\frac{2}{3}}\tau = \arcsin\sqrt{\eta} + \sqrt{\eta(1-\eta)}. \tag{7}$$

In the dimensionless variables, Eq. (2) transforms to

$$\frac{dr}{d\tau} = rh - \alpha, \tag{8}$$

where $r = R/R_0$ and the dimensionless parameter

$$\alpha = \frac{c_T t_0}{R_0} \tag{9}$$

is the ratio of the speed of sound c_T to the characteristic gas flow velocity R_0/t_0. Note that $\alpha = M_0^{-1}$, where M_0 is an integral analog of the Mach number. In this case,

$$\alpha = \sqrt{\frac{\varepsilon_t}{5}}, \tag{10}$$

where

$$\varepsilon_t = \frac{\Pi}{E_g} \tag{11}$$

is the initial ratio of the scalar pressure integral

$$\Pi = \int_V P dV$$

to the absolute value of the cloud gravitational energy E_g. Using (7), we can rewrite Eq. (8) as

$$\frac{dr}{d\eta} + \frac{r}{1-\eta} = -\alpha\sqrt{\frac{3}{2}}\sqrt{\frac{1-\eta}{\eta}}. \tag{12}$$

Integrating this equation yields

$$r(\eta) = (1-\eta)(1-\alpha\sqrt{6}\arcsin\sqrt{\eta}). \tag{13}$$

The solutions obtained for various values of α are shown in Fig. 1.

It is easy to determine the critical value α_* that separates the two types of solution (see [10]). Since the maximum value of the function $\arcsin\eta$ in the $\sqrt{\eta}$ range under consideration is equal to $\pi/2$, we can obtain

$$\alpha_* = \frac{2}{\pi\sqrt{6}} \approx 0.26. \tag{14}$$

The following value corresponds to the critical parameter α_*:

$$\varepsilon_t^* = \frac{10}{3\pi^2} \approx 0.34.$$

At small $\alpha < \alpha_*$ (cold clouds), the second factor in (13) is always positive in the range $0 \le \eta \le 1$. Therefore, the root of the equation $r(\eta) = 0$ that determines the focusing time (the time at which the radius of the

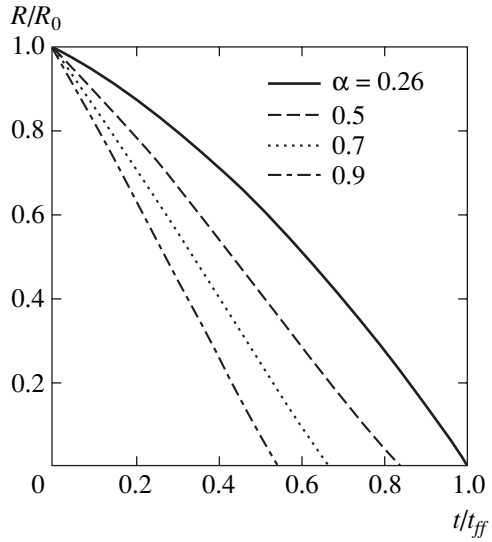


Fig. 1. Rarefaction wave front coordinate versus time for various values of α.

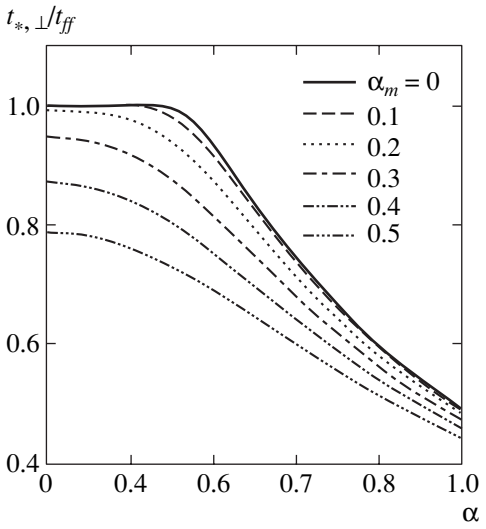


Fig. 2. Rarefaction wave focusing time $t_{*,\perp}$ in the transverse direction versus α for various values of α_m.

rarefaction wave front is zero) is equal to 1. This implies that the rarefaction wave front for such clouds is focused at the time $t_* = t_{ff}$. In this case, as time elapses, the collapse of a protostellar cloud passes to a self-similar regime, because the initial density in the cloud ceases to affect the distributions of quantities in the rarefaction wave region immediately adjacent to its front as the central density increases. The flow in this region "forgets" the initial conditions. Therefore, in the course of time, the solution will be determined only by two dimensional constants—the speed of sound c_T and the gravitational constant G. The density and velocity distributions in this region will be determined by standard self-similar profiles (see, e.g., [4]), $\rho \propto r^{-2}$ and $v \propto -r^{-1}$. It should also be noted that for the values of α under consideration, an opaque core is formed in the central region before the focusing time t_* and the subsequent evolution cannot be considered in terms of the isothermal approximation. After the formation of the opaque core (protostar), the flow of gas in the shell probably passes into an accretion regime with the characteristic density and velocity profiles $\rho \propto r^{-3/2}$ and $v \propto -r^{-1/2}$.

For large $\alpha > \alpha_*$ (hot clouds), the rarefaction wave is focused at the center in a time shorter than the cloud free-fall collapse time t_{ff}. After the reflection of the weak discontinuity from the center, a nonuniform density profile is formed in the cloud. As a result, a pressure gradient significantly affects the subsequent collapse. This case can correspond to the quasi-static contraction of hot clouds or clouds supported by a turbulent pressure. For $\alpha > \alpha_*$, we derive the following expression for the rarefaction wave focusing time from (13):

$$t_* = \frac{2}{\pi}\left(\frac{1}{\alpha\sqrt{6}} + \frac{1}{2}\sin\frac{2}{\alpha\sqrt{6}}\right)t_{ff}. \quad (15)$$

The inferred function $t_*(\alpha)$ is shown in Fig. 2 (solid line). Note that at large α, this function reaches its asymptotic limit

$$\frac{t_*}{t_{ff}} = \sqrt{\frac{8}{3}}\frac{1}{\pi\alpha} \approx \frac{0.52}{\alpha}. \quad (16)$$

3. THE EFFECT OF A MAGNETIC FIELD

Consider a nonrotating protostellar cloud threaded by a uniform magnetic field. In this case, a centered fast MHD rarefaction wave propagating toward the cloud center emerges at the first collapse stage. In the inner region, the magnetic field is uniform (and, hence, force-free) and varies with time as $B \propto \rho^{2/3}$. The boundary R of the rarefaction wave front can be determined from the solution of the equation

$$\frac{dR}{dt} = v - u_f, \quad (17)$$

where

$$u_f = \left\{\frac{c_T^2 + u_A^2}{2} + \frac{1}{2}\left[\left(c_T^2 + u_A^2\right)^2 - 4c_T^2u_A^2\cos^2\theta\right]^{1/2}\right\}^{1/2} \quad (18)$$

is the fast magnetosonic velocity, θ is the angle between the magnetic field vector **B** and the normal vector **n** at a given point of the front surface, and

$$u_A = \frac{|B|}{\sqrt{4\pi\rho}} \tag{19}$$

is the Alfven velocity.

Since the angle θ is zero or π along magnetic field lines, the longitudinal velocity of propagation of the rarefaction wave boundary through the gas is

$$u_\parallel = \max(c_T, u_A). \tag{20}$$

In the transverse direction, $\theta = \pm\pi/2$ and the boundary moves through the gas with the velocity

$$u_\perp = \sqrt{c_T^2 + u_A^2}. \tag{21}$$

Note that $u_\parallel < u_\perp$. Thus, at a given time, the surface of the rarefaction wave front in a magnetic cloud has a shape elongated along magnetic field lines that is similar to the shape of prolate ellipsoid of revolution.

Consider the transverse motion of the rarefaction wave front. Using (21), we can rewrite Eq. (17) in dimensionless form:

$$\frac{dr_\perp}{d\tau} = r_\perp h - \sqrt{\alpha^2 + \alpha_m^2\sigma^{1/3}}, \tag{22}$$

where the dimensionless parameter α is defined by relation (9) and the parameter

$$\alpha_m = \frac{B_0}{\sqrt{4\pi\rho_0}}\frac{t_0}{R_0} \tag{23}$$

is the ratio of the characteristic velocities $B_0/\sqrt{4\pi\rho_0}$ and R_0/t_0. Note that $\alpha_m = M_{A,0}^{-1}$, where $M_{A,0}$ is an integral analog of the Alfven Mach number. In this case,

$$\alpha_m = \sqrt{\frac{2}{5}\varepsilon_m}, \tag{24}$$

where ε_m is the initial ratio of the cloud magnetic energy to the absolute value of its gravitational energy.

Changing from the dimensionless time τ to the variable η using formula (7), we can transform Eq. (22) to

$$\frac{dr_\perp}{d\eta} + \frac{r_\perp}{1-\eta} = -\sqrt{\frac{3}{2}}\sqrt{\frac{1-\eta}{\eta}}\sqrt{\alpha^2 + \frac{\alpha_m^2}{1-\eta}}. \tag{25}$$

Integrating this equation yields the expression

$$r_\perp(\eta) = (1-\eta)(1-\alpha\varphi(\eta, \alpha_m/\alpha)), \tag{26}$$

where the function

$$\varphi(\eta, q) = \sqrt{6}\arctan\sqrt{\frac{\eta}{q^2 + 1 - \eta}} + \sqrt{\frac{3}{2}}q\ln\left(\frac{1+\sqrt{\eta}}{1-\sqrt{\eta}}\frac{1-\sqrt{\eta}+q^2+q\sqrt{q^2+1-\eta}}{1+\sqrt{\eta}+q^2+q\sqrt{q^2+1-\eta}}\right). \tag{27}$$

It should be noted that the function $\varphi(\eta, q)$ increases without limit as $\eta \longrightarrow 1$ for any positive value of q. Therefore, the rarefaction wave focusing time in the transverse direction $t_{*,\perp}$ is always shorter than t_{ff}. In Fig. 2, the rarefaction wave focusing time $t_{*,\perp}$ in the transverse direction is plotted against α. Different curves in the figure correspond to different values of α_m. The focusing time is seen to decrease sharply with increasing α_m.

The longitudinal motion of the rarefaction wave front cannot be determined analytically, because the pressure gradient affects the dynamics of the gas flow immediately adjacent to the rarefaction wave region. However, for qualitative estimates, we can assume that in this case, the gas velocity in Eq. (17) can also be determined from the solution of the problem of free-fall collapse. Thus, for the longitudinal direction, we obtain the equation

$$\frac{dr_\parallel}{d\eta} + \frac{r_\parallel}{1-\eta} = -\sqrt{\frac{3}{2}}\sqrt{\frac{1-\eta}{\eta}}w(\eta), \tag{28}$$

where

$$w(\eta) = \max\left(\alpha, \frac{\alpha_m}{\sqrt{1-\eta}}\right). \tag{29}$$

If $\alpha_m \geq \alpha$ (a strong magnetic field), then $w(\eta) = \alpha_m/\sqrt{1-\eta}$, implying that in this case, the rarefaction wave front in the longitudinal direction always moves with the Alfven velocity. Integrating (28) yields

$$r_\parallel(\eta) = (1-\eta)\left(1 - \sqrt{\frac{3}{2}}\alpha_m\ln\frac{1+\sqrt{\eta}}{1-\sqrt{\eta}}\right). \tag{30}$$

The derived law of motion does not depend on α_m. The focusing time is defined by the expression

$$t_{*,\parallel} = \frac{2}{\pi}\left(\arcsin\tanh\frac{1}{\sqrt{6}\alpha_m} + \tanh\frac{1}{\sqrt{6}\alpha_m}\Big/\cosh\frac{1}{\sqrt{6}\alpha_m}\right)t_{ff}. \tag{31}$$

Consider $\alpha_m < \alpha$. The function

$$w(\eta) = \begin{cases} \alpha, & \eta \leq \eta_1 \\ \dfrac{\alpha_m}{\sqrt{1-\eta}}, & \eta > \eta_1, \end{cases} \tag{32}$$

where $\eta_1 = 1 - \alpha_m^2/\alpha^2$. Integrating (28) yields

$$r_{\|}(\eta) = (1-\eta)(1-\sqrt{6}\alpha\arcsin\sqrt{\eta}), \quad \eta \le \eta_1, \tag{33}$$

$$r_{\|}(\eta) = (1-\eta)\left[1-\sqrt{6}\alpha\arcsin\sqrt{\eta_1} - \sqrt{\frac{3}{2}}\alpha_m \ln\left(\frac{1+\sqrt{\eta}}{1+\sqrt{\eta_1}}\frac{1-\sqrt{\eta_1}}{1-\sqrt{\eta}}\right)\right], \quad \eta > \eta_1. \tag{34}$$

The value η_1 corresponds to the time

$$t_1 = \frac{2}{\pi}\left(\arcsin\sqrt{1-\frac{\alpha_m^2}{\alpha^2}} + \frac{\alpha_m}{\alpha}\sqrt{1-\frac{\alpha_m^2}{\alpha^2}}\right)t_{ff}. \tag{35}$$

At this time, the Alfven velocity at the rarefaction wave front is equal to the isothermal speed of sound.

Assume that the focusing occurs before the time t_1. In this case, the longitudinal velocity of the rarefaction wave front is always equal to the isothermal speed of sound c_T. This regime takes place under the following conditions (a weak magnetic field):

$$\alpha_m \le \alpha\left|\cos\frac{1}{\sqrt{6}\alpha}\right|, \quad \alpha > \alpha_*. \tag{36}$$

For intermediate magnetic field strengths,

$$\alpha < \alpha_m < \alpha\left|\cos\frac{1}{\sqrt{6}\alpha}\right|,$$

the focusing occurs after the time t_1. In this case, the rarefaction wave front initially moves with the speed of sound and, after t_1, with the Alfven velocity.

Let us discuss the solutions obtained. In the limit of weak magnetic fields ($\alpha_m \longrightarrow 0$), the derived expressions for $r(\eta)$ in the longitudinal and transverse directions transform into the solution for the spherically symmetric case (13). This corresponds to the kinematic approximation where the magnetic field is so weak that the electromagnetic force is negligible compared with the gravitational force and the pressure gradient. In this limit, the magnetic field acts as a passive admixture and its configuration can be calculated independently from the induction equation using a given velocity field.

The focusing time t_* decreases with increasing magnetic field strength. Even for moderately strong magnetic fields ($\alpha_m \approx 0.1$), the electromagnetic force begins to affect the collapse dynamics. This effect can manifest itself, for example, in a small delay between the longitudinal focusing of the rarefaction wave and its transverse focusing. Under the electromagnetic force, the collapsing cloud itself (the rarefaction wave region) assumes a shape flattened along magnetic field lines. It should be particularly emphasized that the shape of the rarefaction wave surface does not coincide with the shape of the central flattened condensations formed at late cloud collapse stages.

For sufficiently strong magnetic fields, intense MHD waves resulting from the reflection of the fast magnetosonic rarefaction wave that arrives at the center must be generated at the isothermal collapse stage of protostellar clouds. This conclusion was first drawn by Dudorov and Sazonov [13] from numerical simulations of MHD collapse in the one and a half dimensional approximation. We also observed such effects in two-dimensional numerical computations of the collapse of magnetic protostellar clouds [14]. In some cases (large α), the unloading of matter in the central region can probably even be accompanied by bipolar plasma ejection parallel to the initial magnetic field. This conclusion is directly corroborated by two-dimensional numerical computations.

Figure 3 shows two-dimensional numerical simulations of the collapse of a magnetized protostellar cloud. The initial parameters correspond to $\alpha = 0.245$ and $\alpha_m = 0.283$. This figure shows the distributions of the logarithm of density in cylindrical coordinates r and z in the cloud for four consecutive times: $0.468t_{ff}$, $0.755t_{ff}$, $0.870t_{ff}$, and $1.044t_{ff}$. The panels corresponding to the first three times demonstrate the evolution of the MHD rarefaction wave at the initial collapse stages when its characteristic shape elongated along magnetic field lines is formed. The last panel corresponds to the time after the rarefaction wave focusing. Through the focusing and reflection of the rarefaction wave front, the central region of the cloud assumed the shape of a thin disk.

For strong magnetic fields, the focusing time is so short that at the time when the rarefaction wave is reflected from the center, the cloud matter does not gain enough kinetic energy to generate MHD waves capable of affecting the collapse dynamics. In this case, after the reflection of the rarefaction wave, the cloud passes into a state of quasi-magnetostatic contraction with a relatively small density difference from the periphery to the center. In this state, the cloud evolves not on the dynamic scale but on the diffusion scale. The latter is determined by turbulence dissipation and by ambipolar and ohmic magnetic field diffusion.

Note the significant difference between the patterns of evolution of the rarefaction wave in nonmagnetic rotating and magnetic nonrotating clouds. In the former case (see [11]), the centrifugal force causes the collapsing gas to slow down in the central region of uniformly contracting gas in the transverse direction with respect to the symmetry axis. The velocity of propagation of the weak discontinuity through the gas is equal to the speed of sound c_T. As a result, the rarefaction wave assumes a shape flattened along the rotation axis. For a magnetic nonrotating cloud, the magnetic field remains forcefree in the inner region. Therefore, the electromagnetic forces have no effect on the gas velocity (kinematics). However, the velocities of propagation of

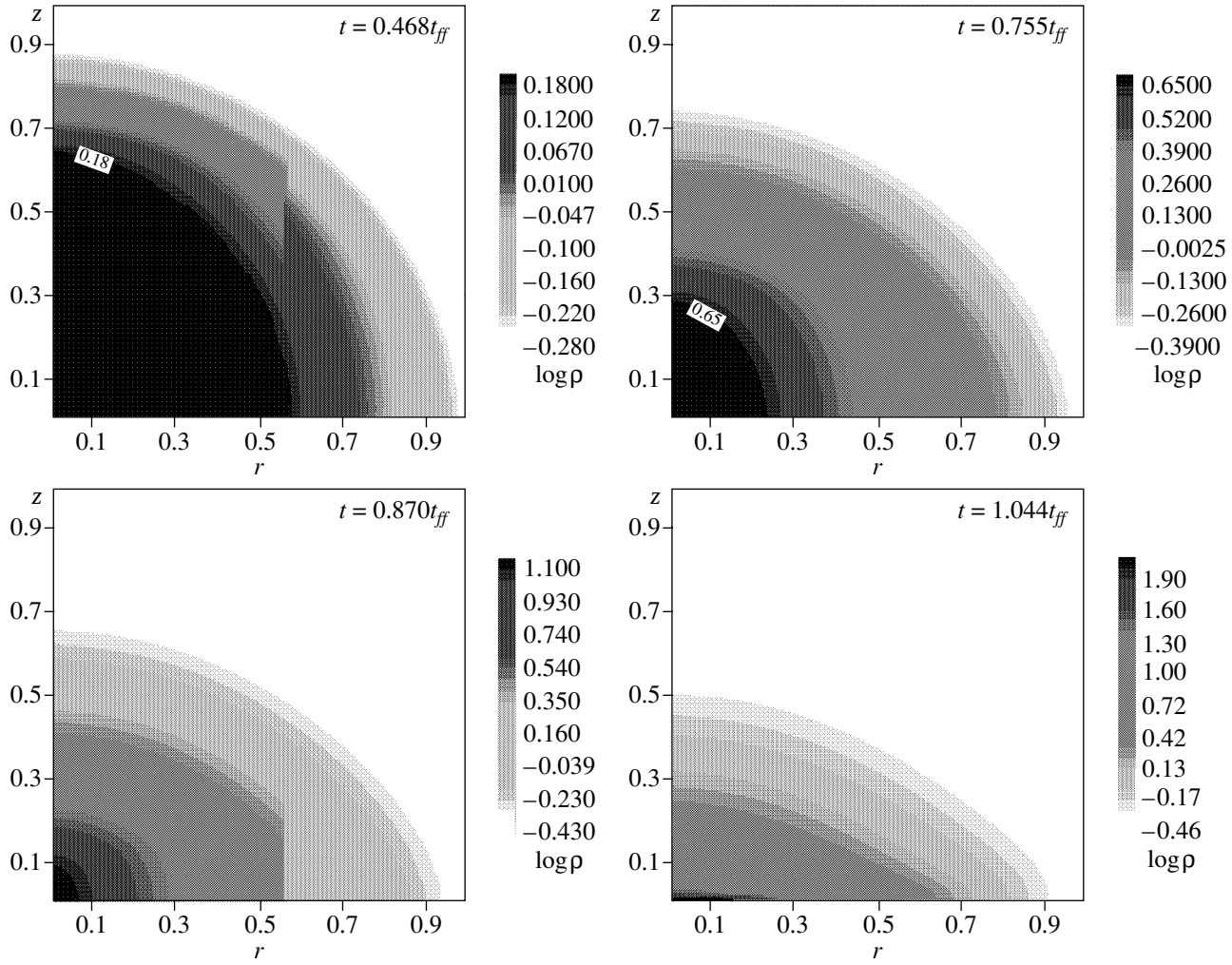


Fig. 3. Distribution of the logarithm of density and the evolution of the MHD rarefaction wave in a collapsing magnetized protostellar cloud for various times. The initial parameters correspond to $\alpha = 0.245$ and $\alpha_m = 0.283$.

the weak discontinuity through the gas in a magnetic cloud along and across magnetic field lines are different. As a result, the surface of the MHD rarefaction wave front assumes a shape elongated (rather than flattened) along magnetic field lines. In the region of the rarefaction wave itself, the cloud assumes a flattened shape under the electromagnetic force.

4. CONCLUSIONS

Our main results are as follows.

The dynamics of the spherically symmetric collapse of isothermal protostellar clouds in the classical statement of the problem of cloud contraction from a given volume is characterized by the formation of a centered rarefaction wave at the outer boundary at the initial time and by its subsequent propagation toward the cloud center. The front of this wave separates the cloud into two regions. In the inner region, the matter is homogeneous and collapses freely. In the rarefaction wave region, a nonuniform density profile is formed.

The rarefaction wave focusing time is determined by the dimensionless parameter α (9). In cold clouds ($\alpha \le \alpha_* \approx 0.26$), the rarefaction wave is focused at the time $t_* = t_{ff}$. In this case, the characteristic self-similar density, $\rho \propto r^{-2}$, and velocity, $v \propto -r^{-1}$, profiles are formed in the rarefaction wave region immediately adjacent to the front. Initially, this is a narrow region, but it expands as the central density increases. After the separation of an opaque core (a protostar), the gas motion in its vicinity passes into an accretion regime with the characteristic density profile $\rho \propto r^{-3/2}$.

In hot clouds ($\alpha > \alpha_*$), the focusing takes place before the free-fall collapse time ($t_* < t_{ff}$). After the reflection of the weak discontinuity from the center, a nonuniform density profile is formed in the cloud and a pressure gradient significantly affects its subsequent

collapse. Since the collapse of such clouds is appreciably slower, this case can correspond to the quasi-static contraction of hot clouds or clouds supported by turbulent pressure.

The effect of magnetic field causes the velocities of the weak discontinuity along and across magnetic field lines to differ. As a result, the surface of the rarefaction wave front assumes a shape elongated along magnetic field lines similar to the shape of an prolate ellipsoid of revolution. This conclusion suggests that the patterns of evolution of the rarefaction waves in magnetic nonrotating and rotating nonmagnetic collapsing protostellar clouds significantly differ. In a nonmagnetic rotating cloud, the surface of the rarefaction wave front assumes a shape flattened (not elongated) along the rotation axis.

The dynamics of the rarefaction wave in a magnetic cloud is described by two dimensionless parameters, α and α_m (23). For nonzero values of α_m, the rarefaction wave is always focused in the transverse direction in time $t_{*,\perp} < t_{ff}$, although this difference can be negligibly small for weak magnetic fields. The focusing time $t_{*,\perp}$ decreases with increasing magnetic field strength.

In the longitudinal direction, the rarefaction wave is also always focused in time $t_{*,\parallel} < t_{ff}$. In this case, however, $t_{*,\parallel} > t_{*,\perp}$. The pattern of longitudinal motion of the rarefaction wave front significantly depends on the relationship between α and α_m.

If $\alpha_m \geq \alpha$ (a strong magnetic field), then the weak discontinuity moves in the longitudinal direction with the Alfven velocity. As the initial magnetic field strength decreases until a time t_1 (35), the isothermal speed of sound c_T at the rarefaction wave front can exceed the Alfven velocity. As a result, the rarefaction wave front will propagate through the gas in the longitudinal direction initially with the velocity c_T and subsequently, starting from time t_1, with the Alfven velocity. This scenario takes place if conditions (36) are satisfied. In hot clouds ($\alpha > \alpha_*$) with a weak magnetic field, the rarefaction wave front can propagate through the moving gas in the longitudinal direction with the velocity c_T from the very beginning until the focusing time. In all cases, the focusing takes place before the free-fall collapse time t_{ff}.

Note that all our main conclusions drawn from analytical calculations are in good agreement with numerical simulations of the collapse of protostellar clouds in the sesqui- and two-dimensional approximations.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 99-02-16938 and 02-02-17642).

REFERENCES

1. L. I. Sedov, *Similarity and Dimensional Methods in Mechanics*, 9th ed. (Nauka, Moscow, 1981; CRC Press, Boca Raton, 1993).
2. R. B. Larson, Mon. Not. R. Astron. Soc. **145**, 271 (1969).
3. M. V. Penston, Mon. Not. R. Astron. Soc. **144**, 425 (1969).
4. F. H. Shu, Astrophys. J. **214**, 488 (1977).
5. C. Hunter, Astrophys. J. **218**, 834 (1977).
6. A. Whitworth and D. Summers, Mon. Not. R. Astron. Soc. **214**, 1 (1985).
7. S. Tereby, F. H. Shu, and P. Cassen, Astrophys. J. **286**, 529 (1984).
8. D. Galli and F. H. Shu, Astrophys. J. **417**, 220 (1993).
9. Ya. B. Zel'dovich and Ya. M. Kazhdan, Astrofizika **6**, 109 (1970).
10. K. Truelove, R. I. Klein, C. F. McKee, *et al.*, Astrophys. J. **495**, 821 (1998).
11. T. Tsuribe and S. Inutsuka, Astrophys. J. **526**, 307 (1999).
12. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics*, 4th ed. (Pergamon, Oxford, 1987; Nauka, Moscow, 1988).
13. A. E. Dudorov and Yu. V. Sazonov, Nauchn. Inf. Astrosov. Akad. Nauk SSSR **50**, 98 (1982).
14. A. E. Dudorov, A. G. Zhilkin, and O. A. Kuznetsov, Mat. Model. **11**, 109 (1999).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 172–179.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 203–210.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Investigation of Liquid Fluoropolymers as Possible Materials for Low-Temperature Liquid-Wall Chambers for Ultracold Neutron Storage¶

Yu. N. Pokotilovski

Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia
**e-mail: pokot@nf.jinr.ru*

Received April 27, 2002

Abstract—Several hydrogen-free liquid low-temperature fluoropolymers are investigated from the point of view of their possible use as the material for walls of ultracold neutron traps with low losses. Viscosity was measured in the temperature range 150–300 K, and neutron scattering cross sections were measured in the temperature range 10–300 K and in the neutron wavelength range 1–20 Å. Some conclusions are made for their possible ultracold neutron bottle properties. Quasi-elastic neutron reflection from the surface of a viscous liquid is considered in the framework of the Maxwell dynamic model. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The goal of present-day investigations of free neutron decay is to reach a precision better than 0.1% for parameters of the V–A theory of neutron beta decay via measurement of the asymmetry of neutron decay and the neutron lifetime. Neutron lifetime measurements are necessary for determining the fundamental coupling constants of weak interactions. Additionally, it is an important parameter in astrophysical calculations concerning the creation of nuclei in the early stage after the Big Bang, including the abundance of helium in the Universe, the number of species of light stable neutrinos, and solar neutrino flux calculations.

An improvement by almost an order of magnitude in the accuracy of measuring the neutron lifetime has been achieved in recent years [1]. This progress was reached entirely due to the application of ultracold neutron (UCN) storage in closed volumes. A review of neutron lifetime measurements up to 1990 was published by Schreckenbach and Mampe [2] (see also the review by Pendlebury [3] on wider fundamental applications of UCNs). For a recent review of investigations of free neutron decay see [4, 5]. The most precise measurements of neutron lifetime were carried out with very low surface temperature beryllium and solid oxygen traps [6], or with traps covered with Fomblin oil [7–10].

The experiments with very cold (near 10 K) beryllium UCN traps led [11, 12] to the observation of surprisingly large UCN losses in the traps, exceeding theoretical predictions by two orders of magnitude. At room temperature, the wall losses exceed theoretical ones by an order of magnitude [12]. The reason for this anomaly is not yet understood and is a serious impediment to further significant progress in the precision of neutron lifetime measurements by this method.

The losses of UCNs stored in traps with walls covered with solid oxygen [6, 12] or cooled graphite [13] also exceed those calculated according to cold-neutron transmission cross sections by two to three orders of magnitude.

Application of the hydrogen-free perfluoropolyether (PFPE) oil (Fomblin) was first proposed by Bates [14] and was tested for neutron lifetime measurement in [15]. Fomblin has the chemical formula $CF_3(C_3F_6O)_n(OCF_2)_mOCF_3$, with m/n = 20–40 and a molecular weight near 3000 [16].

The UCN losses caused by Fomblin wall collisions in a large UCN trap (volume up to 72 l) were around 20% (at 10°C) and 10% (at 4°C) of the beta decay in [15] and subsequent experiments [7, 8] and were the most important source of systematic errors when corrections for these losses were made. There were plans [17] to continue neutron lifetime experiments with an upgraded installation in the same way as in [7, 8].

Careful measurements of UCN losses in Fomblin traps with the aim of obtaining the UCN energy dependence of the loss coefficient were performed in [18]. In the expression for the UCN loss probability averaged over the isotropic angular distribution,

$$\bar{\mu}(E) = 2\eta\left[\frac{V}{E}\arcsin\left(\sqrt{\frac{E}{V}}\right) - \sqrt{\frac{V-E}{E}}\right], \qquad (1)$$

where E is the neutron energy and V is the boundary potential of the trap, the experimental UCN reduced loss coefficient η was found to be $\eta = 2.35(0.10) \times 10^{-5}$ for Fomblin oil and $\eta = 1.85(0.10) \times 10^{-5}$ for Fomblin

¶This article was submitted by the author in English.

grease at 21°C. It is interesting to note that these values of the measured UCN loss coefficient η in Fomblin are significantly lower than the limiting anomalous loss coefficient $\eta_{anom} = 3.3 \times 10^{-5}$ in beryllium traps at 10 K [12]. At a lower temperature of Fomblin, 4°C, the loss coefficient was as low as $\eta = 1.3(1) \times 10^{-5}$, the best one ever obtained experimentally and 2.5 times lower than the anomalous loss coefficient in experiments [12]. We use the standard formalism for the complex potential U describing the UCN interaction with walls,

$$\eta = \frac{\mathrm{Im}U}{\mathrm{Re}U}, \quad U = \frac{4\pi\hbar^2}{2m}\sum_i N_i b_i, \quad \mathrm{Im}b = \frac{\sigma}{2\lambda}, \quad (2)$$

where m is the neutron mass, N_i is the number of nuclei in a unit volume of the wall material, b_i is the coherent scattering length on a bound nucleus of the wall, and σ is the cross section of inelastic processes for neutrons with wavelength λ; this formalism allows calculating the part of the wall loss coefficient attributed to neutron capture in Fomblin, which is as low as 3.6×10^{-7} (the experimental value $\mathrm{Re}U = 106.5$ neV [19] was used in this calculation). This implies that the main component of UCN losses is the inelastic scattering. For Fomblin, a strong dependence of UCN wall losses on temperature was observed: the losses decrease by about 3% as the temperature is lowered by 1°C [7]. This requires a very good wall temperature uniformity over the trap surface for precision measurement of the neutron lifetime. However, the use of Fomblin at lowered temperatures proved unpromising in view of crumbling of the Fomblin surface in solid and near-solid states [7, 9], resulting in a significant increase in UCN losses. The authors of [7] found that below 0°C, Fomblin oil becomes too viscous for respraying over the wall surface.

However, for a liquid Fomblin surface, contrary to solid surfaces, the difference between the measured UCN loss coefficients and the ones calculated from transmission measurements is not large.

Except for the indication [7] that transmission measurements with $\lambda = 60$ Å neutrons at 20°C Fomblin agree with the cross section of UCN losses within a factor of 1.5, there is no (to the author's knowledge) quantitative experimental information on neutron inelastic scattering in PFPE; the possibility of calculating it suffers from uncertainties about the dynamics of thermal motion in this liquid polymer. It is interesting to compare the total inelastic UCN cross section in PFPE extracted from [7, 18] in accordance with Eq. (2) and extrapolated to the thermal point in accordance with the inverse-velocity law $\sigma_{extr}^{PFPE} = 8.65$ b (≈0.51 b per atom of PFPE) with the experimental data for a "similar" compound, Teflon $(CF_2)_n$ [20]. For the lowest neutron energy used in these measurements, 0.376 meV, at which the elastic coherent scattering effects are believed to be negligible (below all the Bragg peaks in the cross section measured in [20]), the cross section per atom extrapolated to the thermal point is 0.63 b, which satisfactorily agrees with 0.51 b.

In [21], it was proposed to try low-temperature liquid wall chambers for storing UCN. This proposal was prompted by publication [22], which reported on new, recently synthesized, fully fluorinated compounds whose physical properties are promising from the standpoint of using them in liquid-wall UCN traps. These new compounds—perfluorinated poly formaldehydes—have a large liquid range and good low-temperature properties. Their formula is $C_4F_9(OCF_2)_nC_4F_9$. Depending on the number n (between 6 and 10), the boiling temperature of this polymer varies between 125 and 200°C; the melting point is between –145 and –152°C.

There is hope that at lower temperatures—in the vicinity of the melting point (e.g., –100°C)—inelastic upscattering, which is believed to be the most important component of UCN losses upon collisions with chamber walls, decreases by several times. Corrections for the UCN loss when inferring the neutron lifetime from the UCN storage data correspondingly decrease.

It is not yet clear to what extent the recently observed [23] small UCN cooling and heating during storage in Fomblin traps is an important component of UCN losses from liquid-wall traps.

The lack of any information on the dynamical properties of this new compound (as well as for Fomblin) only allows a very approximate estimate of gain in UCN losses (a UCN upscattering cross section) with decreasing temperature from 10 to –120°C. It is assumed in this estimate that dynamical properties of PFPE and this new polymer are similar, and the frequency distribution is described by the Debye model. From the known specific heat of Fomblin, 0.24 cal/(g grad) [24], the Debye temperature of Fomblin can be calculated in the standard way as 750 K; it in turn allows calculating the UCN upscattering cross section in the incoherent approximation. The result of this calculation is that the upscattering cross section decreases by five times from room temperature to –120°C, and changes by 0.7% per 1°C at room temperature. The latter figure is four times lower than was reported [7], and our estimate therefore seems to be a lower bound of the possible decrease of UCN upscattering with decreasing temperature.

To our regret, we were unable to obtain this substance [22] at our disposal, but it turned out that low-temperature fluoropolymers with similar properties are produced in Russia.[1]

[1] The investigated substances were produced by the Perm Branch of the Russian Scientific Center of Applied Chemistry and by the State Scientific Institute for Organic Chemistry and Technology, Moscow.

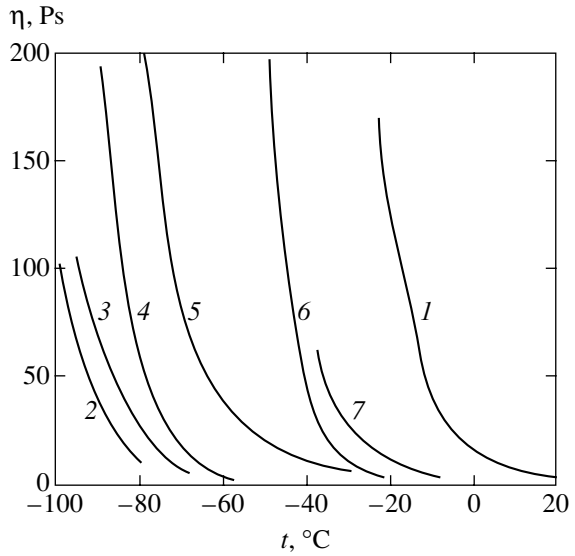


Fig. 1. Viscosity of various liquid fluoropolymers as a function of temperature: 1—Fomblin; 2—POM-310; 3—POM; 4–7—several other liquid fluoropolymers with a different chemical content.

The goal of this work was to investigate several possible low-temperature fluoropolymers that are candidates for a low-temperature liquid-wall UCN bottle. Viscosity as a function of temperature was measured, because viscosity determines the temperature range at which the liquid wall can be used practically, and cold neutron cross sections were measured as a function of temperature.

2. VISCOSITY AND NEUTRON CROSS-SECTION MEASUREMENTS

Viscosity was measured with the simplest possible method, that of a "sinking ball" [25]. The results are presented in Fig. 1. As can be seen, POM-310 and POM possess the most promising properties. They are the mixtures of complex fluoropolyoxymethylenes with the general formula $CF_3O(CF_2O)_n(CF_2CF_3O)_m(OCF_2CF_2O)_lCF_3$ with $n : m : l = 65.8 : 3.1 : 0.2$ and a molecular weight of 4883 for POM-310; and $n : m : l = 30.3 : 1.5 : 0.2$ and a molecular weight of 2354 for POM.

It can be seen from Fig. 1 that the viscosity of the most promising liquids, POM and POM-310, at –90°C is close to the viscosity of Fomblin at a temperature near 0°C, at which it can still be used in experiments on neutron lifetime measurement [8, 15]. This means that from the experimental standpoint, they are appropriate for similar use as in experiments [7].

Neutron cross section measurements were performed at channel 6B of reactor IBR-2. The main goal was to measure the total cross sections at the lowest possible energies in a wide temperature range that would allow the upscattering part of the cross section from these measurements to be inferred. This can be used to estimate possible UCN losses in storage experiments.

In view of an overwhelming flux of delayed thermal neutrons in the direct-beam time-of-flight spectrum be low a neutron energy of about 1 meV, diffraction from the stack of mica (artificial fluorophlogopite [26, 27]) with a lattice parameter of 9.97 Å for the (001) plane was used for neutron monochromatization. A significant suppression of the delayed thermal neutron component in the diffracted neutron beam was achieved in this way. The spectrum of neutrons diffracted from the stack of mica (a thickness of about 1.8 mm) is shown in Fig. 2.

The total cross sections were measured at several Bragg angles of neutron diffraction from mica, from 45° to 83.5°. In Fig. 3, we show the total neutron cross section three room-temperature fluoropolymers as a function of neutron wavelength. Strong coherent

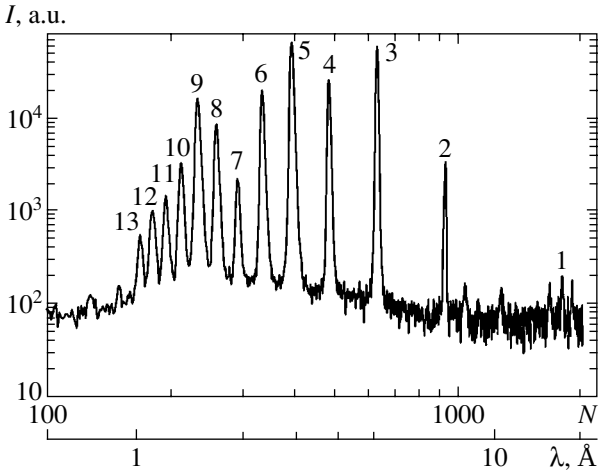


Fig. 2. Time-of-flight spectrum of neutrons diffracted from the stack of artificial fluorophlogopite at a Bragg angle of 45°. Numbers indicate the order of diffraction.

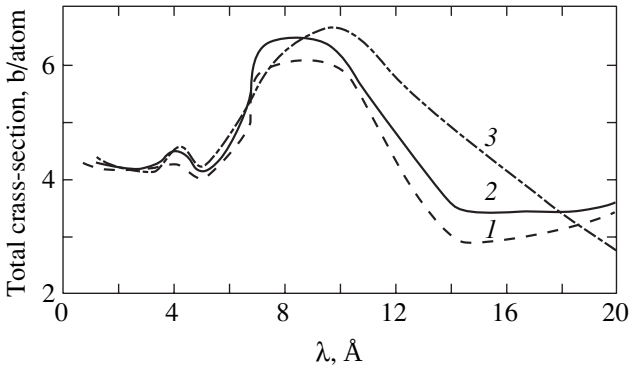


Fig. 3. The total neutron cross section of liquid fluoropolymers at 290 K as a function of neutron wavelength: *1*—Fomblin; *2*—POM-310; *3*—POM.

effects can be seen in a wide wavelength range up to the largest wave length reached in these measurements, 20 Å. This behavior is similar to that of the cross section for another well-known solid fluoropolymer, Teflon. Our measurements of the total cross section for Teflon agree well with the previous one [20].

To obtain the value of the upscattering probability for UCNs in traps with walls covered with these liquids, the measurements of the total neutron cross section were performed in a wide temperature range from 290 down to 10 K and in a wavelength range from 1 to 20 Å. In Figs. 4 and 5, we show the temperature dependence of the total cross section of long-wavelength neutrons for Fomblin and POM-310. The cross section for POM (not shown) is close to the cross section for POM-310.

In the temperature range of interest (200–300 K) and with a wavelength larger than 10 Å, upscattering is a strongly dominant component of inelastic scattering (for neutron energies of $E_n \ll kT$). At low temperatures (10–100 K), the cross section is independent of temperature and is one of entirely elastic scattering. The upscattering cross section can therefore be obtained after subtracting this elastic component from the total cross section. In Fig. 6, we show the result of this operation for a neutron wavelength of 20 Å. It is seen that in these measurements, we did not reach the neutron energy range where the upscattering cross section behaves according to the inverse-velocity law because of a strong coherent in elastic contribution (at $T \geq 100$ K) in a wide range of wavelengths around 10 Å. Measurements of the total cross section these substances at room temperature for neutrons in the wavelength range 200–800 Å demonstrated good agreement with the inverse-velocity-law dependence of the cross section and yielded coincident results with the present measurements at 20 Å for the value of the upscattering cross section extrapolated to the thermal point at room temperature [28].

At large wavelengths, the upscattering cross section as a function of temperature can be taken as a starting point for calculating the possible upscattering contribution to the UCN loss in liquid-wall traps.

From the upscattering cross section 3 b/atom for Fomblin at $\lambda = 20$ Å and room temperature, it follows that $\mathrm{Im}\, b = \sigma/2\lambda = 0.75 \times 10^{-17}$ cm, which leads to a loss coefficient of $\eta \approx 1.25 \times 10^{-5}$ (the mean value of the real part of the scattering length for Fomblin was calculated as $\mathrm{Re}\, b \approx 5.9$ F). The experimental value $\eta = 2.35 \times 10^{-5}$ [18] inferred from UCN storage experiments is almost twice the value inferred from the transmission data of the present measurements. Some additional loss processes possibly occur when UCNs are reflected from the liquid surface—due to some hydrogen contamination of the surface that increases the probability of upscattering, or a significant influence of quasi-elastic scattering (small heating [23]) due to the surface excitation of a viscous liquid [29].

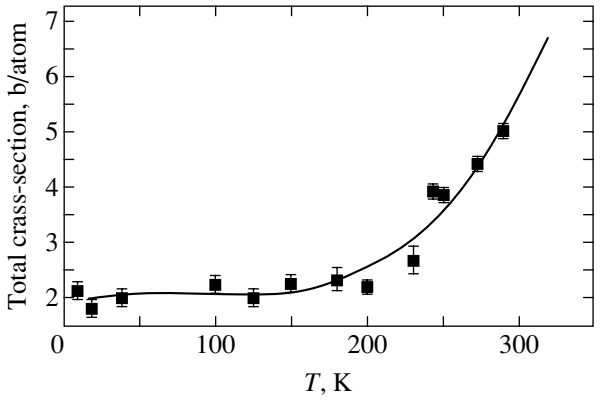


Fig. 4. The total neutron cross section of Fomblin versus temperature for a neutron wavelength of 14.1 A.

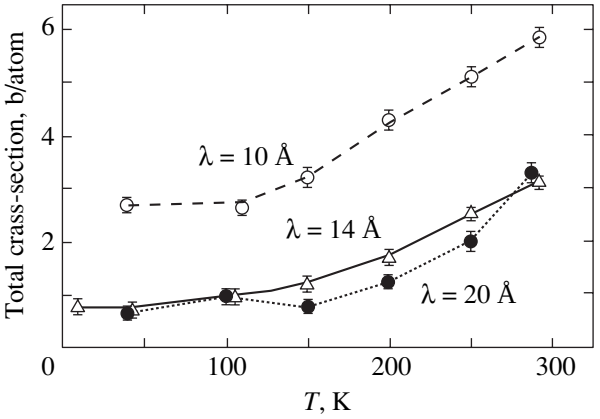


Fig. 5. The total neutron cross section of POM-310 versus temperature for several neutron wavelengths.

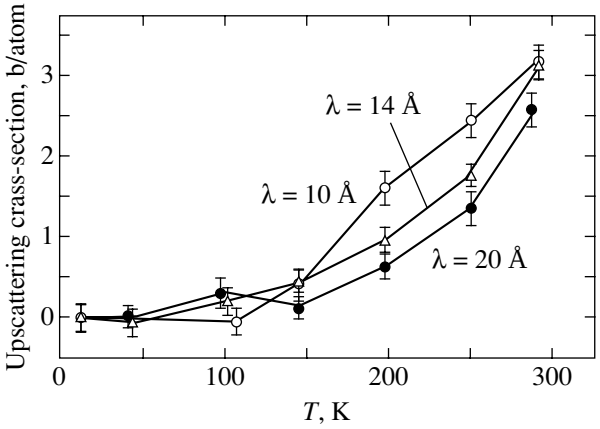


Fig. 6. The upscattering neutron cross section for POM-310 versus temperature obtained as a result of subtracting the elastic scattering contribution from the total cross section.

Upscattering cross sections for other liquid fluoropolymers are close to the Fomblin one. Most important is the temperature behavior of the cross section for low-temperature liquids. It follows from Fig. 6 that at a temperature around 190 K, where viscous proper ties are appropriate for experiments with a UCN bottle, the upscattering cross section at wavelength 20 Å is 0.5–0.6 b. For the expected UCN loss coefficient due to upscattering at this temperature, this gives $\eta \approx 2 \times 10^{-6}$. This figure is an order of magnitude lower than the experimental one for Fomblin [8, 18]. We can therefore hope to decrease the UCN loss correction in the neutron lifetime experiment approaching a value of 3×10^{-4} for the precision neutron lifetime measurements.

3. INFLUENCE OF QUASI-ELASTIC UPSCATTERING

It was shown in [29] that another possible source of UCN losses in liquid-wall traps—small neutron heating during collisions with the walls [23]—can be explained as a result of neutron interaction with thermal surface excitations of the liquid surface. Quantitative experimental data on UCN quasi-elastic scattering on liquid surfaces are scarce: the differential probability of scattering as a function of the incident UCN energy and of the change in energy has not yet been measured; the integral probability is known very approximately because absolute calibration of the detecting apparatus is difficult and the range of integration over the incident and final neutron energy is not determined precisely. It is possible, however, to determine the parameters of the dynamic model of a viscous liquid with some precision using even this scant information.

First, it is easy to show that in typical experimental situations of interest to us, the capillary wave contribution to a small UCN change in energy is insignificant. For practical needs, the viscosity range of liquid polymers used for UCN storage is between one and tens of poise, and the surface tension of our liquids is about 20 din/cm^2. As is well known [30, 31], capillary waves are not damped when the dispersion curve $\omega = \sqrt{\sigma/\rho}\, q^{3/2}$ for capillary waves lies above the line $\omega = (2\eta/\rho)q^2$. Here, ω is the frequency of surface oscillations, σ is the surface tension, η is the viscosity, q is the surface wave vector, and ρ is the density of the liquid. The critical wave vector $q_c = \sigma\rho/4\eta^2$ (about 0.4 cm^{-1} at η = 5 Ps) corresponds to the critical capillary frequency $\omega_c = \sigma^2\rho/8\eta^3$. It is as low as 1 s^{-1} and outside measurement capacity in any neutron experiments. The capillary waves in the energy range of our interest, 0.1–100 neV ($10^5 \le \omega \le 10^8$ s^{-1}), are strongly overdamped.

However, at small time scales ($t \le 10^{-5}$ s), viscous liquids demonstrate elastic properties: they deform under the influence of an external force, and shear stress then relaxes with a characteristic time of τ. Viscoelastic properties can be described introducing [32] the complex viscosity via the Maxwell formula

$$\eta = \eta_0/(1 - i\omega\tau), \tag{3}$$

which is simply the Fourier transform of the exponential time dependence of elastic deformations in a viscoelastic liquid.

As shown in [33, 34], taking elastic properties into account leads to significant variability of the spectrum of surface fluctuations: the spectrum of capillary waves is suppressed and narrowed, and energy dissipation of surface vibrations is increased at the expense of arising at higher frequencies of the wave motion of a phonon nature. These elastic effects prevail totally in the frequency range of interest.

To find the effect of surface fluctuations on UCN quasi-elastic interaction with a liquid surface, we use the results of [35] for the dynamic structure factor of a viscous liquid,

$$S(\mathbf{q}, \omega) = (2\pi)^3 \langle \xi^2_{\mathbf{q}, \omega_\mathbf{q}} \rangle, \tag{4}$$

where $\langle \xi^2_{\mathbf{q}, \omega_\mathbf{q}} \rangle$ is the mean squared surface fluctuation with wave vector $\mathbf{q}$ and frequency $\omega_\mathbf{q}$. The structure factor is given by

$$S(q, \omega) = \frac{8k_B T \mathrm{Re}[\eta(\omega)]q^3}{|D(q, \omega)|^2} \times \left\{1 + \frac{1}{2\mathrm{Re}[\alpha(q, \omega)]} - 2\mathrm{Re}\left(1 + \frac{1}{1 + \alpha(q, \omega)}\right)\right\}, \tag{5}$$

where the surface-mode dispersion relation [31, 33, 36–38] is $D(q, \omega) = 0$, with

$$D(q, \omega) = [i\omega + 2\nu(\omega)q^2]^2 - 4\nu(\omega)^2 q^4 \alpha(q, \omega) + \sigma q^3/\rho. \tag{6}$$

Here, $\nu(\omega) = \eta(\omega)/\rho$ is the kinematic viscosity; the complex frequency-dependent viscosity $\eta(\omega)$ in Eq. (3) crosses from the viscous behavior at low frequencies to the elastic behavior at high frequencies with $\eta_0 = G\tau$ [32, 35] and

$$\alpha(q, \omega) = \sqrt{1 + \frac{i\omega}{\nu(\omega)q^2}}. \tag{7}$$

In expressions (5)–(7), k_B is the Boltzmann constant, T is the temperature, σ is the surface tension, ρ is the density, τ is the liquid-polymer stress relaxation time,

and G is the frequency-independent shear modulus of the polymer network.

In this simple Maxwell model, the surface dynamics of the liquid and the interaction of neutrons with the liquid surface are determined for any ($\mathbf{q}$, ω) by four parameters, T, ρ, G, and τ.

The probability of quasi-elastic reflection of neutrons from liquid surface fluctuations is given by [29]

$$w(\mathbf{k}_0 \longrightarrow E) = \frac{16k_{0\perp}}{(2\pi)^3}\frac{1}{2\hbar}\int S\left(q, \frac{\Delta E}{\hbar}\right)k_\perp q(dq)d\phi. \quad (8)$$

In Fig. 7, we show typical results of calculations according to Eq. (8) for the probability of quasielastic upscattering of a neutron with an energy of 50 neV, incident at the angle $\pi/4$ on the surface of the room-temperature liquid with a surface tension of 20 din/cm^2.

Apart from the possibility (very restricted in view of the lack of experimental information mentioned above) of a quantitative comparison of the calculated probabilities with the experimental data [23], it is now possible to infer some parameters of the dynamic model of a liquid. It follows from the recently measured [39] viscosity dependence of quasi-elastic upscattering from the Fomblin surface demonstrating a close proportionality of the upscattering probability to the inverse viscosity, at least in the viscosity range between 0.5 and 15 Ps. A detailed analysis of the calculated (Eq. (8)) probability of UCN upscattering as a function of stress relaxation

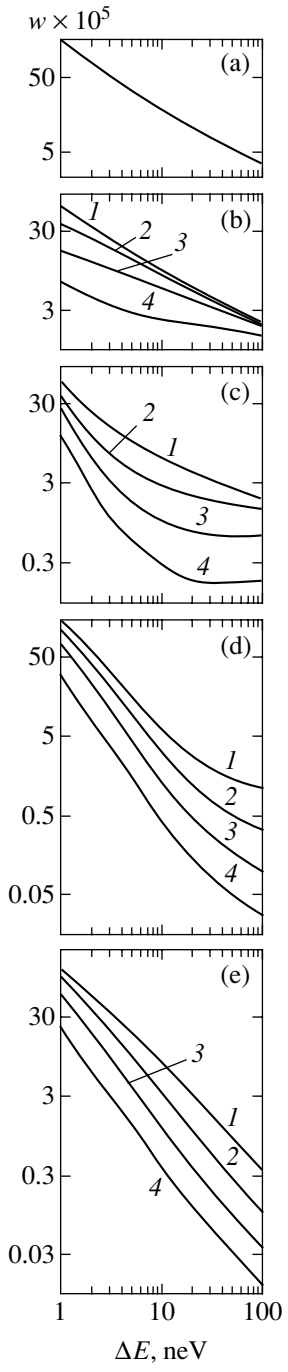


Fig. 7. Results of calculations of the probability (per 1 neV) of quasi-elastic upscattering of a neutron with an energy of 50 neV, incident at an angle of $\pi/4$ on the surface of the room-temperature liquid with a surface tension of 20 din/cm^2. The model of a liquid with a complex viscosity was assumed (Eq. (3)) with $\tau = 10^{-3}$ (a), 10^{-5} (b), 10^{-6} (c), 10^{-7} (d) and 10^{-12} (e); $\eta_0 = 0.3$ (*1*), 1.0 (*2*), 3.0 (*3*), and 10.0 Ps (*4*).

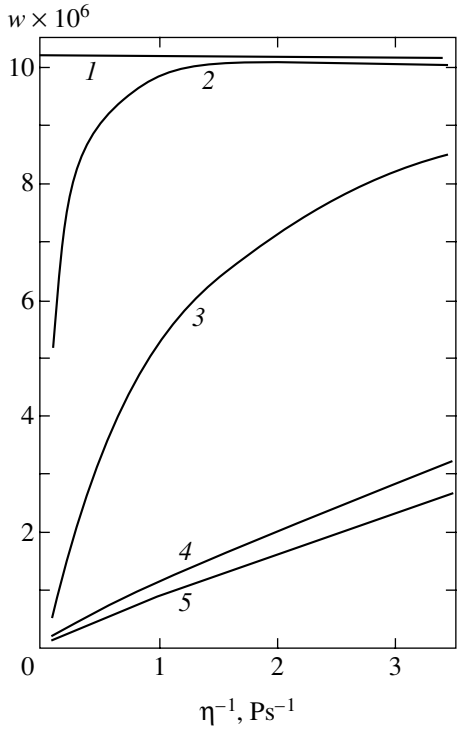


Fig. 8. The probability (per 1 neV) of quasi-elastic upscattering of a neutron with an energy of 50 neV incident on the liquid surface at an angle of $\pi/4$ with a change in energy of $\Delta E = 50$ neV as a function of inverse viscosity η_0^{-1}; each curve for constant $\tau = 10^{-3}$ (*1*), 10^{-5} (*2*), 10^8 (*3*), 10^{-7} (*4*), and 10^{-8} s (*5*).

time τ and shear modulus G shows that this proportionality is observed only in the τ range $(1–3) \times 10^{-7}$ s and the G range 10^6–10^8 din/cm^2. This behavior is demonstrated in Fig. 8, showing the probability of quasi-elastic upscattering of a neutron with an energy of 50 neV incident on the liquid surface at an angle of $\pi/4$ with a change in energy of ΔE = 50 neV as a function of the inverse viscosity η_0^{-1}; each curve corresponds to a constant τ. A similar picture is observed when the probability of upscattering is presented as curves of constant G.

Computational analysis shows that the data for the probability of UCN quasi-elastic scattering on a liquid surface as a function of incident neutron energy and energy transfer allow determination of the Maxwell model dynamic parameters of a viscous liquid.

The author wishes to thank V.V. Nitz for his kind introduction on the use of channel 6B of reactor IBR-2 and for valuable consultations; V.G. Simkin for help in the viscosity measurements; G.F. Syrykh, who presented the author with mica samples; and A. Steyerl, B.G. Erozolimsky, and P. Geltenbort for their interest in this work and discussions.

REFERENCES

1. Review of Particle Physics, Eur. Phys. J. C **15**, 54 (2000).
2. K. Schreckenbach and W. Mampe, J. Phys. G **18**, 1 (1992).
3. J. M. Pendlebury, Annu. Rev. Nucl. Part. Sci. **43**, 687 (1993).
4. B. G. Yerozolimsky, Nucl. Instrum. Methods Phys. Res. A **440**, 491 (2000).
5. H. Abele, Nucl. Instrum. Methods Phys. Res. A **440**, 499 (2000).
6. V. P. Alfimenkov, V. P. Varlamov, V. E. Vasil'ev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **52**, 984 (1990) [JETP Lett. **52**, 373 (1990)]; V. V. Nesvizhevsky, A. P. Serebrov, R. R. Tal'daev, *et al.*, Zh. Éksp. Teor. Fiz. **102**, 740 (1992) [Sov. Phys. JETP **75**, 405 (1992)].
7. W. Mampe, P. Ageron, C. Bates, *et al.*, Nucl. Instrum. Methods Phys. Res. A **284**, 111 (1989); Phys. Rev. Lett. **63**, 593 (1989).
8. A. Pichlmaier, J. Butterworth, P. Geltenbort, *et al.*, Nucl. Instrum. Methods Phys. Res. A **440**, 517 (2000).
9. W. Mampe, L. N. Bondarenko, V. I. Morozov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **57**, 77 (1993) [JETP Lett. **57**, 82 (1993)].
10. S. Arzumanov, L. Bondarenko, S. Chernyavsky, *et al.*, in *Proceedings of International Seminar on Interaction of Neutrons with Nuclei: Neutron Spectroscopy, Nuclear Structure, Related Topics* (Dubna, 1997), p. 53; Phys. Lett. B **483**, 15 (2000); Nucl. Instrum. Methods Phys. Res. A **440**, 511 (2000).
11. P. Ageron, W. Mampe, and A. I. Kilvington, Z. Phys. B **59**, 261 (1985).
12. V. P. Alfimenkov, V. V. Nesvizhevski, A. P. Serebrov, *et al.*, Preprint No. 1729, LIYaF AN SSSR (Leningrad Inst. of Nuclear Physics, USSR Academy of Sciences, Gatchina, 1991); Pis'ma Zh. Éksp. Teor. Fiz. **55**, 92 (1992) [JETP Lett. **55**, 84 (1992)].
13. V. I. Morozov, *Talk at Mini-Workshop on UCN Anomalies: Where Do We Stand?* (ILL, Grenoble, 2000).
14. J. C. Bates, Phys. Lett. A **88**, 427 (1982); Nucl. Instrum. Methods Phys. Res. A **216**, 535 (1983).
15. P. Ageron, W. Mampe, J. C. Bates, and J. M. Pendlebury, Nucl. Instrum. Methods Phys. Res. A **249**, 261 (1986).
16. Ausimont Corporation (see www.ausimont.com).
17. A. Pichlmaier, V. Nesvizhevsky, S. Neumaier, *et al.*, in *Proceedings of International Seminar on Interaction of Neutrons with Nuclei: Neutron Spectroscopy, Nuclear Structure, Related Topics* (Dubna, 1997), p. 34.
18. D. J. Richardson, J. M. Pendlebury, P. Iaydjiev, *et al.*, Nucl. Instrum. Methods Phys. Res. A **308**, 568 (1991).
19. F. Tervisidis and N. Tsagas, Nucl. Instrum. Methods Phys. Res. A **305**, 433 (1991).
20. G. J. Cuello, J. R. Sanstisteban, R. E. Mayer, and T. R. Granada, Nucl. Instrum. Methods Phys. Res. A **357**, 519 (1995).
21. Yu. N. Pokotilovski, Nucl. Instrum. Methods Phys. Res. A **425**, 320 (1999).
22. K. S. Sung and R. J. Lagow, J. Am. Chem. Soc. **117**, 4276 (1995); K. S. Sung and R. J. Lagow, Synth. Commun. **26**, 375 (1996).
23. A. Steyerl, S. S. Malik, P. Geltenbort, *et al.*, *ILL Annual Report 1996* (Grenoble, 1997), p. 51; J. Phys. III **7**, 1941 (1997); A. Steyerl, in *Proceedings of International Seminar on Interaction of Neutrons with Nuclei ISINN-6: Neutron Spectroscopy, Nuclear Structure, Related Topics* (Dubna, 1998), p. 74; T. Bestle, P. Geltenbort, H. Yust, *et al.*, Phys. Lett. A **244**, 217 (1998); L. N. Bondarenko, V. V. Morozov, E. Korobkina, *et al.*, *ILL Experimental Report No. 3-14-44* (Grenoble, 1997); in *Proceedings of International Seminar on Interaction of Neutrons with Nuclei ISINN-6: Neutron Spectroscopy, Nuclear Structure, Related Topics* (Dubna, 1998), p. 101; L. Bondarenko, P. Geltenbort, E. Korobkina, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 663 (1998) [JETP Lett. **68**, 691 (1998)]; V. V. Nesvizhevsky, A. V. Strelkov, P. Geltenbort, and P. Iaydjiev, *ILL Annual Report 1997* (Grenoble, 1998), p. 62; Yad. Fiz. **62**, 832 (1999) [Phys. At. Nucl. **62**, 776 (1999)].
24. V. A. Ponomarenko, S. P. Krukovskii, and A. Yu. Alybina, *Fluorine Containing Heterochain Polymers* (Nauka, Moscow, 1973).
25. G. Barr, *A Monograph of Viscometry* (Oxford Univ. Press, London, 1931; Gl. Red. Khim. Lit., Leningrad, 1938).
26. F. M. Zelenyuk, K. H. Zaĭtsev, A. V. Timakov, *et al.*, Prib. Tekh. Éksp., No. 2, 57 (1973).
27. F. M. Zelenyuk, V. E. Zhitarev, S. B. Stepanov, and A. V. Timakov, Prib. Tekh. Éksp., No. 4, 43 (1973).
28. Yu. N. Pokotilovski, M. I. Novopoltsev, and P. Geltenbort, Nucl. Instrum. Methods Phys. Res. A (in press).

29. Yu. N. Pokotilovski, Phys. Lett. A **255**, 173 (1999).

30. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics*, 3rd ed. (Nauka, Moscow, 1986; Pergamon, New York, 1987).

31. V. G. Levich, *Physicochemical Hydrodynamics* (Fizmatgiz, Moscow, 1959).

32. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 7: *Theory of Elasticity*, 4th ed. (Pergamon, New York, 1986; Nauka, Moscow, 1987).

33. Yu. A. Bykovsky, E. A. Manykin, I. E. Nakhutin, *et al.*, Zh. Tekh. Fiz. **46**, 2211 (1976) [Sov. Phys. Tech. Phys. **21**, 1302 (1976)].

34. C. F. Tejero, M. J. Rodriguez, and M. Baus, Phys. Lett. A **98**, 371 (1983).

35. J. L. Harden, H. Pleiner, and P. A. Pincus, J. Chem. Phys. **94**, 5208 (1991).

36. G. Platero, V. R. Velasco, and F. Garcia-Moliner, Phys. Scr. **23**, 1108 (1981).

37. H. Pleiner, J. L. Harden, and P. Pincus, Europhys. Lett. **7**, 383 (1988).

38. U-Ser Jeng, L. Esibov, L. Crow, and A. Steyerl, J. Phys.: Condens. Matter **10**, 4955 (1998).

39. A. P. Serebrov, D. Butterworth, V. E. Varlamov, *et al.*, Preprint No. 2438, PIYaF RAN (St. Petersburg Inst. of Nuclear Physics, Russian Academy of Sciences, Gatchina, 2001).

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 180–192.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 211–223.

NUCLEI, PARTICLES, AND THEIR INTERACTION

The Vacuum–Vacuum Amplitude and Bogoliubov Coefficients¶

A. I. Nikishov

Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 119991 Russia
e-mail: nikishov@lpi.ru

Received July 18, 2002

Abstract—We consider the problem of fixing the phases of Bogoliubov coefficients in quantum electrodynamics such that the vacuum–vacuum amplitude can be expressed via them. For a constant electric field and particles with spins of 0 and 1/2, this is done starting from the definition of these coefficients. Using the symmetry between electric and magnetic fields, we extend the result to a constant electromagnetic field. It turns out that for a constant magnetic field, it is necessary to distinguish the in- and out-states, although they differ only by a phase factor. For a spin-1 particle with a gyromagnetic of ratio $g = 2$, this approach fails and we reconsider the problem using the proper-time method. *© 2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Even if the electromagnetic field does not create pairs, virtual pairs lead to the appearance of a phase in the vacuum–vacuum amplitude. This makes it necessary to distinguish the in- and out-solutions even when it is commonly assumed that there is only one complete set of solutions as, e.g., in the case of a constant magnetic field. The in- and out-solutions then differ only by a phase factor, that is, in essence, the Bogoliubov coefficient. In terms of the in- and out-states, the propagator takes the same form as for pair-creating fields.

We use the solutions with conserved quantum numbers and do not consider radiation processes. Then the events in a cell with quantum numbers n are independent of the events in cells with different quantum numbers. In other words, we work in the diagonal representation. Knowledge of the Bogoliubov coefficients is sufficient for obtaining the probability of any process in the external field (disregarding the radiation processes) [1–3], but the real part of the action integral W that defines the vacuum–vacuum amplitude,

$$\langle 0_{\rm out}|0_{\rm in}\rangle = e^{iW}, \quad W = \int d^4x\mathscr{L}, \tag{1}$$

is not directly expressed via Bogoliubov coefficients. At the same time, some effects related to ReW are observable. Thus, the Lagrange function $\mathscr{L}$ of a slowly varying field determines the dielectric permittivity and magnetic permeability of the field [4, 5].

The Lagrange function of a constant electromagnetic field was obtained in [6–8] in the one-loop approximation and in [9] in the two-loop approximation. Studying a model of particle production, De Witt noted that ReW can be expressed via Bogoliubov coefficients with the natural choice of their phases [10]. Our purpose is to choose these phases such that ReW can be expressed via them. We show that for the constant electric field and particles with spins 0 and 1/2, the natural choice would be sufficient if it were not for the necessity of making renormalizations. For a vector boson with the gyromagnetic ratio $g = 2$, the situation is more complicated even for a constant electric field.

We note that the transition amplitude for an electron to go from an in-state to an out-state is equal to unity. To show this, we write the Bogoliubov transformations and the relation between $\langle 0_{n\,\rm out}|$ and $\langle 0_{n\,\rm in}|$ [2] (where n is the set of quantum numbers)

$$\begin{aligned} a_{n\,\rm out} &= c_{1n}a_{n\,\rm in} - c_{2n}^*b_{n\,\rm in}^+, \\ b_{n\,\rm out}^+ &= c_{2n}a_{n\,\rm in} + c_{1n}^*b_{n\,\rm in}^+, \\ \langle 0_{n\,\rm out}| &= \langle 0_{n\,\rm in}|(c_{1n}^* - c_{2n}a_{n\,\rm in}b_{n\,\rm in}), \end{aligned} \tag{1'}$$

where

$$|c_{1n}|^2 + |c_{2n}|^2 = 1.$$

Here, $a_{n\,\rm in}(b_{n\,\rm in}^+)$ is the particle (antiparticle) annihilation (creation) operator, $a_{n\,\rm in}|0_{n\,\rm in}\rangle = 0$, and similarly for the out-states; $|0_{n\,\rm in}\rangle$ is the vacuum state in the cell with the quantum number n, c_{1n} and c_{2n} are the Bogoliubov coefficients, and the asterisk denotes complex conjugation.

The third relation in (1') implies Eq. (28) below, and the first relation implies that

$$a_{n\,\rm in}^+ = c_{1n}^{*-1}[a_{n\,\rm out}^+ + c_{2n}b_{n\,\rm in}].$$

Using this relation and the anticommutator

¶This article was submitted by the author in English.

$\{a_{n'\,\mathrm{out}}, a^+_{n\,\mathrm{in}}\} = \delta_{n',n}$, we find [2]

$$\langle 0_{n\,\mathrm{out}}|a_{n\,\mathrm{out}}a^+_{n\,\mathrm{in}}|0_{n\,\mathrm{in}}\rangle = c^{*-1}_{1n}\langle 0_{n\,\mathrm{out}}|0_{n\,\mathrm{in}}\rangle = 1. \tag{1"}$$

The Pauli principle prohibits virtual pair creation in the state occupied by the electron. Therefore, even the phase of the scattering amplitude remains unchanged. In particular, $c_{2n} = 0$ for the constant magnetic field, but we cannot assume that $c_{1n} = 1$ without violating Eq. (1") and Eqs. (28), (29) below because $W \neq 0$ [4, 5]. In other words, even if $c_{2n} = 0$, the in- and out-vacua are different. (This is in contrast to the remark after Eq. (15) in [10].) The Bogoliubov coefficient c_{1n} must therefore be coordinated with the vacuum–vacuum amplitude. For the constant electromagnetic field, we represent the action integral as a sum over the set of quantum numbers n,

$$W = \int d^4x\mathcal{L}(x) = \Sigma_n W_n.$$

Then W_n define the phase of the Bogoliubov coefficient (in general, complex).

In Sections 2 and 3, starting from the definition of the Bogoliubov coefficients, we consider the phase fixing for particles with respective spins of 0 and 1/2. In Sections 4–6, we reconsider the problem using a more general proper-time method for spins 0, 1/2, and 1.

2. SCALAR PARTICLE IN THE CONSTANT ELECTROMAGNETIC FIELD

For a set of wave functions with conserved quantum numbers n, the Bogoliubov transformation is given by

$$\begin{aligned} {}_+\psi_n &= c_{1n}{}^+\psi_n + c_{2n}{}^-\psi_n, \\ {}_-\psi_n &= c^*_{2n}{}^+\psi_n + c^*_{1n}{}^-\psi_n, \end{aligned} \tag{2}$$

where

$$|c_{1n}|^2 - |c_{2n}|^2 = 1$$

and ${}_+\psi_n$ (${}^+\psi_n$) is the positive-frequency in- (out-) solution, and similarly for the negative-frequency states.

We are free to choose the phase of c_{1n} by redefining ψ_n. Indeed, if we substitute

$${}_\pm\psi_n = e^{\pm if}{}_\pm\psi^{\mathrm{new}}_n, \quad {}^\pm\psi_n = e^{\mp if}{}^\pm\psi^{\mathrm{new}}_n,$$

$$c_{1n} = e^{i2f}c^{\mathrm{new}}_{1n},$$

then Eq. (2) and the propagator [2, 11]

$$G_0(x, x') = i\Sigma_n c^{*-1}_{1n}\begin{cases} {}^+\psi_n(x){}_+\psi^*_n(x'), & t > t' \\ {}_-\psi_n(x){}^-\psi^*_n(x'), & t < t' \end{cases} \tag{2'}$$

retain their form in terms of the redefined quantities.

For definiteness, we assume that the particle charge is $e' = -e$, $e = |e|$. For a constant electric field, we then have [2] ($n = (p_1, p_2, p_3)$, $A_\mu = -\delta_{\mu 3}Et$)

$$\begin{aligned} c_{1n} &= \frac{\sqrt{2\pi}}{\Gamma\left(\frac{1}{2} - i\kappa\right)}\exp\left(-\frac{\pi\kappa}{2} + i\frac{\pi}{4}\right), \\ c_{2n} &= \exp\left(-\pi\kappa - i\frac{\pi}{2}\right), \quad \kappa = \frac{m^2 + p_1^2 + p_2^2}{2eE}. \end{aligned} \tag{3}$$

We note that in a weak electric field, $|c_{2n}|$ is exponentially small and can be neglected. The in- and out-states then differ only by a phase factor. The same must be true for the magnetic field, where $c_{2n} = 0$ exactly and $\ln c^*_{1n}$ is to be determined.

The amplitude of probability that the vacuum in the state n remains vacuum is [2]

$$\langle 0_{n\,\mathrm{out}}|0_{n\,\mathrm{in}}\rangle = c^{*-1}_{1n}. \tag{4}$$

The total vacuum–vacuum amplitude is

$$\begin{aligned} \langle 0_{\mathrm{out}}|0_{\mathrm{in}}\rangle &= \prod_n c^{*-1}_{1n} = e^{iW_0}, \\ W_0 &= \sum_n W_{0n}, \quad W_{0n} = i\ln c^*_{1n}. \end{aligned} \tag{5}$$

As we see below, c^*_{1n} must be replaced by $C^{*\mathrm{ren}}_{1n}$ in (4) and (5). This is the renormalization of c^*_{1n}. From (3), we have

$$\ln c^*_{1n} = \frac{1}{2}\ln 2\pi - \frac{\pi\kappa}{2} - \frac{i\pi}{4} - \ln\Gamma\left(\frac{1}{2} + i\kappa\right). \tag{6}$$

As shown in [2], the vacuum–vacuum probability $|\langle 0_{\mathrm{out}}|0_{\mathrm{in}}\rangle|^2$ obtained from (5) and (3) agrees with the Schwinger result [8]. This implies that $\mathrm{Im}\,W_0$ is correctly given by (5) and (3). To find $\mathrm{Re}\,W_0$, we first consider the asymptotic representation (see Eq. (1.3.12) in [12])

$$\begin{aligned} \ln\Gamma\left(\frac{1}{2} + i\kappa\right) &= i\kappa[\ln(i\kappa) - 1] + \frac{1}{2}\ln 2\pi \\ &+ \sum_{k=1}\frac{B_{2k}(1/2)}{2k(2k-1)}(i\kappa)^{1-2k}. \end{aligned} \tag{7}$$

(Letting k range to ∞, we can say that the right-hand side of (7) represents the left-hand side in a certain sense exactly; the information encoded in the right-

hand side can be decoded [13].) From (6) and (7), it follows that

$$\ln c_{1n}^* = -i\left[\kappa(\ln\kappa - 1) + \frac{\pi}{4} + \sum_{k=1}\frac{(-1)^k B_{2k}(1/2)}{2k(2k-1)\kappa^{2k-1}}\right]. \quad (8)$$

This asymptotic expansion contains only the imaginary part of $\ln c_{1n}^*$ or only the real part of W_{0n}. It can be seen from (8) that, as the first step, we must pass from $\ln c_{1n}^*$ to

$$\ln C_{1n}^* = \ln c_{1n}^* + i\left[\kappa(\ln\kappa - 1) + \frac{\pi}{4}\right] \quad (9)$$

in order to have $\ln C_{1n}^* \longrightarrow 0$ as $\kappa \longrightarrow \infty$ (i.e., as $E \longrightarrow 0$). Because charge renormalization is necessary, we must make the second step and introduce

$$\ln C_{1n}^{*\mathrm{ren}} = \ln c_{1n}^* + i\left[\kappa(\ln\kappa - 1) + \frac{\pi}{4} + \frac{1}{24\kappa}\right]. \quad (10)$$

In other words, we also let $\ln C_{1n}^{*\mathrm{ren}}$ contain the term with $k = 1$ in (8). We then have the asymptotic representation

$$\ln C_{1n}^{*\mathrm{ren}} = -i\sum_{k=2}\frac{(-1)^k B_{2k}(1/2)}{2k(2k-1)\kappa^{2k-1}}. \quad (11)$$

Summing (11) over n as

$$\sum_k \longrightarrow \int\frac{d^3pL^3}{(2\pi)^3}, \quad \int dp_3 \longrightarrow eET, \quad (12)$$

and making renormalization [8], we obtain the correct asymptotic representation for $\mathrm{Re}\,\mathscr{L}_0$,

$$\mathrm{Re}\,\mathscr{L}_0 = \frac{1}{2}E^2 + \frac{(eE)^2}{16\pi^2}\sum_{k=2}\frac{(-1)^k B_{2k}(1/2)}{k(k-1)(2k-1)\kappa_0^{2k-2}}, \quad \kappa_0 = \frac{m^2}{2eE}. \quad (13)$$

To simplify formulas and minimize confusion with T in Eq. (50), we often set $L = T = 1$ in expressions like (12). In addition, we drop the Maxwell part of the Lagrangian in what follows ($E^2/2$ in this case).

We now show that expression (9) can be brought to the form suggested by the proper-time formalism,

$$\ln C_{1n}^* \equiv \ln\sqrt{2\pi} + \eta(\ln\eta - 1) - \ln\Gamma\left(\frac{1}{2} + \eta\right) = -F(\eta),$$
$$F(\eta) = \frac{1}{2}\int_0^\infty\frac{d\theta}{\theta}e^{-2\eta\theta}\left(\frac{1}{\sinh\theta} - \frac{1}{\theta}\right), \quad \eta = i\kappa. \quad (14)$$

Differentiating (14) with respect to η and using Eq. (2.4.22.5) in [14], we see that the results in the left- and the right-hand sides coincide. In addition, both sides have the same asymptotic behavior as $\eta \longrightarrow \infty$. We therefore have

$$\ln C_{1n}^* = -\frac{1}{2}\int_0^\infty\frac{ds}{s\sinh\theta}\exp[-is(m^2 + p_\perp^2)]\left[1 - \frac{\sinh\theta}{\theta}\right], \quad \theta = eEs, \quad p_\perp^2 = p_1^2 + p_2^2. \quad (15)$$

Next, we note that the term $i/24\kappa$ in (10) can be written as

$$\frac{i}{24\kappa} = -\frac{1}{12}\int_0^\infty d\theta e^{-2i\kappa\theta}, \quad (16)$$

and therefore,

$$\ln C_{1n}^{*\mathrm{ren}} = -\frac{1}{2}\int_0^\infty\frac{ds}{s\sinh\theta}\exp[-is(m^2 + p_\perp^2)]R(\theta), \quad R(\theta) = 1 - \left(\frac{1}{\theta} - \frac{\theta}{6}\right)\sinh\theta. \quad (17)$$

Here, $R(\theta)$ is a "regulator." It is independent of the quantum numbers n and is the same as in the proper-time representation of the Lagrange function [8].

We now consider the case where a constant magnetic field is collinear with a constant electric field. Then

$$\ln C_{1n}^{*\mathrm{ren}}(E, H) = -\frac{1}{2}\int_0^\infty\frac{ds}{s\sinh\theta} \times \exp\{-is[m^2 + eH(2l+1)]\}R(\theta, \tau), \quad \tau = eHs, \quad l = 0, 1, \ldots, \quad (18)$$

and we assume that $R(\theta, \tau)$ can be obtained by the same reasoning as in [8] (or simply taken from [8]),

$$R(\theta, \tau) = 1 - \left(\frac{1}{\theta\tau} + \frac{1}{6}\frac{H^2 - E^2}{EH}\right)\sinh\theta\sin\tau, \quad \tau = eHs, \quad \theta = eEs. \quad (19)$$

Integrating over p_3, we obtain (see (12) with $T = 1$)

$$\int dp_3 \ln C_{1n}^{*\mathrm{ren}}(E, H) = -\frac{1}{2}eE\int_0^\infty\frac{ds}{s\sin\theta} \times \exp\{-is[m^2 + eH(2l+1)]\}R(\theta, \tau). \quad (20)$$

In this expression, we can turn the electric field off,

$$\int dp_3 \ln C_{1n}^{*\,\mathrm{ren}}(E=0,H) = -\frac{1}{2}\int_0^\infty \frac{ds}{s^2}\exp\{-is[m^2+eH(2l+1)]\}R(0,\tau), \quad (21)$$

$$R(0,\tau) = 1-\left(\frac{1}{\tau}+\frac{\tau}{6}\right)\sin\tau.$$

To remove the integration over p_3, we write the factor s^{-2} as $s^{-3/2}s^{-1/2}$ and note that $1/\sqrt{s}$ must arise from the integration over p_3,

$$\frac{1}{\sqrt{s}} = \frac{e^{i\pi/4}}{\sqrt{\pi}}\int_{-\infty}^{\infty} dp_3 \exp(-isp_3^2). \quad (22)$$

Therefore,

$$\ln C_{1n}^{*\,\mathrm{ren}}(E=0,H) = -\frac{e^{i\pi/4}}{2\sqrt{\pi}}\int_0^\infty \frac{ds}{s^{3/2}} \times \exp\{-is[m^2+eH(2l+1)+p_3^2]\}R(0,\tau). \quad (23)$$

(Substituting $s \longrightarrow -it$, we see that expression (23) is purely imaginary.) From here, or from (21), we obtain

$$i\sum_n \ln C_{1n}^{*\,\mathrm{ren}}(E=0,H) = i\int\frac{dp_2}{2\pi}\int\frac{dp_3}{2\pi} \times \sum_{l=0}^{\infty}\ln C_{1n}^{*\,\mathrm{ren}}(E=0,H) = \mathcal{L}_0 = -\frac{eH}{16\pi^2}\int_0^\infty \frac{ds}{s^2\sin\tau}\exp(-ism^2)R(0,\tau) \quad (L=T=1), \quad (24)$$

which agrees with [8, 9]. Relation (39) below was used here, and the sum over l was performed with the help of the formula

$$\sum_{l=0}^{\infty}\exp[-iseH(2l+1)] = \frac{1}{2i\sin(eHs)}. \quad (25)$$

3. AN ELECTRON IN THE CONSTANT ELECTROMAGNETIC FIELD

The Bogoliubov transformation is given by

$$\begin{aligned} {}_+\psi_n &= c_{1n}\,{}^+\psi_n + c_{2n}\,{}^-\psi_n, \\ {}_-\psi_n &= -c_{2n}^*\,{}^+\psi_n + c_{1n}^*\,{}^-\psi_n, \end{aligned} \quad (26)$$

where

$$|c_{1n}|^2 + |c_{2n}|^2 = 1.$$

For the constant electric field, we have

$$c_{1n}^* = -i\sqrt{\frac{2\pi}{\kappa}}\frac{e^{-\pi\kappa/2}}{\Gamma(i\kappa)}, \quad c_{2n} = e^{-\pi\kappa}, \quad n=(p_1,p_2,p_3,r). \quad (27)$$

These Bogoliubov coefficients are independent of the spin state index $r = 1, 2$.

As in the scalar case, we start with the relations [2]

$$\langle 0_{n\,\mathrm{out}}|0_{n\,\mathrm{in}}\rangle = c_{1n}^*, \quad (28)$$

and

$$\langle 0_{\mathrm{out}}|0_{\mathrm{in}}\rangle = \prod_n c_{1n}^* = e^{iW_{1/2}}, \quad W_{1/2} = \sum_n W_{1/2;\,n}, \quad W_{1/2;\,n} = -i\ln c_{1n}^*. \quad (29)$$

It follows from (27) that

$$\ln c_{1n}^* = -\frac{i\pi}{2}+\frac{1}{2}\ln\frac{2\pi}{\kappa}-\frac{\pi\kappa}{2}-\ln\Gamma(i\kappa). \quad (30)$$

The asymptotic expansion for $\Gamma(i\kappa)$ is

$$\ln\Gamma(i\kappa) = \left(i\kappa-\frac{1}{2}\right)\ln(i\kappa)-i\kappa+\frac{1}{2}\ln 2\pi + i\sum_{k=1}(-1)^k\frac{B_{2k}}{2k(2k-1)}(\kappa)^{1-2k} \quad (31)$$

(see Eq. (8.344) in [15] or Eq. (6.1.40) in [16]). From (30) and (31), we obtain

$$\ln C_{1n}^* \equiv \ln c_{1n}^* + i\left(\kappa\ln\kappa-\kappa+\frac{\pi}{4}\right) = -i\sum_{k=1}(-1)^k\frac{B_{2k}}{2k(2k-1)}(\kappa)^{1-2k}, \quad (32)$$

$$\ln C_{1n}^{*\,\mathrm{ren}} \equiv \ln C_{1n}^* - \frac{i}{12\kappa} = -i\sum_{k=2}(-1)^k\frac{B_{2k}}{2k(2k-1)}(\kappa)^{1-2k}. \quad (33)$$

As in the scalar case, we find that

$$\ln C_{1n}^* = -\frac{1}{2}\int_0^\infty \frac{dx}{x}e^{-2i\kappa x}\left(\coth x-\frac{1}{x}\right), \quad (34)$$

$$\ln C_{1n}^{*\mathrm{ren}} = -\frac{1}{2} \times \int_0^\infty \frac{dx}{x} e^{-2i\kappa x}\left[1-\left(\frac{1}{x}+\frac{x}{3}\right)\tanh x\right]\coth x. \quad (35)$$

Equation (2.4.22.6) in [14] was used to verify (34) (cf. the text before Eq. (15)).

The generalization of (35) to the presence of a constant magnetic field is straightforward. We rewrite it as ($x = \theta = eEs$)

$$\ln C_{1n}^{*\mathrm{ren}}(E, H) = -\frac{1}{2}\int_0^\infty \frac{d\theta}{\theta} \times \exp[-is(m^2 + 2eHl)]R(\theta, \tau)\coth\theta, \quad (36)$$

where $n = (p_1, p_2, p_3, r)$; $l = l_{\min}, l_{\min}+1, \ldots$; $l_{\min} = 0$ for $r = 1$, $l_{\min} = 1$ for $r = 2$, and $R(\theta, \tau)$ can be taken from the Lagrange function [8, 9] ($\tau = eHs$),

$$R(\theta, \tau) = 1 - \left(\frac{1}{\theta\tau} + \frac{E^2 - H^2}{3EH}\right)\tan\tau\tanh\theta. \quad (37)$$

Integrating over p_3 using the second equation in (12), we find

$$\int \frac{dp_3}{2\pi}\ln C_{1n}^{*\mathrm{ren}} = -\frac{eE}{4\pi}\int_0^\infty \frac{d\theta}{\theta} \times \exp[-is(m^2 + 2eHl)]R(\theta, \tau)\coth\theta. \quad (38)$$

The subsequent integration over p_2 is performed using a formula similar to (12) [2],

$$\int dp_2 = eHL. \quad (39)$$

To sum over r and l in (36), we use the formula

$$\sum_{r=1}^{2}\sum_{l_{\min}}^{\infty} e^{-2iseHl} = -i\cot(eHs) \quad (40)$$

that follows from (25). In agreement with the Lagrange function for the constant electromagnetic field [8, 9], we therefore have

$$\sum_n \ln C_{1n}^{*\mathrm{ren}} = i\frac{e^2EH}{8\pi^2}\int_0^\infty \frac{d\theta}{\theta} \times \exp(-ism^2)R(\theta, \tau)\coth\theta\cot\tau \quad (L = T = 1). \quad (41)$$

Returning to (38), we can switch the electric field off,

$$\int \frac{dp_3}{2\pi}\ln C_{1n}^{*\mathrm{ren}} = -\frac{1}{4\pi}\int_0^\infty \frac{ds}{s^2} \times \exp[-is(m^2 + 2eHl)]R(0, \tau), \quad R(0, \tau) = 1 - \left(\frac{1}{\tau} - \frac{\tau}{3}\right)\tan\tau. \quad (42)$$

Here, l are given in (36). As in the scalar case, using (22), we obtain

$$\ln C_{1n}^{*\mathrm{ren}}(E = 0, H) = -\frac{e^{i\pi/4}}{2\sqrt{\pi}}\int_0^\infty \frac{ds}{s^{3/2}} \times \exp[-is(m^2 + p_3^2 + 2eHl)]R(0, \tau), \quad (43)$$

where $n = (p_1, p_2, p_3, r)$, $l = 0, 1, 2, \ldots$ for $r = 1$, and $l = 1, 2, \ldots$ for $r = 2$.

In the subsequent sections, we give a heuristic derivation of $\ln C_{1n}^{*\mathrm{ren}}$ not resorting to c_{1n}^*, but using the proper-time method. The main problem occurring here is that renormalizations must be made. We know how to renormalize $\mathscr{L}$ as a whole, but we must renormalize the contribution to it from a particular state n. To do this, we assume, as before, that the regulator is independent of n.

4. SCALAR PARTICLE

We take the vector potential of a constant electromagnetic field in the form

$$A_\mu = \delta_{\mu 2}Hx_1 - \delta_{\mu 3}Et, \quad (44)$$

but start with the particle in a constant magnetic field, $E = 0$ in (44). The propagator with coinciding x and x' is given by (see, e.g., [11])

$$G_0(x, x|E = 0, H) = i\sqrt{\frac{eH}{\pi}}\sum_{l=0}^\infty \int_{-\infty}^\infty \frac{dp_2}{2\pi}\int_{-\infty}^\infty \frac{dp_3}{2\pi}\int_{-\infty}^\infty \frac{dp^0}{2\pi} \times \int_0^\infty ds \frac{D_l^2(\zeta)}{l!}\exp\{-is[m^2 + eH(2l+1) + p_3^2 - p_0^2]\}, \quad (45)$$

$$\zeta = \sqrt{2eH}\left(x_1 + \frac{p_2}{eH}\right).$$

In accordance with (1), we must integrate $\mathscr{L}_0$ and hence

$G_0(x, x)$ over d^4x. The integration over x_1 is done using the formula

$$\int_{-\infty}^{\infty} d\zeta D_l^2(\zeta) = \sqrt{2\pi} l!, \quad \text{or} \quad \int_{-\infty}^{\infty} dx_1 D_l^2(\zeta) = \sqrt{\frac{\pi}{eH}} l!. \tag{46}$$

Integrating over p^0 and x_1, we obtain

$$\int_{-\infty}^{\infty} dx_1 G_0(x, x) = \frac{\exp(3\pi i/4)}{2\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{dp_2}{2\pi} \int_{-\infty}^{\infty} \frac{dp_3}{2\pi} \times \sum_{l=0}^{\infty} \int_0^{\infty} \frac{ds}{\sqrt{s}} \exp\{-is[m^2 + eH(2l+1) + p_3^2]\}. \tag{47}$$

As noted in [3] (see Eq. (2.12) therein), it follows from Schwinger results [8] that for a scalar particle (boson),

$$-i\frac{\partial W_b}{\partial m^2} = \int d^4x G_b(x, x), \quad \text{or} \quad W_b = -i\int_{m^2}^{\infty} d\tilde{m}^2 \int d^4x G_b(x, x|\tilde{m}^2). \tag{48}$$

This implies that $\mathcal{L}_0$ can be obtained from (47) by inserting $-1/s$ in the integrand. Also inserting the regulator from (21), we obtain

$$\begin{aligned} iW_0(E=0, H) &= i\mathcal{L}_0(E=0, H) \\ &= \frac{\exp(\pi i/4)}{2\sqrt{\pi}} \int_{-\infty}^{\infty} \frac{dp_2}{2\pi} \int_{-\infty}^{\infty} \frac{dp_3}{2\pi} \sum_{l=0}^{\infty} \int_0^{\infty} \frac{ds}{s^{3/2}} \\ &\times \exp\{-is[m^2 + eH(2l+1) + p_3^2]\} R(0, \tau) \\ &= -\int_{-\infty}^{\infty} \frac{dp_2}{2\pi} \int_{-\infty}^{\infty} \frac{dp_3}{2\pi} \sum_{l=0}^{\infty} \ln C_{1n}^{*\mathrm{ren}} \quad (L = T = 1). \end{aligned} \tag{49}$$

For the constant electromagnetic field described by vector potential (44), we now insert the expressions for the wave functions in (2') (see [2] with the modifications for $e' = -e = -|e|$) and use relation (93) in [11] (or a relation similar to (96) below). We then find

$$\begin{aligned} G_0(x, x|E, H) &= \frac{e^{3\pi i/4}}{2\sqrt{\pi eE}} \int_{-\infty}^{\infty} \frac{dp_2}{2\pi} \int_{-\infty}^{\infty} \frac{dp_3}{2\pi} \sum_{l=0}^{\infty} \sqrt{\frac{eH}{\pi}} \frac{D_l^2}{l!} \sqrt{2} \\ &\times \int_0^{\infty} \frac{d\theta}{\sqrt{\sinh 2\theta}} \exp\left(-2i\kappa\theta - i\frac{T^2}{2\coth\theta}\right), \\ \theta &= eEs, \quad T = \sqrt{2eE}\left(t - \frac{p_3}{eE}\right). \end{aligned} \tag{50}$$

Integrating over x_1 (see (46)) and t, we obtain

$$\int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dt G_0(x, x|E, H) = \frac{i}{2} \int_{-\infty}^{\infty} \frac{dp_2}{2\pi} \int_{-\infty}^{\infty} \frac{dp_3}{2\pi} \times \sum_{l=0}^{\infty} \int_0^{\infty} \frac{ds}{\sinh(eEs)} \exp\{-is[m^2 + eH(2l+1)]\}. \tag{51}$$

Passing from $G_0(x, x)$ to $\mathcal{L}_0$ is realized by inserting the factor $-1/s$ in the integrand in (51). Also inserting the regulator $R(\tau, \theta)$, see Eq. (19), we obtain

$$\begin{aligned} W_0(E, H) &= i\int_{-\infty}^{\infty} \frac{dp_2}{2\pi} \int_{-\infty}^{\infty} \frac{dp_3}{2\pi} \sum_{l=0}^{\infty} \ln C_{1n}^{*\mathrm{ren}} \\ &= -\frac{i}{2} \int_{-\infty}^{\infty} \frac{dp_2}{2\pi} \int_{-\infty}^{\infty} \frac{dp_3}{2\pi} \sum_{l=0}^{\infty} \int_0^{\infty} \frac{ds}{s\sinh\theta} \\ &\times \exp\{-is[m^2 + eH(2l+1)]\} R(\tau, \theta). \end{aligned} \tag{52}$$

5. SPINOR PARTICLE

We first consider the electron in the constant magnetic field, $E = 0$ in (44). The squared Dirac equation can be brought to the form

$$\left\{\frac{d^2}{d\zeta^2} - \frac{\zeta^2}{4} + \frac{p_0^2 - p_3^2}{2eH} - \frac{1}{2}\Sigma_3\right\} Z = 0, \quad \Sigma_3 = \begin{pmatrix} \sigma_3 & 0 \\ 0 & \sigma_3 \end{pmatrix}, \tag{53}$$

where ζ is the same as in (45). We see that Z can be written as

$$Z = \mathrm{diag}(f_1, f_2, f_1, f_2) \times \exp[i(p_2x_2 + p_3x_3 - p^0t)], \tag{54}$$

and f_1 and f_2 must satisfy the equation

$$\left\{\frac{d^2}{d\zeta^2} - \frac{\zeta^2}{4} + \frac{p_0^2 - p_3^2}{2eH} \mp \frac{1}{2}\right\} f_{1,2} = 0. \tag{55}$$

We choose $f_1 = D_{l-1}(\zeta)$ and $f_2 = D_l(\zeta)$ in order that $p_\perp^2 = 2eHl$ in both cases. The solutions of the Dirac equation are obtained as the columns of the matrix [2],

$$(m - i\hat{\Pi})Z, \tag{56}$$

where $\hat{\Pi} = \gamma_\mu \Pi_\mu$, $\Pi_\mu = -i\partial_\mu + eA_\mu$.

Using the γ matrices in the standard representation [4], we have

$$m - i\hat{\Pi} = \begin{pmatrix} m+\Pi^0 & 0 & -\Pi_3 & -\Pi_1 + i\Pi_2 \\ 0 & m+\Pi^0 & -\Pi_1 - i\Pi_2 & \Pi_3 \\ \Pi_3 & \Pi_1 - i\Pi_2 & m-\Pi^0 & 0 \\ \Pi_1 + i\Pi_2 & -\Pi_3 & 0 & m-\Pi^0 \end{pmatrix}. \quad (57)$$

In terms of ζ, we obtain

$$\Pi_1 + i\Pi_2 = -i\sqrt{2eH}\left(\frac{d}{d\zeta} - \frac{\zeta}{2}\right),$$
$$\Pi_1 - i\Pi_2 = -i\sqrt{2eH}\left(\frac{d}{d\zeta} + \frac{\zeta}{2}\right). \quad (58)$$

Also using the relations

$$\left(\frac{d}{d\zeta} + \frac{\zeta}{2}\right)D_l(\zeta) = lD_{l-1}(\zeta),$$
$$\left(\frac{d}{d\zeta} - \frac{\zeta}{2}\right)D_l(\zeta) = -D_{l+1}(\zeta), \quad (59)$$

we find (with the exponential factor in (54) omitted for brevity)

$$(m - i\hat{\Pi})Z = \begin{pmatrix} (m+p^0)D_{l-1}(\zeta) & 0 & -p_3 D_{l-1}(\zeta) & il\sqrt{2eH}D_{l-1}(\zeta) \\ 0 & (m+p^0)D_l(\zeta) & -i\sqrt{2eH}D_l(\zeta) & p_3 D_l(\zeta) \\ p_3 D_{l-1}(\zeta) & -il\sqrt{2eH}D_{l-1}(\zeta) & (m-p^0)D_{l-1}(\zeta) & 0 \\ i\sqrt{2eH}D_l(\zeta) & -p_3 D_l(\zeta) & 0 & (m-p^0)D_l(\zeta) \end{pmatrix}. \quad (60)$$

Choosing the second and the first columns as ψ_1 and ψ_2 (with the subscripts 1 and 2 indicating spin states) and normalizing them, we obtain

$${}_+\psi_1 = N_n \begin{bmatrix} 0 \\ (m+p^0)D_l(\zeta) \\ -il\sqrt{2eH}D_{l-1}(\zeta) \\ -p_3 D_l(\zeta) \end{bmatrix} e^{iq\cdot x},$$
$$N_n = \left(\frac{eH}{\pi}\right)^{1/4}\sqrt{\frac{1}{2p^0(p^0+m)l!}}, \quad (61)$$
$$p^0 = \sqrt{m^2 + 2eHl + p_3^2}, \quad q\cdot x = p_2 x_2 + p_3 x_3 - p^0 t,$$
$$n = (p_2, p_3, l, r), \quad \zeta = \sqrt{2eH}\left(x_1 + \frac{p_2}{eH}\right),$$

$${}_+\psi_2 = N_n\sqrt{l}\begin{bmatrix} (m+p^0)D_{l-1}(\zeta) \\ 0 \\ p_3 D_{l-1}(\zeta) \\ i\sqrt{2eH}D_{l-1}(\zeta) \end{bmatrix} e^{iq\cdot x}, \quad (62)$$
$$l = 0, 1, 2, \ldots .$$

As can be seen from (62), l actually begins with unity in this state. The negative-frequency solutions ${}_-\psi_n$ are obtained from (61) and (62) by the substitution $q \longrightarrow -q$. We note that Eqs. (61) and (62) differ from Eq. (10.5.9) in [4] because the authors there assumed the charge of a spinor particle to be positive.

Having obtained the wave functions, next we find the contribution to $\mathscr{L}_{1/2}$ from each state ψ_n. For the field that does not create pairs, the propagator has the standard form

$$G_{1/2}(x, x') = i\Sigma_n \begin{cases} {}_+\psi_n(x){}_+\bar{\psi}_n(x'), & t > t' \\ -{}_-\psi_n(x){}_-\bar{\psi}_n(x'), & t < t', \end{cases} \quad (63)$$
$$\bar{\psi}_n = \psi_n^*\beta.$$

In the standard representation, we have

$$\beta = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (64)$$

From (61) and (64), we find

$$\mathrm{tr}\,{}_+\psi_1(x){}_+\bar{\psi}_1(x) = N_n^2\{[(m+p^0)^2 - p_3^2]D_l^2(\zeta) - 2eHl^2 D_{l-1}^2(\zeta)\}. \quad (65)$$

Integrating over x_1, we obtain, see (46),

$$\int_{-\infty}^{\infty} dx_1 \mathrm{tr}\,{}_+\psi_1(x){}_+\bar{\psi}_1(x) = \frac{m}{p^0}, \quad (66)$$
$$p^0 = \sqrt{m^2 + 2eHl + p_3^2}, \quad l = 0, 1, \ldots .$$

From (62), we obtain, similarly,

$$\int_{-\infty}^{\infty} dx_1 \mathrm{tr}_+ \psi_2(x)_+ \bar{\psi}_2(x) = \frac{m}{p^0}, \quad l = 1, 2, \ldots . \tag{67}$$

For the negative-frequency states, we must substitute $p^0 \longrightarrow -p^0$. We can then write

$$\frac{1}{|p^0|} = \frac{e^{i\pi/4}}{\sqrt{\pi}} \int_0^{\infty} \frac{ds}{\sqrt{s}} \exp[-is(m^2 + 2eHl + p_3^2)], \tag{68}$$

incorporating both lines in (63). It thus follows from (63) and (66)–(68) that

$$\int_{-\infty}^{\infty} dx_1 \mathrm{tr} G_{1/2}(x, x) = \sum_n \frac{e^{3i\pi/4}}{\sqrt{\pi}} m \int_0^{\infty} \frac{ds}{\sqrt{s}} \times \exp[-is(m^2 + 2eHl + p_3^2)], \tag{69}$$

where $l = 0, 1, \ldots$ for $r = 1$ and $l = 1, 2, \ldots$ for $r = 2$. We next use the analogue of (48) for the electron,

$$W_{1/2} = i \int_m^{\infty} d\tilde{m} \mathrm{Tr} G_{1/2}(x, x|\tilde{m}), \tag{70}$$

where Tr means the integration over d^4x and the trace over spin indices; we set $VT = 1$ as above. Because

$$i \int_m^{\infty} d\tilde{m} \tilde{m} \exp(-is\tilde{m}^2) = \frac{e^{-ism^2}}{2s}, \tag{70'}$$

we see that $W_{1/2}$ can be obtained from (69) by inserting the factor $1/2ms$ in the integrand. We therefore find

$$\mathscr{L}_{1/2} = \Sigma_n \frac{e^{3i\pi/4}}{2\sqrt{\pi}} \int_0^{\infty} \frac{ds}{s^{3/2}} \times \exp[-i(m^2 + 2eHl + p_3^2)] R(0, \tau). \tag{71}$$

This is in agreement with (43) and (29). To check this result, we integrate over $dp_2/2\pi$ with the help of (39), over $dp_3/2\pi$ with the help of (22), and use (40). Then, as expected, we obtain

$$\mathscr{L}_{1/2}(E = 0, H) = \frac{eH}{8\pi^2} \int_0^{\infty} \frac{ds}{s^2} \times \exp(-ism^2) R(0, \tau) \cot\tau, \tag{72}$$

see Eq. (47) in Chapter 1 in the last reference in [9] for $E = 0$.

Passing over to the constant electromagnetic field described by vector potential (44), we use γ matrices in the spinor representation because both α_3 and Σ_3 are then diagonal. The squared Dirac equation has the form

$$(\Pi^2 + m^2 + g)Z = 0, \quad g = e \begin{pmatrix} (H - iE)\sigma_3 & 0 \\ 0 & (H + iE)\sigma_3 \end{pmatrix}, \tag{73}$$

with Π_μ defined in (56). Hence,

$$Z = \mathrm{diag}(f_1, f_2, f_3, f_4) \times \exp[i(p_2 x_2 + p_3 x_3)]. \tag{74}$$

In terms of ζ and T (see (45) and (50)), we obtain the equation

$$\left\{ 2eH\left[-\frac{\partial^2}{\partial\zeta^2} + \frac{\zeta^2}{4} \pm \frac{1}{2} \right] + 2eE\left[\frac{\partial^2}{\partial T^2} + \frac{T^2}{4} \mp \frac{i}{2} \right] + m^2 \right\} f_{1,2} = 0 \tag{75}$$

for f_1 and f_2 and, similarly,

$$\left\{ 2eH\left[-\frac{\partial^2}{\partial\zeta^2} + \frac{\zeta^2}{4} \pm \frac{1}{2} \right] + 2eE\left[\frac{\partial^2}{\partial T^2} + \frac{T^2}{4} \pm \frac{i}{2} \right] + m^2 \right\} f_{3,4} = 0 \tag{76}$$

for f_3 and f_4. From these equations, it follows that

$$^+Z = \mathrm{diag}\{D_{l-1}(\zeta)D_{-i\kappa-1}(\chi), D_l(\zeta)D_{-i\kappa}(\chi), D_{l-1}(\zeta)D_{-i\kappa}(\chi), D_l(\zeta)D_{-i\kappa-1}(\chi)\} \times \exp[i(p_2 x_2 + p_3 x_3)], \quad \chi = e^{i\pi/4} T. \tag{77}$$

Solutions of the Dirac equation with γ matrices in the spinor representation are obtained as the columns of the matrix

$$(m - i\hat{\Pi})Z = \begin{pmatrix} m & 0 & \Pi^0 + \Pi_3 & \Pi_1 - i\Pi_2 \\ 0 & m & \Pi_1 + i\Pi_2 & \Pi^0 - \Pi_3 \\ \Pi^0 - \Pi_3 & -\Pi_1 + i\Pi_2 & m & 0 \\ -\Pi_1 - i\Pi_2 & \Pi^0 + \Pi_3 & 0 & m \end{pmatrix} Z. \tag{78}$$

In terms of χ, we have

$$\Pi^0 \pm \Pi_3 = -e^{-i\pi/4} \sqrt{2eE} \left(\frac{\partial}{\partial\chi} \pm \frac{\chi}{2} \right), \quad \chi = e^{i\pi/4} T. \tag{79}$$

Also taking (58), (59), and the relations

$$(\Pi^0+\Pi_3)D_\nu(\chi) = -e^{-i\pi/4}\nu\sqrt{2eE}D_{\nu-1}(\chi),$$
$$(\Pi^0-\Pi_3)D_\nu(\chi) = e^{-i\pi/4}\sqrt{2eE}D_{\nu+1}(\chi) \tag{80}$$

into account, we find four columns of the matrix $(m - i\hat{\Pi})^+Z$,

$$\begin{bmatrix} mD_{l-1}(\zeta)D_{-i\kappa-1}(\chi) \\ 0 \\ e^{-i\pi/4}\sqrt{2eE}D_{l-1}(\zeta)D_{-i\kappa}(\chi) \\ -i\sqrt{2eH}D_l(\zeta)D_{-i\kappa-1}(\chi) \end{bmatrix},$$

$$\begin{bmatrix} 0 \\ mD_l(\zeta)D_{-i\kappa}(\chi) \\ il\sqrt{2eH}D_{l-1}(\zeta)D_{-i\kappa}(\chi) \\ e^{i\pi/4}\kappa\sqrt{2eE}D_l(\zeta)D_{-i\kappa-1}(\chi) \end{bmatrix},$$

$$\begin{bmatrix} e^{i\pi/4}\kappa\sqrt{2eE}D_{l-1}(\zeta)D_{-i\kappa-1}(\chi) \\ i\sqrt{2eH}D_l(\zeta)D_{-i\kappa}(\chi) \\ mD_{l-1}(\zeta)D_{-i\kappa}(\chi) \\ 0 \end{bmatrix}, \tag{81}$$

$$\begin{bmatrix} -il\sqrt{2eH}D_{l-1}(\zeta)D_{-i\kappa-1}(\chi) \\ e^{-i\pi/4}\sqrt{2eE}D_l(\zeta)D_{-i\kappa}(\chi) \\ 0 \\ mD_l(\zeta)D_{-i\kappa-1}(\chi) \end{bmatrix}.$$

Here and below, $\exp[i(p_2x_2 + p_3x_3)]$ is dropped for brevity. We let ${}^+\psi_1$ (${}^+\psi_2$) denote the fourth (first) column multiplied by the normalization factor ${}^+N_n$ (${}^+N_n\sqrt{l}$),

$${}^+N_n = \exp\left(-\frac{\pi\kappa}{4}\right)(l!2eE)^{-1/2}\left(\frac{eH}{\pi}\right)^{1/4}({}^+N_n\sqrt{l}). \tag{82}$$

We next consider the positive-frequency solution of (73) as $t \longrightarrow -\infty$,

$${}_+Z = \mathrm{diag}\{D_{l-1}(\zeta)D_{i\kappa}(\tau), D_l(\zeta)D_{i\kappa-1}(\tau), D_{l-1}(\zeta)D_{i\kappa-1}(\tau), D_l(\zeta)D_{i\kappa}(\tau)\}, \tag{83}$$

where $\tau = -e^{-i\pi/4}T$. In terms of this variable, we have

$$\Pi^0 \pm \Pi_3 = -e^{i\pi/4}\sqrt{2eE}\left(\frac{\partial}{\partial\tau} \mp \frac{\tau}{2}\right). \tag{84}$$

Similarly to (80), we find

$$(\Pi^0+\Pi_3)D_\nu(\tau) = e^{i\pi/4}\sqrt{2eE}D_{\nu+1}(\tau),$$
$$(\Pi^0-\Pi_3)D_\nu(\tau) = -e^{i\pi/4}\nu\sqrt{2eE}D_{\nu-1}(\tau). \tag{85}$$

Using these relations, we obtain the four columns of the matrix $(m - i\hat{\Pi})_+Z$ in (78) and (83),

$$\begin{bmatrix} mD_{l-1}(\zeta)D_{i\kappa}(\tau) \\ 0 \\ e^{-i\pi/4}\kappa\sqrt{2eE}D_{l-1}(\zeta)D_{i\kappa-1}(\tau) \\ -i\sqrt{2eH}D_l(\zeta)D_{i\kappa}(\tau) \end{bmatrix},$$

$$\begin{bmatrix} 0 \\ mD_l(\zeta)D_{i\kappa-1}(\tau) \\ il\sqrt{2eH}D_{l-1}(\zeta)D_{i\kappa-1}(\tau) \\ e^{i\pi/4}\sqrt{2eE}D_l(\zeta)D_{i\kappa}(\tau) \end{bmatrix},$$

$$\begin{bmatrix} e^{i\pi/4}\sqrt{2eE}D_{l-1}(\zeta)D_{i\kappa}(\tau) \\ i\sqrt{2eH}D_l(\zeta)D_{i\kappa-1}(\tau) \\ mD_{l-1}(\zeta)D_{i\kappa-1}(\tau) \\ 0 \end{bmatrix}, \tag{86}$$

$$\begin{bmatrix} -il\sqrt{2eH}D_{l-1}(\zeta)D_{i\kappa}(\tau) \\ e^{-i\pi/4}\kappa\sqrt{2eE}D_l(\zeta)D_{i\kappa-1}(\tau) \\ 0 \\ mD_l(\zeta)D_{i\kappa}(\tau) \end{bmatrix}.$$

Using the fourth and the first columns again, we have

$${}_+\psi_1(x) = {}_+N_n\begin{bmatrix} -il\sqrt{2eH}D_{l-1}(\zeta)D_{i\kappa}(\tau) \\ e^{-i\pi/4}\kappa\sqrt{2eE}D_l(\zeta)D_{i\kappa-1}(\tau) \\ 0 \\ mD_l(\zeta)D_{i\kappa}(\tau) \end{bmatrix} \times \exp[i(p_2x_2 + p_3x_3)], \tag{87}$$

$${}_+\psi_2(x) = {}_+N_n\sqrt{l}\begin{bmatrix} mD_{l-1}(\zeta)D_{i\kappa}(\tau) \\ 0 \\ e^{-i\pi/4}\kappa\sqrt{2eE}D_{l-1}(\zeta)D_{i\kappa-1}(\tau) \\ -i\sqrt{2eH}D_l(\zeta)D_{i\kappa}(\tau) \end{bmatrix} \times \exp[i(p_2x_2 + p_3x_3)], \tag{88}$$

where ${}_+N_n = {}^+N_n/\sqrt{\kappa}$, see (82).

We note that ${}_-Z$ (${}^-Z$) can be obtained from ${}^+Z$ (${}_+Z$) by the substitution $\chi \longrightarrow -\chi$ ($\tau \longrightarrow -\tau$). To obtain ${}_-\psi$ functions from the corresponding ${}^+\psi$ functions, we also change the sign of $\sqrt{2eE}$ in the columns in addition to

these substitutions; this is because of the relations (see (79) and (80))

$$(\Pi^0 + \Pi_3)D_\nu(\pm\chi) = \mp e^{-i\pi/4}\nu\sqrt{2eE}D_{\nu-1}(\pm\chi),$$
$$(\Pi^0 - \Pi_3)D_\nu(\pm\chi) = \pm e^{-i\pi/4}\sqrt{2eE}D_{\nu+1}(\pm\chi). \quad (89)$$

Thus,

$$_-\psi_1(x) = {}_-N_n \begin{bmatrix} -il\sqrt{2eH}D_{l-1}(\zeta)D_{-i\kappa-1}(-\chi) \\ -e^{-i\pi/4}\sqrt{2eE}D_l(\zeta)D_{-i\kappa}(-\chi) \\ 0 \\ mD_l(\zeta)D_{-i\kappa-1}(-\chi) \end{bmatrix} \times \exp[i(p_2x_2 + p_3x_3)], \quad (90)$$

$$_-\psi_2(x) = {}_-N_n\sqrt{l} \begin{bmatrix} mD_{l-1}(\zeta)D_{-i\kappa-1}(-\chi) \\ 0 \\ -e^{-i\pi/4}\sqrt{2eE}D_{l-1}(\zeta)D_{-i\kappa}(-\chi) \\ -i\sqrt{2eH}D_l(\zeta)D_{-i\kappa-1}(-\chi) \end{bmatrix} \times \exp[i(p_2x_2 + p_3x_3)], \quad {}_-N_n = {}^+N_n, \quad (91)$$

and, similarly, for $^-\psi_1$ and $^-\psi_2$.

We note in passing that the wave functions for the electron in a constant electric field were written in [2] using γ matrices in the standard representation. Acting on these functions by the operator

$$U = \frac{1}{\sqrt{2}}\begin{pmatrix} I & I \\ I & -I \end{pmatrix},$$

we obtain the solutions in the spinor representation. The taking of the magnetic field into account is realized by the substitutions

$$\exp(ip_2x_2)\{1, p_1 - ip_2, p_1 + ip_2\} \longrightarrow \left(\frac{eH}{\pi}\right)^{1/4} \times \frac{1}{\sqrt{l!}}\{D_l(\zeta), -il\sqrt{2eH}D_{l-1}(\zeta), i\sqrt{2eH}D_{l+1}(\zeta)\}$$

for $r = 1$. For $r = 2$, we must replace l by $l - 1$ in these substitutions.

The electron propagator is given by

$$G_{1/2}(x, x') = i\sum_n c_{1n}^{*-1} \times \begin{cases} {}^+\psi_n(x)_+\bar\psi_n(x'), & t > t', \\ -{}_-\psi_n(x)^-\bar\psi_n(x'), & t < t', \end{cases} \quad (92)$$

where $\bar\psi = \psi^*\beta$, $n = (p_2, p_3, l, r)$ for the constant electromagnetic field, and c_{1n}^* is given in (27), where $\rho_\perp^2 = 2eHl$ in the expression for κ, see (69) and (95). In the spinor representation,

$$\beta = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}, \quad (93)$$

and therefore,

$$(a_1, a_2, a_3, a_4)\beta = (a_3, a_4, a_1, a_2).$$

Using (81), (82), and (87), we now obtain

$$\mathrm{tr}({}^+\psi_1(x)_+\bar\psi_1(x)) = \sqrt{\frac{eH}{\pi}}\frac{me^{-\pi\kappa/2}}{l!\sqrt{2eE\kappa}} \times D_l^2(\zeta)\{e^{-i\pi/4}D_{-i\kappa}(\chi)D_{-i\kappa}(-\chi) + e^{i\pi/4}\kappa D_{-i\kappa-1}(\chi)D_{-i\kappa-1}(-\chi)\}. \quad (94)$$

Integrating over x_1, we obtain, see (46),

$$\int_{-\infty}^{\infty} dx_1 \mathrm{tr}({}^+\psi_1(x)_+\bar\psi_1(x)) = \frac{me^{-\pi\kappa/2}}{\sqrt{2eE\kappa}} \times \{e^{-i\pi/4}D_{-i\kappa}(\chi)D_{-i\kappa}(-\chi) + e^{i\pi/4}\kappa D_{-i\kappa-1}(\chi)D_{-i\kappa-1}(-\chi)\}, \quad \kappa = \frac{m^2 + 2eHl}{2eE}, \quad l = 0, 1, \ldots . \quad (95)$$

For $r = 2$, we obtain the same expression, but with $l = 1, 2, \ldots$.

We next multiply (95) with i/c_{1n}^* according to (92) and use the relation (see Eq. (93) in [11] with $-i\kappa \longrightarrow -i\kappa + 1/2$)

$$\Gamma(i\kappa)D_{-i\kappa}(\chi)D_{-i\kappa}(-\chi) = \sqrt{2}\int_0^\infty \frac{d\theta}{\sqrt{\sinh 2\theta}} \times \exp\left(-2i\kappa\theta + \theta - \frac{i}{2}T^2\tanh\theta\right), \quad \theta = eEs, \quad (96)$$

and the relation obtained from this by the substitution $i\kappa \longrightarrow i\kappa + 1$.

We now discover from (95) and (96) that

$$\int_{-\infty}^{\infty} dx_1 \frac{i}{c_{1n}^*}\mathrm{tr}({}^+\psi_1(x)_+\bar\psi_1(x)) = -\frac{me^{-i\pi/4}}{\sqrt{2\pi eE}} \times \int_0^\infty \frac{d\theta}{\sqrt{\sinh 2\theta}} 2\cosh\theta \exp\left(-2i\kappa\theta - \frac{i}{2}T^2\tanh\theta\right), \quad (97)$$

$$T = \sqrt{2eE}\left(t - \frac{p_3}{eE}\right).$$

Integrating this expression over t, we obtain

$$\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dx_1 \frac{i}{c_{1n}^*} \mathrm{tr}({}^{+}\psi_1(x)_{+}\bar{\psi}_1(x)) = im\int_0^{\infty} ds \coth(eEs)\exp[-is(m^2 + 2eHl)], \quad l = 0, 1, 2, \ldots . \tag{98}$$

For $r = 2$, we have the same expression, but with $l = 1, 2, \ldots$.

Taking the remark after Eq. (70') into account and inserting the regulator $R(\theta, \tau)$ into the integrand, we obtain the contribution to $\mathscr{L}_{1/2}$ from the state $n = (p_2, p_3, l, r)$. Summing over l and r (see (40)) and integrating over $dp_2/2\pi$ and $dp_3/2\pi$ (see (39) and (12)), we obtain, in agreement with (41), that

$$\mathscr{L}_{1/2} = \frac{e^2 HE}{8\pi^2}\int_0^{\infty}\frac{ds}{s} \times \exp(-ism^2) R(\theta, \tau)\coth\theta\cot\tau. \tag{99}$$

We finally note that for $H = 0$, we have

$$\int_{-\infty}^{\infty} dt \frac{i}{c_{1n}^*}\mathrm{tr}({}_{+}\psi_1(x)^{+}\bar{\psi}_1(x)) = im\int_0^{\infty} ds \coth(eEs) \times \exp[-is(m^2 + p_1^2 + p_2^2)], \quad l = 0, 1, 2, \ldots \tag{100}$$

instead of (98). Inserting $1/2ms$ and $R(\theta, 0)$, we see that this agrees with (35) and (29).

6. VECTOR BOSON

The propagator and the effective Lagrange function for the vector boson with the gyromagnetic ratio $g = 2$ in a constant electromagnetic field were obtained by Vanyashin and Terentyev [17]. In another form, the propagator was found by the author [11]. In the latter paper, there is a misprint in Eq. (73), where the argument of sin and cos should be 2τ, not τ. In addition, the statement that the divergence term in the expression for the current in Eq. (38) makes no contribution is not true when the magnetic field is present; this, however, is of no consequence because the expression was used only for the normalization of wave functions.

The results of Vanyashin and Terentyev imply that relation (48) in the present paper also holds for the vector boson if we take $G_b = G_\mu^\mu$. Using (48), we can reproduce the expression for $\mathscr{L}_1$ in [17] starting from our propagator. Indeed, our result for

$$G_\mu^\mu = \frac{e^2 EH}{16\pi^2}\int_C \frac{ds}{\sinh\theta\sin\tau}\exp(-ism^2) \times\left\{2\cos 2\tau + 2\cosh 2\theta - \frac{i}{m^2}[eH\cot\tau + eE\coth\theta]\right\} \tag{101}$$

can be written in a simpler form if we note that

$$\frac{d}{ds}\frac{1}{\sinh\theta\sin\tau} = -\frac{1}{\sinh\theta\sin\tau}[eH\cot\tau + eE\coth\theta], \quad \tau = eHs, \quad \theta = eEs. \tag{102}$$

We can then integrate the term in square brackets in (101) by parts,

$$-\frac{ie^2 EH}{16\pi^2 m^2}\int_C \frac{ds}{\sinh\theta\sin\tau}\exp(-ism^2) \times [eH\cot\tau + eE\coth\theta] \longrightarrow -\frac{e^2 EH}{16\pi^2}\int_C \frac{ds}{\sinh\theta\sin\tau}\exp(-ism^2), \tag{103}$$

where we discarded a divergent term independent of E and H. Expression (101) is therefore equivalent to

$$\frac{e^2 EH}{16\pi^2}\int_C \frac{ds}{\sinh\theta\sin\tau}\exp(-ism^2) \times (2\cos 2\tau + 2\cosh 2\theta - 1). \tag{104}$$

Inserting $-1/s$ in the integrand, we obtain Eq. (21) in [17]; we agree with the subsequent formulas in that paper.

Returning to our present problem, we note that for a constant electric field, c_{1n}^* is independent of the polarization state of the vector boson and is the same as in the scalar case [11]. Nevertheless, $\mathrm{Im}\mathscr{L}_1$ is not simply $3\mathrm{Im}\mathscr{L}_0$ [17]. The knowledge of c_{1n}^* is therefore not useful in obtaining $\ln C_{1n}^{*\mathrm{ren}}$. Resorting to the proper-time method, we find that the problem is more difficult than in the previous cases. As seen already from (101), the dependence on m^2 is more complicated here and the contributions from the electric and magnetic fields are not factorized in the proper-time integrand. For these reasons, we here consider only the constant magnetic field.

It follows from [11] that for the spin states $r = 1, 2, 3$,

$$
\begin{aligned}
&{}_+\psi_1^\mu(x)\,{}_+\psi_{1\mu}^*(x) \\
&= \sqrt{\frac{eH}{\pi}}\frac{1}{2|p^0|l!}\frac{1}{(l+1)(m^2+eHl)} \\
&\times\{-(l+1)^2 eHD_l^2(\zeta) + [m^2+eH(2l+1)]D_{l+1}^2(\zeta)\},
\end{aligned}
\tag{105}
$$

$$
{}_+\psi_2^\mu(x)\,{}_+\psi_{2\mu}^*(x) = \sqrt{\frac{eH}{\pi}}\frac{1}{2|p^0|l!}D_l^2(\zeta), \tag{106}
$$

$$
\begin{aligned}
{}_+\psi_3^\mu(x)\,{}_+\psi_{3\mu}^*(x) &= \sqrt{\frac{eH}{\pi}}\frac{1}{2|p^0|l!}\frac{l}{2m^2(m^2+eHl)} \\
&\times\{-2eH[m^2+eH(2l+1)]D_l^2(\zeta) \\
&+ [eHD_{l+1}(\zeta) - (m^2+eHl)D_{l-1}(\zeta)]^2 \\
&+ [eHD_{l+1}(\zeta) + (m^2+eHl)D_{l-1}(\zeta)]^2\}.
\end{aligned}
\tag{107}
$$

Integrating the expressions in (105)–(107) over x_1 with the help of (46), we obtain $1/2|p^0|$ in all three cases, but

$$
\begin{aligned}
l &= l_{\min}, l_{\min}+1, \ldots, \\
l_{\min} &= \begin{cases} -1, & r = 1 \\ 0, & r = 2 \\ 1, & r = 3. \end{cases}
\end{aligned}
\tag{108}
$$

The vector boson propagator is given by [11]

$$
\begin{aligned}
G_1^{\mu\nu}(x,x') &= i\int_{-\infty}^{\infty}\frac{dp_2}{2\pi}\int_{-\infty}^{\infty}\frac{dp_3}{2\pi} \\
&\times\sum_{r=1}^{3}\sum_{l_{\min}}^{\infty}\begin{cases} {}_+\psi_n(x)^\mu\,{}_+\psi_n^{*\nu}(x'), & t > t' \\ {}_-\psi_n(x)^\mu\,{}_-\psi_n^{*\nu}(x'), & t < t'. \end{cases}
\end{aligned}
\tag{109}
$$

We see from (68) and (48) and the above results that the contribution to $\mathscr{L}_1$ from the state with the quantum numbers $n = (p_2, p_3, l, r)$ is

$$
\begin{aligned}
i\ln c_{1n}^* &= -i\frac{e^{i\pi/4}}{2\sqrt{\pi}} \\
&\times\int_0^\infty\frac{ds}{s^{3/2}}\exp\{-is[m^2+eH(2l+1)+p_3^2]\}.
\end{aligned}
\tag{110}
$$

The sum over r and l is performed using the formula

$$
\sum_{r=1}^{3}\sum_{l_{\min}}^{\infty}\exp[-iseH(2l+1)] = \frac{1+2\cos(2eHs)}{2i\sin(eHs)} \tag{111}
$$

that can be obtained from (25). To integrate over $dp_2/2\pi$ and $dp_3/2\pi$, we use (39) and (12). Inserting $R(\tau)$, we then obtain

$$
\begin{aligned}
\sum_n W_{\text{spin}1,\,n} &= -\frac{eH}{16\pi^2}\int_0^\infty\frac{ds}{s^2\sin\tau} \\
&\times\exp(-ism^2)(3-4\sin^2\tau)R(\tau).
\end{aligned}
\tag{112}
$$

In accordance with [17], $R(\tau)$ is defined as

$$
\begin{aligned}
\frac{3-4\sin^2\tau}{\sin\tau} &\longrightarrow 3\left(\frac{1}{\sin\tau}-\frac{1}{\tau}-\frac{\tau}{6}\right) \\
-4(\sin\tau-\tau) &= \frac{3-4\sin^2\tau}{\sin\tau}R(\tau).
\end{aligned}
\tag{113}
$$

This implies that

$$
\begin{aligned}
R(\tau) &= 1-\frac{\sin\tau}{3-\sin^2\tau}\left(\frac{3}{\tau}-\frac{7}{2}\tau\right), \\
R(\tau)|_{\tau\ll 1} &= \frac{29}{120}\tau^4.
\end{aligned}
\tag{114}
$$

From (110), we therefore have

$$
\begin{aligned}
i\ln C_{1n}^{*\text{ren}} &= -i\frac{e^{i\pi/4}}{2\sqrt{\pi}}\int_0^\infty\frac{ds}{s^{3/2}} \\
&\times\exp\{-is[m^2+eH(2l+1)+p_3^2]\}R(\tau),
\end{aligned}
\tag{115}
$$

where l is given in (108). Substituting $\tau \longrightarrow -it$ and rotating the integration contour, we see that $\ln C_{1n}^{*\text{ren}}$ is real, as it should be for the magnetic field.

7. CONCLUSION

We have shown how the renormalized phase of the vacuum–vacuum amplitude in quantum electrodynamics can be expressed via the properly fixed phases of the Bogoliubov coefficients; a nonzero phase of the former indicates nonzero phases of the latter. In general, knowledge of the Bogoliubov coefficients alone is insufficient for obtaining the phase of the vacuum–vacuum amplitude. Additional information is needed. Thus, in the case of constant magnetic and electromagnetic fields, we have used the symmetry between the electric and magnetic fields in the Lagrange function. In the case of a vector boson, the knowledge of the Bogoliubov coefficients is not useful in fixing their phases. Resorting to the proper-time method shows that the expressions for the phases are in general more complicated than that for lower-spin particles. For this reason, we have presented the results only for the constant magnetic field, where they turned out to be as simple as expected.

ACKNOWLEDGMENTS

I am grateful to V.I. Ritus for discussions that led to the appearance of this paper.

This work was supported in part by the Russian Foundation for Basic Research (project nos. 00-15-96566 and 02-02-16944).

REFERENCES

1. N. B. Narozhny and A. I. Nikishov, Yad. Fiz. **11**, 1072 (1970) [Sov. J. Nucl. Phys. **11**, 596 (1970)].
2. A. I. Nikishov, Tr. Fiz. Inst. im. P.N. Lebedeva, Akad. Nauk SSSR **111**, 152 (1979) [J. Sov. Laser Res. **6**, 619 (1985)].
3. A. A. Grib, S. G. Mamaev, and V. M. Mostepanenko, *Vacuum Quantum Effects in Strong Fields* (Énergoatomizdat, Moscow, 1988).
4. A. I. Akhiezer and V. B. Berestetskiĭ, *Quantum Electrodynamics*, 3rd ed. (Nauka, Moscow, 1969; Wiley-Interscience, New York, 1965).
5. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 3rd ed. (Nauka, Moscow, 1989; Pergamon, Oxford, 1982).
6. W. Heisenberg and H. Euler, Z. Phys. **98**, 714 (1936).
7. V. Weisskopf, K. Dan. Vidensk. Selsk. Mat. Fys. Medd. **14** (6), 39 (1936).
8. J. Schwinger, Phys. Rev. **82**, 664 (1951).
9. V. I. Ritus, Zh. Éksp. Teor. Fiz. **69**, 1517 (1975) [Sov. Phys. JETP **42**, 774 (1975)]; Zh. Éksp. Teor. Fiz. **73**, 807 (1977) [Sov. Phys. JETP **46**, 423 (1977)]; Tr. Fiz. Inst. im. P. N. Lebedeva Akad. Nauk SSSR **168**, 52 (1986); *Issues in Intense-Field Quantum Electrodynamics*, Ed. by V. L. Ginzburg (Nova Science, Commack, 1987), p. 3.
10. B. S. De Witt, Phys. Rep. C **19**, 227 (1975).
11. A. I. Nikishov, Zh. Éksp. Teor. Fiz. **120**, 227 (2001) [JETP **93**, 197 (2001)].
12. Y. L. Luke, *Mathematical Functions and Their Approximations* (Academic, London, 1975; Mir, Moscow, 1980).
13. R. B. Dingle, *Asymptotic Expansions: Their Derivation and Interpretation* (Academic, London, 1973).
14. A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series*, Vol. 1: *Elementary Functions* (Nauka, Moscow, 1981; Gordon and Breach, New York, 1986).
15. I. S. Gradstein and I. M. Ryzhik, *Tables of Integrals, Sums, Series, and Products*, 4th ed. (Gostekhizdat, Moscow, 1962; Academic, New York, 1980).
16. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (National Bureau of Standards, Washington, 1964; Nauka, Moscow, 1979).
17. V. S. Vanyashin and M. V. Terentyev, Zh. Éksp. Teor. Fiz. **48**, 565 (1965) [Sov. Phys. JETP **21**, 375 (1965)].

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 193–201.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 224–232.

NUCLEI, PARTICLES, AND THEIR INTERACTION

(Quasi)elastic Large-Angle Electron–Muon Scattering in the Two-Loop Approximation: Contributions of the Eikonal Type¶

V. V. Bytev[a], E. A. Kuraev[a,*], and B. G. Shaikhatdenov[a,b]

[a]*Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia*
[b]*Departamento de Física, Cinvestav del IPN, 07000 México D.F., Mexico*
**e-mail: kuraev@thsun1.jinr.ru*

Received October 16, 2002

Abstract—A part of the eikonal-type contributions to the $e\mu$ large-angle high-energy scattering cross section is considered in a quasielastic experimental setup. In addition to virtual corrections, we examine inelastic processes with emission of one and two soft real photons and soft lepton and pion pairs. Virtual photon contributions are given within a logarithmic accuracy. Box-type Feynman amplitudes with leptonic and a hadronic vacuum polarization insertion and double-box ones are considered explicitly. Wherever appropriate, the analytic expressions obtained are compared with those predicted by the structure function approach. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The need for evaluating radiative corrections at the two-loop order is dictated by experimental data on observables for a collider calibration process of electron–positron scattering that has reached an impressive level of accuracy. Inspired by this, we consider as our ultimate goal the determination of the second-order radiative corrections to the cross section of Bhabha scattering. At the same time, because the task of two-loop calculus is rather involved, it appears easier to consider the electron–muon scattering first, despite different masses of interacting particles. The latter process is also important in itself because it forms a background to the rare processes, in particular those violating the lepton number (for more details, see [1] and references therein). Improving theoretical predictions on its observables could therefore impose more stringent bounds on the physics beyond the standard model.

The aim of this investigation is to calculate the next-to-leading order contributions to the large-angle electron–muon high-energy cross section,

$$e^-(p_1) + \mu^-(p_2) \longrightarrow e^-(p_1') + \mu^-(p_2'), \qquad (1)$$

in a quasielastic experimental setup,

$$\frac{2\varepsilon - \varepsilon_1' - \varepsilon_2'}{2\varepsilon} = \frac{\Delta\varepsilon}{\varepsilon} \equiv \Delta \ll 1, \quad \Delta\varepsilon \gg m_\mu(m_\pi), \qquad (2)$$

where ε, ε_1', and ε_2' are the energies of the initial and scattered leptons in the center-of-mass reference frame and the Mandelstam variables are much larger than the squared mass of any particle involved in the process. The quantity $\Delta\varepsilon$ indicates the energy resolution of detectors that are supposed to track final particles. In the leading logarithmic approximation, the cross section is that of the Drell–Yan process [2],

$$d\sigma(s,t) = \int\prod_{i=1}^{4} dx_i \mathcal{D}(x_i, \rho_t) \times d\sigma_0(sx_1x_2, tx_1x_3)\left(1 + \frac{\alpha}{\pi}K\right), \qquad (3)$$

where

$$\rho_t = \ln\frac{-t}{m_e m_\mu}, \quad t = (p_1 - p_1')^2, \quad s = (p_1 + p_2)^2, \quad u = (p_1 - p_2')^2. \qquad (4)$$

In the above expression, the quantities $\mathcal{D}(x_i, \rho_t)$ are the nonsinglet structure functions that satisfy the renormalization group (RG) evolution equations. Their expansion in the leading logarithmic approximation

$$(\alpha/\pi) \ll 1, \quad (\alpha/\pi)\rho_t \sim 1$$

can be written as

$$\mathcal{D}(x, \rho_t) = \delta(1-x) + \sum_{n=1}^{\infty}\frac{1}{n!}\left(\frac{\alpha\rho_t}{2\pi}\right)^n \mathcal{P}^{(n)}(x). \qquad (5)$$

¶This article was submitted by the authors in English.

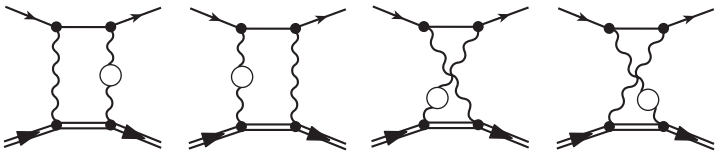


Fig. 1. Box-type graphs with a vacuum polarization insertion.

In a quasielastic setup, it is appropriate to use only the δ part of the splitting function $\mathcal{P}^{(n)}(x)$, denoted by $\mathcal{P}_{\Delta}^{(n)}(x)$:

$$\mathcal{P}^{(n)}(x) = \int_x^1 \frac{dy}{y}\mathcal{P}^{(1)}(y)\mathcal{P}^{(n-1)}\left(\frac{x}{y}\right), \quad n \geq 2,$$

$$\mathcal{P}^{(1)}(x) = \left(\frac{1+x^2}{1-x}\right)_+ = \lim_{\Delta\to 0}[\mathcal{P}_{\Delta}^{(1)}(x) + \mathcal{P}_{\theta}^{(1)}(x)], \tag{6}$$

$$\mathcal{P}_{\Delta}^{(1)}(x) = \mathcal{P}_{\Delta}^{(1)}\delta(1-x), \quad \mathcal{P}_{\Delta}^{(1)} = 2\ln\Delta + \frac{3}{2},$$

$$P_{\theta}^{(1)}(x) = \frac{1+x^2}{1-x}\Theta(1-x-\Delta).$$

The structure function then becomes

$$\mathcal{D}(x, \rho_t) = \delta(1-x)\left[1 + \sum_{n=1}^{\infty}\frac{1}{n!}\left(\frac{\alpha\rho_t}{2\pi}\right)^n \mathcal{P}_{\Delta}^{(n)}\right]. \tag{7}$$

Because the structure function approach outlined is capable of providing only the leading logarithmic corrections, we need to explicitly calculate the so-called *K* factor entering Eq. (3) in the one- and two-loop approximations.

Broadly speaking, the radiative corrections to the differential cross section in the adopted mass regularization scheme are of two types. The first ones are those arising from the virtual photon emission up to the second order of perturbation theory, which requires calculating, among others, the real two-loop Feynman amplitudes. They suffer from infrared divergences, which are regularized by assigning the photon a negligibly small mass λ that is set to zero at the end of the calculations. Contributions of the second type come from the emission of soft real photons and charged particle pairs.

The general structure of the correction to the cross section can be represented as a sum of three types: vertex, eikonal, and decorated box type. Each of them contains virtual and real soft photon contributions, is free of infrared divergences, and preserves the structure of the leading log correction predicted on the basis of RG ideas through the contributions of individual diagrams containing up to the fourth power of the large logarithm ρ_t at the two-loop order. In this regard, we recall that in our previous paper [1], it was shown that the vertex contributions already provide a result consistent with the RG approach. Because the first-order radiative corrections coming from box-type diagrams are given in our previous work devoted to evaluating of vertex-type contributions [1], we here concentrate on the investigation of some eikonal box-type diagrams at the second order of perturbation theory. In the case of elastic processes, they correspond to graphs with one, two (box diagram), and three (double-box diagram) virtual photons mediated between interacting leptons. Box-type graphs with a vacuum polarization insertion of either of the virtual exchange photons into the Green function must also be taken into account (see Fig. 1). A single soft photon approximation must be applied to the one-loop corrected Feynman amplitudes in order to obtain another set of contributions. Finally, the emission of two soft photons (pairs of charged particles) must also be taken into account at this order.

We briefly describe the contents of the paper. In Section 2, we consider the vacuum polarization effects in box-type Feynman amplitudes with lepton ($\mu\bar{\mu}$, $e\bar{e}$) and pion ($\pi^-\pi^+$) pairs running a loop. Also in this section we consider the corresponding contribution coming from a soft lepton pair and a soft charged pion pair production with one soft photon emission (see Fig. 2) associated with the one-loop self-energy amplitudes of the virtual exchange photon. In Section 3, the results of evaluation of the corrections corresponding to a single and double soft photon emission (see Fig. 3) and to a square of box-type diagrams are presented; they are fol-

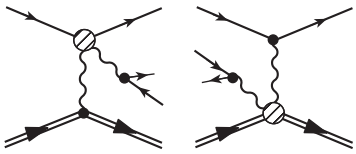


Fig. 2. Soft lepton and pion pair production.

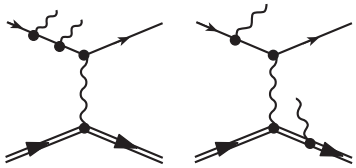


Fig. 3. Sample diagrams pertaining to double soft photon emission.

lowed by brief concluding remarks. In Appendix A, we present a set of scalar integrals for box-type diagrams with a vacuum polarization insertion. In Appendix B, we give some details of the derivation of radiative corrections coming from the squared box-type diagrams and all the integrals encountered during the calculation.

2. BOX-TYPE DIAGRAMS WITH A VACUUM POLARIZATION INSERTION

Vacuum polarization effects in the box-type Feynman amplitudes can be taken into account by replacing one of the photon propagators by the vacuum polarization insertion (see [3]). In the case where leptons with mass μ run a loop, it is given by

$$\frac{1}{k^2} \longrightarrow \frac{\alpha}{3\pi}\int_0^1 \frac{\phi(v)dv}{(1-v^2)(k^2-M^2(v))},$$
$$M^2 = \frac{4\mu^2}{1-v^2}, \quad \phi(v) = 2-(1-v^2)(2-v^2), \tag{8}$$

and for a pion–antipion pair in the loop, it is

$$\frac{1}{k^2} \longrightarrow \frac{\alpha}{3\pi}\int\limits_{4m_\pi^2}^{\infty} \frac{dM^2}{M^2}\frac{\mathcal{R}(M^2)}{k^2-M^2},$$
$$\mathcal{R}(M^2) = \frac{\sigma^{e\bar{e}\to hadr}(M^2)}{\sigma^{e\bar{e}\to \mu\bar{\mu}}}. \tag{9}$$

Here, the quantity M is the invariant mass of the hadronic jet produced in single-photon annihilation of a lepton pair and $\mathcal{R}(M^2)$ is the known experimental input ratio [4]. For the matrix element squared, we then obtain

$$\delta|\mathcal{M}|^2_{vp(lept)} = \frac{2^8\alpha^4}{3t}\int_0^1 dv\frac{\phi(v)}{1-v^2}[S(s,t,M^2)-S(u,t,M^2)] \tag{10}$$

for the vacuum polarization induced by leptons, and

$$\delta|\mathcal{M}|^2_{vp(hadr)} = \frac{2^8\alpha^4}{3t}\int\limits_{4m_\pi^2}^{\infty}\frac{dM^2}{M^2}\mathcal{R}(M^2)[S(s,t,M^2)-S(u,t,M^2)] \tag{11}$$

for the hadronic vacuum polarization contribution.

The quantity $S(s, t, M^2)$ is universal, irrespective of the virtual pair running a self-energy loop, and is given by

$$S(s,t,M^2) = \int\frac{d^4k}{i\pi^2}\frac{\mathrm{Tr}(e)\mathrm{Tr}(\mu)}{(1)(2)(3)(4)}, \tag{12}$$

where

$$(1) = k^2-2kp_1, \quad (2) = k^2+2kp_2,$$
$$(3) = k^2-2kq+\tilde{t}, \quad (4) = k^2-\lambda^2,$$
$$\mathrm{Tr}(e) = \frac{1}{4}\mathrm{Sp}\{p_1\gamma_\mu p_1'\gamma_\nu(p_1-k)\gamma_\lambda\},$$
$$\mathrm{Tr}(\mu) = \frac{1}{4}\mathrm{Sp}\{p_2\gamma_\mu p_2'\gamma_\nu(p_2+k)\gamma_\lambda\}, \tag{13}$$
$$p_1^2 = m_e^2, \quad p_2^2 = m_\mu^2,$$
$$\tilde{t} = t-M^2, \quad q = p_1-p_1'.$$

Using the set of scalar, vector, and tensor box-type integrals given in Appendix A, we can express the quantity $S(s, t, M^2)$ through several basic integrals,

$$S(s,t,M^2) = u\left[\ln\frac{s}{-\tilde{t}}+\frac{M^2}{t}\ln\frac{-\tilde{t}}{M^2}\right]$$
$$-\left(s(s-u)+\frac{t\tilde{t}}{2}\right)[I_{134}+I_{234}] \tag{14}$$
$$+s(s^2+u^2)I+s\left(u+\frac{\tilde{t}}{2}\right)[-I_{123}-I_{124}+\tilde{t}I],$$

where

$$I_{ijl} = \int\frac{d^4k}{i\pi^2}\frac{1}{(i)(j)(l)},$$
$$I = \int\frac{d^4k}{i\pi^2}\frac{1}{(1)(2)(3)(4)}. \tag{15}$$

Performing loop-momentum integration and neglecting terms of the order of $m_\mu^2/(-t) \ll 1$, we find, in the limit of large invariant variables,

$$[S(s,t,M^2)-S(u,t,M^2)]\big|_{|t|\gg M^2} = \frac{s^2+u^2}{t}L_{us}(\rho_m-2\rho_t-\rho_\lambda)$$
$$+(u-s)\left(\frac{1}{2}\pi^2+\rho_m^2-\frac{1}{2}L_{st}^2-\frac{1}{2}L_{ut}^2+\ln^2\frac{m_\mu}{m_e}\right) \tag{16}$$
$$+uL_{st}-sL_{ut},$$

$$\rho_m = \ln\frac{M^2}{m_e m_\mu}, \quad \rho_\lambda = \ln\frac{m_e m_\mu}{\lambda^2},$$

$$L_{st} = \ln\frac{s}{-t}, \quad L_{ut} = \ln\frac{u}{t}, \quad L_{us} = \ln\frac{-u}{s}.$$

In the opposite limit, the result is found to be

$$\begin{aligned}&[S(s,t,M^2) - S(u,t,M^2)]\big|_{M^2 \gg |t|}\\ &= \frac{1}{M^2}\Big[\frac{s^2+u^2}{2}L_{us}(\rho_s+\rho_u+2\rho_\lambda)\\ &+\frac{3}{2}(u^2\rho_s - s^2\rho_u) + t^2 L_{us} + t(u-s)\\ &\times\Big(\frac{3}{2}\rho_m + \frac{7}{4}\Big) + \frac{s^2+u^2}{2}\pi^2\Big].\end{aligned} \tag{17}$$

For the leptonic vacuum polarization with a mass of

$$M^2 = \frac{4\mu^2}{1-v^2}$$

(where both cases $\mu = m_e, m_\mu$ are taken into account), further integration leads to the following expression within logarithmic accuracy:

$$\begin{aligned}\frac{d\sigma_{vp}^{box}}{d\sigma_0} &= \frac{2\alpha^2}{3\pi^2}\rho_t\Big\{2L_{su}\Big(\frac{3}{2}\rho_t + \rho_\lambda - \frac{10}{3}\Big)\\ &-\frac{s^2-u^2}{s^2+u^2}(L_{st}^2 + L_{ut}^2 - 2\pi^2)\\ &+\frac{2t}{s^2+u^2}[(t-s)L_{st} - (t-u)L_{ut}]\Big\}.\end{aligned} \tag{18}$$

To finalize this result, we must remove infrared divergences. For this, the interference between the soft photon emission tree-level amplitudes and those bearing a leptonic vacuum polarization insertion must be taken into account, with the result

$$\begin{aligned}\frac{d\sigma_{vp}^{\gamma}}{d\sigma_0} &= -\frac{4\alpha^2}{3\pi^2}\Big(\rho_t - \frac{5}{3}\Big)\Big[(2\ln\Delta + \rho_\lambda)L_{su} + \rho_t L_{su}\\ &-\frac{1}{2}(L_{ut}^2 - L_{st}^2) - \mathrm{Li}_2\Big(\frac{1-c}{2}\Big)\Big],\end{aligned} \tag{19}$$

where Δ is given in Eq. (2), $c = \cos\widehat{p_1, p_1'}$ is the cosine of the scattering angle in the center-of-mass reference frame and the dilogarithm function is defined by the standard formula

$$\mathrm{Li}_2(x) = -\int_0^x \frac{\ln(1-t)}{t}dt. \tag{20}$$

Next, we must consider the contribution coming from the soft lepton pair production with the total pair energy not exceeding $\Delta\varepsilon$ ($2\mu \ll \Delta\varepsilon \ll \varepsilon$). This can be read off, e.g., from [5],

$$\begin{aligned}\frac{d\sigma_{sp}}{d\sigma_0} &= -\frac{2\alpha^2}{3\pi^2}\rho_t\Big\{L_{su}\Big[\rho_t + L_{st} + L_{ut}\\ &+2\Big(2\ln\Delta - \frac{5}{3}\Big)\Big] - 2\mathrm{Li}_2\Big(\frac{1-c}{2}\Big)\Big\}.\end{aligned} \tag{21}$$

The final logarithmically accurate result for the total correction given by the leptonic vacuum polarization and the soft $e\bar{e}$, $\mu\bar{\mu}$ pair production is then takes the form (see Eqs. (18), (19), (21))

$$\begin{aligned}\frac{d\sigma_{vp+sp}}{d\sigma_0} &= \frac{2\alpha^2}{3\pi^2}\rho_t\Big\{2(L_{ut}^2 - L_{st}^2) + 8L_{us}\ln\Delta\\ &-\frac{s^2-u^2}{s^2+u^2}(L_{ut}^2 + L_{st}^2 - 2\pi^2)\\ &+\frac{2t}{s^2+u^2}[tL_{su} - sL_{st} + uL_{ut}] + 4\mathrm{Li}_2\Big(\frac{1-c}{2}\Big)\Big\}.\end{aligned} \tag{22}$$

This expression is seen to contain only a next-to-leading term (of the order of $\alpha^2\rho_t$) and to be free of infrared divergences.

We now consider the soft pion pair production with the total pair energy below $\Delta\varepsilon$ and the invariant mass squared M^2 bounded as

$$4m_\pi^2 \ll M^2 < (\Delta\varepsilon)^2 \ll \varepsilon^2 = s/4. \tag{23}$$

The corresponding contribution to the differential cross section arises from the interference of "up–down" pair production, which refers to pairs created by virtual photons emitted from the electron line and the muon line,

$$\begin{aligned}&\frac{d\sigma}{dM^2 d\sigma_0}\Big|_{\pi^\pm}\\ &= 2\Big(\frac{4\pi\alpha}{M^2}\Big)^2\frac{d^4q}{M^2}\int\frac{d^3\mathbf{q}_+ d^3\mathbf{q}_-}{2\varepsilon_+ 2\varepsilon_-}\delta^4(q_+ + q_- - q)\\ &\times\Big(\frac{Qp_1'}{qp_1'} - \frac{Qp_1}{qp_1}\Big)\Big(\frac{Qp_2'}{qp_2'} - \frac{Qp_2}{qp_2}\Big),\\ &q^2 = M^2, \quad Q = q_+ - q_-.\end{aligned} \tag{24}$$

We first perform the invariant pion pair phase space integration,

$$\int\frac{d^3\mathbf{q}_+d^3\mathbf{q}_-}{2\varepsilon_+2\varepsilon_-}\delta^4(q_+ + q_- - q)Q_\mu Q_\nu = \frac{1}{3}\frac{\pi\beta}{2}\left(g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2}\right)Q^2, \quad \beta = \sqrt{1-\frac{4m_\pi^2}{q^2}}. \tag{25}$$

Upon rearranging the phase volume,

$$\int\frac{d^4q}{dq^2} = \frac{1}{2}\int\limits_{\sqrt{q^2}}^{\Delta\varepsilon} dq_0\sqrt{q_0^2-q^2}\int d\Omega_q, \tag{26}$$

the right-hand side of Eq. (24) can be recast in the form

$$\frac{\alpha^2}{3\pi^2}\int\limits_M^{\Delta\varepsilon} dq_0\sqrt{q_0^2-M^2}\int\frac{d\Omega_q}{4\pi}\left(\frac{p_1p_2}{p_1q\cdot p_2q} - \frac{p_1p_2'}{p_1q\cdot p_2'q}\right) = \frac{\alpha^2}{3\pi^2}\left(L_{su}\ln\frac{2\Delta\varepsilon}{M} + \mathbb{O}(1)\right). \tag{27}$$

The final result is then given by

$$\left.\frac{d\sigma}{dM^2d\sigma_0}\right|_{\pi^\pm} = \frac{\alpha^2}{6\pi^2M^2}[L_{su}(\rho_t-\rho_m) + \mathbb{O}(1)]. \tag{28}$$

Obviously, the contribution coming from the box-type diagrams with the hadronic vacuum polarization cannot be obtained in analytic form because of the presence of the quantity $\mathcal{R}(M^2)$.

3. SQUARED BOX AND THE CORRESPONDING SOFT PHOTON CORRECTIONS

The up–down interference of the soft photon emission from the electron line and the muon line can be evaluated using the expression

$$I_{p_Ap_B} = \frac{1}{4\pi}\int\frac{d^3\mathbf{k}}{\omega}\frac{p_Ap_B}{p_Ak\cdot p_Bk}\Bigg|_{\omega<\Delta\varepsilon} = \left(\ln\Delta + \frac{1}{2}\rho_\lambda\right)L_{AB} + \frac{1}{4}\left(L_{AB}^2 - \ln^2\frac{m_\mu}{m_e}\right) - \frac{\pi^2}{6} + \frac{1}{2}\mathrm{Li}_2\left(\frac{1+c}{2}\right), \tag{29}$$

where

$$L_{AB} = \ln\left(\frac{2p_Ap_B}{m_em_\mu}\right), \quad p_Ap_B = \varepsilon^2(1-c),$$

$$p_A^2 = m_e^2, \quad p_B^2 = m_\mu^2, \quad \varepsilon_A = \varepsilon_B \equiv \varepsilon,$$

and the quantity ω is the soft photon energy. Using the known results for interference of Born- and box-type elastic amplitudes (see Appendix B), we discover that in the soft photon approximation, the single soft photon emission contribution is given by

$$\frac{d\sigma_{box}^\gamma}{d\sigma_0} = \left(\frac{\alpha}{\pi}\right)^2\left[2L_{su}(\rho_t+\rho_\lambda) + \frac{t^2}{s^2+u^2}\times\left(\frac{u}{t}L_{st} - \frac{s}{t}L_{ut} + \frac{s-u}{2t}(\pi^2 + L_{ut}^2 + L_{st}^2)\right)\right] \times\left[-L_{su}\rho_t + \frac{1}{2}(L_{ut}^2 - L_{st}^2) - 2L_{su}\left(\ln\Delta + \frac{1}{2}\rho_\lambda\right) + \mathrm{Li}_2\left(\frac{1-c}{2}\right)\right]. \tag{30}$$

In the case of the emission of two soft photons with the total energy not exceeding $\Delta\varepsilon$, we have

$$\frac{d\sigma^{\gamma\gamma}}{d\sigma_0} = \left(\frac{2\alpha}{\pi}\right)^2\left\{\left[\frac{1}{2}\rho_tL_{su} + \frac{1}{4}(L_{st}^2 - L_{ut}^2) + L_{su}\left(\ln\Delta + \frac{1}{2}\rho_\lambda\right) - \frac{1}{2}\mathrm{Li}_2\left(\frac{1-c}{2}\right)\right]^2 - \frac{\pi^2}{6}L_{su}^2\right\}. \tag{31}$$

Finally, from the evaluation of the squared box-type graphs in Appendix B, we infer the logarithmic contribution

$$\frac{d\sigma_{BB}}{d\sigma_0} = \frac{\alpha^2}{\pi^2}\frac{t^2}{s^2+u^2}\rho_t[A\rho_t + B], \tag{32}$$

where the coefficients are given by

$$A = 2\frac{s^2+u^2}{t^2}(L_{us}^2 + \pi^2),$$

$$B = 4\frac{s^2+u^2}{t^2}(L_{us}^2+\pi^2)\rho_\lambda + 2L_{us}\left(\frac{s}{t}L_{ut} - \frac{u}{t}L_{st}\right) + \frac{s-u}{t}[\pi^2(2L_{st}-L_{us}) - L_{us}(L_{ut}^2+L_{st}^2)] + \frac{8u}{t}\pi^2.$$

4. SUMMARY

This paper is devoted to determining part of the second-order radiative corrections to the cross section of the process of large-angle quasielastic $e\mu$ scattering, namely, those corresponding to eikonal box-type diagrams. For box-type diagrams with a vacuum polarization insertion, we obtain the formulas in Eqs. (16), (17), and (28), which imply that the contributions coming from the interference between the tree-level diagram and those (bearing a vacuum polarization insertion)

with straight and crossed "legs" become, in fact, equal when we exchange $s \rightleftarrows u$ (with accuracy up to terms on the order of π^2) and alternate the overall sign of the contribution. This is indeed a manifestation of the well-known symmetry relation between amplitudes corresponding to different channels of a given reaction.

The main results of this work are analytic formulas given in the logarithmic approximation, but intermediate formulas presented to a power accuracy allow at least a numerical evaluation of the impact of subleading terms on the overall value of the corrections. For example, in Section 2, we obtain two limiting cases of the leptonic vacuum polarization contribution, for a small (Eq. (16)) and large (Eq. (17)) lepton pair invariant mass M with constant accuracy.

As a consistency check of the calculation, the auxiliary infrared parameter λ is expected to completely cancel in the final results. Within the gauge invariant set of amplitudes considered in Section 2, we show that, integrating over v and then adding the contribution given by the soft lepton pair production, we indeed obtain a result free of infrared divergences (Eq. (22)). The structure of this correction is in agreement with the RG predictions and does not contain large logarithms raised to a power higher than two, but the same cannot be done for the contributions calculated in Section 3 because the analysis there is in fact incomplete. We also give the expression for the cross section of soft pion pair production (Eq. (28)). Here, we cannot explicitly show the cancellation of the occurrence of leading or next-to-leading logarithms when the expression is combined with the corresponding virtual correction. This is because of a partially nonanalytic form of the expression for the radiative corrections caused by the hadronic vacuum polarization insertion.

In Section 3, we examined the contribution coming from squared box-type diagrams (see Eq. (32)) supplied by the corresponding one and two soft photon emission contributions with the explicit expressions given in Eqs. (30) and (31). To complete the picture, we must take the radiative corrections caused by genuine two-loop eikonal-type amplitudes into account. Keeping in mind the validity of the RG approach in the leading logarithmic approximation and the effect of cancellation of large logarithms in the expression for the lowest order radiative corrections to eikonal-type diagrams (see [1]), we expect the interference between them and the Born-level amplitude to completely cancel when added to the contributions in Eqs. (30)–(32). An explicit evaluation of them will be the subject of a forthcoming paper.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research (project no. 01-02-17437) and INTAS (grant no. 00366). We are also grateful for a Heisenberg–Landau 2001-02 grant for support. The work of B.G.S. was supported by Conacyt (México).

APPENDIX A

In this appendix, we give a set of scalar integrals encountered in dealing with box-type diagrams with a vacuum polarization insertion in one of the exchange virtual photon propagators. Clearly, in this case, we need integrals with a virtual exchange photon endowed with a mass M. In evaluating vector and tensor integrals, we therefore use the technique presented in Appendix B with the only change that all scalar integrals with three (I_{ijk}) and four (I) denominators are replaced by the following:

(1) in the case of a large mass M ($M^2 \gg s \sim -t$),

$$I_{123} = \frac{1}{M^2}\left\{-\ln\frac{M^2}{s} - 1 + \frac{s}{M^2}\left[\frac{1}{2}\ln\frac{M^2}{s} + \frac{1}{4}\right]\right\},$$

$$I_{134} = -\frac{1}{M^2}\left\{\ln\frac{M^2}{m_e^2} + 1 + \frac{t}{M^2}\left[\frac{1}{2}\ln\frac{M^2}{m_e^2} + \frac{1}{4}\right]\right\},$$

$$I_{234} = -\frac{1}{M^2}\left\{\ln\frac{M^2}{m_\mu^2} + 1 + \frac{t}{M^2}\left[\frac{1}{2}\ln\frac{M^2}{m_\mu^2} + \frac{1}{4}\right]\right\},$$

$$I = -\frac{1}{2sM^2}\left\{2\rho_s\rho_\lambda + \rho_s^2 - \frac{4\pi^2}{3}\right\}, \qquad \text{(A.1)}$$

$$I_3 = \tilde{t}I - I_{124} = \frac{1}{M^2}\left\{-\frac{2t}{s}\rho_s + 1 - \ln\frac{s}{M^2}\right.$$

$$\left. + \frac{s}{M^2}\left[\frac{1}{2}\ln\frac{s}{M^2} - \frac{1}{4} + \left(\frac{t}{s}\right)^2\rho_s\right]\right\},$$

(2) in the opposite limit $-t \gg M^2$, we must use the integrals

$$I_{134} = \frac{1}{t}\left[\ln\frac{M^2}{m_e^2}\ln\frac{-t}{M^2} + \frac{\pi^2}{6} + \frac{1}{2}\ln^2\frac{-t}{M^2}\right],$$

$$I_{234} = \frac{1}{t}\left[\ln\frac{M^2}{m_\mu^2}\ln\frac{-t}{M^2} + \frac{\pi^2}{6} + \frac{1}{2}\ln^2\frac{-t}{M^2}\right],$$

$$I_{123} = -\frac{1}{2s}\left[2\rho_s\rho_m + \frac{4\pi^2}{3} - \rho_s^2 + \ln^2\frac{m_\mu}{m_e}\right], \qquad \text{(A.2)}$$

$$I = \frac{1}{st}\rho_s[\rho_\lambda + 2\rho_t - \rho_m],$$

$$I_3 = \frac{1}{s}\left[2\rho_s \ln\frac{-t}{M^2} + \rho_s\rho_m - \frac{1}{2}\rho_s^2 + \frac{1}{2}\ln^2\frac{m_\mu}{m_e} + \frac{2\pi^2}{3}\right].$$

APPENDIX B

Here, we give the details of the box–box contribution calculation. First of all, we must distinguish three cases: two box squares with straight and crossed legs and one case with the interference of amplitudes with crossed and straight legs.

To calculate the contributions, we must evaluate tensor, vector, and scalar integrals with four and three denominators. We first consider the integral for the box with straight legs. The vector integral can be written as

$$\int\frac{d^4kk^\mu}{i\pi^2(1)(2)(3)(4)} = Ap_{1\mu} + Bp_{2\mu} + Cq_\mu, \quad \text{(B.1)}$$

where quantities (1), (2), and (4) were defined in (13), and we use the notation $m = m_e$, $M = m_\mu$, and (3) is $k^2 - 2kq + t$ with

$$q = p_1 - p_1' = p_2' - p_2, \quad q^2 = t. \quad \text{(B.2)}$$

The coefficients A, B, and C are determined as

$$\begin{aligned} A &= \frac{1}{2stu}[-t^2a - t(2s+t)b - stc], \\ B &= \frac{1}{2stu}[-t(2s+t)a - t^2b + stc], \\ C &= \frac{1}{2stu}[-sta + stb - s^2c], \\ a &= I_{123} - I_{234}, \quad b = I_{134} - I_{123}, \quad c = tI. \end{aligned} \quad \text{(B.3)}$$

The scalar integrals I and I_{ijk} are given by

$$I = \int\frac{d^4k}{i\pi^2(1)(2)(3)(4)} = \frac{2}{st}\left[\ln\frac{s}{mM} - i\pi\right]\ln\frac{-t}{\lambda^2},$$

$$\begin{aligned} I_{123} &= \int\frac{d^4k}{i\pi^2(1)(2)(3)} \\ &= -\frac{1}{2s}\left[2\left[\ln\frac{s}{mM} - i\pi\right]\ln\frac{\lambda^2}{mM}\right. \\ &\left. + \frac{\pi^2}{3} - \left[\ln\frac{s}{mM} - i\pi\right]^2 + \ln^2\frac{M}{m}\right], \end{aligned} \quad \text{(B.4)}$$

$$I_{134} = \int\frac{d^4k}{i\pi(1)(3)(4)} = \frac{1}{t}\left[\frac{1}{2}\ln^2\frac{-t}{m^2} + \frac{2\pi^2}{3}\right],$$

$$I_{234} = \int\frac{d^4k}{i\pi(2)(3)(4)} = \frac{1}{t}\left[\frac{1}{2}\ln^2\frac{-t}{M^2} + \frac{2\pi^2}{3}\right].$$

To consider the tensor integral, we use the algebraic method,

$$\begin{aligned} \int\frac{d^4kk_\mu k_\nu}{i\pi^2(1)(2)(3)(4)} &= a_g g_{\mu\nu} + a_{11}p_{1\mu}p_{1\nu} \\ &+ a_{22}p_{2\mu}p_{2\nu} + a_{12}(p_{1\mu}p_{2\nu} + p_{1\nu}p_{2\mu}) \\ &+ a_{1q}(p_{1\mu}q_\nu + p_{1\nu}q_\mu) + a_{2q}(p_{2\mu}q_\nu + p_{2\nu}q_\mu) \\ &+ a_{qq}q_\mu q_\nu. \end{aligned} \quad \text{(B.5)}$$

Multiplying the above equation with four-vectors p_1, p_2, and q, we obtain a system of algebraic equations, whence the quantities a_{ij} are expressed through the scalar integrals

$$\begin{aligned} a_{22} &= \frac{1}{s}(A_2 - ta_{2q}), \\ a_{11} &= \frac{1}{s}(A_4 + ta_{1q}), \\ a_{12} &= \frac{1}{s}(A_1 - 2a_g - ta_{1q}), \\ a_g &= \frac{1}{2}(A_9 - 2ta_{qq} - ta_{1q} + ta_{2q}), \\ a_{1q} &= \frac{1}{t}(A_1 - A_5 - ta_{2q}), \\ a_{2q} &= \frac{1}{s+t}(A_3 + A_{10} - A_5 - A_9), \\ a_{qq} &= \frac{1}{t(s+t)}(tA_3 + s(A_5 + A_9 - A_{10})). \end{aligned} \quad \text{(B.6)}$$

The quantities A_j are given by

$$\begin{aligned} A_1 &= \frac{1}{s}(I_{13} - I_{12}), \\ A_2 &= I_{234} + \frac{1}{s}(I_{12} - I_{23}) + \frac{1}{t}(2I_{34} - I_{23} - I_{24}), \\ A_3 &= I_{123} + \frac{1}{s}(2I_{12} - I_{13} - I_{23}) + \frac{1}{t}(-I_{23} + I_{34}), \\ A_4 &= I_{134} + \frac{1}{s}(I_{12} - I_{13}) + \frac{1}{t}(2I_{34} - I_{13} - I_{14}), \\ A_5 &= \frac{1}{s}(I_{23} - I_{12}), \\ A_9 &= tC + I_{123} + \frac{1}{s}(2I_{12} - I_{13} - I_{23}), \\ A_{10} &= I_{123}, \end{aligned} \quad \text{(B.7)}$$

where I_{ij} denote scalar integrals with two denominators,

$$I_{12} = \int\frac{d^4k}{i\pi^2(1)(2)} = \ln\frac{\Lambda^2}{M^2} - \ln\frac{s}{M^2} + i\pi + 1,$$

$$I_{13} = \int\frac{d^4k}{i\pi^2(1)(3)} = I_{14} = \int\frac{d^4k}{i\pi^2(1)(4)}$$

$$= \ln\frac{\Lambda^2}{M^2} + \ln\frac{M^2}{m^2} + 1, \qquad \text{(B.8)}$$

$$I_{23} = \int\frac{d^4k}{i\pi^2(2)(3)} = I_{24} = \int\frac{d^4k}{i\pi^2(2)(4)} = \ln\frac{\Lambda^2}{M^2} + 1,$$

$$I_{34} = \int\frac{d^4k}{i\pi^2(3)(4)} = \ln\frac{\Lambda^2}{M^2} - \ln\frac{-t}{M^2} + 1,$$

and I_{ijk} and I are determined above. For crossed legs in a box-type diagram, we must evaluate the integrals

$$\int\frac{d^4kk^\mu}{i\pi^2(1)(\tilde{2})(3)(4)} = \tilde{A}p_{1\mu} - \tilde{B}p'_{2\mu} + \tilde{C}q_\mu, \qquad \text{(B.9)}$$

where $(\tilde{2}) = k^2 + p'_2k$ and

$$\tilde{A} = \frac{1}{2stu}[-t^2\tilde{a} - t(2u+t)\tilde{b} - ut\tilde{c}],$$
$$\tilde{B} = \frac{1}{2stu}[-t(2u+t)\tilde{a} - t^2b + ut\tilde{c}],$$
$$\tilde{C} = \frac{1}{2stu}[-ut\tilde{a} + ut\tilde{b} - u^2\tilde{c}], \qquad \text{(B.10)}$$
$$\tilde{a} = I_{1\tilde{2}3} - I_{\tilde{2}34}, \quad \tilde{b} = I_{134} - I_{1\tilde{2}3}, \quad \tilde{c} = t\tilde{I}.$$

The integrals are given by

$$\tilde{I} = \int\frac{d^4k}{i\pi^2(1)(\tilde{2})(3)(4)} = \frac{2}{ut}\ln\frac{-u}{mM}\ln\frac{-t}{\lambda^2},$$

$$I_{1\tilde{2}3} = \int\frac{d^4k}{i\pi^2(1)(\tilde{2})(3)}$$
$$= -\frac{1}{2u}\left[2\ln\frac{-u}{mM}\ln\frac{\lambda^2}{mM} + \frac{\pi^2}{3} - \ln^2\frac{-u}{mM} + \ln^2\frac{M}{m}\right], \qquad \text{(B.11)}$$

$$I_{\tilde{2}34} = \int\frac{d^4k}{i\pi(\tilde{2})(3)(4)} = \frac{1}{t}\left[\frac{1}{2}\ln^2\frac{-t}{M^2} + \frac{2\pi^2}{3}\right],$$

and I_{134} is given in (B.4).

For the tensor integral, we have

$$\int\frac{d^4kk_\mu k_\nu}{i\pi^2(1)(\tilde{2})(3)(4)} = \tilde{a}_g g_{\mu\nu} + \tilde{a}_{11}p_{1\mu}p_{1\nu}$$
$$+ \tilde{a}_{22}p'_{2\mu}p'_{2\nu} - \tilde{a}_{12}(p_{1\mu}p'_{2\nu} + p_{1\nu}p'_{2\mu}) \qquad \text{(B.12)}$$
$$+ \tilde{a}_{1q}(p_{1\mu}q_\nu + p_{1\nu}q_\mu) - \tilde{a}_{2q}(p'_{2\mu}q_\nu + p'_{2\nu}q_\mu)$$
$$+ \tilde{a}_{qq}q_\mu q_\nu,$$

where we use

$$\tilde{a}_{22} = \frac{1}{u}(\tilde{A}_2 - t\tilde{a}_{2q}),$$

$$\tilde{a}_{11} = \frac{1}{u}(\tilde{A}_4 + t\tilde{a}_{1q}),$$

$$\tilde{a}_{12} = \frac{1}{u}(\tilde{A}_1 - 2\tilde{a}_g - t\tilde{a}_{1q}),$$

$$\tilde{a}_g = \frac{1}{2}(\tilde{A}_9 - 2t\tilde{a}_{qq} - t\tilde{a}_{1q} + t\tilde{a}_{2q}), \qquad \text{(B.13)}$$

$$\tilde{a}_{1q} = \frac{1}{t}(\tilde{A}_1 - \tilde{A}_5 - t\tilde{a}_{2q}),$$

$$\tilde{a}_{2q} = \frac{1}{u+t}(\tilde{A}_3 + \tilde{A}_{10} - \tilde{A}_5 - \tilde{A}_9),$$

$$\tilde{a}_{qq} = \frac{1}{t(u+t)}(t\tilde{A}_3 + u(\tilde{A}_5 + \tilde{A}_9 - \tilde{A}_{10})).$$

The quantities $\tilde{A}_j$ are given by

$$\tilde{A}_1 = \frac{1}{u}(I_{13} - I_{1\tilde{2}}),$$

$$\tilde{A}_2 = I_{\tilde{2}34} + \frac{1}{u}(I_{1\tilde{2}} - I_{\tilde{2}3}) + \frac{1}{t}(2I_{34} - I_{\tilde{2}3} - I_{\tilde{2}4}),$$

$$\tilde{A}_3 = I_{1\tilde{2}3} + \frac{1}{u}(2I_{1\tilde{2}} - I_{13} - I_{\tilde{2}3}) + \frac{1}{t}(-I_{\tilde{2}3} + I_{34}),$$

$$\tilde{A}_4 = I_{134} + \frac{1}{u}(I_{1\tilde{2}} - I_{13}) + \frac{1}{t}(2I_{34} - I_{13} - I_{14}), \qquad \text{(B.14)}$$

$$\tilde{A}_5 = \frac{1}{u}(I_{\tilde{2}3} - I_{1\tilde{2}}),$$

$$\tilde{A}_9 = t\tilde{C} + I_{1\tilde{2}3} + \frac{1}{u}(2I_{1\tilde{2}} - I_{13} - I_{\tilde{2}3}),$$

$$\tilde{A}_{10} = I_{1\tilde{2}3},$$

where I_{ij} denote scalar integrals with two denominators,

$$I_{1\tilde{2}} = \int \frac{d^4k}{i\pi^2(1)(\tilde{2})} = \ln\frac{\Lambda^2}{M^2} - \ln\frac{-u}{M^2} + 1,$$
$$I_{\tilde{2}3} = \int \frac{d^4k}{i\pi^2(\tilde{2})(3)} = I_{\tilde{2}4} = \int \frac{d^4k}{i\pi^2(\tilde{2})(4)} = \ln\frac{\Lambda^2}{M^2} + 1, \quad \text{(B.15)}$$

and I_{ijk} and I are determined above. The other integrals, I_{13}, I_{14}, and I_{34} are given in (B.8).

With all these integrals, we can straightforwardly obtain the final result for the squared box-type diagrams. With the intention of realizing subsequent numeric calculations, we give it in the form where all terms not enhanced by large logarithms are retained,

$$\sum |\mathcal{M}_{box}|^2 = 16\alpha^4 \mathcal{B}(s,t,u),$$

$$\mathcal{B}(s,t,u) = \frac{8(s^2+u^2)}{t^2}(L_{us}^2 + \pi^2)\ln^2\left(\frac{-t}{\lambda^2}\right)$$
$$- 4\ln\left(\frac{-t}{\lambda^2}\right)L_{us}\left[\frac{s-u}{t}(L_{ut}^2 + L_{st}^2 - L_{ut} - L_{st}) + L_{us}\right]$$
$$+ \frac{(s-u)^2}{2}\left[\frac{1}{s^2}L_{ut}^4 + \frac{1}{u^2}L_{st}^4\right]$$
$$+ 2(s-u)\left[-\frac{1}{s}L_{ut}^3 + \frac{1}{u}L_{st}^3\right] + 2[L_{ut}^2 + L_{st}^2] \quad \text{(B.16)}$$
$$+ \pi^2\left\{4\ln\left(\frac{-t}{\lambda^2}\right)\left[\frac{s-u}{t}(2L_{st} - L_{us}) + \frac{2u}{t}\right]\right.$$
$$\left. + \left[L_{ut}\left(1 - \frac{u}{s}\right) - 1\right]^2 + 2\left[L_{st}\left(1 - \frac{s}{u}\right) - 1\right]^2 - 1\right\}$$
$$+ \frac{\pi^4}{2}\left(1 - \frac{u}{s}\right)^2.$$

For completeness, we here present a formula for the interference of a tree-level and a box-type diagram amplitude,

$$2\sum \mathcal{M}_{\text{Born}}^* \mathcal{M}_{box}$$
$$= \sum |\mathcal{M}_{\text{Born}}|^2 \frac{\alpha}{\pi}\left\{2L_{su}(\rho_t + \rho_\lambda)\right. \quad \text{(B.17)}$$
$$\left. + \frac{t^2}{s^2+u^2}\left[\frac{u}{t}L_{st} - \frac{s}{t}L_{ut} + \frac{s-u}{2t}(\pi^2 + L_{ut}^2 + L_{st}^2)\right]\right\}.$$

Adding to this expression the contribution arising from the up–down interference of a soft photon emission by electron and muon lines, we arrive at the expression for the radiative corrections given in Eq. (16) in [1].

REFERENCES

1. V. Bytev, E. Kuraev, and B. Shaikhatdenov, E-print archives, hep-ph/0203127; Zh. Éksp. Teor. Fiz. **122**, 472 (2002) [JETP **95**, 404 (2002)].
2. E. Kuraev and V. Fadin, Yad. Fiz. **41**, 733 (1985) [Sov. J. Nucl. Phys. **41**, 466 (1985)]; Yad. Fiz. **47**, 1593 (1988) [Sov. J. Nucl. Phys. **47**, 1009 (1988)].
3. J. Schwinger, Phys. Rev. **76**, 790 (1949); R. Barbieri, J. Mignaco, and E. Remiddi, Nuovo Cimento A **11**, 865 (1972).
4. S. Eidelman and F. Jegerlehner, Z. Phys. C **67**, 585 (1995).
5. A. Arbuzov *et al.*, Yad. Fiz. **60**, 673 (1997) [Phys. At. Nucl. **60**, 591 (1997)].
6. A. B. Arbuzov, E. A. Kuraev, and B. G. Shaikhatdenov, Zh. Éksp. Teor. Fiz. **115**, 392 (1999) [JETP **88**, 213 (1999)].

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 202–221.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 233–254.

ATOMS, SPECTRA, RADIATION

Two-Wave Stimulated Raman Scattering in the Field of Intense Radiation Resonant to the Raman Transition

V. P. Kochanov* and Yu. V. Bogdanova

Institute of Atmospheric Optics, Siberian Division, Russian Academy of Sciences, Akademicheskiĭ pr. 1, Tomsk, 634055 Russia

**e-mail: koch@iao.ru*

Received April 17, 2002

Abstract—The influence of a strong field resonant to the Raman transition on the behavior of the Stokes harmonic in saturated two-wave stimulated Raman scattering is theoretically considered based on the description of a gas of three-level atoms by medium density matrix equations. The vibrational mode of energy redistribution between laser and scattered radiation, threshold with respect to the resonant field intensity, is predicted. The action of optical fields on the absorption and amplification of the resonant field in a Raman medium is described. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known that, in simple quantum systems involved in resonant interactions with two stationary laser radiation fields, linking up a third field whose frequency is equal to the difference of the frequencies of these two fields introduces a qualitatively new interaction factor, namely, the relation between wave phases. This factor can have a substantial influence on the distribution of level populations and nonlinear interference effects that act on the absorbing ability of the system and the shape of absorption and spontaneous emission lines. It has, for instance, been shown in [1] that, in the absence of radiative and collisional relaxation, population distribution and absorption in three- and four-level systems are sensitive to the summed phase of the fields. Based on this sensitivity, the atomic interferometry method has been suggested. The use of phase relations under the conditions of ring composition of fields in the problem of deep cooling of atoms, which are modeled by the *W*-scheme of levels, allows more substantial coolings to be attained [2, 3]. It is shown in [4] that the distribution of populations in a three-level Λ-system that interacts with three resonant fields can vary in a wide range depending on the summed phase and field value at the forbidden transition. This effect is observed under intensities of the order of or larger than saturation intensities and results in noise pollution of absorption and spontaneous emission lines caused by fluctuations of field phases.

Phase relations under the ring composition of fields should also influence the conditions of nonresonant nonlinear interactions of optical fields with atoms and molecules. One of such processes is the generation of the Stokes harmonic in two-wave stimulated Raman scattering in the presence of a strong field whose frequency is resonant to the transition responsible for scattering. The interaction of a field with a Raman transition can be magnetodipole, quadrupole, or electrodipole because of the removal of parity selection rules as a result of intramolecular interactions or the presence of external constant fields. We will for definiteness consider magnetodipole interactions, because the solutions to the electro- and magnetodipole interaction problems are identical up to the initial notation (the magnetic dipole moment and magnetic field strength are replaced by the corresponding electric quantities), whereas the quadrupole interaction problem should be solved separately. One of the possible and easily implementable methods for closing stimulated Raman scattering is two-photon resonance at the Raman transition frequency.

The most significant consequences of linking up the resonant field should be expected for nonlinear interactions of optical fields with atoms (molecules) [4]. Therefore, generally, we must consider saturated stimulated Raman scattering in a strong resonant field. Here, saturation is understood as both leveling of atomic populations during scattering by one atom (saturation proper) and various parametric, Raman, and interference processes [5, 6], which are related to the higher orders of the expansion of macroscopic polarization in powers of field amplitudes, that simultaneously come into effect.

The purpose of this work was to analyze the influence of a strong field resonant to the Raman transition on the generation of the Stokes component of saturated two-wave stimulated Raman scattering and to determine the character of damping and amplification of the resonant wave in the presence of two intense optical waves. In other words, we consider the nonlinear mixing of three waves caused by their ring composition under stimulated Raman scattering conditions.

1063-7761/03/9602-0202$24.00

So far as we know, no consistent theory of saturated stimulated Raman scattering has been suggested, although an approach to exact calculations of the nonlinear susceptibility of a three-level system involved in resonant interactions with three fields has been formulated in classic monograph [5]. Namely, in [5], stationary equations for the medium density matrix in the general case of three strong fields were obtained and used to classify the types of possible nonlinear processes, and several conditions that controlled the ratio between these types were obtained for the particular case of a single strong optical field. Phase relations were not taken into consideration in [5]. For this reason, we present a detailed elaboration of this approach, which is also described in monograph [6] in general terms. The approach is based on analytical calculation of polarization and magnetization nonlinear with respect to the field with the use of the equations for the medium density matrix (gas of three-level atoms) for the case of three strong fields. We use the simplest stimulated Raman scattering model, according to which the predominant scattered radiation component is the first Stokes harmonic, whereas the generation of the anti-Stokes and higher Stokes harmonics and also back scattering are suppressed under certain experimental conditions. The determination of these conditions and calculations of the total spectrum of harmonics is a separate topical problem, which has repeatedly been handled by theoretical and experimental physicists (e.g., see [7–9]). In addition, in what follows, the laser radiation pulse width is assumed to far exceed the medium transverse relaxation time, which allows calculations of stimulated Raman scattering to be limited to analyzing the development of the process in space.

2. NONLINEAR MEDIUM RESPONSE

In conformity with the accepted stimulated Raman scattering model, the field interacting with three-level atoms is written as

$$\begin{aligned} \mathscr{E}(t,z) &= E_1(z)\cos\Psi_1 + E_2(z)\cos\Psi_2, \\ \mathscr{H}(t,z) &= H(z)\cos\Psi_3, \\ \Psi_j &= \omega_j t - k_j z + \varphi_j, \quad j = 1,2,3, \\ \omega_3 &= \omega_1 - \omega_2 > 0, \quad k_3 = k_1 - k_2. \end{aligned} \tag{1}$$

Here, $E_1(z)$, ω_1, φ_1 and $E_2(z)$, ω_2, φ_2 are the amplitudes of the electric fields and the frequencies and amplitudes of laser (1) and Stokes (2) radiation; $H(z)$, ω_3, and φ_3 are the amplitude of the magnetic field and the frequency and phase of radiation resonant to the Raman transition; and k_j are the corresponding wave numbers. Note that the last equality in (1) is only approximately satisfied under wave mismatching conditions; this circumstance will be taken into account when necessary.

An atom (molecule) is represented by a three-level system whose states 1 (lower) and 2 are involved in the Raman transition, whereas the higher level (3) is the excited electronic state level that makes the major contribution to the Raman scattering tensor. The equations for the medium density matrix in the model of relaxation constants corresponding to homogeneous line broadening have the form

$$\begin{aligned} \dot\rho_1 + \gamma(\rho_1 - \rho_1^0) &= 2\mathrm{Re}\frac{i(d_1\mathscr{E}\rho_{31} + \mu\mathscr{H}\rho_{21})}{\hbar}, \\ \dot\rho_2 + \gamma(\rho_2 - \rho_2^0) &= 2\mathrm{Re}\frac{i(d_2\mathscr{E}\rho_{32} - \mu\mathscr{H}\rho_{21})}{\hbar}, \\ \rho_1 + \rho_2 + \rho_3 &= 1, \quad \rho_1^0 + \rho_2^0 = 1, \\ \dot\rho_{31} + (\Gamma_1 + i\omega_{31})\rho_{31} &= \frac{i}{\hbar}[d_1\mathscr{E}(\rho_1 - \rho_3) + d_2\mathscr{E}\rho_{21} - \mu\mathscr{H}\rho_{32}], \\ \dot\rho_{32} + (\Gamma_2 + i\omega_{32})\rho_{32} &= \frac{i}{\hbar}[d_2\mathscr{E}(\rho_2 - \rho_3) + d_1\mathscr{E}\rho_{21}^* - \mu\mathscr{H}\rho_{31}]; \\ \dot\rho_{21} + (\Gamma + i\omega_{21})\rho_{21} &= \frac{i}{\hbar}[\mathscr{E}(d_2\rho_{31} - d_1\rho_{32}^*) + \mu\mathscr{H}(\rho_1 - \rho_2)]. \end{aligned} \tag{2}$$

Here, ρ_j are the populations of levels $j = 1$, 2, and 3; ρ_{31} and ρ_{32} are the off-diagonal elements of the density matrix (up to the polarization factor) of the allowed nonresonant 1–3 and 2–3 transitions; ρ_{21} is the polarization of the 1–2 forbidden Raman transition; d_1 and d_2 are the dipole moment matrix elements of the 1–3 and 2–3 transitions, respectively; μ is the magnetic dipole moment matrix element of the 1–2 transition; ω_{ij} are the natural frequencies of the j–i transitions; γ and Γ, Γ_1, and Γ_2 are the collisional diagonal and off-diagonal relaxation constants, respectively; and ρ_1^0 and ρ_2^0 are the equilibrium populations of levels 1 and 2. The relaxation scheme accepted in (2) corresponds to the situation in which fairly many levels are situated close to levels 1 and 2, and collisional population transfer from the working transition levels predominantly occurs to neighboring states. Radiative relaxation is considered to be insignificant compared with collisional relaxation and is ignored.

The sought medium response containing both linear and nonlinear components without the expansion of the latter in powers of field amplitudes is polarization P and magnetization M,

$$P = N\mathrm{Sp}\hat\rho\hat d, \quad M = N\mathrm{Sp}\hat\rho\hat\mu, \tag{3}$$

where N is the density of atoms interacting with radiation.

We assume that laser pulses are sufficiently wide and gas pressure is sufficiently high for the Doppler broadening of lines to be ignored. The transition from (2) to the stationary equations can then be performed by separating the "fast" dependences of the polarizations on time and coordinate in the standard way [5],

$$\rho_{31} = R_1 e^{-i\Psi_1}, \quad \rho_{32} = R_2 e^{-i\Psi_2}, \quad \rho_{21} = re^{-i\Psi_3}. \quad (4)$$

Substituting (4) into (2) and introducing the notation

$$\Delta = \omega_{31} - \omega_1 \approx \omega_{32} - \omega_2 > 0, \quad \varepsilon^2 = \frac{\Gamma}{\Delta},$$

$$\Phi = \varphi_1 - \varphi_2 - \varphi_3, \quad (5)$$

$$V_1 = \frac{d_1 E_1}{2\hbar\sqrt{\Delta\Gamma}}, \quad V_2 = \frac{d_2 E_2}{2\hbar\sqrt{\Delta\Gamma}}, \quad G = \frac{\mu H}{2\hbar\Gamma}$$

and the natural simplifying assumptions

$$\gamma = \Gamma, \quad \omega_1 - \omega_2 = \omega_{21}, \quad \Gamma_{1,2} \ll \Delta, \quad (6)$$

we obtain

$$\rho_1 = \rho_1^0 - \frac{2V_1 R_1''}{\varepsilon} - 2Gr'',$$

$$\rho_2 = \rho_2^0 - \frac{2V_2 R_2''}{\varepsilon} + 2Gr'', \quad (7)$$

$$R_1 + Ge^{i\Phi}R_2 = \varepsilon V_1(2\rho_1 + \rho_2 - 1) + \varepsilon V_2 e^{i\Phi} r,$$

$$R_2 + Ge^{-i\Phi}R_1 = \varepsilon V_2(\rho_1 + 2\rho_2 - 1) + \varepsilon V_1 e^{-i\Phi} r^*,$$

$$r = ie^{-i\Phi}\frac{V_2 R_1 - V_1 R_2^*}{\varepsilon} + iG(\rho_1 - \rho_2),$$

$$R_{1,2} = R_{1,2}' + iR_{1,2}'', \quad r = r' + ir''.$$

The system of stationary equations (7) describes the exact one-photon resonance of external field G with the Raman transition and is valid under the following conditions:

(1) the "rotating wave" resonance approximation holds,

$$\Gamma \ll \omega_{21}, \quad \omega_{21} \gg V_{1,2}^0 = \frac{d_{1,2}E_{1,2}}{2\hbar}, \quad G^0 = \frac{\mu H}{2\hbar};$$

(2) the inequality $|2\omega_{21} - \omega_{31}| \gg V_{1,2}^0$ is satisfied for unequally spaced levels;

(3) hyper-Raman scattering is absent, $|2\omega_1 - \omega_2 - \omega_{31}| \gg V_{1,2}^0$, $|2\omega_2 - \omega_1 - \omega_{32}| \gg V_{1,2}^0$, or $|\Delta - \omega_{21}| \gg V_{1,2}^0$;

(4) parametric processes initiated by magnetic field are absent, $(\mu H)^2/(2\hbar\Delta)^2 \ll 1$;

(5) stimulated Raman scattering is quasi-resonant, $\Delta < \omega_{1,2}$;

(6) the higher Stokes and anti-Stokes scattering harmonics are absent.

If the corresponding frequency detunings in conditions (1)–(3) are assumed to be equal to the $V_{1,2}^0$ Rabi frequencies, the optical field intensities at $\delta \geq 100$ cm^{-1} and $d_{1,2} = 1$ D are limited by the inequality

$$I_{1,2} \leq \frac{c(\hbar\delta/d_{1,2})^2}{2\pi} = 30 \text{ GW/cm}^2.$$

Here, c is the velocity of light and $\hbar$ is the Planck constant. If the magnetic dipole moment is on the order of one Bohr magneton (μ_B), condition (4) at the same intensity values is satisfied if $\Delta \geq 1$ cm^{-1}. Note that limitation (5) is not strict, because if it is violated, the necessary taking into account of complex conjugate exponents in (4) leads to equations that can be reduced to (7) by redefining the $d_{1,2}$ and Δ values. The strictest limitation is the last one, according to which (7) is only applicable if stimulated Raman scattering field intensities are of the order of saturating intensities. Nevertheless, in spite of this limitation, it is expedient to obtain a formally exact solution to (7) in which all three fields are assumed to be strong, because such a solution gives a compact form of the representation of wave equations and allows trends of saturating stimulated Raman scattering to be correctly described.

It follows from (5) and (7) that the dimensionless electric and magnetic field amplitudes or dimensionless Rabi frequencies $V_{1,2}$ and G, if they equal one, determine the intensities of stimulated Raman scattering saturation and magnetic field G absorption saturation at the Raman transition in a natural way,

$$I_{1,2}^E = \frac{c\Gamma\Delta(\hbar/\mu)^2}{2\pi}, \quad I^H = \frac{c(\hbar\Gamma/\mu)^2}{2\pi}.$$

Assuming that $\Delta = 10^3$–10^4 cm^{-1}, $\Gamma = 0.1$ cm^{-1}, $d_{1,2} = 1$ D, and $\mu = \mu_B = 0.93 \times 10^{-20}$ CGS units, we obtain $I_{1,2}^E = 0.3$–3 GW/cm^2 and $I^H = 50$ MW/cm^2. It follows that, formally, system (7) gives a correct description of the process under consideration under typical experimental conditions at intensities 1–2 orders of magnitude higher than saturation intensities. In reality, the intensities of pumping and of the Stokes harmonic should not exceed $I_{1,2}^E$ to prevent the generation of the higher harmonics, but field G can be arbitrary within the limits specified above.

In the accepted notation, the sought polarization and magnetization [Eq. (3)] have the form

$$\begin{aligned} P &= 2N[d_1(R_1'\cos\Psi_1 + R_1''\sin\Psi_1) \\ &+ d_2(R_2'\cos\Psi_2 + R_2''\sin\Psi_2)], \\ M &= 2N\mu(r'\cos\Psi_3 + r''\sin\Psi_3). \end{aligned} \tag{8}$$

The exact algebraic solutions to (7) are somewhat cumbersome (the complete solutions to analogous equations for level populations at arbitrary frequency detunings can be found in [4]). For this reason, we will use simpler approximate solutions written accurate to $O(\varepsilon^3)$. Numerical calculations show that these solutions give errors less than 1% for dimensionless field amplitudes $V_{1,2}$, $G < 10$,

$$\begin{aligned} R_1'' &= -\frac{V_2 R_2''}{V_1} = \frac{\varepsilon V_2(\rho_1^0 - \rho_2^0)}{D} \\ &\times \{V_1V_2 + G[\cos\Phi + (V_1^2 - V_2^2)\sin\Phi]\}, \\ r'' &= \frac{\rho_1^0 - \rho_2^0}{D}\{G + V_1V_2[\cos\Phi - (V_1^2 - V_2^2)\sin\Phi]\}, \\ R_1' &= \frac{\varepsilon}{D}\{V_1[V_2^2(V_1^2 + V_2^2) + 2G^2] \\ &+ V_2G[(3V_1^2 + V_2^2)\cos\Phi + \sin\Phi] \\ &+ [V_1(1 + V_1^4 - V_2^4) \\ &+ 2V_2G((V_1^2 - V_2^2)\cos\Phi - \sin\Phi)]\rho_1^0\}, \\ R_2' &= \frac{\varepsilon}{D}\{V_2[1 + V_2^2(V_1^2 + V_2^2) + 2G^2] \\ &+ V_1G[(5V_2^2 - V_1^2)\cos\Phi + \sin\Phi] \\ &- [V_2(1 - V_1^4 + V_2^4) \\ &- 2V_1G((V_1^2 - V_2^2)\cos\Phi - \sin\Phi)]\rho_1^0\}, \\ r' &= \frac{\rho_1^0 - \rho_2^0}{D}\{G(V_1^2 - V_2^2) \\ &+ V_1V_2[(V_1^2 - V_2^2)\cos\Phi + \sin\Phi]\}, \\ D &= 1 + (V_1^2 + V_2^2)^2 + 4G^2 + 8V_1V_2\cos\Phi. \end{aligned} \tag{9}$$

Note that, in this approximation, the sum of the populations of two lower levels is exactly equal to one at arbitrary field intensities, and the higher level remains unpopulated. An analysis of the exact solution shows that the higher level can only be populated at much higher field intensities. Note also that the imaginary parts of the polarizations of optical transitions responsible for the nonlinear mixing of fields are equal to within a factor. It is shown in the next section that the Manley–Rowe theorem [10] is satisfied thanks precisely to this circumstance. Generally, the exact solution does not give such an equality, and this theorem is therefore violated at ultimately high intensities.

Importantly, the polarizations of the allowed and forbidden transitions [Eq. (9)] contain interference terms in the form of the products of field amplitudes raised to the first power and factors depending on the summed phase Φ in the numerators and denominator D. It follows that nonlinear interference effects can be controlled under the ring composition of fields by varying field phases.

Equations (8) and (9) give a solution to the problem of a nonlinear medium response at a definite point in space (for one atom). We will use them in the problem of Raman mixing of waves to trace the development in space of the generation of the Stokes harmonic and the absorption or amplification of the resonant wave.

3. WAVE EQUATIONS FOR SLOW AMPLITUDES

The generation of stimulated Raman scattering in the presence of a variable magnetic field is described by the wave equations

$$\begin{aligned} \left(\frac{\partial^2}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathscr{E}(t, z) &= \frac{4\pi}{c^2}\frac{\partial^2}{\partial t^2}P(t, z), \\ \left(\frac{\partial^2}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathscr{H}(t, z) &= \frac{4\pi}{c^2}\frac{\partial^2}{\partial t^2}M(t, z). \end{aligned} \tag{10}$$

Performing standard transformations, which include the substitution of (8) and (9) into (10), the neglect of small second derivatives of field amplitudes with respect to the coordinate, and the equating to zero of the sums of the coefficients of $\cos\Psi_j$ and $\sin\Psi_j$, $j = 1$–3, we obtain algebraic equations for wave numbers k_j and a system of three first-order differential equations for the coupled electric and magnetic field amplitudes. An analysis of (9) for the real parts of the polarizations, which determine wave numbers, shows that, for arbitrary field intensities, the contributions to the polarizations nonlinear in field values are limited from above by a value on the order of the linear contribution to the polarization. For this reason and because fine effects related to the intensity dependence of refractive indexes are beyond the scope of this paper, we will restrict ourselves, to the linear approximation, to optical polarizations, namely, $R_{1,2}' \approx \varepsilon V_{1,2}\rho_{1,2}^0$, and ignore the dispersion of the resonant wave. This leads to the following

simple equations for wave numbers k_j and refractive indexes n_j in terms of the parameters introduced above:

$$k_{1,2} \approx \frac{\omega_{1,2} n_{1,2}}{c}, \quad n_{1,2}^2 \approx 1 + \frac{4\pi N d_{1,2}^2}{\hbar\Delta}\rho_{1,2}^0, \qquad (11)$$
$$k_3 \approx k_1 - k_2.$$

Note that writing the refractive index in terms of the d (effective dipole moment) and Δ (effective pumping frequency detuning from the electronic transition frequency) three-level model parameters allows the polarizability of the molecule (Raman scattering tensor), which is $\alpha = d^2/\hbar\Delta$ in terms of our approach, to be estimated. Indeed, (11) gives $\alpha \approx (n-1)/2\pi N$. Substituting $N = 2.5 \times 10^{19}$ cm^{-3}, which corresponds to atmospheric conditions, and $n - 1 \approx 3 \times 10^{-4}$ [11] (the nitrogen molecule) into this result yields $\alpha \approx 1.9 \times 10^{-24}$ cm^3. This is close to $\alpha = 1.76 \times 10^{-24}$ cm^3 obtained from the data on Raman scattering (radiation wavelength 0.366 μm) by the 2331 cm^{-1} nitrogen vibrational mode [12]. At $d = 1$ D, the latter value gives the effective detuning $\Delta = 2860$ cm^{-1}. The stimulated Raman scattering saturation intensity is then estimated as

$$I^E = \frac{c\hbar\Gamma}{2\pi\alpha} = 0.86 \text{ GW/cm}^2$$

at $\Gamma = 0.1$ cm^{-1}.

If energy dissipation is ignored, the equations for the coupled dimensionless field amplitudes obtained from (10) as described above have the form

$$\frac{dV_1}{d\zeta} = -\frac{1}{2}V_2 F_1(V_1, V_2, G),$$
$$\frac{dV_2}{d\zeta} = -\frac{1}{2q}V_1 F_1(V_1, V_2, G), \qquad (12)$$
$$\frac{dG}{d\zeta} = -\frac{p}{2}F_2(V_1, V_2, G),$$

$$F_1 = \frac{1}{D}\{V_1 V_2 + G[\cos\Phi + (V_1^2 - V_2^2)\sin\Phi]\},$$

$$F_2 = \frac{1}{D}\{G + V_1 V_2[\cos\Phi - (V_1^2 - V_2^2)\sin\Phi]\},$$

$$\zeta = Qz, \quad Q = \frac{4\pi N d_1^2 \omega_1}{c n_1 \hbar\Delta}(\rho_1^0 - \rho_2^0),$$

$$q = \frac{d_1^2 \omega_1 n_2}{d_2^2 \omega_2 n_1}, \quad p = \frac{\mu^2 \omega_3 \Delta n_1}{d_1^2 \omega_1 \Gamma}.$$

Wave mismatching caused by nonzero wave detuning $\delta k = k_1 - k_2 - k_3 \neq 0$ is taken into account in (12) and (7) via replacing the summed phase of the waves at the entrance to the medium by coordinate-dependent phase Φ_1,

$$\Phi \longrightarrow \Phi_1 = \Phi - \delta k z. \qquad (13)$$

It follows from (12) that, at positive Q and $F_{1,2}$, the amplitudes of laser radiation field (V_1) and low-frequency field (G) decrease while the amplitude of scattered radiation (V_2) increases as z grows. The opposite situation arises in a medium with population inversion, in which $\rho_1^0 < \rho_2^0$ and $Q < 0$. The parameters of (12) are on the order of $Q \sim 10$–10^4 cm^{-1}, $q \sim 1.0$–1.2, and $p \sim 10^{-4}$–1 at $\omega_1 \sim 10^4$–10^5 cm^{-1}, $\omega_3 \sim 10^2$–10^3 cm^{-1}, $\Delta \sim 10^2$–10^4 cm^{-1}, $\Gamma = 0.1$ cm^{-1}, $d_1 \approx d_2 \sim 10^{-18}$ CGS units, $\mu/d_1 \sim 10^{-2}$, and $N = 2.5 \times 10^{19}$ cm^{-3}. Clearly, if $p \ll 1$, field G is virtually not absorbed and can be treated as a set parameter influencing the generation of stimulated Raman scattering. The generation is then only determined by the first two equations from (12). Further, we consider both cases, $p \ll 1$ and $p \sim 1$. Under typical experimental conditions, δk is 1–5 cm^{-1} [13].

Let us introduce the dimensionless intensities of pumping (index 1) and Stokes (index 2) waves, $W_{1,2} = V_{1,2}^2$. The first and second equations in (12) give the integral of motion (the Manley–Rowe equation)

$$W_1 + qW_2 \equiv U = W_1(0) + qW_2(0). \qquad (14)$$

As mentioned in the preceding section, (14) holds within the framework of the accepted approximation of moderately high intensities irrespective of whether or not the resonant field is considered set. It follows from (14) that $0 \le W_1 \le U$ and $0 \le W_2 \le U/q$. The last inequality sets limits to the maximum Stokes wave intensity at large distances when energy is redistributed from the pumping to the Stokes wave, $W_2(z \longrightarrow \infty) = U/q$.

Because of the strict equality (14), it is sufficient to trace the behavior of the Stokes wave only. Taking into account (14) and the limitations on the W_1 and W_2 intensities specified above, let us introduce phase variable Θ, which simultaneously determines the behaviors of pumping and scattered waves,

$$W_1 = \frac{1}{2}U(1 + \cos\Theta), \quad W_2 = \frac{1}{2q}U(1 - \cos\Theta). \qquad (15)$$

As a result, (12) takes the final form

$$\frac{d\Theta}{d\zeta} = \frac{1}{2qD}\{U\sin\Theta + 2\sqrt{q}G[\cos\Phi + q_1 U(q_2 + \cos\Theta)\sin\Phi]\},$$
$$\frac{dG}{d\zeta} = -\frac{p}{4\sqrt{q}D}\{2\sqrt{q}G + U\sin\Theta[\cos\Phi - q_1 U(q_2 + \cos\Theta)\sin\Phi]\}, \qquad (16)$$

$$D = 1 + q_1^2U^2(1 + q_2\cos\Theta)^2 + 4G\left(G + \frac{U\sin\Theta\cos\Phi}{\sqrt{q}}\right),$$

$$q_1 = \frac{q+1}{2q}, \quad q_2 = \frac{q-1}{q+1},$$

$$\frac{W_2(\zeta = 0)}{W_1(\zeta = 0)} \equiv \eta, \quad \Theta(\zeta = 0) \equiv \Theta_0 = 2\arctan\sqrt{q\eta}.$$

4. ANALYTIC SOLUTIONS TO (16)

Let us obtain simple solutions to (16) for several particular cases in the approximation of a set resonant field $G(p = 0)$ and for coupled waves ($p \neq 0$).

(A) $p = 0$, $G \ll 1$, $|q - 1| \ll 1$. Let the dimensionless Stokes wave intensity be expanded in powers of the small G and $|q - 1|$ parameters to first-order terms,

$$W_2 = W_0 + (q-1)W_q + GW_G. \tag{17}$$

Solutions to (15) and (16) for W_2 of type (17) are

$$W_0 = \frac{\eta U e^{K\zeta}}{1 + \eta e^{K\zeta}}, \quad K = \frac{U}{1 + U^2},$$

$$W_q = \frac{UW_0\Lambda}{U - 2W_0}\left(1 - \frac{2W_0(U - W_0)}{1 + U^2}\right),$$

$$W_G = \Lambda\sqrt{W_0(U - W_0)} \tag{18}$$

$$\times\left\{\cos\Phi\left[\frac{8W_0(U - W_0)}{1 + U^2} - 1\right] - \sin\Phi\right\},$$

$$\Lambda = 1 - \exp\left(\frac{U - 2W_0}{1 + U^2}\zeta\right).$$

The formula for W_0 in (18) coincides up to the denotations with the expression for Stokes radiation [6], which takes into account pump depletion under unsaturated stimulated Raman scattering conditions. This formula contains only a different amplification factor K value; according to [6], $K = U$ in [6] (in our notation). Equation (18) shows that the direct consequence of the saturation effect is an increase in amplification factor K to its maximum value $K_{max} = 1/2$ as the summed intensity of pumping radiation and Stokes wave U increases to $U_{max} = 1$; further, K decreases as $1/U$. It can be shown using (7) that a decrease in K as U increases is caused by a shift of the Raman transition lines induced by the dynamic Stark effect for optical fields. As a result, stimulated Raman scattering goes out of resonance. This circumstance should be taken into account in selecting the optimal laser radiation power at which the length of its maximum transformation into the first Stokes harmonic is the smallest.

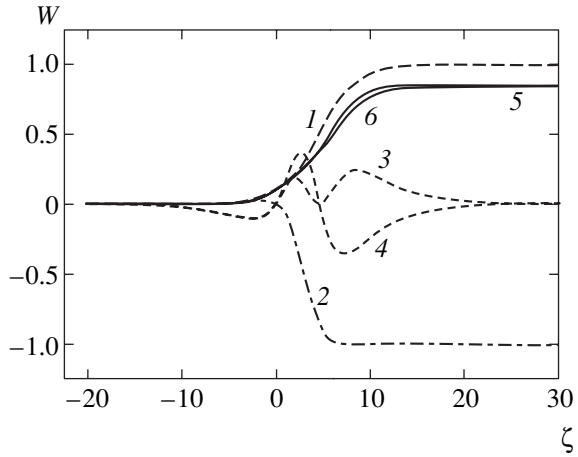


Fig. 1. Dependences of (*5*, *6*) Stokes wave intensity W_2 (17) and its components (*1*) W_0, (*2*) W_q, and (*3*, *4*) W_G on dimensionless length ζ for summed phase values Φ = (*3*, *5*) 0 and (*4*, *6*) $\pi/2$; $U = 1$, $q = 1.15$, $G = 0.1$, and $\eta = 0.1$.

The formula for W_0 in (18) shows that, in the absence of a resonant field and at $q = 1$, the Stokes wave intensity monotonically increases to its limiting value U as dimensionless coordinate ζ grows from zero toward positive values. In a medium with population inversion ($Q < 0$), W_0 monotonically decreases from its initial value at $z = 0$ to zero as z increases (ζ then changes toward negative values). The influence of q deviations from 1 and of field G on W_2 [Eq. (17)] is shown in Fig. 1. The role played by the W_q contribution reduces to a decrease in the limiting energy transfer from the pumping to the Stokes wave to U/q at large z. The W_G component alternates and depends on the summed phase of the waves Φ. These corrections are maximum for media without population inversion.

Note that the representation of the Stokes wave intensity in form (17) and (18) is sufficient for describing many experimental situations because it takes into account stimulated Raman scattering saturation, and the admissible $G < 0.1$–0.2 values correspond to fairly large (~5–10 MW/cm^2) resonant field intensities.

(B) $p = 0$, $\cos\Phi = 0$, $q = 1$. We will use the denotations $\sin\Phi \equiv \sigma = 1$ ($\Phi = \pi/2$), -1 ($\Phi = -\pi/2, 3\pi/2, \ldots$). The first equation from (16) has an exact solution at $G = \text{const}$. Taking into account (15) and the notation introduced above, we obtain

$$W_2 = \frac{U[1 + X + s(X-1)]^2}{2s[X^2 - 1 + s(X^2 + 1)]},$$

$$X = \frac{s + 2\sigma G\sqrt{\eta} - 1}{s - 2\sigma G\sqrt{\eta} + 1}\exp\left[\frac{sU\zeta}{2(s^2 + U^2)}\right], \tag{19}$$

$$s = \sqrt{1 + 4G^2}.$$

The W_2 Stokes signal [Eq. (19)] is a monotonic function of the coordinate, similar mainly to (17). Unlike (17), (19) is, however, valid at large (in terms of the accepted approach) resonant field intensities. The most important differences from the situation with small G values [from the W_0 function given by (18)] are those in the limiting Stokes wave intensities at large lengths and in the effective amplification factors. Namely, the limits of energy transfer into the Stokes harmonic are determined by the equation

$$W_2(\zeta \longrightarrow \pm\infty) = \frac{U}{2}\left(1 \pm \frac{1}{\sqrt{1+4G^2}}\right). \qquad (20)$$

If $G \gg 1$, it follows from (20) that, no matter what the sign of ζ, the limit for energy transfer decreases two times in comparison with the situation when field G is absent. A new qualitative feature is the possibility of resonant field-induced energy transfer from laser to scattered radiation in a medium with population inversion. At small G, such a transfer is directly proportional to the set field intensity, $W_2 = UG^2$.

The effective amplification factor similar to (18), K_{eff}, can be defined for (19) as

$$K_{\text{eff}} = \frac{U\sqrt{1+4G^2}}{1+U^2+4G^2}. \qquad (21)$$

Its maximum value is 1/2, as with low fields G [Eq. (18)]. As opposed to K from (18), K_{eff} (21), however, reaches a maximum at larger U values, $U_{\max} = (1 + 4G^2)^{1/2}$. In other words, the resonant field impedes stimulated Raman scattering saturation.

(C) $\sin\Phi = 0$, $q = 1$. Set $\cos\Phi \equiv \sigma = 1$ ($\Phi = 0, \pm 2\pi, \ldots$), -1 ($\Phi = \pm\pi, \pm 3\pi, \ldots$). At $p = 0$ (constant G), the first equation from (16) takes the form

$$\frac{d\Theta}{d\zeta} = \frac{\sigma}{8G}\frac{U\sin\Theta + 2\sigma G}{U\sin\Theta + (1+U^2+4G^2)/4\sigma G}. \qquad (22)$$

Equation (22) has an analytic solution, which is, however, transcendental with respect to the sought $\Theta(\zeta)$ function. It is, however, easy to see that the numerator and denominator in (22) are cancelled if the equality

$$G = \frac{1}{2}\sqrt{1+U^2} \qquad (23)$$

is satisfied. As a result, the solution to (23) gives a linear dependence of phase Θ on the coordinate,

$$\Theta = \frac{\sigma\zeta}{4\sqrt{1+U^2}} + \Theta_0. \qquad (24)$$

It follows that the intensity of the Stokes wave is an oscillating function of the coordinate in both normal medium and medium with population inversion ($\zeta < 0$),

$$W_2 = \frac{U}{2}\left[1 - \cos\left(\frac{\sigma\zeta}{4\sqrt{1+U^2}} + \Theta\right)\right]. \qquad (25)$$

This nontrivial result (space-periodic energy redistribution between pumping and scattered radiation waves) is caused by nonlinear interference of polarizations in the atom, which is strongest at $|\cos\Phi| = 1$. The effect is present at arbitrary, including low U, intensities. Equation (23) sets the boundary of resonant field G intensities above which oscillations begin. It follows that periodic energy transfer between pumping and Stokes waves is a threshold phenomenon. Numerical calculations, however, show that threshold (23) is smoothly smeared as the intensities of stimulated Raman scattering fields decrease and is virtually absent at $U < 0.05$. The behavior of the threshold resonant field amplitude G_{thr} as a function of U is approximated by the equation

$$G_{\text{thr}} = 10^{-1/(0.48 + 2.36U)} \qquad (26)$$

obtained in numerical calculations.

It follows from (25) that the shortest period of oscillations corresponds to low field intensities $U \ll 1$ and equals 8π. At high intensities $U \gg 1$, the period of oscillations ($8\pi U$) linearly increases as the intensity grows.

If the fields are coupled ($p \neq 0$), Eqs. (16) in the approximation that we use have the integral of motion

$$\Theta + \frac{2\sigma G}{p} = \text{const} \equiv \Theta_1 = \Theta_0 + \frac{2\sigma G_0}{p}. \qquad (27)$$

Here, the Θ angular variable is a linear function of the "controlling" resonant field G. Substituting the $\Theta(G)$ dependence that follows from (27) into the second equation from (16) allows a transcendental solution to this equation to be obtained at small amplitudes $G \ll 1$. It follows from this solution that, because of the interference effect, the resonant field at large distances does not decay completely but obeys the equation

$$G(\zeta \longrightarrow \infty) = -\frac{\sigma p U \sin\Theta_1}{2(p - U\cos\Theta_1)}, \quad |G| \ll 1. \qquad (28)$$

It follows that the influence of stimulated Raman scattering on resonant field absorption under the conditions under consideration reduces to interference-induced blooming of the medium.

(D) $U \ll 1$, $|G|$. At $p = 0$ and low laser and scattered radiation (unsaturated stimulated Raman scattering) intensities, a solution to (16) similar to that given in (C)

also leads to an oscillatory dependence of waves on the distance, because the threshold with respect to field G (23), (26) is then obviously exceeded,

$$\Theta = \frac{G\cos\Phi}{\sqrt{q}(1+4G^2)}\zeta + \Theta_0, \quad (29)$$

$$W_2 = \frac{U}{2q}\left[1 - \cos\left(\frac{G\cos\Phi}{\sqrt{q}(1+4G^2)}\zeta + \Theta_0\right)\right]. \quad (30)$$

Oscillations exist at $\cos\Phi \neq 0$; their frequency is a linear function of amplitude G at small G values and is inversely proportional to the amplitude at $G \gg 1$.

The wave mismatch effect on spatial oscillations is taken into account by (13). In the approximation that we use, (16) at $\delta k \neq 0$ and constant G reduces to

$$\frac{d\Theta}{d\zeta} = \frac{1}{\sqrt{q}}\frac{G}{1+4G^2}\cos(\kappa\zeta - \Phi), \quad \kappa \equiv \frac{\delta k}{Q}. \quad (31)$$

The solution to (31) is

$$\Theta = \Theta_0 + \frac{1}{\kappa\sqrt{q}}\frac{G}{1+4G^2}[\sin\Phi - \sin(\Phi - \kappa\zeta)]. \quad (32)$$

This solution determines the oscillating behavior of the Stokes wave under the combined influence of the initial phase difference Φ and wave detuning δk. As $\delta k \longrightarrow 0$, (32) transforms into (29). It follows from (32) that the ratio between the periods of oscillations of the two specified types, ζ_Φ and ζ_κ, is given by

$$R \equiv \frac{\zeta_\Phi}{\zeta_\kappa} = \frac{\kappa(1+4G^2)}{\sqrt{q}G} \quad (33)$$

at $\kappa\zeta \ll 1$.

According to (33), the R value can change in a wide range, from $R \ll 1$ to $R \gg 1$, and the ratio between the ζ_Φ and ζ_κ periods is determined by particular experimental conditions.

Note that (29), (30), and (32), which give a very structurized oscillatory dependence of the Stokes harmonic intensity on the coordinate, can be used in quantitative studies of the phase relations between the waves introduced into the system. In other words, they can be used to develop a technique implementing the atomic interferometry method suggested in [1].

Next, consider loss-free field G amplification in a medium with population inversion at $p \neq 0$. According to the first equation for the Θ angular variable from (16), the derivative of Θ with respect to the coordinate tends to zero as $1/G$ at fairly high fields G ($G \gg 1$). It follows that phase Θ becomes constant, and an increase in the G^2 resonant field intensity is determined only by the second equation from (16), which, in the case under consideration, coincides with the loss-free transfer equation in an amplifying medium with saturation [14],

$$\frac{dG^2}{d|\zeta|} = \frac{pG^2}{1+4G^2}. \quad (34)$$

At $G \gg 1$, (34) gives the dependence of amplification linear in the coordinate (lethargic),

$$G^2 = \frac{p|\zeta|}{4}, \quad (35)$$

which is caused by the saturation effect for the resonant field at the Raman transition. Note that nonlinear interference effects can in certain situations cause a still slower root dependence of the intensity on length [15].

5. NUMERICAL CALCULATIONS

Equations (16) for coupled waves contain six parameters, namely, ζ, U, G (or p), Φ, q, and η. We will trace the development of generation in space by determining $W_2(\zeta)$ and $G(\zeta)$. For this purpose, (16) will be numerically solved depending on the U, G, and Φ principal parameters. The q and η parameters will be fixed in all calculations, $q = 1.15$ and $\eta = 0.1$.

First, consider Stokes wave generation in the presence of resonant field G constant over the length ($p = 0$, Figs. 2–4). The calculated $W_2(\zeta, G)$ dependences for two summed Φ phases are shown in Fig. 2. In conformity with the analysis performed in the preceding section, an oscillatory dependence of the Stokes harmonic intensity on the distance is observed at zero phase Φ (Fig. 2a). The threshold character of the effect and an increase in the period of oscillations as G increases [see Eqs. (23) and (25)] are clearly seen. Note the symmetrical character of the oscillatory behavior of $W_2(\zeta)$ in normal ($\zeta > 0$) and inverse ($\zeta < 0$) media. The role played by resonant radiation at $\Phi = \pi/2$ (Fig. 2b) largely reduces to providing a possibility of energy transfer from laser to scattered radiation in a medium with population inversion as opposed to an inversion-free medium. The transfer effectiveness increases as G grows [see (20)]. Figure 2c illustrates a decrease in and smearing of the threshold for oscillations as the intensity of stimulated Raman scattering waves decreases.

The $W_2(\zeta, U)$ dependences at fixed G for two Φ phase values $\Phi = 0$ and $\Phi = \pi/2$ are shown in Figs. 3a and 3b. Figures 3a and 3b correspond to analytic cases (C) and (B) considered above, respectively. In Fig. 3a, the period of oscillations at small U is exactly equal to the period calculated by (30) with $G = 0.794$ and $q = 1.15$ selected to construct this figure. Close agreement is also observed between the boundary of oscillations calculated by (26) and the $U \approx 1$ value

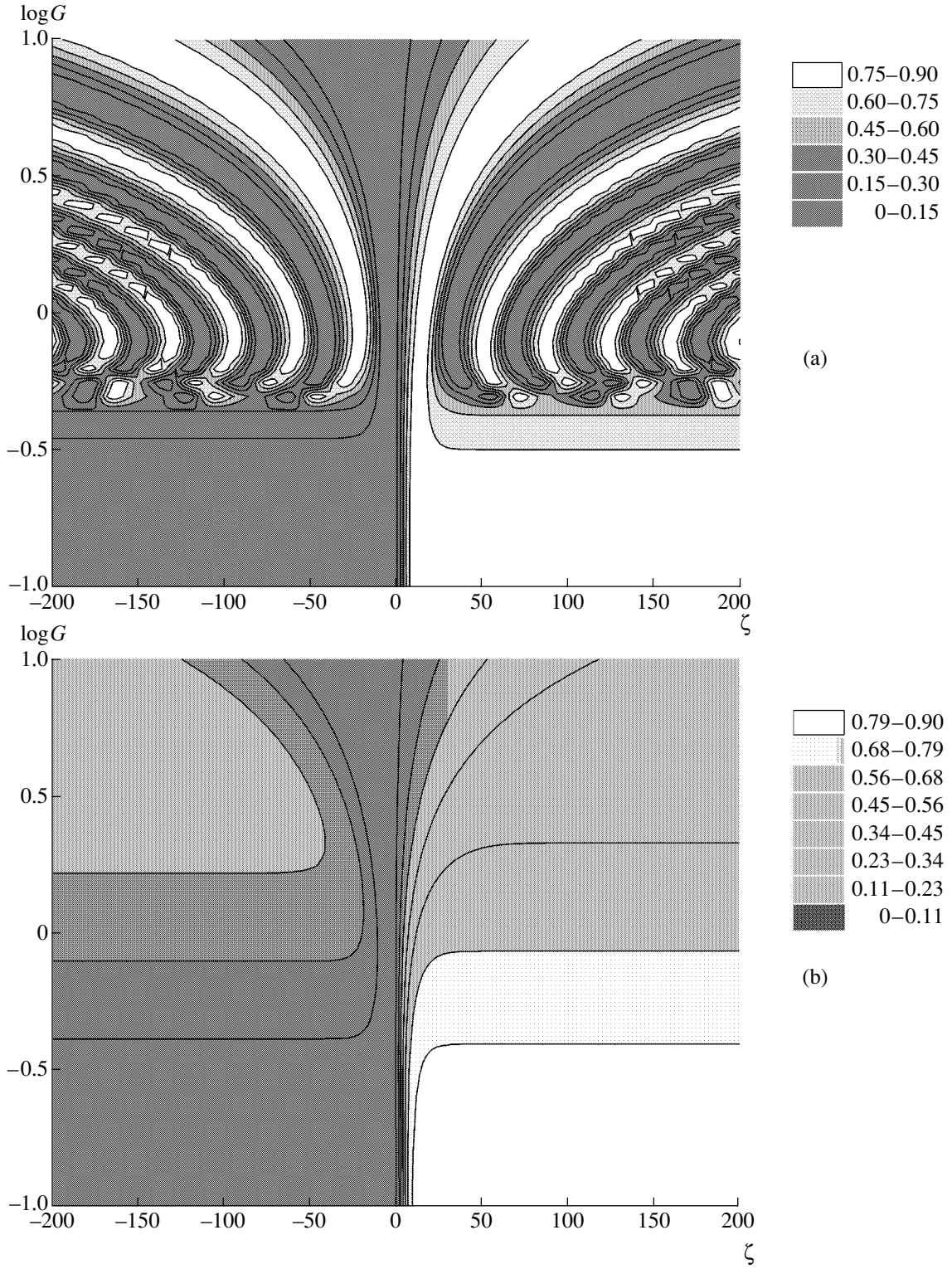


Fig. 2. Dependence $W_2(\zeta, G)$ in the approximation of fixed field G ($p = 0$) for (a) $\Phi = 0$ and $U = 1$, (b) $\Phi = \pi/2$ and $U = 1$, and (c) $\Phi = 0$ and $U = 0.1$; $q = 1.15$ and $\eta = 0.1$.

determined from Fig. 3a. The symmetrical character of the plots shown in Fig. 3b along the U coordinate is determined by the character of the dependence of the effective amplification factor K_{eff} (21) on the summed intensity of optical fields, which is one of the manifestations of stimulated Raman scattering saturation.

The dependence of Stokes harmonic intensity on the summed phase of waves is shown in Fig. 4. A compar-

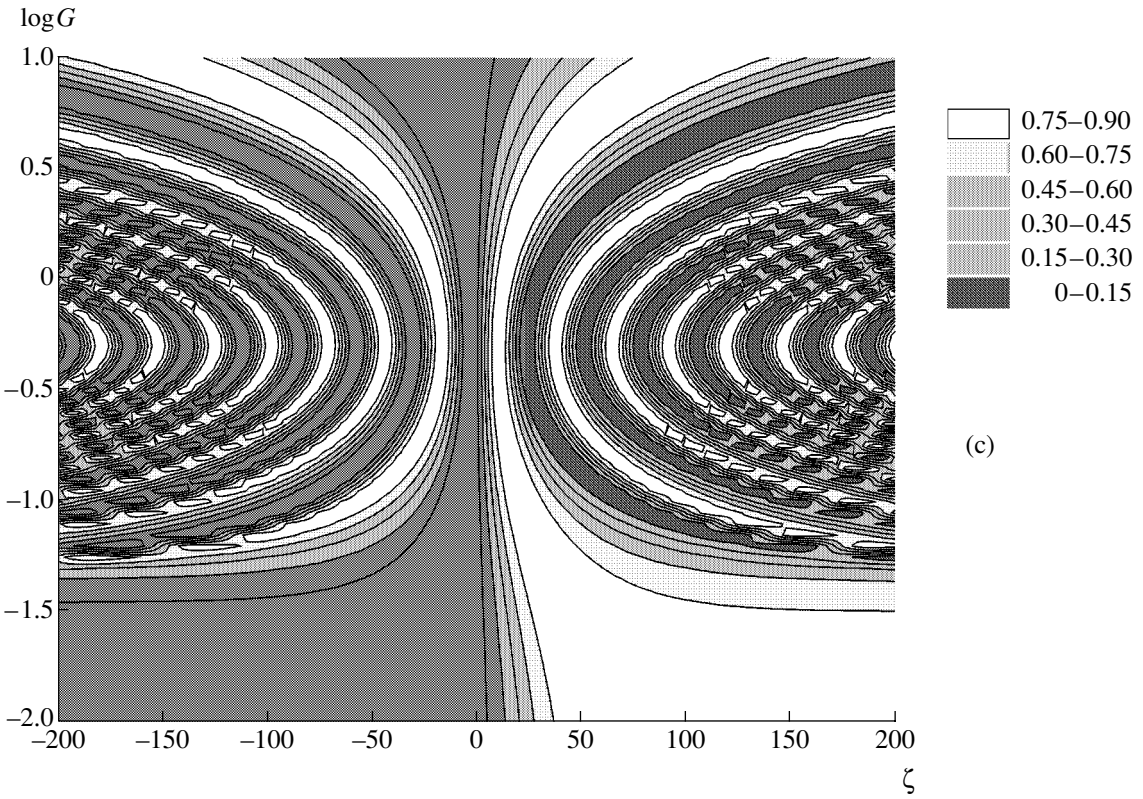


Fig. 2. (Contd.)

ison of the shapes of the dependences corresponding to various U values shows that the oscillatory generation mode is most obvious at low $U = 0.1$ (Fig. 4a) and absent at large $U = 10$ (Fig. 4b). This is related to the threshold character of the arising oscillations and corresponds with the qualitative analysis of this effect performed in Section 4.

Next, consider the mixing of three waves under the conditions of strong coupling of the optical and resonant fields by selecting $p = 1$. We will solve two coupled equations (16) to trace Stokes wave $W_2(\zeta)$ intensity and field $G(\zeta)$ damping (or amplification in a medium with population inversion) depending on the summed wave intensity U and phase Φ at a fixed initial amplitude value $G(0) \equiv G_0 = 10^{-0.1} = 0.794$.

The dependences of the Stokes wave intensity and resonant wave amplitude on summed intensity U are shown in Fig. 5 for two phase values $\Phi = 0$ and $\pi/2$. According to Fig. 5a, the behavior of the Stokes signal at low $U < 1$ sharply changes compared with the case of a constant G amplitude; namely, oscillations completely disappear in the $\zeta > 0$ region, whereas in a medium with population inversion ($\zeta < 0$), their period increases and becomes irregular. This behavior is explained by rapid damping of amplitude G, which decreases below the threshold for the arising of oscillations, in the region of positive ζ values, see Fig. 5c. In contrast, the resonant field grows stronger in the region of negative ζ values, which, in agreement with (30), decreases the frequency of oscillations. At a $\Phi = \pi/2$ phase value, there are much fewer changes in the behavior of W_2 and G compared with the case of a constant G value (cf. Figs. 5c and 3b), because, as follows from (9), interference effects responsible for oscillations are then suppressed. A definitely different behavior of $W_2(\zeta)$ is observed for a medium with population inversion at $U \approx 1$ because of a sharp change in the behavior of $G(\zeta)$ (Fig. 5d) induced by the saturation effect.

The phase picture, which at the same time illustrates the action of nonlinear interference of polarizations, is shown in Fig. 6 for the optimal $U = 1$ value. According to Fig. 6a, the special feature of Stokes harmonic generation at $\zeta > 0$ is the absence of complete energy transfer from laser to scattered radiation at phase $\Phi = \pi$. In a medium with population inversion, the oscillatory mode alternates with monotonic variations in the Stokes radiation intensity under phase changes. Of great interest is the absorption of field G at $\zeta > 0$ (Fig. 6b). As follows from the calculations, field amplitude G at large distances $\zeta > 0$ becomes negative and does not decay in the limit of large lengths at the $\Phi = 0$, π, and 2π phase values, at which interference effects are the strongest. It follows that, in conformity with (28), we observe interference-induced medium blooming at the Raman transition frequency for these phase values.

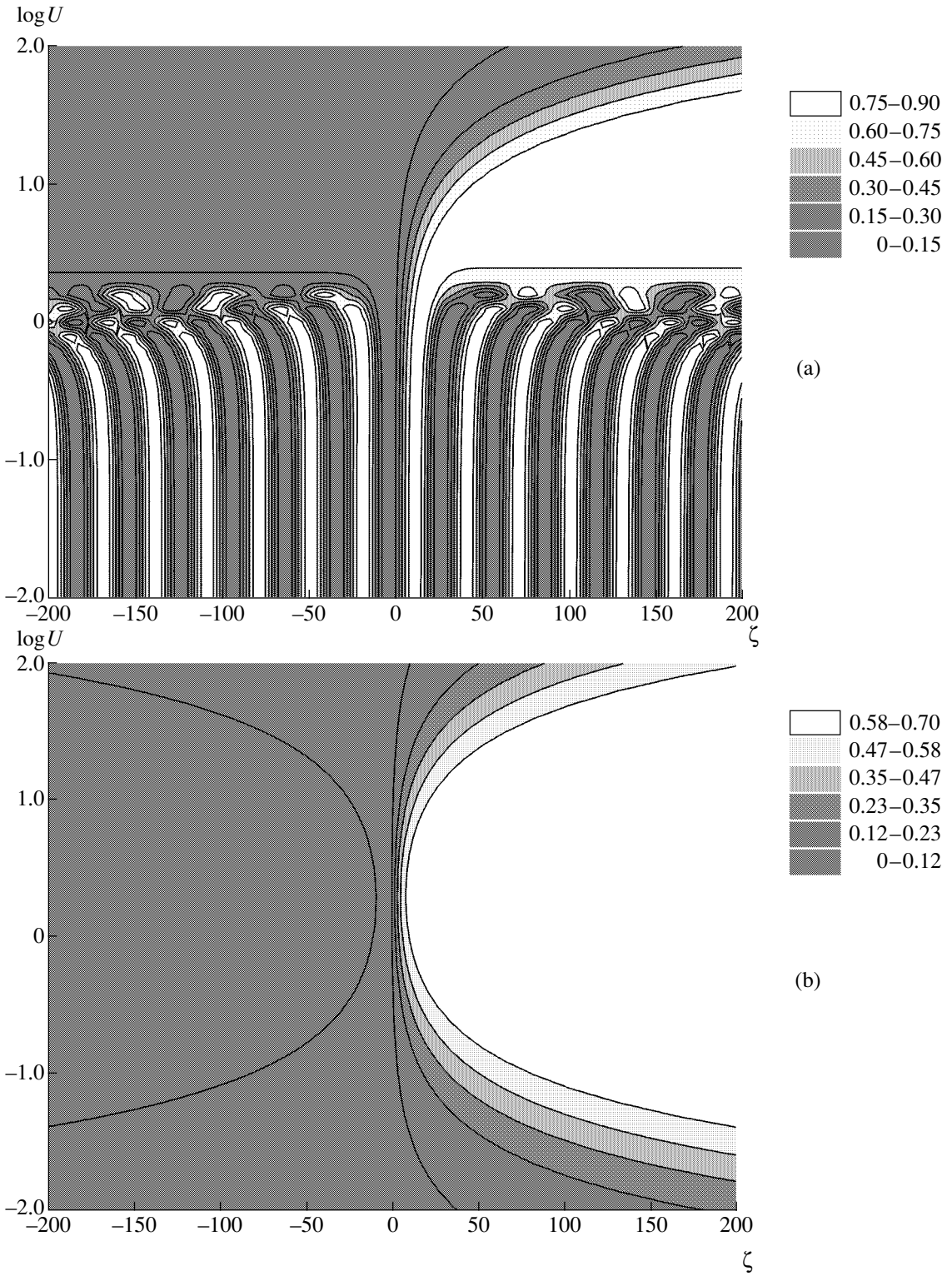


Fig. 3. Function $W_2(\zeta, U)/U$ for Φ = (a) 0 and (b) $\pi/2$; $p = 0$, $G = 10^{-0.1}$, $q = 1.15$, and $\eta = 0.1$.

Numerical calculations show that an increase in resonant field intensity G^2 in a medium with population inversion weakly depends on the phase and is almost linear in the coordinate at low intensities $U < 1$. The coordinate dependence of dimensionless intensity G^2 averaged over phases and $U < 1$ values is approximated by the equation $G^2 = -(0.247 \pm 0.001)\zeta$, which is in close agreement with (31) if the equality $p = 1$ is fulfilled.

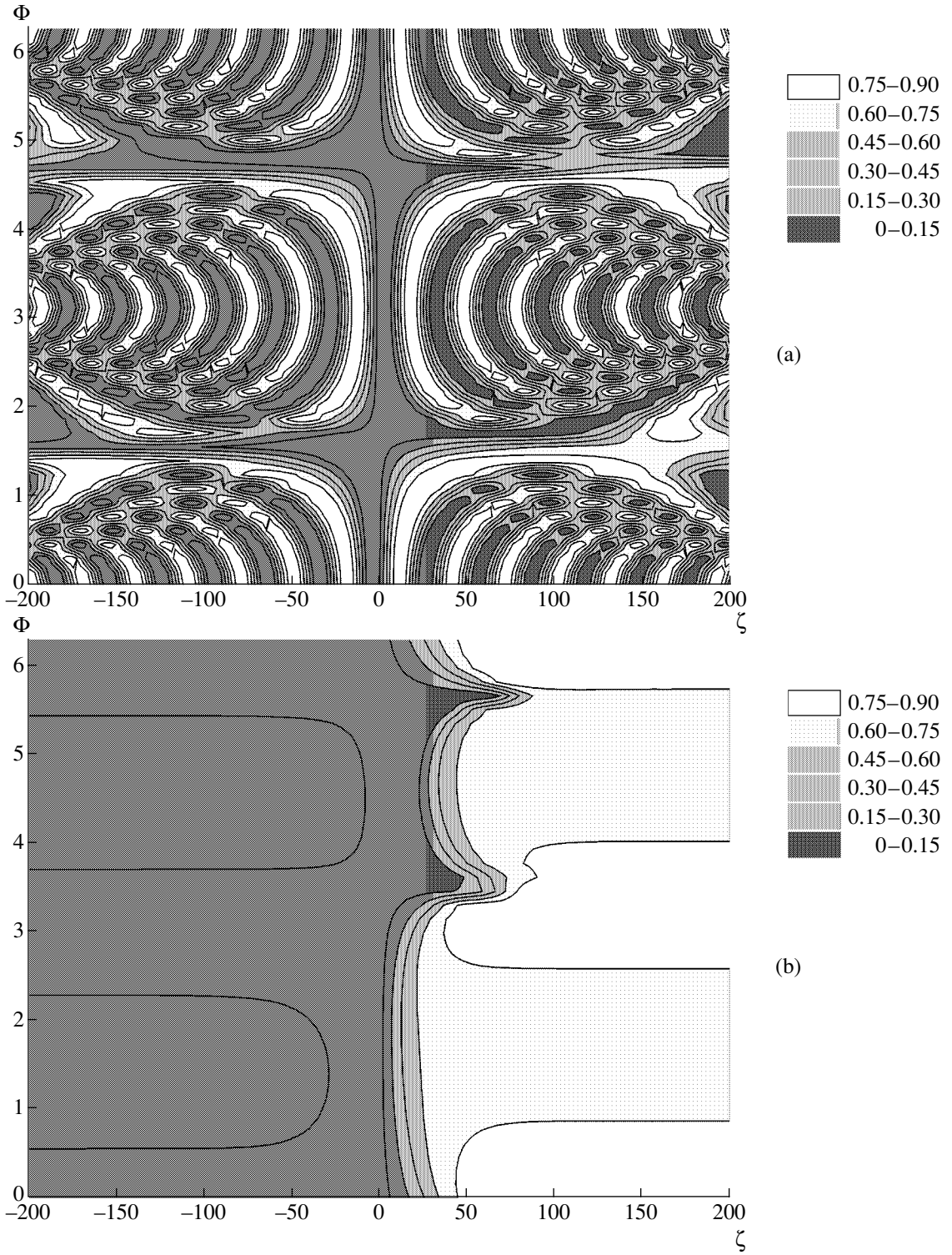


Fig. 4. Dependence $W_2(\zeta, \Phi)/U$ for U = (a) 0.1 and (b) 10; $p = 0$, $G = 10^{-0.1}$, $q = 1.15$, and $\eta = 0.1$.

In conclusion, let us analyze the situation characterized by strong fluctuations of coupled wave phases. We assume that random phase Φ is uniformly distributed over the $[0, 2\pi)$ interval. Calculations of the phase-average dependences of Stokes and resonant wave intensities on the summed intensity of optical waves and on the coordinate show that, for an inversion-free medium, the situation with phase fluctuations that we are considering is close to the $\Phi = \pi/2$ case (Fig. 5c); that is, nonlinear interference effects are suppressed as a result of averaging over phases. However, in a medium with population inversion, no complete inter-

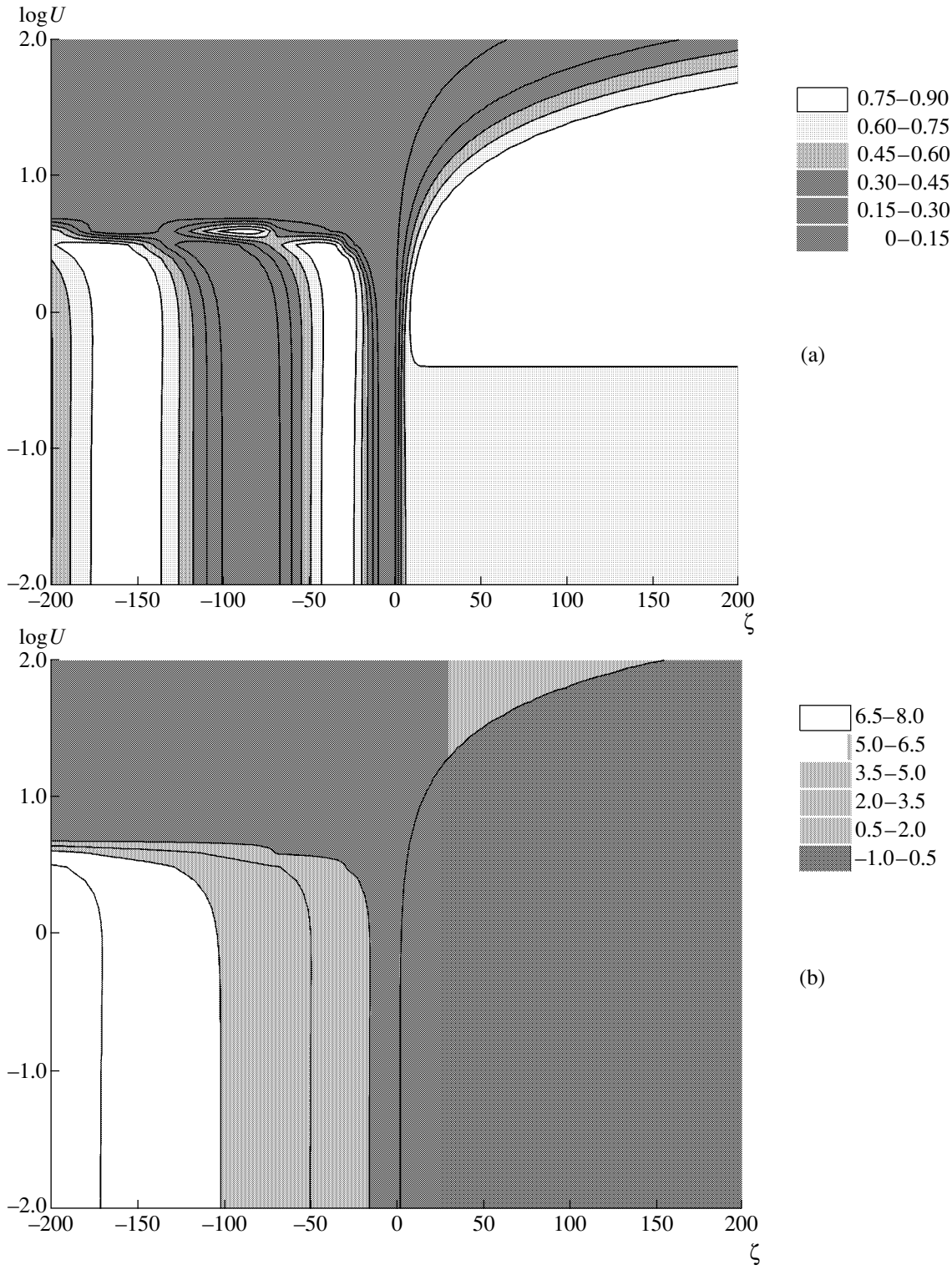


Fig. 5. (a, c) Relative Stokes wave intensity $W_2(\zeta, U)/U$ and (b, d) consistent field absorption (amplification) $G(\zeta, U)$ as functions of total intensity U for phases Φ = (a, b) 0 and (c, d) $\pi/2$; $p = 1$, $G_0 = 10^{-0.1}$, $q = 1.15$, and $\eta = 0.1$.

ference suppression occurs and the behavior of the phase-average Stokes wave is closer to the $\Phi = 0$ case (Fig. 5a). As has been mentioned above, the dependence of the averaged resonant field intensity on the coordinate is linear for a medium with population inversion at $U < 1$ and nonlinearity only comes into effect at high optical field intensities. The effect of interference-induced inversion-free medium bloom-

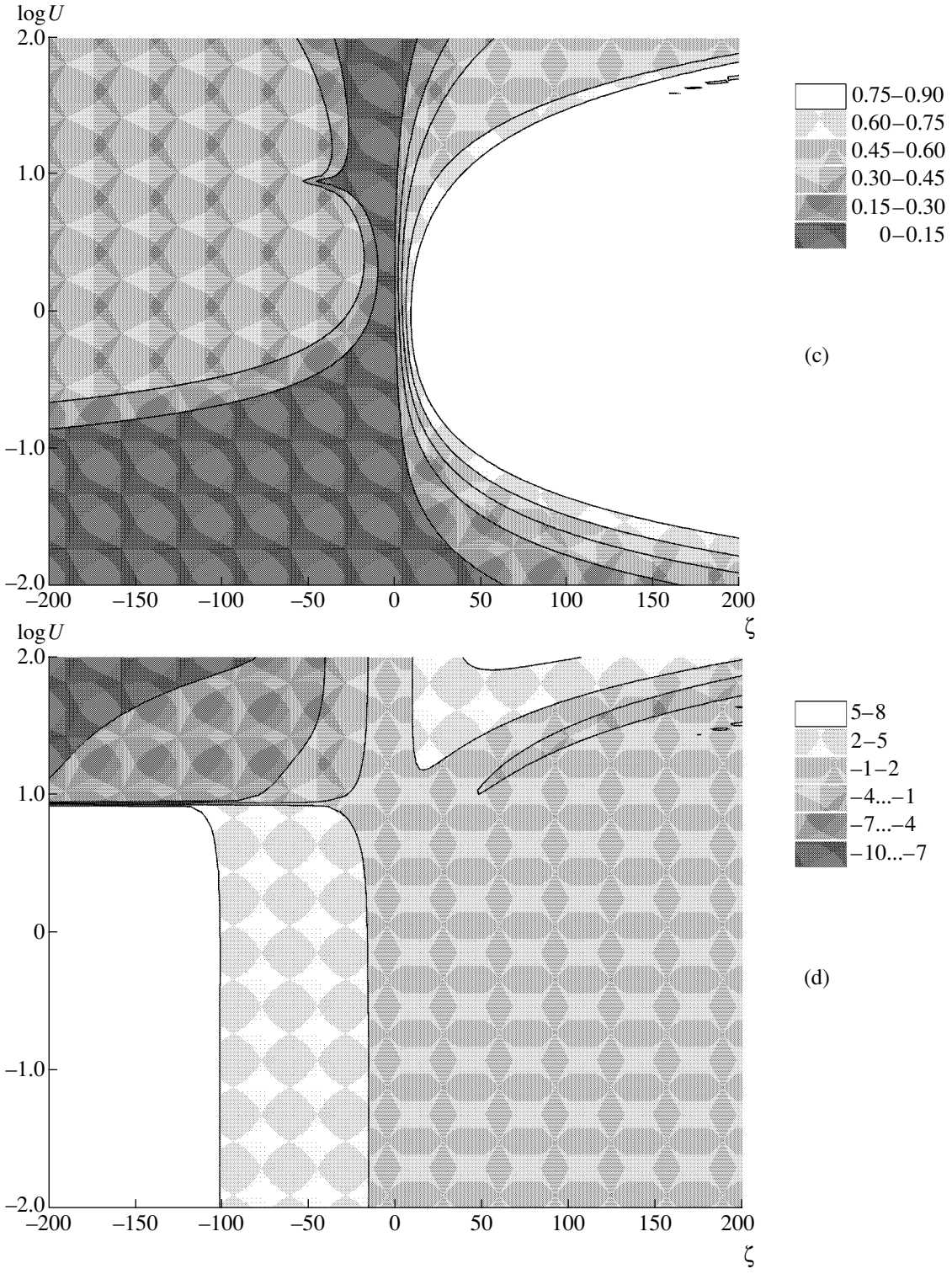


Fig. 5. (Contd.)

ing then disappears because of strong phase fluctuations. The mean-square deviations, which characterize noisiness of the Stokes and resonant waves, are minimum in a narrow ζ region close to zero, where energy transfer is maximum, and amount to about 30% at low $U < 1$ values for the Stokes wave and at large $U > 1$ values for the resonant field.

6. TWO-PHOTON ABSORPTION

The ring interaction of fields considered in Sections 2–5 can be achieved in the stimulated Raman

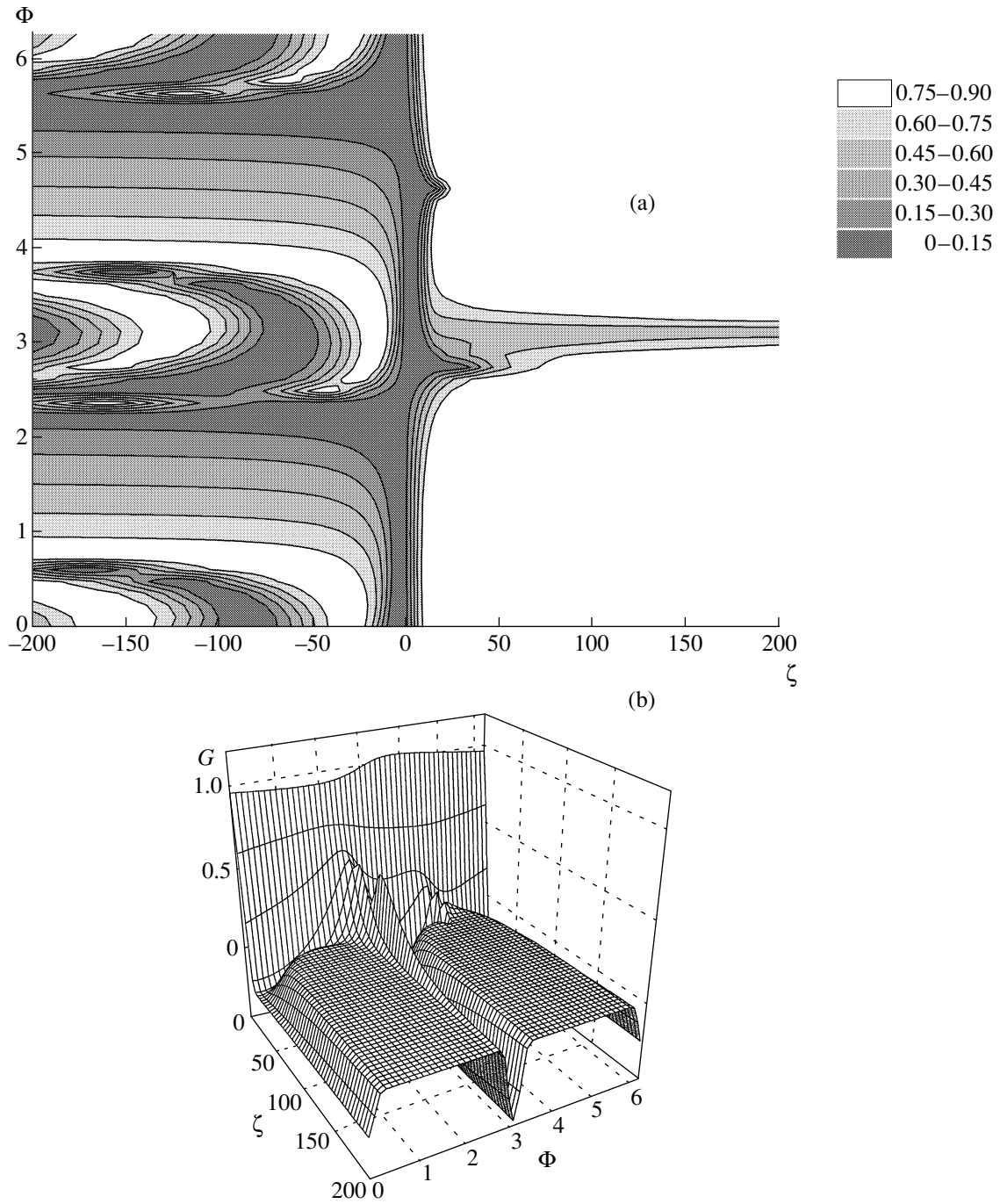


Fig. 6. Influence of the summed phase of waves and related nonlinear interference effects on (a) Stokes signal intensity $W_2(\zeta, \Phi)$ and (b) resonant field amplitude $G(\zeta, \Phi)$ in a medium without population inversion at $p = 1$, $G_0 = 10^{-0.1}$, $U = 1$, $q = 1.15$, and $\eta = 0.1$.

scattering scheme at vibration-rotation transitions in molecules or at excited forbidden transitions in atoms in the visible and infrared spectral regions. In certain instances, in particular, for alkali metal vapors, the ring composition of fields can be more conveniently effected by two-photon absorption at the transition from the ground to the first excited S state. The corresponding generalization of the results will be given based on (2) and field representation in the form

$$\begin{gathered}\mathscr{E}(t, z) = \sum_{j=1}^{3} E_j(z)\cos\Psi_j, \\ \Psi_j = \omega_j t - k_j z + \varphi_j, \quad j = 1, 2, 3, \\ 2\omega_3 = \omega_1 - \omega_2, \quad 2k_3 = k_1 - k_2.\end{gathered} \tag{36}$$

Here, index $j = 3$ refers to the field that acts at the Raman transition; the other notation is as introduced previously.

Instead of (4), we assume that

$$\begin{aligned}\rho_{31} &= R_1 e^{-i\Psi_1} + R_{13} e^{-i\Psi_3},\\ \rho_{32} &= R_2 e^{-i\Psi_2} + R_{23} e^{i\Psi_3},\\ \rho_{21} &= r e^{-2i\Psi_3}\end{aligned} \tag{37}$$

in the transition from (2) to stationary equations.

The substitution of (37) into (2) and the elimination of R_{13} and R_{23} from the system of equations yields stationary equations for the density matrix similar to (7),

$$\begin{aligned}\rho_1 &= \rho_1^0 - \frac{2V_1 R_1''}{\varepsilon} - 2Gr'',\\ \rho_2 &= \rho_2^0 - \frac{2V_2 R_2''}{\varepsilon} + 2Gr'',\\ R_1 &= \sigma\varepsilon[V_1(2\rho_1 + \rho_2 - 1) + V_2 e^{i\Phi} r],\\ R_2 &= \sigma\varepsilon[V_2(\rho_1 + 2\rho_2 - 1) + V_1 e^{-i\Phi} r^*],\\ (1 + i\beta G)r &= ie^{-i\Phi}\frac{V_2 R_1 - V_1 R_2^*}{\varepsilon} + iG(\rho_1 - \rho_2),\end{aligned} \tag{38}$$

where all denotations are retained except

$$\begin{aligned}G &= \frac{d_1 d_2 E_3^2}{4\hbar^2\Gamma\Delta}, \quad V_{1,2} = \frac{d_{1,2}E_{1,2}}{2\hbar\sqrt{\Gamma|\Delta|}},\\ \beta &= \frac{d_1^2 - d_2^2}{d_1 d_2},\\ \Delta_1 &= \omega_{31} - \omega_3, \quad \Phi = \varphi_1 - \varphi_2 - 2\varphi_3.\end{aligned} \tag{39}$$

In addition, we explicitly introduce sign $\sigma = \operatorname{sgn}\Delta$ of effective detuning $\Delta = \omega_{31} - \omega_1$ into (38), because the main perturbing state 3 in alkali metals is the lower p level situated below the first excited s level. As previously, wave detuning in (38) and in wave equations is performed by redefining phase Φ (13), where $\delta k = k_1 - k_2 - 2k_3$.

The differences between (38) and (7) are as follows. The meaning of the G parameter is changed; for two-photon absorption, this parameter is proportional to field intensity E_3 rather than the field amplitude. The interference of polarizations R_1 and R_2 directly caused by the resonant field is absent [in (38), the corresponding equations are only coupled through r]. The $i\beta G$ phase term in the multiplier in the left-hand side of the equation for polarization r at the forbidden transition, which is absent in (7), is present in (38); this term describes the field shift of the line caused by the difference of dipole moments d_1 and d_2.

The nonlinear polarization P [Eq. (3)] components for fields 1–3 expressed through the forbidden transition polarization are

$$\begin{aligned}P_{1,2}^{NL} &= \pm 2\sigma\varepsilon N d_{1,2} V_{1,2}\\ &\times (r'\sin\Phi + r''\cos\Phi)\sin\Psi_{1,2},\\ P_3^{NL} &= \frac{2N d_1 d_2 E_3}{\hbar\Delta_1} r''\sin\Psi_3.\end{aligned} \tag{40}$$

The exact solution to (38) with respect to r has the form

$$\begin{aligned}r' &= \frac{n_0}{D}\{\beta G^2 + \sigma G(V_1^2 - V_2^2)\\ &+ V_1 V_2[(\sigma\beta G + V_1^2 - V_2^2)\cos\Phi - \sigma\sin\Phi]\},\\ r'' &= \frac{n_0}{D}\{G + V_1 V_2\\ &\times[\sigma\cos\Phi - (\sigma\beta G + V_1^2 - V_2^2)\sin\Phi]\},\\ D &= 1 + (V_1^2 + V_2^2)^2 + (4 + \beta^2)G^2\\ &+ 2\sigma G[\beta(V_1^2 - V_2^2) + 4V_1 V_2\cos\Phi],\\ n_0 &= \rho_1^0 - \rho_2^0.\end{aligned} \tag{41}$$

Using (40) and (41) and denotations (12), (13), and (15) introduced in Section 3, we obtain equations similar to (16) that determine the intensities of three coupled waves,

$$\begin{aligned}\frac{d\Theta}{d\zeta} &= \frac{1}{2qD}\{\sigma U\sin\Theta + 2\sqrt{q}G\\ &\times[\cos\Phi + (\beta G + \sigma q_1 U(q_2 + \cos\Theta))\sin\Phi]\},\\ \frac{dG}{d\zeta} &= -\frac{pG}{2\sqrt{q}D}\{2\sqrt{q}G + \sigma U\sin\Theta\\ &\times[\cos\Phi - (\beta G + \sigma q_1 U(q_2 + \cos\Theta))\sin\Phi]\},\\ D &= 1 + q_1^2 U^2(1 + q_2\cos\Theta)^2 + (4 + \beta^2)G^2\\ &+ 2\sigma G U\left[\frac{2\sin\Theta\cos\Phi}{\sqrt{q}} + \beta q_1(q_2 + \cos\Theta)\right],\\ p &= \frac{2d_2\omega_3|\Delta|n_1}{d_1\omega_1\Delta_1 n_3}, \quad Q = \frac{4\pi n_0 d_1^2\omega_1}{\hbar c|\Delta|n_1},\\ G(0) &\equiv G_0, \quad \Theta(0) \equiv \Theta_0.\end{aligned} \tag{42}$$

As $\omega_3 \sim \omega_1$, $|\Delta| \geq |\Delta_1|$, and $d_1 \sim d_2$ for atoms, we have $p \geq 1$, which results in strong coupling of the waves.

Let us analytically consider the case of low Raman scattering field intensities, $U \ll 1$, G, taking into

account that $p \geq 1$. In this approximation, Eqs. (42) reduce to

$$\frac{dG}{d\zeta} = \frac{pG^2}{1 + (4 + \beta^2)G^2}, \tag{43}$$

$$\frac{d\Theta}{d\zeta} = \frac{G}{\sqrt{q}} \frac{\cos\Phi + \beta G \sin\Phi}{1 + (4 + \beta^2)G^2}. \tag{44}$$

Equation (43) is independent of fields participating in stimulated Raman scattering and is the Bouguer–Lambert law for two-photon absorption taking into account its saturation (it seems, performed for the first time). As opposed to one-photon absorption, for which radiation intensity in the saturation mode is given by a transcendental equation, the solution to (43) for two-photon absorption is a simple algebraic function of coordinates,

$$\begin{gathered} G = \frac{2(\sqrt{x^2 + B^2} - x)}{B^2} = \frac{2}{\sqrt{x^2 + B^2} + x}, \\ x \equiv p\zeta + c_1, \quad c_1 = \frac{1}{G_0} - (4 + \beta^2)G_0, \\ B^2 \equiv 4(4 + \beta^2). \end{gathered} \tag{45}$$

For a medium without population inversion ($\zeta > 0$), hyperbolic attenuation of intensity as a function of length follows from (45) at $x^2 \gg B^2$,

$$G = \frac{1}{p\zeta + c_1}. \tag{46}$$

Equation (46) coincides with formula (12.8) from [6] obtained for unsaturated absorption. Under the conditions of strong two-photon absorption saturation, when $G^2 \gg 1/(4 + \beta^2)$, intensity G linearly lowers as distance increases,

$$G = G_0 - \frac{p\zeta}{4 + \beta^2}. \tag{47}$$

Interestingly, two-photon absorption saturation obeys the same attenuation law [Eq. (47)] as that for unsaturated absorption (or absorption at small lengths) [cf. (12.9) from [6]). Formally, this is explained by large c_1 values in (45) at both $G_0 \gg 1$ and $G_0 \ll 1$. The transition from one of these limiting cases to the other is accompanied by sign reversal of c_1, which determines the transformation of (45) from (46) to (47). As with stimulated Raman scattering, the physical mechanism of this feature of two-photon absorption saturation is related to radiation detuning from resonance with the transition as a result of the dynamic Stark shift of levels.

In a medium with amplification ($\zeta < 0$), intensity G grows linearly as length increases, cf. (35),

$$G = \frac{p|\zeta|}{4 + \beta^2}. \tag{48}$$

The solution to (44) with the use of (45) is

$$\begin{gathered} \Theta(x) = \Theta_0 + f(x) - f(c_1), \\ p\sqrt{q}\, f(x) = \frac{2}{B^2}(x + B - \sqrt{x^2 + B^2})\sin\Phi \\ + \ln\left(\frac{x + \sqrt{x^2 + B^2}}{B}\right)\cos\Phi. \end{gathered} \tag{49}$$

Substituting Θ determined by (49) into (15) for wave detuning $\delta k = 0$ gives the $W_2(\zeta, G_0, \Phi, p, q, \beta, \eta)$ function. According to (48), (49), and numerical calculation results, its special features are spatial oscillations in media with population inversion, similar to those shown in Fig. 5a but, as opposed to them, regular and structurized nonmonotonic behavior depending on ζ, Φ, and G_0 at $\zeta > 0$. In particular, at the $\zeta = \zeta_{ext}$ extremum point, where

$$\begin{gathered} p\zeta_{ext} = \frac{4 + \beta^2}{\beta}\cot\Phi - \beta\tan\Phi \\ - \frac{1}{G_0} + (4 + \beta^2)G_0, \end{gathered} \tag{50}$$

the dependence on Φ, β, and G_0 can exhibit a peak or a dip or, at this point, $W_2(\zeta)$ can experience the largest change. Such a transformation provides possibilities for solving the inverse problem of quantitatively determining the parameters present in W_2. For instance, if we know the β and G_0 values, we can use experimental $W_2(\zeta)$ dependences to measure the difference of wave phases Φ and thereby perform atomic interferometry [1]. The values to be determined can also be the relative difference of dipole moments β [Eq. (39)] and dimensionless intensity G_0 at the entrance to the medium, which depends on effective intraatomic parameters d_1, d_2, and Δ_1.

7. DISCUSSION AND CONCLUSIONS

The description of stimulated Raman scattering closed by resonant radiation with the use of a simple three-level atomic model and the complete system of kinetic equations for the density matrix allowed us to trace the action of saturation and nonlinear interference of polarizations induced by three interacting fields. The special feature of interference in the scheme of the ring composition of fields is its dependence on the relation of phases at the entrance to the medium.

The effects considered above can be observed by recording the intensity of the Stokes or pumping wave as a function of Raman interaction length (or of gas

pressure in the absorbing cell) or by measuring the transmission of radiation resonant to the Raman transition. The first variant corresponds to stimulated Raman scattering experiments (the Stokes wave can then be generated in the scattering process) and active Raman scattering measurements, when three collinear waves are present at the entrance to the medium and the intensity of pumping can be low.

The most significant manifestations of interference are length-periodic redistribution of energy between pumping and Stokes radiation waves, that is, spatial oscillations controlled by the difference of wave phases, and medium blooming at the Raman transition frequency. As stimulated Raman scattering is by its nature a process nonlinear in field, the oscillations of optical waves arise when they have low intensities. Accordingly, the optimal conditions for the occurrence of spatial oscillations are weak Raman scattering fields ($U \ll 1$, G) and resonant field intensity comparable with saturating intensity, $G \sim 1$. This corresponds with usual active Raman scattering spectroscopy conditions. Note that oscillations can also be observed at low fields $G \ll 1$ if $G \gg U$. As follows from (30), the G value is limited from below by the oscillation period proportional to $1/G$, which should not exceed values of about 30 cm accessible to observations. Numerical estimates given below show that precisely such conditions exist in molecular gases.

An increase in the pumping field intensity suppresses oscillations and qualitatively changes the situation because of sequential generation of the higher Stokes and anti-Stokes harmonics, which effectively occurs at $U \geq 1$. Equations (7), (16), (38), and (42) should then be modified to take into account the higher scattering harmonics in field representations (1) and (36). In particular, the addition of the second Stokes and first anti-Stokes harmonics complicates phase dependences; namely, in addition to phase Φ (5), several other combinations of initial field phases appear in stationary equations (7). It follows that, under stimulated Raman scattering saturation conditions ($U \geq 1$), the developed theory is only applicable at small Raman interaction lengths $\zeta < U + 1/U$, at which the higher scattering harmonics are still absent. At the same time, there are no limitations on field value G at $U \ll 1$, because this field then does not cause the generation of the higher harmonics, and its role reduces to the initiation of interference phenomena depending on the summed phase of waves Φ.

The feasibility of observing spatial oscillations and medium blooming will be discussed based on the experiment [16] in which infrared radiation at the $\nu = 0 \longrightarrow \nu = 1$ dipole-forbidden vibrational transition in hydrogen was generated under the conditions of biharmonic pumping in a constant electric field. We will also obtain estimates for the experimental conditions of two-photon absorption by potassium vapor reported in [17].

In [16], a CASR (coherent anti-Stokes Raman scattering) spectrometer was used. Pumping at frequency ω_1 ($\lambda_1 = 532$ nm) was effected using the second neodymium laser harmonic with a 12 mJ pulse energy and a 10 ns pulse width. A dye laser generated radiation at frequency ω_2 ($\lambda_2 = 683$ nm) with a 1 mJ pulse energy; this radiation provided Raman resonance $\omega_1 - \omega_2 = \omega_{21}$. At the exit from the cell, into which focused collinear beams with frequencies ω_1 and ω_2 were directed, coherent radiation with frequency ω_3 ($\lambda_3 = 2.4$ μm) was recorded. Voltage creating constant electric field E_c of 0 to 45 kV/cm was applied to capacitor plates in the cell filled with hydrogen up to a 10 atm pressure.

For definiteness, set H_2 pressure equal to 8 atm, $E_c = 40$ kV/cm, and the diameter of focused beams to 2 mm. Dipole moment d_3 induced by the constant field at the ω_{21} transition can be estimated at $d_3 \equiv \mu \approx 10^{-22}$ CGS units based on the results obtained in [18]. The collisional width under these conditions is $\Gamma = 0.014$ cm^{-1} [18]. Infrared radiation power was not reported in [16]; we assume that this power at the entrance to the cell was 25 kW. Radiation was generated in cesium vapor as a result of stimulated Raman scattering [19]. In agreement with the estimates made above, we put $\Delta = 3 \times 10^3$ cm^{-1} and $d_{1,2} = 1$ D.

Using the accepted parameter values and (5) and (12), we obtain the estimates $p = 0.5 \times 10^{-3}$, $q = 1.3$, $Q = 240$ cm^{-1}, $W_1 = 0.05$, $W_2 = 0.004$, and $G = 1.5 \times 10^{-3}$. As $p \ll 1$, field G can be considered constant over length ($G = G_0$). Decreasing W_1 and W_2 to satisfy the condition $U \ll G$, we obtain case (D), Section 4. Equation (28) allows us to estimate of the period of spatial oscillations, $z_0 \approx 2\pi q^{1/2}/GQ = 20$ cm, which is a value convenient for observations.

Next, let us estimate the period of oscillations related to wave detuning δk (13). The δk value at a 8 atm hydrogen pressure is 0.6 cm^{-1} [13], and the period $z_1 = 2\pi/\delta k = 10$ cm. These estimates show that spatial oscillations caused both by interference of waves determined by their initial phases and by wave mismatch can be observed under the specified conditions. Note that the nonmonotonic behavior of the Stokes harmonic intensity as a function of the Raman interaction length and oscillations related to wave detuning have been predicted in [17] based on the theory developed for nonsaturating fields.

It follows from (28) and the W_1, p, and G_0 values accepted above that the relative blooming value $G(z \longrightarrow \infty)/G_0$ equals 10%, which is quite a measurable value. The z_∞ length of the attainment of this value can be estimated at $z_\infty \sim (20–50)/Q \sim 0.1–0.2$ cm; that is, the blooming mode is established virtually as soon as the waves enter the medium.

As mentioned in [16], molecules that can also be studied using this experimental scheme are N_2, CO_2, H_2O, CO, etc.

The experimental situation in potassium vapor considered in [17] presupposes two-photon absorption of radiation of wave length $\lambda_3 = 951$ nm at the 4*s*–5*s* transition and stimulated Raman scattering with $\lambda_1 = 455$ nm, $\lambda_2 = 10.6$ μm, and frequency ω_1 close to resonance with the 4*s*–5*p* transition. Dye lasers, CO_2 and Ti^{3+} : Al_2O_3 lasers, and an $LiNbO_3$ parametric light oscillator [20] can be used as radiation sources. The parameter values reported in [17] are $I_1 = 2.5 \times 10^7$ W/cm^2, $I_3 = 2.5 \times 10^8$ W/cm^2, $I_2 = 10^{-6} I_1$, where I_j ($j = 1, 2, 3$) are radiation intensities at transitions 1–3, and $\delta k = 1.25$ cm^{-1} at a 15 Torr potassium vapor pressure, which corresponds to temperature 772 K [21].

Calculations of effective dipole moments d_1 and d_2 and frequency detunings Δ and Δ_1 are performed on the assumption that the perturbing levels are the *np* potassium levels with $n = 2$–8. Generalizing (2) to the situation with many perturbing states gives

$$d_1 = \sum_j d_{1j}, \quad d_2 = \sum_j d_{2j},$$
$$\Delta = \frac{d_1 d_2}{\sum_j d_{1j} d_{2j}/\Delta_j},$$
$$\Delta_1 = \frac{d_1 d_2}{\sum_j d_{1j} d_{2j}/(\Delta_j + \omega_1 - \omega_3)}, \qquad (51)$$
$$\Delta_j = \omega_{j1} - \omega_1,$$

where index j denotes *np* levels and indices 1 and 2 refer to the 4*s* and 5*s* states, respectively.

The d_{1j} and d_{2j} values in (51) were calculated from the known oscillator strengths [22, 23] and the formulas that determine their relation to dipole moments taking into account the degeneracy of levels [24]. The level positions were found from the Grothrian diagrams [25]. The calculated values were $d_1 = 13$ D, $d_2 = 23$ D, $\Delta = -35510$ cm^{-1}, and $\Delta_1 = 7102$ cm^{-1}.

Set $\Gamma = 2.8 \times 10^9$ s^{-1}, which corresponds to a pressure of 15 Torr and a broadening factor of 30 MHz/Torr. From this value and the values specified above, we obtain estimates for all the significant model parameters, namely, $\beta = -1.3$, $p = 8.8$, $q = 6.8$, $Q = 7.4$ cm^{-1}, $W_1 = 0.4$, and $G = 36$. It follows that strong coupling of waves ($p > 1$) considered in Section 6 and strong two-photon absorption saturation can be achieved in potassium vapor. Extremum coordinate (50) then substantially depends on phase Φ and varies from fractions of a centimeter to dozens of centimeters.

The estimates made above are evidence of the feasibility of experimentally studying the effects predicted in this work in molecular gases and metal vapors. In our opinion, such studies as a variant of active Raman scattering spectroscopy based on the ring composition of fields and their phases offer promise for quantitative determination of the dipole moments of forbidden and allowed transitions and other parameters of atoms and molecules and for studying the phase characteristics of radiation. A comparison of experimental results with theory variants in which a field interacts with the Raman transition by the magnetodipole and electric quadrupole mechanisms would allow conclusions about the relative contributions of these interaction types to be drawn and magnetic dipole and electric quadrupole moments, which are rarely reported in the literature, to be measured. These interactions to a great extent determine rich spectra of lines corresponding to weakly allowed microwave and infrared transitions in molecules; as a rule, these spectra are not taken into account in inverse molecular spectroscopy problems [26].

A promising variant of handling the problem with the ring composition of fields is the use of stimulated Raman scattering at a dipole-allowed transition, whose possibility was experimentally proved in [27] for the example of potassium atoms. Fairly intense quasi-resonant stimulated Raman scattering was generated as a result of magnetodipole and electric quadrupole interactions (predominantly quadrupole interactions). As the Raman transition is allowed, it can easily be saturated by external resonant radiation, which would allow us to thoroughly trace the effects of nonlinear interference of polarizations and saturation. Our main interest is, however, the elucidation of the physical characteristics of the interaction (absorption and scattering) of high-power nonresonant radiation under conditions when electric and magnetic field intensities of one light wave simultaneously participate in these processes. Such processes are described based on the field vector potential. There are, however, contradictions not completely resolved for nonresonant fields interacting with quantum systems modeled by a limited number of states [28–30]. For instance, the cross sections of processes calculated with the Hamiltonian of atom–field interactions represented in terms of the vector potential, on the one hand, and of field amplitudes and the multipole expansion, on the other, are strongly different at substantial detunings from the resonance. This fundamental problem has not been solved completely, and conducting the corresponding experimental studies along with testing different theoretical variants should clarify the situation to a certain extent.

REFERENCES

1. S. J. Buckle, S. M. Barnett, P. L. Knight, *et al.*, Opt. Acta **33**, 1129 (1986).
2. D. V. Kosachev, B. G. Matisov, and Yu. V. Rozhdestvenskiĭ, Zh. Éksp. Teor. Fiz. **102**, 8 (1992) [Sov. Phys. JETP **75**, 4 (1992)].
3. D. V. Kosachev and Yu. V. Rozhdestvenskiĭ, Zh. Éksp. Teor. Fiz. **109**, 1169 (1996) [JETP **82**, 628 (1996)].
4. V. P. Kochanov and Yu. V. Mal'tseva, Opt. Atmos. Okeana **13**, 461 (2000).

5. N. Blombergen, *Nonlinear Optics* (Benjamin, New York, 1965; Nauka, Moscow, 1966).
6. Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley, New York, 1984; Nauka, Moscow, 1989).
7. L. L. Losev and A. P. Lutsenko, Kvantovaya Élektron. (Moscow) **20**, 1054 (1993).
8. D. Eimerl and R. S. Hargrove, Phys. Rev. Lett. **46**, 651 (1981).
9. T. Imasaka, S. Kavasaki, and N. Ishibashi, Appl. Phys. B **49**, 389 (1989).
10. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 3rd ed. (Nauka, Moscow, 1982; Pergamon, New York, 1984).
11. *Physical Quantities. Handbook*, Ed. by I. S. Grigor'ev and E. Z. Meĭlikhov (Énergoatomizdat, Moscow, 1991).
12. J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954; Inostrannaya Literatura, Moscow, 1961).
13. M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon, Oxford, 1969; Nauka, Moscow, 1971).
14. N. V. Karlov, *Lectures on Quantum Electronics* (Nauka, Moscow, 1983).
15. V. P. Kochanov and M. S. Zubova, Zh. Éksp. Teor. Fiz. **105**, 499 (1994) [JETP **78**, 266 (1994)].
16. V. P. Gavrilenko, E. B. Kupriyanova, D. P. Okolokulak, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **56**, 3 (1992) [JETP Lett. **56**, 1 (1992)].
17. G. V. Venkin, G. M. Krochik, L. L. Kulyuk, *et al.*, Zh. Éksp. Teor. Fiz. **70**, 1674 (1976) [Sov. Phys. JETP **43**, 873 (1976)].
18. J. V. Foltz, D. H. Rank, and T. A. Wiggins, J. Mol. Spectrosc. **21**, 203 (1966).
19. D. Cotter, D. C. Hanna, and R. Wyatt, Opt. Commun. **16**, 256 (1976).
20. *Handbook on Lasers*, Ed. by A. M. Prokhorov (Sovetskoe Radio, Moscow, 1978), Vols. 1 and 2.
21. *Tables of Physical Data: Reference Book*, Ed. by I. K. Kikoin (Atomizdat, Moscow, 1976).
22. G. A. Kasabov and V. V. Eliseev, *Spectroscopic Tables for Low-Temperature Plasma: Handbook* (Atomizdat, Moscow, 1973).
23. M. Outred, J. Phys. Chem. Ref. Data **7**, 80 (1978).
24. I. I. Sobel'man, *Atomic Spectra and Radiative Transitions* (Fizmatgiz, Moscow, 1963; Springer, Berlin, 1979).
25. A. S. Yatsenko, *Grotrian Diagrams of Neutral Atoms* (Nauka, Novosibirsk, 1993).
26. V. M. Mikhaĭlov, private communication; T. Oka, in *Molecular Spectroscopy: Modern Research*, Ed. by K. N. Rao (Academic, New York, 1976), Vol. 2.
27. D. Cotter and M. A. Yuratich, Opt. Commun. **29**, 307 (1979).
28. N. B. Delone and V. P. Kraĭnov, *Atom in a Strong Light Field* (Énergoatomizdat, Moscow, 1984).
29. S. G. Rautian, Zh. Éksp. Teor. Fiz. **112**, 836 (1997) [JETP **85**, 451 (1997)].
30. V. P. Bykov, Usp. Fiz. Nauk **143**, 657 (1984) [Sov. Phys.–Usp. **27**, 631 (1984)].

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 222–240.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 255–275.

ATOMS, SPECTRA, RADIATION

Transient Scattering of Resonance Radiation in a Two-Level System

A. A. Panteleev, Vl. K. Roerich*, and A. N. Starostin

Troitsk Institute for Innovation and Nuclear Research, State Scientific Center of the Russian Federation, Troitsk, Moscow oblast, 142190 Russia

**e-mail: vroerich@triniti.ru*

Received June 13, 2002

Abstract—A new method is developed for the description of transfer and scattering of a time-dependent laser pulse, which is based on a hierarchy of Bogoliubov–Born–Green–Kirkwood–Yvone (BBGKY) equations for reduced density matrices. The systems of equations are obtained for the atomic and photon density matrices and the correlation operator in the second-order Born approximation. The system of equations describes stimulated and spontaneous scattering of laser radiation in an optically thin medium, where reabsorption of spontaneous radiation can be neglected. The method proposed in the paper yields for the first time the correct expressions for nonstationary resonance fluorescence spectra, which do not give the negative values for fluorescence intensity. The expressions obtained for the resonance emission spectra coincide in limiting cases with those obtained within the framework of the quantum regression theorem. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The spectra of resonance fluorescence excited by a strong resonance laser wave have been extensively studied for many years. These studies attract interest because of a variety of interference and nonlinear effects, which are important both for a general theory and various applications.

The simplest phenomenon, in which, however, most of these effects can be observed, is scattering of light by a two-level atom. Weisskopf and Wigner [1, 2] were among the first to consider this problem. They showed that the fluorescence spectrum of a two-level system excited by a weak monochromatic field is determined by coherent Rayleigh scattering. In this case, a weak intensity of pump radiation is determined by a small ratio of the Rabi frequency of exciting radiation to the radiative relaxation constant or to the detuning from the resonance. In strong fields, spontaneous radiation in the form of two side components is added to the central coherent component of scattered radiation, resulting in the well-known Mollow triplet [3]. It was shown in paper [3] that the spectrum of light scattered by an atomic system in the case of radiative relaxation has a symmetric shape irrespective of the detuning of the exciting field. Further studies [4, 5] showed that the spectrum could become asymmetric due to collisions, the center of gravity of the triplet being shifted to the transition frequency. Note that the results obtained in papers [3–5] were confirmed experimentally [6–8]. Further studies were developed in many directions and were devoted to the investigation of scattering of biharmonic [9, 10] and polyharmonic [11] fields, the analysis of cooperative effects in a dense medium [12, 13], the study of scattering of light in multilevel media [14], etc.

However, analysis of the transient spectra of resonance fluorescence of a medium excited by a short laser pulse and the study of the spectra taking into account the influence of cooperative fields still involve substantial difficulties. At present, there are two alternative approaches for solving these problems. The first approach, which is most popular [3], is based on the quantum regression theorem (see, for example, review [15]). According to this approach, the spectrum of a random process is determined, under certain assumptions, with the help of the Fourier transform of the autocorrelation function [15]:

$$G(\omega, t) = 2\mathrm{Re}\left[\int_{-\infty}^{\infty} e^{-i\omega t}\langle\sigma^{+}(r)\sigma^{-}(t+\tau)\rangle d\tau\right], \qquad (1)$$

where the angle brackets $\langle\ldots\rangle$ denote the autocorrelation function of two operators; and $\sigma^{+}(t)$ and $\sigma^{-}(t)$ are the creation and annihilation operators, respectively, for an atomic system in the Heisenberg representation. The explicit form of the autocorrelation function is determined with the help of the quantum regression theorem. This theorem was first used for the calculation of spontaneous emission spectra by Lax [16]. It is necessary to discuss in detail the conditions and assumptions at which this approach can be used [15]. They are as follows: (i) the system under study should be Markovian—in other words, further behavior of the system is determined by its current state; (ii) the system is stationary, i.e., invariant with respect to the time shift, so that the transition probability should not depend explic-

itly on time; (iii) the system is linear—in other words, a response of the system subjected to an external force should be linear in this force; (iv) the system should be reversible in time, at least from the macroscopic point of view; (v) a random quantity fluctuates near the equilibrium state; (vi) individual variations in random quantities are sufficiently small for the Fokker–Planck condition to be satisfied; and (vii) fluctuations of a random quantity are caused by the action of a random force (the Langevin approximation).

Further investigations showed that some restrictions imposed on the system properties, at which the approach based on the Fourier transform of the autocorrelation function is valid, can be weakened. However, the essence of these restrictions did not change. Namely, the system should be Markovian and quasi-stationary. In other words, the spectrum of a random process can be determined at times that are much longer than the time of the system relaxation to the equilibrium state, and a change in the action of external forces also should occur during a time that is much longer than the relaxation time. The system should be quasi-linear, when the smallness of a nonlinear component allows one to seek the solution in the form of a series using perturbation theory.

Note that these conditions are not satisfied, as a rule, for most problems of quantum optics. The duration of laser pulses is substantially shorter than the times of radiative relaxation and collision decay. Most modern lasers operate in the range of parameters $N \gg 1$, where $N = n\lambda\!\!\!^{-3}_L$ is the number of carriers in a volume on the order of $\lambda\!\!\!^{-3}_L$, $\lambda_L = 2\pi\lambda\!\!\!^{-}_L$ is the wavelength of laser radiation, and n is the concentration. In this case, cooperative fields play a substantial role, which violates the linearity of the system. When high-power femtosecond lasers are used, nonlinear dynamic relaxation can also occur [17], during which the system is no longer Markovian. The nonlinear relaxation of femtosecond pulses in semiconductors was studied theoretically and experimentally in papers [18–20].

Note that the transient spectra of resonance fluorescence were analyzed with the help of the quantum regression theorem using the expression for the so-called "physical spectrum" [22–27]. It was obtained in paper [22] and depends on a finite width of a detector. Note that the authors of paper [23] understood the drawback of this approach.

Note that the calculation of spontaneous emission spectra with the help of quantum regression theorems also involves purely phenomenological problems. The spectra are analyzed "from top to bottom": first, we postulate the nature of the process based on some phenomenological or other considerations (in our case, the radiative relaxation of atomic populations) and then calculate the spectrum of spontaneously emitted photons from the dynamics of this process. Within the framework of this approach, the relationship between the initial Hamiltonian, describing the interaction of an atom with a quantized radiation field, and phenomenological equations for atomic populations and radiative relaxation constants is not obvious.

An alternative approach to the analysis of resonance fluorescence spectra is the solution of equations for the Scully–Lamb atom–field density matrix [29]. This method was first used by Baklanov [30] for calculating fluorescence spectra in a strong field. The method was further developed by Sargent *et al.* In papers devoted to the multiwave interaction [31–33], we also used this method for studying reabsorption of spontaneous emission [34], nonlinear dynamic relaxation [35], and analysis of quantum interference [36, 37]. This method has a substantial advantage over the quantum regression theorem because it allows one to obtain simultaneously radiative relaxation constants and fluorescence spectra without using any additional assumptions [38]. However, a direct application of the method of the atom–field density matrix for analysis of resonance fluorescence spectra excited by short laser pulses leads to a paradoxical result according to which there exist spectral regions where the fluorescence intensity is negative. Note also that the introduction of a cooperative field requires the use of a phenomenological procedure.

In this paper, we attempt to eliminate the contradictions appearing in the study of transient spectra of resonance fluorescence and to develop a unified method describing transient scattering of resonance radiation in a two-level system. We solved this problem by using a hierarchy of statistical Bogoliubov–Born–Green–Kirkwood–Yvone (BBGKY) equations for reduced density matrices [39, 40] with the help of the apparatus of coherent states [41–43]. Note that a hierarchy of statistical Bogoliubov equations was used in [39] (see also references therein) in the study of Bloch equations for semiconductor lasers and analysis of collisions in the field of a monochromatic wave. The Bogoliubov chain was also used in conjunction with the Glauber–Sudarshan representation for deriving the equation for a field density matrix in the problem of micromaser generation [44].

The paper consists of seven main sections. In Section 2, we describe and substantiate the approximation used in the paper. In Section 3, nonstationary stimulated (Rayleigh) scattering of radiation is analyzed and its relation with cooperative fields is studied. Section 4 is devoted to the derivation of equations describing nonstationary spontaneous emission spectra. In Section 5, we discuss the influence of the measuring process and boundary conditions on the spectral properties. In Section 6, we perform a passage to the limit from the obtained equations, describing transient spectra, to the results obtained with the help of the autocorrelation function. Section 7 is devoted to analysis of the transient spectra of resonance fluorescence excited by a cw field with a phase interruption and to a comparison of

the results with theoretical and experimental studies [26, 27].

2. THE APPROXIMATION USED

Consider the interaction of a resonance electromagnetic (laser) wave with a medium of nondegenerate two-level atoms, which is resonant with the $1 \longrightarrow 2$ transition (Fig. 1). We use in the paper the following assumptions.

The exciting field is nonmonochromatic, and the envelope of a laser pulse is nonstationary in time. We assume also that the rate of the envelope variation is rather low, $\omega_L t_p \gg 1$ (where ω_L is the carrier frequency of the laser pulse and t_p is the characteristic duration of the pulse) and that the rotating-wave and resonance approximations are valid. We used a unified quantum description of the exciting field and scattered radiation. We did not consider polarization characteristics. All the fields are treated as scalar quantities, and polarization variables are omitted in the equations.

We use a collisionless approximation, assuming, however, that the density of atoms in the medium is arbitrary. We will also consider the region of parameters $N \gg 1$, when cooperative fields play a substantial role. The interaction of an electromagnetic wave with an ensemble of two-level atoms will be analyzed using a lumped model. We assume also that N atoms located in a volume on the order of λ_L^3 interact with the electromagnetic field without the phase shift caused by the delay. Within the framework of this approximation, we neglect the longitudinal dipole–dipole interaction because the energy of this interaction in the lumped model is infinite. Note that the contribution of the dipole–dipole interaction results in the collision decay of atomic polarization, which we assume weak, and also in the shift of atomic levels [28], which is insignificant for our study.

The Hamiltonian of the system describing in the dipole approximation the interaction of N atoms located in a volume on the order of λ_L^3 with the laser field has the form (in units of $1/\hbar$)

$$\begin{aligned} H &= \sum^{N} H_a + \sum_{\mathbf{k}} H_f + \sum^{N}\sum_{\mathbf{k}} V_{af}, \\ H_a &= \omega_{21}|2\rangle\langle 2|, \quad H_f = \omega_{\mathbf{k}}\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}, \\ V_{af} &= ig_{\mathbf{k}}(|2\rangle\langle 1|\hat{a}_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}} - |1\rangle\langle 2|\hat{a}_{\mathbf{k}}^{+}e^{-i\mathbf{k}\cdot\mathbf{r}}). \end{aligned} \tag{2}$$

Here, H_a is the Hamiltonian of the unperturbed atomic system, where ω_{21} is the transition frequency, and $|j\rangle$ and $\langle j|$ ($j = 1, 2$) are the projection operators of the corresponding state; H_f is the Hamiltonian of the quantized radiation field, where $\hat{a}_{\mathbf{k}}^{+}$ and $\hat{a}_{\mathbf{k}}$ are the creation and annihilation photon operators, and $\omega_{\mathbf{k}}$ is the frequency of photons with the wave vector $\mathbf{k}$. Hereafter, the symbols "*" and "+" denote complex and Hermitian conjugation, respectively.

The operator V_{af} describes the interaction of an atom with a mode of the quantized field, where

$$g_{\mathbf{k}} = \|\mu_{21}\|\sqrt{\frac{2\pi\omega_{\mathbf{k}}}{\hbar W}}$$

is the coupling constant of the interaction, where $\|\mu_{21}\|$ is the transition dipole moment and W is the quantization volume.

Let us discuss in detail the quantum representation of the laser field that we use in the paper. Within the framework of classical description, the strength of the field of a time-dependent laser pulse has the form

$$E(t, \mathbf{r}) = 2P(t, \mathbf{r})\sin(\omega_L t - \theta(t)), \tag{3}$$

where $P(t, \mathbf{r})$ is the pulse envelope and $\theta(t)$ is the phase of the laser field. By expanding $E(t, \mathbf{r})$ in (3) in the field modes, we obtain

$$\begin{aligned} E(t, \mathbf{r}) &= 2\sum_{\mathbf{k}}|R_{\mathbf{k}}|\sin(\omega_{\mathbf{k}}t - \theta_{\mathbf{k}} - \mathbf{k}\cdot\mathbf{r}), \\ R_{\mathbf{k}} &= \int\!\!\int_{W} P(t, \mathbf{r}) \\ &\times \exp[i((\omega_L - \omega_{\mathbf{k}})t - \theta(t) + \mathbf{k}\cdot\mathbf{r})]d\mathbf{r}dt, \\ \theta_{\mathbf{k}} &= \arccos\left(\frac{R_{\mathbf{k}} + R_{\mathbf{k}}^{*}}{2|R_{\mathbf{k}}|}\right), \end{aligned} \tag{4}$$

where $|R_{\mathbf{k}}|$ and $\theta_{\mathbf{k}}$ determine the amplitude and phase of the field in the $\mathbf{k}$ mode. The quantum analog of such a representation is well known [42]:

$$E(t, \mathbf{r}) = \sum_{\mathbf{k}}\langle\alpha_{\mathbf{k}}|\hat{E}_{\mathbf{k}}|\alpha_{\mathbf{k}}\rangle$$

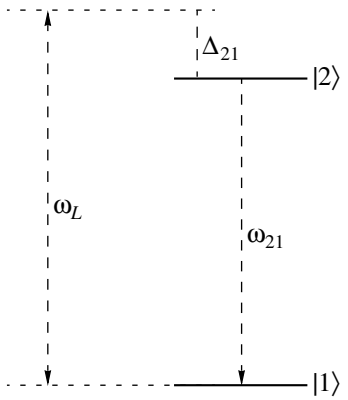


Fig. 1. Energy level diagram of a two-level atom.

$$= 2\sum_{\mathbf{k}}\sqrt{\frac{2\pi\omega_{\mathbf{k}}}{\hbar W}}|\alpha_{\mathbf{k}}|\sin(\omega_{\mathbf{k}}t-\theta_{\mathbf{k}}-\mathbf{k}\cdot\mathbf{r}),$$

$$\hat{E}_{\mathbf{k}} = i\sqrt{\frac{2\pi\omega_{\mathbf{k}}}{\hbar W}} \tag{5}$$

$$\times(\hat{a}_{\mathbf{k}}\exp(-i\omega_{\mathbf{k}}t+i\mathbf{k}\cdot\mathbf{r})-\hat{a}_{\mathbf{k}}^{+}\exp(i\omega_{\mathbf{k}}t-i\mathbf{k}\cdot\mathbf{r})),$$

$$R_{\mathbf{k}} = \sqrt{\frac{2\pi\omega_{\mathbf{k}}}{\hbar W}}\alpha_{\mathbf{k}}.$$

The density matrix ρ_f^L of the photon field of laser radiation has the form

$$\rho_f^L = \prod_k|\alpha_{\mathbf{k}}\rangle\langle\alpha_{\mathbf{k}}|,$$

where $|\alpha_{\mathbf{k}}\rangle$ is the coherent state. Assuming that atoms of the medium and laser radiation do not interact with each other and are not correlated at the initial instant, the equation of motion for the atom–field density matrix can be written in the form

$$i\frac{d\rho}{dt} = [H,\rho],\quad \rho(0) = \prod^{N}\rho_a(0)\prod_k|\alpha_{\mathbf{k}}\rangle\langle\alpha_{\mathbf{k}}|, \tag{6}$$

where ρ_a is the density matrix for an atom.

3. RAYLEIGH SCATTERING AND COOPERATIVE FIELDS

In this section, we consider stimulated Rayleigh scattering and show its relation to cooperative fields. For this purpose, it is convenient for the resonance approximation using the transformation

$$\rho = \exp(-i(\omega_L|2\rangle\langle 2| + (1/2)\Delta_{21}I + \omega_L\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}})t)\rho^{(1)} \times\exp(i(\omega_L|2\rangle\langle 2| + (1/2)\Delta_{21}I + \omega_L\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}})t), \tag{7}$$

where I is the unit operator and $\Delta_{21} = \omega_{21} - \omega_L$ is the detuning of the laser wave from the resonance. In Appendix A, we present the properties of the creation and annihilation operators $\hat{a}_{\mathbf{k}}^{+}$ and $\hat{a}_{\mathbf{k}}$, respectively, and of the operators of a coherent state $D(\alpha_{\mathbf{k}})$ and $D^{+}(\alpha_{\mathbf{k}})$, which are used many times below. Assuming that $\nu_{\mathbf{k}} = \omega_{\mathbf{k}} - \omega_L$ and taking into account (A.5), Eq. (6) can be written in the form

$$i\frac{d\rho^{(1)}}{dt} = [H^{(1)},\rho^{(1)}],$$

$$\rho^{(1)}(0) = \prod^{N}\rho_a(0)\prod_k|\alpha_{\mathbf{k}}\rangle\langle\alpha_{\mathbf{k}}|,$$

$$H^{(1)} = \sum^{N}H_a^r + \sum_{\mathbf{k}}H_f^r + \sum^{N}\sum_{\mathbf{k}}V_{af}, \tag{8}$$

$$H_a^r = \Delta_{21}\sigma_z,\quad H_f^r = \nu_{\mathbf{k}}\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}},$$

$$V_{af} = ig_{\mathbf{k}}(|2\rangle\langle 1|\hat{a}_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}} - |1\rangle\langle 2|\hat{a}_{\mathbf{k}}^{+}e^{-i\mathbf{k}\cdot\mathbf{r}}).$$

At the first stage, we will obtain the operator of interaction of two-level atoms with the laser field. The main idea of this transformation is to exclude the laser field from the initial conditions for the atom–field density matrix. The density matrix, obtained by solving the equation with a changed Hamiltonian, will describe a state of the system, which will contain only the atoms of the medium and the field of scattered radiation. The required transformation has the explicit form

$$\rho^{(1)} = \exp(iH_f^rt)\left(\prod_{\mathbf{k}}D(\alpha_{\mathbf{k}})\right)\exp(-iH_f^rt)\rho^{(2)} \times\exp(iH_f^rt)\left(\prod_{\mathbf{k}}D^{+}(\alpha_{\mathbf{k}})\right)\exp(-iH_f^rt). \tag{9}$$

By using relations (A.5), the Hamiltonian of the system can be written in the form

$$H^{(2)} = \exp(-iH_f^rt)\left[\left(\prod_{\mathbf{k}}D^{+}(\alpha_{\mathbf{k}})\right)\exp(iH_f^rt) \times(H-H_f^r)\exp(-iH_f^rt)\left(\prod_{\mathbf{k}}D(\alpha_{\mathbf{k}})\right) + H_f^r\right]\exp(iH_f^rt)$$

$$= \sum^{N}\Delta_{21}\sigma_z + \sum_{\mathbf{k}}\nu_{\mathbf{k}}\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}} + \sum^{N}\sum_{\mathbf{k}}ig_{\mathbf{k}}(|2\rangle\langle 1|(\hat{a}_{\mathbf{k}}+\alpha_{\mathbf{k}})e^{i\mathbf{k}\cdot\mathbf{r}} - |1\rangle\langle 2|(\hat{a}_{\mathbf{k}}^{+}+\alpha_{\mathbf{k}}^{*})e^{-i\mathbf{k}\cdot\mathbf{r}})$$

$$= \sum^{N}(H_a^r+V_a) + \sum_{\mathbf{k}}H_f^r + \sum^{N}\sum_{\mathbf{k}}V_{af}, \tag{10}$$

where the operator V_a describes the interaction of a two-level atom with the laser field:

$$V_a = i(R(t,\mathbf{r})|2\rangle\langle 1| - R^{*}(t,\mathbf{r})|1\rangle\langle 2|),$$

$$R(t,\mathbf{r}) = \sum_{\mathbf{k}}g_{\mathbf{k}}\alpha_{\mathbf{k}}\exp(-i\nu_{\mathbf{k}}t+i\mathbf{k}\cdot\mathbf{r}). \tag{11}$$

Hamiltonian (10) has the form of a usual Hamiltonian [38, 42] for the interaction of a laser field with a

two-level system when the laser field is described classically. Note that the field of quantized radiation in Eq. (10) in a vacuum state at the initial instant of time. Therefore, the field density matrix obtained by solving Eq. (10) and taking the trace over atomic variables will be the density matrix of scattered radiation. Note, however, that this matrix is not a density matrix if the direction of observation is close to the propagation axis of a laser beam. This matrix can be calculated by performing a transformation inverse to (9).

At the next stage, we will analyze the equation for atom–field density matrix (10) using a hierarchy of BBGKY equations [39] for reduced density matrices. In this section, we will use the Hartree–Fock approximation, which neglects the second- and higher order correlations [39]. By using such an approximation of total Eqs. (10), we neglect the radiative decay of excited states of the atoms and reabsorption of spontaneous emission and analyze only stimulated (Rayleigh) scattering of radiation. Taking into account Hamiltonian $H^{(2)}$, the system of BBGKY equations, consisting of two equations for the atomic and photon density matrices, has the form

$$\begin{aligned} i\frac{d\rho_a}{dt} - [H_a^r + V_a, \rho_a] - [V_{af}, \rho_a\rho_f]^f &= 0, \\ i\frac{d\rho_f}{dt} - [H_f^r, \rho_f] - [V_{af}, \rho_a\rho_f]^a &= 0, \\ \rho_a(0) = \rho_a^0, \quad \rho_f(0) &= |0_{\mathbf{k}}\rangle\langle 0_{\mathbf{k}}|. \end{aligned} \tag{12}$$

Here, the square brackets $[\ldots]^f$ and $[\ldots]^a$ are the traces of the commutators for photon and atomic variables, respectively; ρ_a and $\rho_f = \rho_f(\mathbf{k})$ are the one-particle density matrices for a two-level atom and a mode of the photon field, respectively. The initial conditions are kept invariable during the following transformations, unless otherwise specified.

Note that a hierarchy of BBGKY equations in the approximation used here consists of the Bloch equation for a two-level atom in the field of a laser wave with the term $[V_{af}, \rho_a\rho_f]^f$, which takes into account the influence of scattered radiation on the atom, and the wave equation for the field written in the Schrödinger representation. The component $[V_{af}, \rho_a\rho_f]^a$ of the equation represents a quantum description of the medium polarization induced by the laser field.

At the next step, using two successive transformations, we obtain in explicit form the solution of the equation for the field density matrix. For this purpose, we pass to the wave representation in the equation for the field density matrix:

$$\rho_f = \exp(-iH_f^r t)\rho_f^{(1)}\exp(iH_f^r t).$$

By changing variables, we can write the system of equations (12) in the form

$$\begin{aligned} i\frac{d\rho_a}{dt} - [H_a^r + V_a, \rho_a] - [\tilde{V}_{af}, \rho_a\rho_f^{(1)}]^f &= 0, \\ i\frac{d\rho_f^{(1)}}{dt} - [\tilde{V}_{af}, \rho_a\rho_f^{(1)}]^a &= 0, \\ \tilde{V}_{af} = ig_{\mathbf{k}}[|2\rangle\langle 1|\hat{a}_{\mathbf{k}}\exp(-i\nu_{\mathbf{k}}t + i\mathbf{k}\cdot\mathbf{r}) & \\ - |1\rangle\langle 2|\hat{a}_{\mathbf{k}}^+\exp(i\nu_{\mathbf{k}}t - i\mathbf{k}\cdot\mathbf{r})]. & \end{aligned} \tag{13}$$

Let us write explicitly the operator describing the polarization of the medium:

$$\begin{aligned} [\tilde{V}_{af}, \rho_a\rho_f^{(1)}]^a &= [P_f, \rho_f^{(1)}] \\ = \sum_N ig_{\mathbf{k}}[(\rho_a)_{12}\hat{a}_{\mathbf{k}}\exp(-i\nu_{\mathbf{k}}t + i\mathbf{k}\cdot\mathbf{r}) & \\ - (\rho_a)_{21}\hat{a}_{\mathbf{k}}^+\exp(i\nu_{\mathbf{k}}t - i\mathbf{k}\cdot\mathbf{r}), \rho_f^{(1)}], & \end{aligned} \tag{14}$$

where $(\rho_a)_{12}$ and $(\rho_a)_{21}$ are the nondiagonal elements of the atomic density matrix. Because we use the lumped model, the summation over N should be performed by assuming that the radius vector $\mathbf{r}$ of atoms is equal to a constant for the entire volume on the order of λ_L^3. Note that, neglecting fluctuations in the medium density in the dipole approximation, this assumption is exactly valid. Taking this into account, we can represent the polarization operator of the medium in the form

$$\begin{aligned} P_f = iNg_{\mathbf{k}}((\rho_a)_{12}\hat{a}_{\mathbf{k}}\exp(-i\nu_{\mathbf{k}}t + i\mathbf{k}\cdot\mathbf{r}) & \\ - (\rho_a)_{21}\hat{a}_{\mathbf{k}}^+\exp(i\nu_{\mathbf{k}}t - i\mathbf{k}\cdot\mathbf{r})). & \end{aligned} \tag{15}$$

It follows from expression (15) that the operator exponential $\exp(-iP_f)$ is a coherent state [43] and is described by the expression

$$\begin{aligned} \exp(-iP_f)(t) &= \exp(-i\phi(t))D(\beta_{\mathbf{k}}(t)), \\ \frac{d\beta_{\mathbf{k}}}{dt} &= -Ng_{\mathbf{k}}\rho_{21}\exp(i\nu_{\mathbf{k}}t - i\mathbf{k}\cdot\mathbf{r}) = \zeta_{\mathbf{k}}(t), \\ \frac{d\phi_{\mathbf{k}}}{dt} &= -\frac{i(\zeta_{\mathbf{k}}\beta_{\mathbf{k}}^* - \zeta_{\mathbf{k}}^*\beta_{\mathbf{k}})}{2}. \end{aligned} \tag{16}$$

In this case, the operator relations

$$\begin{aligned} \exp(i\phi(t))D^+(\beta_{\mathbf{k}}(t))\hat{a}_{\mathbf{k}}\exp(-i\phi(t))D(\beta_{\mathbf{k}}(t)) & \\ = (\hat{a}_{\mathbf{k}} + \beta_{\mathbf{k}}(t)), & \\ \exp(i\phi(t))D^+(\beta_{\mathbf{k}}(t))\hat{a}_{\mathbf{k}}^+\exp(-i\phi(t))D(\beta_{\mathbf{k}}(t)) & \\ = (\hat{a}_{\mathbf{k}}^+ + \beta_{\mathbf{k}}^*(t)) & \end{aligned} \tag{17}$$

are satisfied, which are analogous to (A.3).

By using equalities (17) and performing the transformation

$$\rho_f^{(1)} = \exp(-iP_f)(t)\rho_f^{(2)}(\exp(-iP_f)(t))^{-1},$$

we reduce the system (13) to the form

$$\begin{gathered} i\frac{d\rho_a}{dt} - [H_a^r + V_a, \rho_a] - [\tilde{V}_{af} + \tilde{V}_{af}^p, \rho_a\rho_f^{(2)}]^f = 0, \\ i\frac{d\rho_f^{(2)}}{dt} = 0, \\ \tilde{V}_{af}^p = ig_{\mathbf{k}}(|2\rangle\langle 1|\beta_{\mathbf{k}}(t)\exp(-i\nu_{\mathbf{k}}t + i\mathbf{k}\cdot\mathbf{r}) \\ - |1\rangle\langle 2|\beta_{\mathbf{k}}^*(t)\exp(-i\nu_{\mathbf{k}}t - i\mathbf{k}\cdot\mathbf{r})). \end{gathered} \tag{18}$$

By performing the summation over field variables (see Appendix B), we can write the expression $[\tilde{V}_{af}^p, \rho_a\rho_f^{(2)}]^f$ in the form

$$\begin{gathered} [\tilde{V}_{af}^p, \rho_a\rho_f^{(2)}]^f = [V_a^p, \rho_a], \\ V_a^p = -(iN\gamma/2)(|2\rangle\langle 1|(\rho_a)_{21} - |1\rangle\langle 2|(\rho_a)_{12}), \end{gathered} \tag{19}$$

where γ is the radiative relaxation constant (see Appendix B). The operator V_a^p describes the total contribution of radiation scattered by atoms of the medium to the equation of motion for atomic variables. Note that the explicit form of the functions appearing in atomic projection operators in the operator V_a^p of the cooperative field is completely determined by the method of introducing the medium polarization in the model. However, the general structure of the operator V_a^p is preserved for any method of polarization introduction and is determined by the fact that the polarization operator P_f is linear in photon operators. Within the framework of the lumped model that we use here, the physical meaning of the operator of the cooperative field can be interpreted as follows. An atom located in a laser field produces an additional field, which is caused by the induced polarization. In turn, the cooperative field is determined as a direct sum of polarization fields produced by individual atoms. Operator (19) describes the Lorentz field. Note that operator (19) was used in paper [46] in the classical approximation. A similar operator was obtained in paper [47] in the Heisenberg representation by using the adiabatic exclusion of field variables.

Note that operator V_a^p does not describe the Dicke superradiance [48]. The contribution of this process will be substantial only when all the dipole moments of the atoms located in the volume on the order of λ_L^3 are strongly correlated. This imposes the condition on the possible value of the transverse relaxation constant, $\gamma_2 \ll N\gamma$, at which superradiance can be observed [49]. Note that the collision approximation used in our paper is much weaker ($\gamma_2 \ll \gamma$).

Using the expression for the operator of the cooperative field and taking into account that

$$\frac{d\rho_f^{(2)}}{dt} \equiv 0,$$

we have

$$\rho_f^{(2)} \equiv |0_{\mathbf{k}}\rangle\langle 0_{\mathbf{k}}|, \quad [\tilde{V}_{af}^p, \rho_a\rho_f^{(2)}]^f \equiv 0,$$

so that system (18) can be written in the form

$$\begin{gathered} i\frac{d\rho_a}{dt} - [H_a^r + V_a + V_a^p, \rho_a] = 0, \\ \rho_f^{(2)}(t) = |0_{\mathbf{k}}\rangle\langle 0_{\mathbf{k}}|. \end{gathered} \tag{20}$$

By performing transformations that are inverse to the changes of variables, which we used in passing from Eq. (12) to (20), and taking into account the properties of the operators of the coherent state (A.3) and (A.5), we can represent the density matrix of the photon field of the observed radiation in the form

$$\begin{gathered} \rho_f^n = D(\alpha_{\mathbf{k}})\exp(-i\nu_{\mathbf{k}}\hat{a}_{\mathbf{k}}^+\hat{a}_{\mathbf{k}}t)D(\beta_{\mathbf{k}}(t)) \\ \times |0_{\mathbf{k}}\rangle\langle 0_{\mathbf{k}}|D^+(\beta_{\mathbf{k}}(t))\exp(i\nu_{\mathbf{k}}\hat{a}_{\mathbf{k}}^+\hat{a}_{\mathbf{k}}t)D^+(\alpha_{\mathbf{k}}), \\ \rho_f^n = D(\alpha_{\mathbf{k}} + \beta_{\mathbf{k}}(t)e^{-i\nu_{\mathbf{k}}t}) \\ \times |0_{\mathbf{k}}\rangle\langle 0_{\mathbf{k}}|D^+(\alpha_{\mathbf{k}} + \beta_{\mathbf{k}}(t)e^{-i\nu_{\mathbf{k}}t}). \end{gathered} \tag{21}$$

By using the transformation

$$\beta_{\mathbf{k}}(t)e^{-i\nu_{\mathbf{k}}t} = \beta'_{\mathbf{k}}(t),$$

we finally obtain

$$\begin{gathered} \frac{d\beta'_{\mathbf{k}}}{dt} = -i\nu_{\mathbf{k}}\beta'_{\mathbf{k}} - Ng_{\mathbf{k}}\rho_{21}e^{-i\mathbf{k}\cdot\mathbf{r}}, \\ \rho_f^n = |\alpha_{\mathbf{k}} + \beta'_{\mathbf{k}}(t)\rangle\langle\alpha_{\mathbf{k}} + \beta'_{\mathbf{k}}(t)|. \end{gathered} \tag{22}$$

We can readily find from (22) the average value of the field strength [42]

$$\begin{gathered} \hat{E} = ig_{\mathbf{k}}(\hat{a}_{\mathbf{k}}\exp(-i\omega_{\mathbf{k}}t + i\mathbf{k}\cdot\mathbf{r}) \\ - \hat{a}_{\mathbf{k}}^+\exp(i\omega_{\mathbf{k}}t - i\mathbf{k}\cdot\mathbf{r})), \\ E(t, \mathbf{r}) = \sum_{\mathbf{k}}\langle\alpha_{\mathbf{k}} + \beta'_{\mathbf{k}}(t)|\hat{E}|\alpha_{\mathbf{k}} + \beta'_{\mathbf{k}}(t)\rangle \\ = 2\sum_{\mathbf{k}}|R_{\mathbf{k}}^r(t)|\sin(\omega_{\mathbf{k}}t - \theta_k(t) - \mathbf{k}\cdot\mathbf{r}), \\ R_{\mathbf{k}}^r(t) = \sqrt{\frac{2\pi\omega_{\mathbf{k}}}{\hbar W}}(\alpha_{\mathbf{k}} + \beta'_{\mathbf{k}}(t)), \\ \theta_{\mathbf{k}}(t) = \arccos\left(\frac{R_{\mathbf{k}}^r(t) + R_{\mathbf{k}}^{r*}(t)}{2|R_{\mathbf{k}}^r|}\right). \end{gathered} \tag{23}$$

In the last part of this section, we derive the equations for stimulated scattering of radiation in the (t, $\mathbf{r}$) space. Note that, because we use the lumped model and dipole approximation, the field density matrix in Eqs. (20) has the region of definition on the order of $\varkappa_L^3$. Therefore, it is necessary to consider the dependence of the field density matrix on the spatial coordinates $\rho_f(t, \mathbf{r})$. Note that the equation for the field density matrix written in Eq. (18) represents the transfer equation for an electromagnetic field in vacuum, the photon-field density matrix displaced by $\mathbf{n}dr$ having the form

$$\begin{aligned}&\rho_f^n(t + dr/c, \mathbf{r} + \mathbf{n}dr)\\ &= |\alpha_\mathbf{k}^r(t, \mathbf{r}) + \beta_\mathbf{k}'(t, \mathbf{r})\rangle\langle\alpha_\mathbf{k}^r(t, \mathbf{r}) + \beta_\mathbf{k}'(t, \mathbf{r})|,\end{aligned} \tag{24}$$

where $\mathbf{n} = \mathbf{k}/|k|$, c is the speed of light, and $\alpha_\mathbf{k}^r(t, \mathbf{r})$ describes the field that is external with respect to the medium volume under study. By performing successive displacements ($\mathbf{n}dr_0$, $\mathbf{n}dr_1$, …, $\mathbf{n}dr_i$) in the space, we obtain

$$\begin{aligned}&\alpha_\mathbf{k}^r(t + dr/c, \mathbf{r} + \mathbf{n}dr)[i]\\ &= \alpha_\mathbf{k}^r(t, \mathbf{r})[i-1] + \beta_\mathbf{k}'(t, \mathbf{r})[i-1],\end{aligned} \tag{25}$$

where i is the iteration step. Taking into account that a product of coherent states is the coherent state of their sum multiplied by some phase (A.3) and assuming that $\mathbf{n}dr_i \longrightarrow 0$, we finally obtain the field density matrix

$$\begin{aligned}\rho_f^n(t, \mathbf{r}) &= \left|\alpha_\mathbf{k} + \int_{\mathbf{r}_0}^{\mathbf{r}_0 + \mathbf{n}L} \beta_\mathbf{k}'(t - r/c, \mathbf{r})dr\right\rangle\\ &\times\left\langle\alpha_\mathbf{k} + \int_{\mathbf{r}_0}^{\mathbf{r}_0 + \mathbf{n}L} \beta_\mathbf{k}'(t - r/c, \mathbf{r})dr\right|.\end{aligned} \tag{26}$$

Here, $L = |\mathbf{r} - \mathbf{r}_0|$, where $\mathbf{r}_0$ is the coordinate of intersection of a beam emerging from the point $\mathbf{r}$ and directed along $\mathbf{n}$ with the boundary of the excited part of a two-level medium. By using the expression for the field density matrix (26), we can write the system of equations, describing stimulated scattering of radiation in the (t, $\mathbf{r}$) space in the form

$$\begin{aligned}&i\frac{d\rho_a(t, \mathbf{r})}{dt}\\ &- [H_a^r + V_a(t, \mathbf{r}) + V_a^p(t, \mathbf{r}), \rho_a(t, \mathbf{r})] = 0,\\ &\rho_a(0, \mathbf{r}) = \rho_a^0(\mathbf{r}), \quad H_a^r = \Delta_{21}\sigma_z,\\ &V_a(t, \mathbf{r}) = |2\rangle\langle 1|R(t, \mathbf{r}) - |1\rangle\langle 2|R^*(t, \mathbf{r}),\\ &V_a^p(t, \mathbf{r}) = -(iN\gamma/2)(|2\rangle\langle 1|\rho_{21}(t, \mathbf{r}) - |1\rangle\langle 2|\rho_{12}(t, \mathbf{r})).\end{aligned} \tag{27}$$

The field $R(t, \mathbf{r})$ at the point (t, $\mathbf{r}$) has the form

$$\begin{aligned}R(t, \mathbf{r}) &= \sum_\mathbf{k} g_\mathbf{k} \exp\{i(\mathbf{k}\cdot\mathbf{r} - \nu_\mathbf{k}t)\}\\ &\times\left(\alpha_\mathbf{k} + \int_{\mathbf{r}_0}^{\mathbf{r}_0 + \mathbf{n}L} \beta_\mathbf{k}'(t - r/c, \mathbf{r})dr\right),\end{aligned} \tag{28}$$

$$\frac{d\beta_\mathbf{k}'(t, \mathbf{r})}{dt} = -i\nu_\mathbf{k}\beta_\mathbf{k}'(t, \mathbf{r}) - Ng_\mathbf{k}e^{-i\mathbf{k}\cdot\mathbf{r}}\rho_{21}(t, \mathbf{r}).$$

Frequency $\nu_\mathbf{k}$ was substituted into expression (28) because we used the resonance approximation in this equation.

Let us analyze the structure of Eqs. (27) and (28). Expression (28) for the electromagnetic field consists of two separated components. The first component of the field appearing at $\alpha_\mathbf{k}$ describes the transfer and diffraction of the laser field in vacuum. The second component containing $\beta_\mathbf{k}'$ has a more complicated structure and describes the field of stimulated scattered radiation. Note that expression (28) for the field at a given point describes the contribution of scattered radiation from the entire light cone $r - ct < 0$, including the field scattered backward or to the side with respect to the direction of propagation of the laser beam. The Hamiltonian of equation (27) for the density matrix also can be divided into two main components. The first component $[H_a^r + V_a, \rho_a]$ is linear in atomic transitions and corresponds to the classical Bloch equations for a two-level atom in the field of the laser wave. The second component $[V_a^p, \rho_a]$ describes the contribution of scattered (cooperative) radiation to the acting field. Unlike the first component, it is quadratic in atomic variables, and this component is responsible for the appearance of bistability in the system under study.

In concluding this section, we discuss briefly the use of the term "stimulated Rayleigh (elastic) component of scattered radiation" as applied to the process considered in this part of the paper. When the efficiency of a process is proportional to the number of photons, the process is usually called stimulated; on the other hand, the Rayleigh or elastic component corresponds to scattering of radiation without the frequency change. In our case, these concepts are peculiarly mixed, because both processes are described by the same equation (28) in different limiting cases: stimulated scattering corresponds to $R(t) \ll \gamma$ or $R(t) \ll \Delta$, while Rayleigh scattering corresponds to $t \longrightarrow \infty$. In this case, in the saturation regime upon excitation by short laser pulses, the scattered radiation will be neither stimulated nor elastic in the direct meaning of these terms. Its intensity is limited from above by the quantity $|N\gamma/2\rho_{21}|^2$, and it will

contain a multicomponent structure. Below, we will use the term "stimulated emission," which is, in our opinion, more appropriate and points out the fact that this component of scattered radiation is caused by the medium polarization induced by the laser field. Note that scattered radiation of this type will exist after the shutdown of the laser field during the relaxation of the dipole moment of the medium to the equilibrium state.

4. RESONANCE FLUORESCENCE

In this section, we add to Eq. (12) the processes describing the spontaneous decay of the upper level and obtain the expressions describing the resonance fluorescence spectrum of a two-level medium. As in Section 3, we will use a hierarchy of the BBGKY equations to obtain the corresponding equations [39]. We will assume that the intensity of spontaneous emission is weak and will neglect reabsorption of spontaneous emission when analyzing the equations. Such an approach has two substantial drawbacks. The first one is that the total energy will not be conserved in the system of equations obtained [39]. The second disadvantage is that resonance fluorescence is considered as "a post-process" and does not contribute to the constitutive equations for a two-level medium, which may play a substantial role in the analysis of radiation transfer in an optically dense medium. However, these effects can be neglected in optically thin media.

To obtain the expressions for the operator of the radiative decay and spectrum of resonance fluorescence, we will use a hierarchy of the BBGKY equations, third-correlation inclusive and neglecting higher order correlations. In this case, the equation for the second correlation will give the spectrum of spontaneous emission and the radiative decay rate in the equation for the atomic density matrix. The equation for the third correlation will give the radiative decay rate in the equation for the second correlation, which will determine the radiative linewidth.

By using expression (10) for the Hamiltonian obtained in Section 3, we will represent the system of equations in the form

$$i\frac{d\rho_a}{dt} - [H_a^r + V_a, \rho_a] - [V_{af''}, \rho_a\rho_{f''}]^{f''} = [V_{af''}, g_{af''}]^{f''},$$

$$i\frac{d\rho_f}{dt} - [H_f^r, \rho_f] - [V_{a''f}, \rho_{a''}\rho_f]^{a''} = [V_{a''f}, g_{a''f}]^{a''},$$

$$i\frac{dg_{af}}{dt} - [H_a^r + V_a + H_f^r + V_{af}, g_{af}] - [V_{a''f}, \rho_{a''}g_{af}]^{a''} - [V_{af''}, \rho_{f''}g_{af}]^{f''} = [V_{af}, \rho_a\rho_f] + [V_{af''}, \rho_a g_{ff''}]^{f''} + [V_{a''f}, \rho_f g_{aa''}]^{a''} + [V_{af''}, g_{aff''}]^{f''} + [V_{a''f}, g_{aa''f}]^{a''},$$

$$i\frac{dg_{ff'}}{dt} - [H_f^r + H_{f'}^r, g_{ff'}] - [V_{a''f} + V_{a''f}, \rho_{a''}g_{ff'}]^{a''} = [V_{a''f}, \rho_f g_{a''f'}]^{a''} + [V_{a''f}, \rho_{f'}g_{a''f}]^{a''} + [V_{a''f} + V_{a''f'}, g_{a''ff'}]^{a''},$$

$$i\frac{dg_{aa'}}{dt} - [H_a^r + H_{a'}^r + V_a + V_a', g_{aa'}] - [V_{af''} + V_{a'f''}, \rho_{f''}g_{aa'}]^{f''} = [V_{af''}, \rho_a g_{a'f''}]^{f''} + [V_{a'f''}, \rho_{a'}g_{af''}]^{f''} + [V_{af''} + V_{a'f''}, g_{aa'f''}]^{f''}, \quad (29)$$

$$i\frac{dg_{aff'}}{dt} - [H_a^r + V_a + H_f^r + H_{f'}^r + V_{af} + V_{af'}, g_{aff'}] - [V_{a''f} + V_{a''f'}, \rho_{a''}g_{aff'}]^{a''},$$

$$[V_{af''}, \rho_{f''}g_{aff'}]^{f''} = [V_{af} + V_{af'}, \rho_a\rho_f\rho_f'] + [V_{af'}, \rho_f'g_{af}] + [V_{af}, \rho_f g_{af'}] + [V_{af} + V_{af'}, \rho_a g_{ff'}] + O(4),$$

$$i\frac{dg_{aa'f}}{dt} - [H_a^r + V_a + H_{a'}^r + V_{a'} + H_f^r + V_{af} + V_{a'f}, g_{aa'f}] - [V_{af''} + V_{a'f''}, \rho_{f''}g_{aa'f}]^{f''} - [V_{a''f}, \rho_{a''}g_{aa'f}]^{a''} = [V_{af} + V_{a'f}, \rho_a\rho_{a'}\rho_f] + [V_{a'f}, \rho_a'g_{af}] + [V_{af}, \rho_a g_{a'f}] + [V_{af} + V_{a'f}, \rho_f g_{aa'}] + O(4).$$

Here, as in (12), ρ_a and $\rho_f = \rho_f(\mathbf{k})$ are the one-particle density matrices for a two-level atom and a mode of the phonon field, respectively; g_{af}, $g_{ff'}$, and $g_{aa'}$ are the two-particle operators; $g_{aff'}$ and $g_{aa'f}$ are the three-particle "correlation operators" of correlations of the corresponding variables; and $O(4)$ are the fourth-order terms in ρ, which will be neglected. Taking into account that at the zero instant of time the atoms of the medium and the laser field are uncorrelated, the initial conditions can be written in the form

$$\rho_a = \rho_a(0), \quad \rho_f = |0_\mathbf{k}\rangle\langle 0_\mathbf{k}|, \quad g_{af} = 0, \quad g_{ff'} = 0, \quad g_{aa'} = 0, \quad g_{aff'} = 0, \quad g_{aa'f} = 0. \quad (30)$$

Here, "0" denotes the zero operator of the appropriate dimensionality. After the transformations of the system of equations (29), repeating the changes of variables (17) and (18), which we used in the derivation of the operator of the cooperative field, the system of Eqs. (29) will take the form

$$i\frac{d\rho_a}{dt} - [H_a^r + V_a + V_a^p, \rho_a] - [V_{af''}, \rho_a\rho_{f''}]^{f''} = [V_{af''} + V_{af''}^p, g_{af''}]^{f''},$$

$$
\begin{aligned}
&i\frac{d\rho_f}{dt} - [H_f^r, \rho_f] = [V_{a''f} + V_{a''f}^p, g_{a''f}]^{a''},\\
&i\frac{dg_{af}}{dt} - [H_a^r + V_a + V_a^p + H_f^r + V_{af}, g_{af}]\\
&- [V_{af''}, \rho_{f''}g_{af}]^{f''} = [V_{af} + V_{af}^p, \rho_a\rho_f]\\
&+ [V_{af''} + V_{af''}^p, \rho_a g_{ff''}]^{f''} + [V_{a''f} + V_{a''f}^p, \rho_f g_{aa''}]^{a''}\\
&+ [V_{af''} + V_{af''}^p, g_{af}]^{f''} + [V_{a''f} + V_{a''f}^p, g_{aa''f}]^{a''},\\
&i\frac{dg_{ff'}}{dt} - [H_f^r + H_{f'}^r, g_{ff'}] = [V_{a''f} + V_{a''f}^p, \rho_f g_{a''f'}]^{a''}\\
&+ [V_{a''f'} + V_{a''f'}^p, \rho_{f'}g_{a''f}]^{a''}\\
&+ [V_{a''f} + V_{a''f}^p + V_{a''f'} + V_{a''f'}^p, g_{a''ff'}]^{a''},\\
&i\frac{dg_{aa'}}{dt} - [H_a^r + H_{a'}^r + V_a + V_a^p + V_{a'} + V_{a'}^p, g_{aa'}]\\
&- [V_{af''} + V_{a'f''}, \rho_{f''}g_{aa'}]^{f''} = [V_{af''} + V_{af''}^p, \rho_a g_{a'f''}]^{f''}\\
&+ [V_{a'f''} + V_{a'f''}^p, \rho_{a'}g_{af''}]^{f''}\\
&+ [V_{af''} + V_{af''}^p + V_{a'f''} + V_{a'f''}^p, g_{aa'f''}]^{f''},\\
&i\frac{dg_{aff'}}{dt} - [H_a^r + V_a + V_a^p + H_f^r + H_{f'}^r\\
&+ V_{af} + V_{af}^p + V_{af'} + V_{af'}^p, g_{aff'}] - [V_{af''}, \rho_{f''}g_{aff'}]^{f''}\\
&= [V_{af} + V_{af}^p + V_{af'} + V_{af'}^p, \rho_a\rho_f\rho_{f'}]\\
&+ [V_{af'} + V_{af'}^p, \rho_{f'}g_{af}] + [V_{af} + V_{af}^p, \rho_f g_{af'}]\\
&+ [V_{af} + V_{af}^p + V_{af'} + V_{af'}^p, \rho_a g_{ff'}] + O(4),\\
&i\frac{dg_{aa'f}}{dt} - [H_a^r + V_a + V_a^p + H_{a'}^r + V_{a'} + V_{a'}^p + H_f^r\\
&+ V_{af} + V_{a'f}, g_{aa'f}] - [V_{af''} + V_{a'f''}, \rho_{f''}g_{aa'f}]^{f''}\\
&= [V_{af} + V_{af}^p + V_{a'f} + V_{a'f}^p, \rho_a\rho_{a'}\rho_f]\\
&+ [V_{a'f} + V_{a'f}^p, \rho_{a'}g_{af}] + [V_{af} + V_{af}^p, \rho_a g_{a'f}]\\
&+ [V_{af} + V_{af}^p + V_{a'f} + V_{a'f}^p, \rho_f g_{aa'}] + O(4),
\end{aligned}
\tag{31}
$$

$$V_{af}^p = ig_{\mathbf{k}}(|2\rangle\langle 1|\beta_{\mathbf{k}}(t)e^{i\mathbf{k}\cdot\mathbf{r}} - |1\rangle\langle 2|\beta_{\mathbf{k}}^*(t)e^{-i\mathbf{k}\cdot\mathbf{r}}).$$

The initial conditions also have form (30). By solving system (31), we will neglect the terms that are proportional to $g_{\mathbf{k}}$, $Ng_{\mathbf{k}}$, $g_{\mathbf{k}}^2$, and $(Ng_{\mathbf{k}})^2$ and the higher power terms, but will retain the terms of the type $\sum_{\mathbf{k}} g_{\mathbf{k}}^2$, because the former quantities are inversely proportional to the quantization volume W and, when the gas density is not too high, they are much smaller than $\gamma \propto \sum_{\mathbf{k}} g_{\mathbf{k}}^2$. Therefore, the terms in the left-hand sides of the equations containing the operators V_{af} and V_{af}^p can be omitted. Because we are solving system (31) by neglecting reabsorption of spontaneous emission, the contribution $[V_{a''f} + V_{a''f}^p, g_{a''f}]^{a''}$ from the right-hand side of the equation should be neglected in solving the equation for the field density matrix, which gives $\rho_f(t) \equiv |0_{\mathbf{k}}\rangle\langle 0_{\mathbf{k}}|$. As a result, we can omit in the right-hand sides of all the equations for the correlation operators g_{af}, $g_{ff'}$, $g_{aa'}$, $g_{aa'f}$, and $g_{aff'}$ the terms acting on the field density matrix ρ_f, whose contributions to the equations for ρ_a, ρ_f, and g_{af} are either odd with respect to the corresponding creation and annihilation operators $\hat{a}_{\mathbf{k}}$ and $\hat{a}_{\mathbf{k}}^+$ or contain only components that are proportional to $(\hat{a}_{\mathbf{k}})^2$ and $(\hat{a}_{\mathbf{k}}^+)^2$.

Our analysis showed that, under such assumptions, system (31) can be considerably simplified and written in the form

$$
\begin{aligned}
&i\frac{d\rho_a}{dt} - [H_a^r + V_a + V_a^p, \rho_a] = [V_{af''}, g_{af''}]^{f''},\\
&i\frac{d\rho_f}{dt} - [H_f^r, \rho_f] = [V_{a''f}, g_{a''f}]^{a''},\\
&i\frac{dg_{af}}{dt} - [H_a^r + V_a + V_a^p + H_f^r, g_{af}]\\
&= [V_{af}, \rho_a\rho_f] + [V_{af''}, g_{aff''}]^{f''},\\
&i\frac{dg_{aff''}}{dt} - [H_a^r + V_a + V_a^p + H_f^r + H_{f''}, g_{aff''}]\\
&= [V_{af''}, \rho_{f''}g_{af}].
\end{aligned}
\tag{32}
$$

Note that system of Eqs. (32) has the form of a hierarchy of the BBGKY equations obtained within the framework of the second-order generalized Born approximation [39]. By analyzing the kinetic equations of a plasma within the framework of this approximation, the non-Markovian Landau equation can be obtained with self-energy correction term [39].

By formally solving the equations for $g_{aff''}$ and g_{af}, and substituting the solutions into equations for g_{af} and ρ_a, respectively, we can obtain, using the approach similar to the Wigner–Weisskopf procedure, the radiative relaxation operator in a classical form (see Appendix C). In this case, we neglected the effects of nonlinear dynamic relaxation, which are determined by the ratio of the Rabi frequency to the transition frequency $2R/\omega_{21}$ [17, 35]. Note that nonlinear dynamic relaxation can play a substantial role upon optical transitions

when the laser power density is on the order of 10 GW/cm^2. The two-level approximation is not valid under such conditions, and the model considered in the paper cannot be used. By using the results obtained in Appendix B, we can represent the system of equations (32) in the form

$$
\begin{aligned}
&i\frac{d\rho_a}{dt} - [H_a^r + V_a + V_a^p, \rho_a] = \Gamma^R[\rho_a],\\
&i\frac{d\rho_f}{dt} - [H_f^r, \rho_f] = [V_{a''f}, g_{a''f}]^{a''},\\
&i\frac{dg_{af}}{dt} - [H_a^r + V_a + V_a^p + H_f^r, g_{af}]\\
&\quad = [V_{af}, \rho_a\rho_f] + \Gamma^R[g_{af}],\\
&\Gamma^R[X]\\
&= -i\gamma(|2\rangle\langle 2|X + X|2\rangle\langle 2| - 2|1\rangle\langle 2|X|2\rangle\langle 1|),
\end{aligned}
\tag{33}
$$

where $\Gamma^R[\ldots]$ is the radiative relaxation operator [see Appendix C (C.11)]. System of Eqs. (33) is Markovian. This is explained by the fact that we have neglected nonlinear dynamic relaxation, which led to the independence of the relaxation operator of the prehistory of the system development.

Assuming that the intensity of spontaneous emission is small, we can show that system of Eqs. (33) includes the system of equations for the atom–field density matrix as a particular case [30, 38]. Indeed, because we assume that the intensity of spontaneous emission is small and neglect reabsorption of spontaneous emission, the resonance fluorescence spectra will be determined by the element $\rho_f(1_\mathbf{k}, 1_\mathbf{k})$ of the field density matrix (where $1_\mathbf{k}$ is a state with one photon in a mode). In this case, we should set $\rho_f \equiv |0_\mathbf{k}\rangle\langle 0_\mathbf{k}|$ in the right-hand sides of equations for the field density matrix and the correlation operator. Taking this into account, the explicit form of the equation for $\rho_f(1_\mathbf{k}, 1_\mathbf{k})$ can be considerably simplified by excluding the photon creation and annihilation operators from the equation for the correlation operator g_{af} and dividing it into two equations, as shown below. System (33) takes the form

$$
\begin{aligned}
&i\frac{d\rho_a}{dt} - [H_a^r + V_a + V_a^p, \rho_a] - \Gamma^R[\rho_a] = 0,\\
&i\frac{d\rho_f(1_\mathbf{k}, 1_\mathbf{k})}{dt} = -iNg_\mathbf{k}^2F(\nu, \mathbf{k}, t),\\
&i\frac{dg_\mathbf{k}^{(+)}}{dt} - [H_a^r + V_a + V_a^p, g_\mathbf{k}^{(+)}]\\
&+ \nu_k g_\mathbf{k}^{(+)} - \Gamma^R[g_\mathbf{k}^{(+)}] = -i\rho_a|2\rangle\langle 1|,\\
&i\frac{dg_\mathbf{k}^{(-)}}{dt} - [H_a^r + V_a + V_a^p, g_\mathbf{k}^{(-)}]\\
&- \nu_k g_\mathbf{k}^{(-)} - \Gamma^R[g_\mathbf{k}^{(-)}] = -i|1\rangle\langle 2|\rho_a,\\
&\rho_a(0) = \rho_a^0, \quad \rho_f(0) = |0_\mathbf{k}\rangle\langle 0_\mathbf{k}|,\\
&g^{(+)(3)}(0) = 0, \quad g^{(-)(3)}(0) = 0,\\
&g^{(+)} = \langle 1_\mathbf{k}|g_{af}|0_\mathbf{k}\rangle, \quad g^{(-)} = \langle 0_\mathbf{k}|g_{af}|1_\mathbf{k}\rangle,\\
&F(\nu_\mathbf{k}, \mathbf{k}, t) = \langle 2|g_\mathbf{k}^{(+)}|1\rangle + \langle 1|g_\mathbf{k}^{(-)}|2\rangle.
\end{aligned}
\tag{34}
$$

The third and fourth equations of system (34) are almost completely similar to the well-known equation for the atom–field density matrix and the conjugated equation, which were obtained by a different method and were used in papers [30, 35, 38] to analyze the stationary spectra of resonance fluorescence and the function of reabsorption. The main formal difference in the form of these equations is related to the consideration of the contribution $[V_a^p, g_\mathbf{k}^{(+)}]$, $[V_a^p, g_\mathbf{k}^{(-)}]$ from cooperative fields. Note, however, that final results obtained by solving these equations will differ substantially from the results obtained in papers [30, 38]. This is explained by two reasons. First, by the necessity of taking into account the stimulated component of scattered radiation and, second, by the definition of the quantity treated as the fluorescence spectrum. The latter aspect is related to the detection process, which will be discussed in the next section.

Note that the field density matrix, which was obtained by solving (33), generally speaking, is not the density matrix of observed radiation. When passing from system of Eqs. (29) to (31), we excluded in fact the field of stimulated scattered radiation having made the change of variables. However, this field should be taken into account in the analysis of the experimental spectra. Taking into account the approximations used, the equations for the nonzero components $\rho_f(0_\mathbf{k}, 0_\mathbf{k})$, $\rho_f(1_\mathbf{k}, 1_\mathbf{k})$ of the photon matrix will take the form

$$
\begin{aligned}
&\rho_f(0_\mathbf{k}, 0_\mathbf{k}) \equiv 1,\\
&\frac{d\rho_f(1_\mathbf{k}, 1_\mathbf{k})}{dt} = Ng_\mathbf{k}^2F(\nu_\mathbf{k}, \mathbf{k}, t),\\
&\rho_f(1_\mathbf{k}, 1_\mathbf{k})(0) = 0.
\end{aligned}
\tag{35}
$$

By formally integrating the equation for $\rho_f(1_\mathbf{k}, 1_\mathbf{k})$, we obtain

$$
\rho_f(1_\mathbf{k}, 1_\mathbf{k})(t) = \int_0^t Ng_\mathbf{k}^2F(\nu_\mathbf{k}, \mathbf{k}, \tau)d\tau. \tag{36}
$$

By using inverse transformations described in the previous section, we can represent the observed field

density matrix in the form

$$\rho_f^n = D(\alpha_{\mathbf{k}} + \beta'_{\mathbf{k}}(t))\Bigg(|0_{\mathbf{k}}\rangle\langle 0_{\mathbf{k}}| + \int_0^t N g_{\mathbf{k}}^2 F(\nu_{\mathbf{k}}, \mathbf{k}, \tau) d\tau |1_{\mathbf{k}}\rangle\langle 1_{\mathbf{k}}|\Bigg) D^+(\alpha_{\mathbf{k}} + \beta'_{\mathbf{k}}(t)). \quad (37)$$

If we analyze fluorescence in a direction that does not coincide with the laser beam, we have $\alpha_{\mathbf{k}} = 0$. By neglecting the terms on the order of $g_{\mathbf{k}}^4$, expression (37) can be written in the form

$$\rho_f^n = D(\beta'_{\mathbf{k}}(t))(|0_{\mathbf{k}}\rangle\langle 0_{\mathbf{k}}| D^+(\beta'_{\mathbf{k}}(t)) + \prod_{\mathbf{k}} \int_0^t N g_{\mathbf{k}}^2 F(\nu_{\mathbf{k}}, \mathbf{k}, \tau) d\tau |1_{\mathbf{k}}\rangle\langle 1_{\mathbf{k}}|. \quad (38)$$

We assume that the intensity of scattered radiation in a give mode is equal to the energy flux (neglecting the vacuum energy) in this mode incident on a detector. Such a definition is directly related to the photodetection probability [42] and seems to us the most convenient. The energy flux in the chosen mode incident on a detector is proportional to the Poynting vector:

$$I_{\mathbf{k}} = \sum_n \rho_f \hat{I}_{\mathbf{k}} = \sum_n \rho_f^n \hat{E}_{\mathbf{k}} \times \hat{H}_{\mathbf{k}} = \frac{c^2}{W} \overline{n_{\mathbf{k}}} \hbar \mathbf{k} = \frac{c}{L}\Bigg(|\beta'_{\mathbf{k}}(t)|^2 + \int_0^t N g_{\mathbf{k}}^2 F(\nu_{\mathbf{k}}, \tau) d\tau\Bigg) \hbar \omega_{\mathbf{k}}. \quad (39)$$

We omitted in (39) the geometric factors, which are insignificant in this case, and assumed that the quantization volume is L^3.

Note that the intensity of the emission spectrum was written in papers [30, 38] and our previous papers [35–37] in the form

$$I_{\mathbf{k}} \propto \frac{d\rho_f(1_{\mathbf{k}}, 1_{\mathbf{k}})}{dt} = N g_{\mathbf{k}}^2 F(\nu_{\mathbf{k}}, \mathbf{k}, \tau)$$

with the addition of the Rayleigh component of the spectrum in the form of the delta function. This expression coincided in the limiting cases with theoretical and experimental results obtained for stationary spectra [3, 8]. However, the analysis of transient spectra gave negative intensities for some spectral components. This result can be easily explained using expression (39) for the emission spectrum. The values of the function $N g_{\mathbf{k}}^2 F(\nu_{\mathbf{k}}, \mathbf{k}, \tau)$ determine the rate of change in the number of photons in the given mode. These values can be, generally speaking, negative. However, this function is not a probability for detecting (absorbing) a photon of the mode with a detector. We will show in the next section that these two quantities are identical in the stationary case; however, their difference should be taken into account in the case of transient spectra.

Let us discuss briefly the properties of expression (39) for the intensity of scattered radiation. It consists of two terms. They describe stimulated scattering of emission and spontaneous emission, respectively. Note that the term describing stimulated emission is proportional to $N^2 g_{\mathbf{k}}^2$. However, in the case of low densities ($n\lambda\!\!\!^{-}_L^3 \ll 1$), it is proportional to $g_{\mathbf{k}}^2$ (n is the concentration) and is of the same order of magnitude as the spontaneous emission term. This is explained by the fact that the value of the average density cannot be used in the case of low densities ($n\lambda\!\!\!^{-}_L^3 \approx 1$). It is necessary to solve a series of equations with $N = 0, 1, 2, \ldots$ and to sum up the results according to their statistical weights, which are determined by the distribution function. The case $N = 0$ corresponds to vacuum, where neither scattering nor spontaneous emission appears. For $N = 1$, both the terms are of the same order of magnitude, and the results obtained in this case describe the behavior of scattered radiation in a low-density medium. When $n\lambda\!\!\!^{-}_L^3 \gg 1$, the scattered emission dominates over spontaneous emission, and we can consider only the former emission.

Note that expression (39) for the emission intensity cannot be considered as a final result because, as follows from a simple analysis, the intensity of fluorescence excited by a constant field should increase linearly with time. We will discuss this question in the next section.

5. EFFECT OF A DETECTOR AND BOUNDARY CONDITIONS ON THE RESONANCE FLUORESCENCE SPECTRA

An increase in the intensity of resonance fluorescence, which follows from expression (39), is not paradoxical if we take into account that the procedure of quantization of an electromagnetic field, which we used, describes a resonator (the quantization volume) with perfectly reflecting walls [38, 42]. Therefore, on the one hand, the scattered radiation will not leave the resonator volume, resulting in an increase in the number of photons in modes (36). On the other hand, this radiation will be also absent to an observer. In this sense, expression (39) describes only the energy flux in a mode. To describe the emission spectra correctly, it is necessary to define the boundary conditions describing absorption of radiation in the resonator walls, which is related either to detector operation or any other irreversible radiation loss. It is the photon flux absorbed at the boundaries of the quantization volume that can be detected by an observer.

The consideration of losses results in the appearance of an additional term in Hamiltonian (2), which describes the interaction of the quantized field with a thermostat at zero temperature [43]:

$$K[\rho] = -i\sum_{\mathbf{k}}(\kappa_{\mathbf{k}}/2)(\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}\rho - 2\hat{a}_{\mathbf{k}}\rho\hat{a}_{\mathbf{k}}^{+} + \rho\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}), \quad (40)$$

where the constant $\kappa_{\mathbf{k}}$ determines mode losses. Since we neglected reabsorption of spontaneous emission, we can discard the component of the operator $K[\rho]$, which is proportional to $\hat{a}_{\mathbf{k}}\rho\hat{a}_{\mathbf{k}}^{+}$, because it vanishes in this approximation. This leads to the appearance of additional terms in equations for the field density matrix and the operator g_{af}:

$$(-i\kappa_{\mathbf{k}}/2)(\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}\rho_f + \rho_f\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}),$$

$$(-i\kappa_{\mathbf{k}}/2)(\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}g_{af} + g_{af}\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}).$$

By performing the same transformations as in the previous section, we can represent system (34) and expression (39) in the form

$$i\frac{d\rho_a}{dt} - [H_a^r + V_a + V_a^p, \rho_a] - \Gamma^R[\rho_a] = 0,$$

$$i\frac{d\rho_f(1_{\mathbf{k}}, 1_{\mathbf{k}})}{dt} = -i\kappa_{\mathbf{k}}\rho_f(1_{\mathbf{k}}, 1_{\mathbf{k}})$$

$$- iNg_{\mathbf{k}}^2(\langle 2|g_{\mathbf{k}}^{(+)}|1\rangle + \langle 1|g_{\mathbf{k}}^{(-)}|2\rangle),$$

$$i\frac{dg_{\mathbf{k}}^{(+)}}{dt} - [H_a^r + V_a + V_a^p, g_{\mathbf{k}}^{(+)}] \quad (41)$$

$$+ (\nu_{\mathbf{k}} + i\kappa_{\mathbf{k}}/2)g_{\mathbf{k}}^{(+)} - \Gamma^R[g_{\mathbf{k}}^{(+)(3)}] = -i\rho_a|2\rangle\langle 1|,$$

$$i\frac{dg_{\mathbf{k}}^{(-)}}{dt} - [H_a^r + V_a + V_a^p, g_{\mathbf{k}}^{(-)}]$$

$$- (\nu_{\mathbf{k}} - i\kappa_{\mathbf{k}}/2)g_{\mathbf{k}}^{(-)} - \Gamma^R[g_{\mathbf{k}}^{(-)(3)}] = -i|1\rangle\langle 2|\rho_a,$$

$$\rho_a(0) = \rho_a^0, \quad \rho_f(0) = |0_{\mathbf{k}}\rangle\langle 0_{\mathbf{k}}|,$$

$$g^{(+)}(0) = 0, \quad g^{(-)}(0) = 0,$$

$$\frac{d\beta_{\mathbf{k}}'(t)}{dt} = -(i\nu_{\mathbf{k}} + \kappa_{\mathbf{k}})\beta_{\mathbf{k}}'(t) - Ng_{\mathbf{k}}\rho_{21}(t),$$

$$I_{\mathbf{k}} = \kappa_{\mathbf{k}}\left(|\beta_{\mathbf{k}}'(t)|^2\right. \quad (42)$$

$$\left. + e^{-\kappa_{\mathbf{k}}t}\int_0^t Ng_{\mathbf{k}}^2F(\nu_{\mathbf{k}}, \mathbf{k}, \tau)e^{\kappa_{\mathbf{k}}\tau}d\tau\right)\hbar\omega_{\mathbf{k}}.$$

We can readily show that this expression for the emission intensity is limited and, for $t \longrightarrow \infty$, can be written in the form

$$I_{\mathbf{k}} = N(g_{\mathbf{k}})^2\hbar\omega_{\mathbf{k}}$$

$$\times\left(N|(\rho_a)_{21}(\infty)|^2\frac{\kappa_{\mathbf{k}}}{(\nu_{\mathbf{k}}^2 + \kappa_{\mathbf{k}}^2)} + F(\nu_{\mathbf{k}}, \mathbf{k}, \infty)\right). \quad (43)$$

For $\kappa_{\mathbf{k}} \longrightarrow 0$, expression (43) for the emission intensity coincides with the well-known Mollow spectrum obtained using the atom–field density matrix [38].

By using the system of equations (41) and expression (42), we analyzed numerically both total resonance fluorescence spectra and their separate components corresponding to stimulated and spontaneous emission. The calculations were performed using the DUMKA program packet developed by a group of V.I. Lebedev at the Institute of Atomic Energy, which realizes stable explicit Chebyshev schemes with variable time steps for solving stiff systems of homogeneous differential equations [50, 51]. We performed a great series of calculations of the spectra of scattered radiation in a broad region of possible values of parameters. We found that the resonance fluorescence spectra calculated from (41) and (42) do not contain negative radiation intensities. Figure 2 shows the total spectrum of resonance fluorescence and the spectra of stimulated and spontaneous emission obtained upon excitation of a two-level medium by a short laser pulse with the Gaussian profile

$$R(t) = R_0\exp(-t^2/t_p^2).$$

6. CALCULATION OF THE SPECTRUM WITH THE HELP OF THE AUTOCORRELATION FUNCTION

Let us show that the emission spectrum calculated using the formalism developed in previous sections is identical to the spectrum calculated classically with the help of the quantum regression theorem and the autocorrelation function. Consider system (41), where the right-hand side of the equation for $\rho_f^{(1)}(1_{\mathbf{k}}, 1_{\mathbf{k}})$ is written using the trace over atomic variables:

$$i\frac{d\rho_f(1_{\mathbf{k}}, 1_{\mathbf{k}})}{dt} = -i\kappa_{\mathbf{k}}\rho_f(1_{\mathbf{k}}, 1_{\mathbf{k}})$$

$$- iNg_{\mathbf{k}}^2\mathrm{Sp}_a(|1\rangle\langle 2|g_{\mathbf{k}}^{(+)} + g_{\mathbf{k}}^{(-)}|2\rangle\langle 1|). \quad (44)$$

The formal solution for the operator $g_{\mathbf{k}}^{(+)}$ in the integral form is

$$g_{\mathbf{k}}^{(+)} = \int_0^t \exp(-(i\nu_{\mathbf{k}} + \kappa_{\mathbf{k}}/2)(t-\tau))$$

$$\times\hat{M}(t, \tau)\rho_a(\tau)|2\rangle\langle 1|d\tau, \quad (45)$$

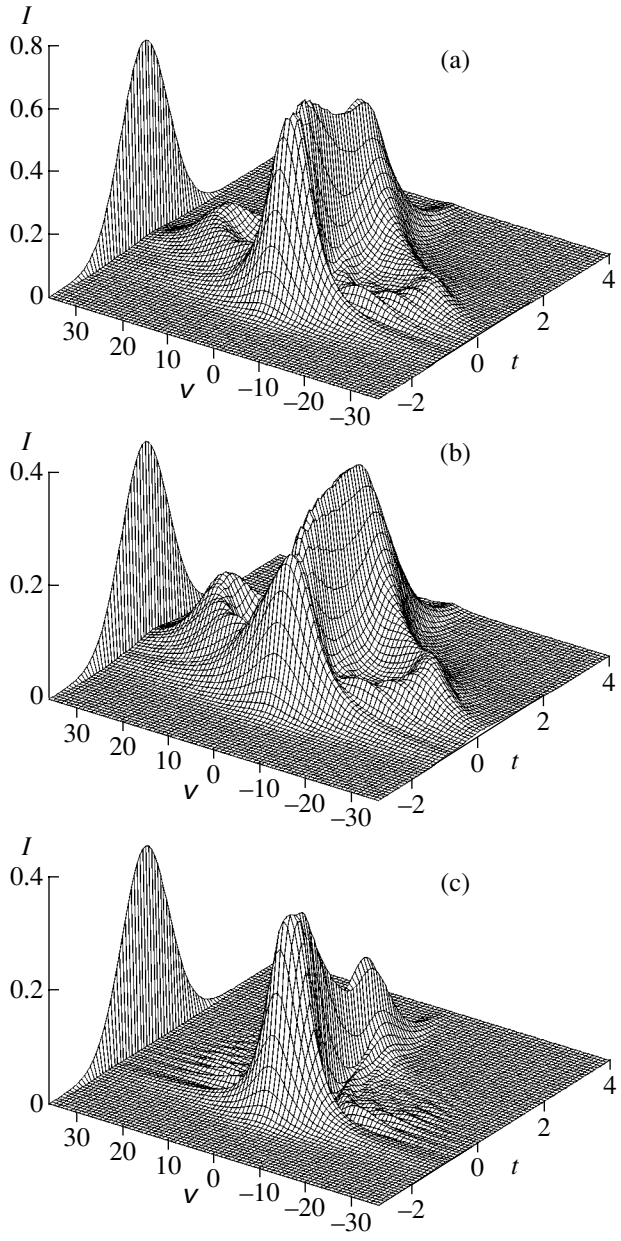


Fig. 2. Transient scattering spectra of a Gaussian laser pulse. The parameters are $t_p = 1$, $R_0 = 12$, $\Delta = -2$, $\gamma = 2$, $\kappa_{\mathbf{k}} = 3$, $N = 3$; (a) total resonance fluorescence spectrum; (b) spontaneous emission spectrum; (c) stimulated emission spectrum. The laser pulse profile is shown at the left.

where the operator $\hat{M}(t, \tau)$ is the fundamental solution of the equation

$$i\frac{dg_{\mathbf{k}}}{dt} - [H_a^r + V_a + V_a^p, g_{\mathbf{k}}] - \Gamma^R[g_{\mathbf{k}}] = 0,$$

$$g_{\mathbf{k}}(0) = g_{\mathbf{k}}^{(0)},$$

$$(g_{\mathbf{k}}(t))_{ji} = \sum_{mn} M_{ijmn}(t, 0)(g_{\mathbf{k}}^{(0)})_{nm}, \tag{46}$$

$$M_{ijmn}(t, \tau) = \sum_{pq} M_{ijqp}(t, 0)M^{-1}_{pqmn}(\tau, 0).$$

By using (44) in the equation for $\rho_f(1_{\mathbf{k}}, 1_{\mathbf{k}})$ and formally integrating it, we obtain

$$\rho_f(1_{\mathbf{k}}, 1_{\mathbf{k}}) = \int_0^t \exp[-\kappa_{\mathbf{k}}(t-\tau_1)] \times \int_0^{\tau_1} \exp[(-i\nu_{\mathbf{k}} - \kappa_{\mathbf{k}}/2)(\tau_1 - \tau_2)] \times \mathrm{Sp}_a[|1\rangle\langle 2|\hat{M}(\tau_1, \tau_2)\rho_a(\tau_2)|2\rangle\langle 1| + \text{c.c.}]d\tau_2 d\tau_1. \tag{47}$$

Note that, if we neglect the effect of a cooperative field, then the operator $\hat{M}(t, 0)$ will also describe the fundamental solution of the equation for the atomic density matrix. Taking this into account, the expression in square brackets represents a two-dimensional correlation function for a product of atomic creation and annihilation operators [41]:

$$\mathrm{Sp}_a[|1\rangle\langle 2|\hat{M}(\tau_1, \tau_2)\rho_a(\tau_2)|2\rangle\langle 1| + \text{c.c.}] = \langle\sigma^+(\tau_2)\sigma^-(\tau_1)\rangle, \tag{48}$$

$$\sigma^+(0) = |2\rangle\langle 1|, \quad \sigma^-(0) = |1\rangle\langle 2|,$$

$$\rho_f(1_{\mathbf{k}}, 1_{\mathbf{k}}) = 2\mathrm{Re}\left[\int_0^t \exp[-\kappa_{\mathbf{k}}(t-\tau_1)] \times \int_0^{\tau_1} \exp[(-i\nu_{\mathbf{k}} - \kappa_{\mathbf{k}}/2)(\tau_1 - \tau_2)]\langle\sigma^+(\tau_2)\sigma^-(\tau_1)\rangle d\tau_2 d\tau_1\right]. \tag{49}$$

For $t \longrightarrow \infty$ and $\kappa_{\mathbf{k}} \longrightarrow 0$ in (49), we obtain a classical expression for the spontaneous emission spectrum [41]:

$$\rho_f(1_{\mathbf{k}}, 1_{\mathbf{k}}) = \lim_{t\to\infty} 2\mathrm{Re}\left[\int_0^\infty e^{-i\nu_{\mathbf{k}}\tau}\langle\sigma^+(t)\sigma^-(t+\tau)\rangle d\tau\right]. \tag{50}$$

Note that expression (49) for the spectrum has virtually the same form as the expression for the spectrum of nonstationary fluorescence obtained in paper [22], with only the upper limit of the inner integral being different.

7. ANALYSIS OF TRANSIENT SPECTRA OF RESONANCE FLUORESCENCE

The first studies of nonstationary resonance fluorescence spectra were performed in the late 1970s. The regime of "switching on" a cw laser field was mainly studied. The early studies [52–54] have shown that the

nonstationary fluorescence spectrum will oscillate at short times due to a change in the population of the upper level at the Rabi frequency. However, the symmetric shape of the Mollow triplet will be conserved. It was shown later theoretically and experimentally [26, 27] that the nonstationary spectrum of spontaneous emission can have a substantially different shape, one of the fluorescence peaks being suppressed when the field is switched on. Such a behavior of an emitting atom is possible if its initial state is prepared in such a way that the atom will be only on one of the levels of a dressed atom. It was shown [27] that such an atomic state can be produced by changing the phase of a laser field at the initial instant of excitation [55]. In this section, we will use the equations describing transient spectra to generalize the results obtained in [26] to the case of an arbitrary detuning from the resonance.

To prepare an atom in one of the states of a dressed atom, we used the following excitation scheme [27] (Fig. 3). During the time interval from 0 to t_{cr} ($t_{cr} \approx 1/2R$, where $2R$ is the Rabi frequency), the atom was in a cw laser field, whose frequency coincided with the transition frequency. At the instant of time t_{cr}, the field phase was changed abruptly by $\pm\pi/2$. The abrupt change in the phase resulted in changes in the wave functions of the dressed states, the projection of the wave function of the atom to one of the "new" states being zero.

Let us use the same scheme for the case of a nonzero detuning and determine the time and magnitude of a change in the field phase. By neglecting spontaneous decay and the influence of the cooperative field, we obtain the evolution of the wave function of the atomic system in our case:

$$\psi(t) = \exp\left[-\frac{i}{\hbar}(H_a^r + V_a^\phi)t\right] \times \exp\left[-\frac{i}{\hbar}(H_a^r + V_a)t_{cr}\right]\psi(0), \quad (51)$$

where $\psi(t)$ is the vector of the wave function of the atomic system. By assuming for definiteness that the initial phase of the laser field is purely imaginary, we obtain the operators of the laser field

$$V_a = R|2\rangle\langle 1| + R|1\rangle\langle 2|,$$

$$V_a^\phi = R\exp(i\phi)|2\rangle\langle 1| + R\exp(-i\phi)|1\rangle\langle 2|.$$

By using the explicit form of the operators H_a^r, V_a, and V_a^ϕ, we can represent the evolution operator for an arbitrary phase of the field in the form

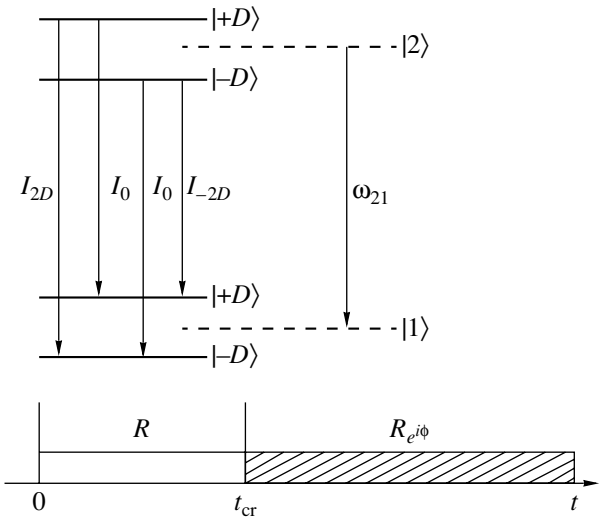


Fig. 3. Energy level diagram of a dressed atom and the scheme of excitation of a two-level atom by a cw laser field with a phase jump used in [27].

$$\exp\left(-\frac{i}{\hbar}(H_a^r + V_a^\phi)t\right) = S_\phi \Lambda S_\phi^{-1},$$

$$S_\phi = \frac{1}{J}\begin{pmatrix} c & -B \\ B^* & c \end{pmatrix},$$

$$S_\phi^{-1} = \frac{1}{J}\begin{pmatrix} c & B \\ -B^* & c \end{pmatrix},$$

$$\Lambda = \begin{pmatrix} \exp(-iDt) & 0 \\ 0 & \exp(iDt) \end{pmatrix},$$

$$A = \Delta/2, \quad B = -iR\exp(-i\phi), \quad (52)$$

$$D = \sqrt{A^2 + |B|^2} = \sqrt{\Delta^2/4 + R^2},$$

$$C = A + D = \Delta/2 + \sqrt{\Delta^2/4 + R^2},$$

$$J^2 = C^2 + |B|^2.$$

The explicit form of the evolution operator is

$$\exp\left(-\frac{i}{\hbar}(H_a^r + V_a^\phi)t\right) = \frac{1}{J^2}\begin{pmatrix} J^2\cos(Dt) - i(C^2 - |B|^2)\sin(Dt) & -i2BC\sin(Dt) \\ -i2B^*C\sin(Dt) & J^2\cos(Dt) + i(C^2 - |B|^2)\sin(Dt) \end{pmatrix}. \quad (53)$$

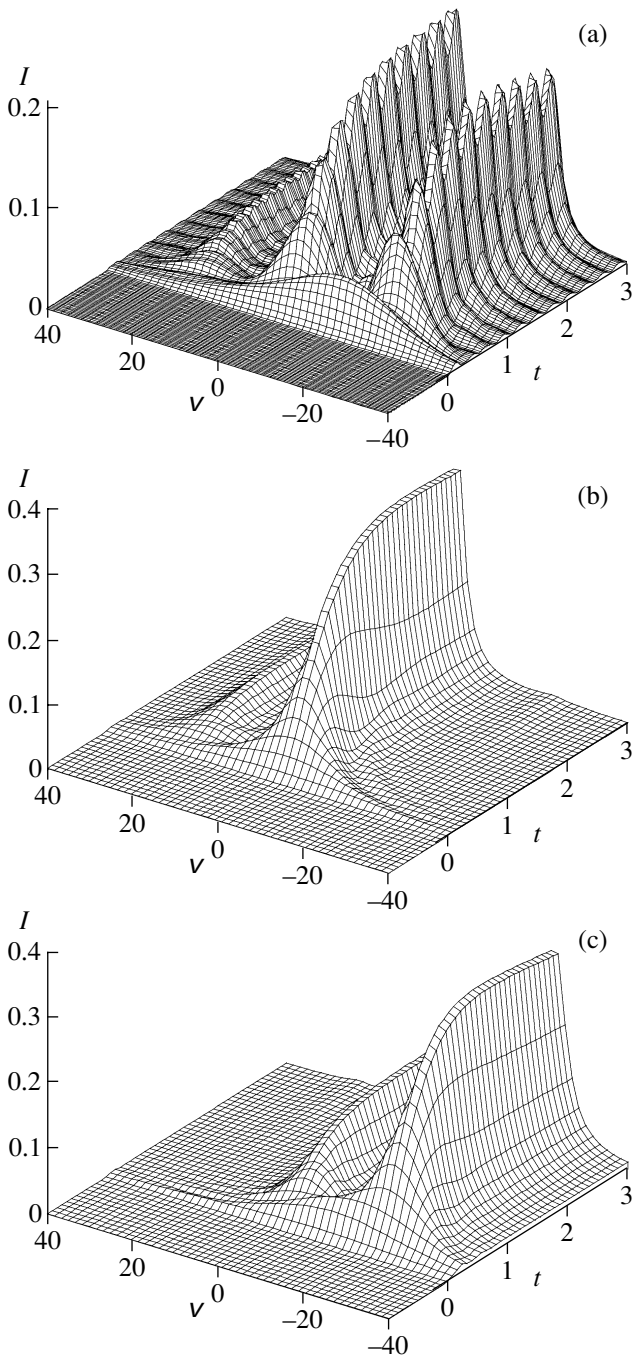


Fig. 4. Transient spectra of resonance fluorescence excited by a cw laser field. The parameters are $R = 10$, $\Delta = (-2/\sqrt{3})R$, $\gamma = 0.1$; (a) excitation without a phase change; (b) an atom is excited to the $|-D\rangle$ state by abruptly changing the field phase, $t_{cr} = \frac{1}{D}\arcsin\frac{\sqrt{C^2+R^2}}{2C}$, $\phi = \arccos\frac{C^2-R^2}{2C^2}$; (c) an atom is excited to the $|+D\rangle$ state by abruptly changing the field phase, $t_{cr} = \frac{1}{D}\arcsin\frac{\sqrt{C^2+R^2}}{2R}$, $\phi = \pi + \arccos\frac{C^2-R^2}{2R^2}$.

Taking into account the matrix expansion for the evolution operator, the equations for the amplitude and time of the phase jump can be written in the form

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} = S_\phi^{-1} \times \exp\left[-\frac{i}{\hbar}(H_a^r + V_a)t_{cr}\right]\psi(0). \tag{54}$$

Depending on the vector chosen, the system will undergo a transition to the state $|-D\rangle$ or $|+D\rangle$ (Fig. 3). By omitting cumbersome calculations, we present the final result:

$$\begin{aligned} |+D\rangle \to \sin(Dt_{cr}) &= \frac{\sqrt{C^2+R^2}}{2R}, \\ \cos(\phi+\pi) &= \frac{C^2-R^2}{2R^2}, \\ |-D\rangle \to \sin(Dt_{cr}) &= \frac{\sqrt{C^2+R^2}}{2C}, \\ \cos\phi &= \frac{C^2-R^2}{2C^2}. \end{aligned} \tag{55}$$

Note that an atomic system can undergo a transition to any of the dressed states only for certain detunings of the laser field. Namely, an atom can undergo a transition to the $|+D\rangle$ state if $(-8/3)R \le \Delta \le (2\sqrt{3})R$, and to the $|-D\rangle$ state if $(-2\sqrt{3})R \le \Delta \le (8/3)R$.

Figure 4 shows the resonance fluorescence spectra calculated for the excitation scheme described above. Figure 4a corresponds to a cw laser field without the phase interruption, while Fig. 4b corresponds to excitation of the system to the states $|+D\rangle$ and $|-D\rangle$. Note that the spectra obtained for the pure states of the dressed atom possess an important feature. They do not exhibit oscillations at the Rabi frequency (unlike the spectrum shown in Fig. 4a) and are in fact stationary for $t \ll \gamma$. This follows directly from the fact that the atom is in a pure state of the dressed atom. Because the wave functions of the atom for the states $|+D\rangle$ and $|-D\rangle$ have the form

$$\begin{aligned} \psi_{|-D\rangle}(t) &= \exp(-iD(t-t_{cr}))S_1, \\ \psi_{|+D\rangle}(t) &= \exp(iD(t-t_{cr}))S_2, \end{aligned} \tag{56}$$

where S_1 and S_2 are the columns of the matrix S, one can see that the atomic density matrix is time-independent.

Note that, unlike the case of the zero detuning [26], the intensities of the side and central components of the spectrum are not identical. Taking into account that the atom is in a stationary state, we will use the explicit form (56) of the wave function and expression (50).

After simple calculations, we obtain the intensities of the side and central lines for the atom in the $|+D\rangle$ state,

$$I_0 = \gamma\rho_{22}\frac{C^2}{C^2 + B^2}, \quad I_{2D} = \gamma\rho_{22}\frac{B^2}{C^2 + B^2}, \qquad (57)$$

and in the $|-D\rangle$ state,

$$I_0 = \gamma\rho_{22}\frac{B^2}{C^2 + B^2}, \quad I_{-2D} = \gamma\rho_{22}\frac{C^2}{C^2 + B^2}. \qquad (58)$$

Note that, although it seems that expressions (57) and (58) are asymmetric with respect to the substitution $\Delta \longrightarrow -\Delta$, the results are coincident when real values are substituted. Taking into account the limitation obtained for the ratio Δ/R, we can show that the maximum possible ratio of the intensities of the side and central components is obtained for the $|\pm D\rangle$ state when $\Delta = (\pm 2/\sqrt{3})R$, which gives $I_0 : I_{\pm 2D} = 1 : 3$.

The asymmetry of transient spectra, which was observed experimentally [27] when a cw laser field was switched on, can be readily explained within the framework of the above analysis. It is explained by the fact that the expansion of the ground state in the states of the dressed atom is not the same in the case of nonzero detuning. Taking into account explicit form (51) of the evolution operator, we represent the populations of the upper pair of states in the form

$$\rho_{|+D\rangle} = \frac{C^2}{C^2 + B^2}, \quad \rho_{|-D\rangle} = \frac{B^2}{C^2 + B^2}. \qquad (59)$$

It follows directly from (59) that the center of gravity of the triplet shifts to the transition frequency during the time of establishment of the equilibrium state.

8. CONCLUSIONS

We have studied the nonstationary resonance fluorescence of a two-level atom and have developed a new method for describing the transfer and scattering of a time-dependent laser pulse, which is based on a hierarchy of Bogoliubov–Born–Green–Kirkwood–Yvone equations for reduced density matrices. A system of equations has been obtained in the Hartree–Fock approximation, which describes stimulated scattering of laser radiation in a medium. By using the apparatus of coherent states for a lumped model, we obtained the operator of a cooperative field. By using the change of variables, we solved the wave equation in a local volume of the order of λ_L^3, obtained the spectrum of stimulated scattering of radiation, and considered its influence on the transfer of a laser field.

At the next stage, using the second-order Born approximation, we obtained the system of equations for the atomic and photon density matrices and the correlation operator. The system of equations describes stimulated and spontaneous scattering of laser radiation in optically thin media, where reabsorption can be neglected. The expressions for the nonstationary spectrum of resonance fluorescence were derived which take into account contributions from spontaneous and stimulated scattering of radiation. The necessity of consideration of the boundary conditions for absorption of an electromagnetic field during quantization is shown, and their influence on kinetic equations is analyzed. Note that it is the consideration of absorption of the field in a detector in kinetic equations for the atomic and photon density matrices that allowed us to derive for the first time the expressions for nonstationary resonance fluorescence spectra, which have no negative intensities.

The results obtained in the paper are compared with the resonance fluorescence spectra calculated using the quantum regression theorem. It is shown that, when a cooperative field is neglected in a stationary case and the relaxation constants of modes tend to zero, the spectrum represents the well-known Mollow triplet.

The transient spectra of resonance fluorescence have been analyzed for arbitrary detunings from the resonance in the regime of switching on of a cw field. It is shown that, by changing abruptly the phase of exciting radiation, the atomic system can be excited to one of the states of a dressed atom. In this case, one of the side components will be absent in the resonance fluorescence spectrum. It is shown that at times that are much shorter than the radiative relaxation time, the resonance fluorescence spectrum of the dressed state will be stationary and no Rabi oscillations will be observed. The ratio of the intensities of the two remaining fluorescence components is been found. It is shown that, for a certain relation between the value of detuning from the resonance and the laser-field power, the intensities of the central and side components can achieve a ratio of 1 : 3.

ACKNOWLEDGMENTS

The authors thank M. Bonitz, A.G. Leonov, and D.I. Chekhov for discussions and comments. We also thank V.I. Lebedev for his help in the use of the DUMKA program package. This work was supported by the Russian Foundation for Basic Research (project no. 02-02-17153), by the support program for scientific schools (project no. 00-15-96539) and young scientists (project no. 02-02-06400), as well as by grant no. 02-15-99309 of the President of the Russian Federation.

APPENDIX A

Properties of the Coherent-State Operator

Recall briefly the properties of photon creation and annihilation operators and the coherent-state operator, which were used in papers [41, 43]. The general expres-

sion for the coherent-state operator $D(\alpha_{\mathbf{k}})$ has the form

$$\begin{aligned} D(\alpha_{\mathbf{k}}) &= \exp(\alpha_{\mathbf{k}}\hat{a}_{\mathbf{k}} - \hat{a}_{\mathbf{k}}^{+}\alpha_{\mathbf{k}}^{*}) \\ &= \exp(-|\alpha_{\mathbf{k}}|^{2}/2)\exp(\hat{a}_{\mathbf{k}}^{+}\alpha_{\mathbf{k}})\exp(-\alpha_{\mathbf{k}}^{*}\hat{a}_{\mathbf{k}}). \end{aligned} \tag{A.1}$$

The coherent state can be represented in the form

$$\begin{aligned} |\alpha_{\mathbf{k}}\rangle &= D(\alpha_{\mathbf{k}})|0_{\mathbf{k}}\rangle \\ &= \exp\left(-\frac{|\alpha_{\mathbf{k}}|^{2}}{2}\right)\sum_{n}\frac{(\alpha_{\mathbf{k}}\hat{a}_{\mathbf{k}}^{+})^{n}}{\sqrt{n!}}|0_{\mathbf{k}}\rangle. \end{aligned} \tag{A.2}$$

Let us present without the proof the properties of the coherent-state operator [43]:

$$\begin{aligned} D^{+}(\alpha_{\mathbf{k}}) &= D(-\alpha_{\mathbf{k}}), \\ D(\alpha_{\mathbf{k}}^{1})D(\alpha_{\mathbf{k}}^{2}) &= \exp(i\mathrm{Im}(\alpha_{\mathbf{k}}^{1}\alpha_{\mathbf{k}}^{2*}))D(\alpha_{\mathbf{k}}^{1} + \alpha_{\mathbf{k}}^{2}), \\ D^{+}(\alpha_{\mathbf{k}})\hat{a}_{\mathbf{k}}D(\alpha_{\mathbf{k}}) &= \hat{a}_{\mathbf{k}} + \alpha_{\mathbf{k}}, \\ D^{+}(\alpha_{\mathbf{k}})\hat{a}_{\mathbf{k}}^{+}D(\alpha_{\mathbf{k}}) &= \hat{a}_{\mathbf{k}}^{+} + \alpha_{\mathbf{k}}^{*}. \end{aligned} \tag{A.3}$$

The coherent-state operators for different modes are commutative:

$$D(\alpha_{\mathbf{k}_1})D(\alpha_{\mathbf{k}_2}) = D(\alpha_{\mathbf{k}_2})D(\alpha_{\mathbf{k}_1}). \tag{A.4}$$

When passing to the wave representation, the photon operators are often transformed as

$$X' = \exp(iH_f t)X_n\exp(-iH_f t).$$

The creation and annihilation operators appearing in the Hamiltonian are transformed as

$$\begin{aligned} &\exp(i\omega_{\mathbf{k}}\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}t)\hat{a}_{\mathbf{k}}\exp(-i\omega_{\mathbf{k}}\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}t) \\ &\quad = \hat{a}_{\mathbf{k}}\exp(-i\omega_{\mathbf{k}}t), \\ &\exp(i\omega_{\mathbf{k}}\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}t)\hat{a}_{\mathbf{k}}^{+}\exp(-i\omega_{\mathbf{k}}\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}t) \\ &\quad = \hat{a}_{\mathbf{k}}^{+}\exp(i\omega_{\mathbf{k}}t). \end{aligned} \tag{A.5}$$

Taking into account (A.5), a similar transformation for the coherent-state operator can be written in the form

$$\begin{aligned} &\exp(i\omega_{\mathbf{k}}\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}t)D(\alpha_{\mathbf{k}})\exp(-i\omega_{\mathbf{k}}\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}t) \\ &= \exp(i\omega_{\mathbf{k}}\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}}t)\sum_{n}\frac{(\alpha_{\mathbf{k}}\hat{a}_{\mathbf{k}} - \alpha_{\mathbf{k}}^{*}\hat{a}_{\mathbf{k}}^{+})^{n}}{n!} \\ &\quad \times \exp(-i\omega_{\mathbf{k}}\hat{a}_{\mathbf{k}}^{+}\hat{a}_{\mathbf{k}})t) \\ &= \sum_{n}\frac{(\alpha_{\mathbf{k}}\hat{a}_{\mathbf{k}}\exp(-i\omega_{\mathbf{k}}t) - \alpha_{\mathbf{k}}^{*}\hat{a}_{\mathbf{k}}^{+}\exp(i\omega_{\mathbf{k}}t))}{n!} \\ &\quad = D(\alpha_{\mathbf{k}}\exp(-i\omega_{\mathbf{k}}t)). \end{aligned} \tag{A.6}$$

APPENDIX B

The Cooperative-Field Operator

Here, we will obtain the cooperative-field operator in the explicit form. We are interested in the expression of type (19):

$$\begin{aligned} &[\tilde{V}_{af}^{p}, \rho_f\rho_a]^{f} \\ &= \sum_{\mathbf{k}} ig_{\mathbf{k}}[(|2\rangle\langle 1|\beta_{\mathbf{k}}(t)\exp(-i\nu_{\mathbf{k}}t + i\mathbf{k}\cdot\mathbf{r}) \\ &\quad - |1\rangle\langle 2|\beta_{\mathbf{k}}^{*}(t)\exp(i\nu_{\mathbf{k}}t - i\mathbf{k}\cdot\mathbf{r}))\rho_f\rho_a \\ &\quad - \rho_f\rho_a(|2\rangle\langle 1|\beta_{\mathbf{k}}(t)\exp(-i\nu_{\mathbf{k}}t + i\mathbf{k}\cdot\mathbf{r}) \\ &\quad - |1\rangle\langle 2|\beta_{\mathbf{k}}^{*}(t)\exp(i\nu_{\mathbf{k}}t - i\mathbf{k}\cdot\mathbf{r}))], \end{aligned} \tag{B.1}$$

where the equation for $\beta_{\mathbf{k}}(t)$ has the form

$$\frac{d\beta_{\mathbf{k}}}{dt} = -Ng_{\mathbf{k}}(\rho_a)_{21}\exp(i(\nu_{\mathbf{k}}t - \mathbf{k}\cdot\mathbf{r})). \tag{B.2}$$

By formally solving the equation for $\beta_{\mathbf{k}}'(t)$, we obtain

$$\beta_{\mathbf{k}}'(t) = -Ng_{\mathbf{k}}\int_{-\infty}^{t}(\rho_a)_{21}(\tau)\exp(i(\nu_{\mathbf{k}}\tau - \mathbf{k}\cdot\mathbf{r}))d\tau. \tag{B.3}$$

Then, we substitute the result into (B.1), and taking into account that

$$\begin{aligned} &\int_{-\omega_L}^{\infty} g_{\mathbf{k}}^{2}\exp(\pm i\nu_{\mathbf{k}}(t-\tau))\omega_L^{3}d\nu_{\mathbf{k}} \\ &\quad = \frac{\gamma}{2}\delta(t-\tau) + O(\Delta_L), \end{aligned} \tag{B.4}$$

where

$$\gamma = \frac{4\omega_{21}^{3}\|\mu_{21}\|^{2}}{3\hbar c^{3}}$$

is the radiative relaxation rate, Δ_L is the Lamb shift of the level, and $\delta(t-\tau)$ is the delta function, we obtain

$$\begin{aligned} [\tilde{V}_{af}^{p}, \rho_f\rho_a]^{f} &= [V_a^{p}, \rho_a], \\ V_a^{p} &= -iN\gamma/2(|2\rangle\langle 1|(\rho_a)_{21} - |1\rangle\langle 2|(\rho_a)_{12}). \end{aligned} \tag{B.5}$$

APPENDIX C

The Radiative-Relaxation Operator

In this section, we obtain the explicit form of the relaxation operator for the atomic density matrix ρ_a. The relaxation operator for the correlation operator g_{af}

can be obtained similarly. We are interested in the explicit form of the expression,

$$[V_{af}, g_{af}]^f,$$
$$V_{af} = ig_{\mathbf{k}}(|2\rangle\langle 1|\hat{a}_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}} - |1\rangle\langle 2|\hat{a}^+_{\mathbf{k}}e^{-i\mathbf{k}\cdot\mathbf{r}}), \quad \text{(C.1)}$$

in the right-hand side of the equation for the density matrix ρ_a in (32). In this case,

$$i\frac{d\rho_a}{dt} - [H^r_a + V_a + V^p_a, \rho_a] = [V_{af}, g_{af}]^f,$$
$$i\frac{dg_{af}}{dt} - [H^r_a + V_a + V^p_a + H^r_f, g_{af}] = [V_{af}, \rho_a\rho_f], \quad \text{(C.2)}$$

where $\rho_f(t) \equiv |0_{\mathbf{k}}\rangle\langle 0_{\mathbf{k}}|$ because we neglected reabsorption of spontaneous emission. Let $U(t)$ and $C(t)$ be the fundamental solutions of equations

$$i\frac{d\rho_a}{dt} - [H^r_a + V_a + V^p_a, \rho_a] = 0,$$
$$\rho_a(t) = U(t)\rho^0_a U^{-1}(t),$$
$$i\frac{dg_{af}}{dt} - [H^r_a + V_a + V^p_a, g_{af}] = 0, \quad \text{(C.3)}$$
$$g_{af}(t) = C(t)g^0_{af}C^{-1}(t).$$

Then, using the changes of variables of the type

$$\rho_a(t) = U(t)\rho'_a U^{-1}(t),$$
$$g_{af}(t) = \exp(-iH^r_f t)C(t)g'_{af}C^{-1}(t)\exp(iH^r_f t)$$

and taking into account that the unitary transformation does not change the operator trace, we can represent the system (C.2) in the form

$$i\frac{d\rho'_a}{dt} = [V^1_{af}, S(t)g'_{af}S^{-1}(t)]^f,$$
$$i\frac{dg'_{af}}{dt} = [V^2_{af}, S^{-1}(t)\rho'_a S(t)\rho'_f],$$
$$V^i_{af} = ig_{\mathbf{k}}(\sigma^+_i(t)\hat{a}_{\mathbf{k}}\exp(-i\nu_{\mathbf{k}}t + i\mathbf{k}\cdot\mathbf{r}) - \sigma^+_i(t)\hat{a}^+_{\mathbf{k}}\exp(i\nu_{\mathbf{k}}t - i\mathbf{k}\cdot\mathbf{r})), \quad \text{(C.4)}$$
$$\sigma^+_1(t) = U^{-1}(t)|2\rangle\langle 1|U(t),$$
$$\sigma^-_1(t) = U^{-1}(t)|1\rangle\langle 2|U(t),$$
$$\sigma^+_2(t) = C^{-1}(t)|2\rangle\langle 1|C(t),$$
$$\sigma^-_2(t) = C^{-1}(t)|1\rangle\langle 2|C(t).$$

By using the result of the formal integration of the equation for g_{af}, we obtain in the equation for the atomic density matrix

$$i\frac{d\rho'_a}{dt} = -i\left[V^1_{af}(t), S(t)\right.$$
$$\left.\times\int_0^t [V^2_{af}(\tau), S^{-1}(\tau)\rho'_a S(\tau)\rho'_f]d\tau S^{-1}(t)\right]^f, \quad \text{(C.5)}$$
$$S(t) = U^{-1}(t)C(t).$$

Because $\rho_f(t) \equiv |0_{\mathbf{k}}\rangle\langle 0_{\mathbf{k}}|$, the nonzero terms appearing upon calculation of the operator trace over field variables will be proportional to the expressions

$$\hat{a}_{\mathbf{k}}\hat{a}^+_{\mathbf{k}}\rho_f,\ \rho_f\hat{a}_{\mathbf{k}}\hat{a}^+_{\mathbf{k}},\ \hat{a}^+_{\mathbf{k}},\ \rho_f\hat{a}_{\mathbf{k}}.$$

By performing a successive summation over $\mathbf{k}$ and integration over τ, returning to the initial variables $\rho'_a \longrightarrow \rho_a$ and using (B.4), we obtain

$$i\frac{d\rho_a}{dt} - [H^r_a + V_a + V^p_a, \rho_a] = \Gamma[\rho_a],$$
$$\Gamma[\rho_a] = -i\gamma/2(|2\rangle\langle 2|\rho_a + \rho_a|2\rangle\langle 2| - 2|1\rangle\langle 2|\rho_a|2\rangle\langle 1|). \quad \text{(C.6)}$$

REFERENCES

1. V. Weisskopf and E. Wigner, Z. Phys. **63**, 54 (1930).
2. V. Weisskopf, Ann. Phys. (Leipzig) **9**, 23 (1931).
3. B. R. Mollow, Phys. Rev. **188**, 1969 (1969).
4. B. R. Mollow, Phys. Rev. A **15**, 1023 (1977).
5. J. L. Carlsten, A. Szoke, and M. G. Raymer, Phys. Rev. A **15**, 1029 (1977).
6. F. Schuda, C. R. Stroud, Jr., and M. J. Hercher, J. Phys. B **7**, L198 (1974).
7. W. Harting, W. Rasmussen, R. Schieder, and H. Walther, Z. Phys. A **278**, 205 (1976).
8. F. Y. Wu, R. E. Crove, and S. Ezekiel, Phys. Rev. A **15**, 227 (1977).
9. S. P. Goreslavskiĭ and V. P. Kraĭnov, Zh. Éksp. Teor. Fiz. **26**, 23 (1979) [Sov. Phys. JETP **49**, 13 (1979)].
10. Y. Zhu, O. Wu, A. Lezama, *et al.*, Phys. Rev. A **41**, 6574 (1990).
11. O. A. Kocharovskaya, Ya. I. Khanin, and V. B. Tsaregradskiĭ, Zh. Éksp. Teor. Fiz. **86**, 423 (1984) [Sov. Phys. JETP **59**, 245 (1984)].
12. I. R. Senitzky, Phys. Rev. Lett. **40**, 1334 (1978).
13. Tran Quang, L. H. Lan, and A. S. Shumovsky, J. Opt. Soc. Am. B **7**, 908 (1990).
14. A. G. Leonov, A. A. Panteleev, A. N. Starostin, and D. I. Chekhov, Zh. Éksp. Teor. Fiz. **105**, 1536 (1994) [JETP **78**, 827 (1994)].
15. M. Lax, Rev. Mod. Phys. **32**, 25 (1960).
16. M. Lax, Phys. Rev. **129**, 2342 (1963).
17. É. G. Pestov, Tr. Fiz. Inst. im. P. N. Lebedeva Akad. Nauk SSSR **187**, 60 (1988).

18. L. Banyai *et al.*, Phys. Rev. Lett. **78**, 879 (1997).
19. F. X. Camacasse *et al.*, Phys. Rev. Lett. **77**, 5429 (1996).
20. M. Bonitz, D. Kremp, D. C. Scott, *et al.*, J. Phys.: Condens. Matter **8**, 6057 (1996).
21. R. Binder, S. H. Kohler, and M. Bonitz, Phys. Rev. B **55**, 5110 (1997).
22. J. H. Eberly and K. Wodkiewicz, J. Opt. Soc. Am. **67**, 1255 (1977).
23. J. J. Sanchez-Mondragon, N. B. Narozhny, and J. H. Eberly, Phys. Rev. Lett. **51**, 550 (1983).
24. T. Ho and H. Rabitz, Phys. Rev. A **37**, 1576 (1988).
25. T. Ho and H. Rabitz, Phys. Rev. A **37**, 4184 (1988).
26. N. Lu, P. R. Berman, Y. S. Bai, *et al.*, Phys. Rev. A **34**, 319 (1986).
27. J. E. Golub and T. W. Mossberg, Phys. Rev. Lett. **59**, 2149 (1987).
28. P. W. Milloni and P. L. Knight, Phys. Rev. A **10**, 1096 (1974).
29. M. O. Scully and W. E. Lamb, Jr., Phys. Rev. **159**, 208 (1967).
30. E. V. Baklanov, Zh. Éksp. Teor. Fiz. **65**, 2203 (1973) [Sov. Phys. JETP **38**, 1100 (1974)].
31. D. A. Holm, M. Sargent, III, and S. Stenholm, J. Opt. Soc. Am. B **2**, 243 (1986).
32. P. Meystre and M. Sargent III, *Elements of Quantum Optics* (Springer, Berlin, 1990).
33. A. A. Panteleev, Zh. Éksp. Teor. Fiz. **111**, 440 (1997) [JETP **84**, 241 (1997)].
34. A. A. Panteleev and A. N. Starostin, Zh. Éksp. Teor. Fiz. **102**, 441 (1992) [Sov. Phys. JETP **55**, 223 (1992)].
35. A. A. Panteleev and A. N. Starostin, Zh. Éksp. Teor. Fiz. **106**, 1606 (1994) [JETP **79**, 866 (1994)].
36. A. A. Panteleev and Vl. K. Roerich, Zh. Éksp. Teor. Fiz. **118**, 312 (2000) [JETP **91**, 273 (2000)].
37. A. A. Panteleev and Vl. K. Roerich, Zh. Éksp. Teor. Fiz. **119**, 243 (2001) [JETP **92**, 210 (2001)].
38. S. Stenholm, *Foundations of Laser Spectroscopy* (Wiley, New York, 1984; Mir, Moscow, 1987).
39. M. Bonitz, *Quantum Kinetic Theory* (Teubner, Stuttgart, 1998).
40. A. I. Akhiezer and S. V. Peletminskiĭ, *Methods of Statistical Physics* (Nauka, Moscow, 1977; Pergamon, Oxford, 1981).
41. L. Mandel and E. Volf, *Optical Coherence and Quantum Optics* (Cambridge Univ. Press, Cambridge, 1995; Nauka, Moscow, 2000).
42. R. Loudon, *The Quantum Theory of Light* (Clarendon Press, Oxford, 1973; Mir, Moscow, 1976).
43. A. M. Perelomov, *Generalized Coherent States and Their Applications* (Nauka, Moscow, 1987; Springer, New York, 1986).
44. V. N. Gorbachev and A. I. Zhiliba, Quantum Opt. **5**, 193 (1993).
45. H. Haug and A. P. Jauho, *Quantum Kinetics in Transport and Optics of Semiconductors* (Springer, Heidelberg, 1996).
46. A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Phenomena in Optics* (Nauka, Moscow, 1988).
47. Y. Ben-Aryeh, C. M. Bowden, and J. C. Englund, Phys. Rev. A **34**, 3917 (1986).
48. R. H. Dicke, Phys. Rev. **93**, 99 (1954).
49. R. Jodoin and L. Mandel, Phys. Rev. A **9**, 873 (1974).
50. V. I. Lebedev, *Computing Processes and Systems* (Nauka, Moscow, 1991), Vol. 8.
51. V. I. Lebedev, in *Numerical Method and Application*, Ed. by G. I. Marchuk (CRC Press, Boca Raton, 1994).
52. J. H. Eberly, C. V. Kunasz, and W. Wodkiewicz, J. Phys. B **13**, 217 (1980).
53. J. H. Eberly, in *Quantum Electrodynamics and Quantum Optics*, Ed. by A. O. Barut (Plenum, New York, 1984).
54. B. Renaud, R. M. Whitley, and C. R. Stroud, Jr., J. Phys. B **10**, 19 (1977).
55. Y. S. Bia, A. G. Yodh, and T. W. Mossberg, Phys. Rev. Lett. **55**, 1277 (1985).

Translated by M. Sapozhnikov

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 241–246.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 276–282.

ATOMS, SPECTRA, RADIATION

Generation of Intense Pulsed Low-Kinetic-Energy Molecular Beams

G. N. Makarov

Institute of Spectroscopy, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia
**e-mail: g.makarov@isan.troitsk.ru*

Received August 29, 2002

Abstract—A method for the generation of intense pulsed low-kinetic-energy molecular beams is described. The method is based on the formation of a cold (≈77 K) pressure shock as a result of interaction between an intense pulsed gas-dynamically cooled molecular beam with a solid surface. The pressure shock is used as a source of a secondary beam for generating low-energy molecules. The suggested method was used to obtain intense molecular beams of H_2, He, CH_4, N_2, and Kr with kinetic energies lower than or equal to 10 meV and H_2/Kr and He/Kr molecular beams with kinetic energies of H_2 and He molecules lower than 1 meV. The energy (velocity) of molecules in low-energy beams can be controlled by varying the intensity of the initial beam or temperature in the pressure shock. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Currently, molecular (atomic) beams [1], including low-energy beams [2], in which the kinetic energy of molecules (atoms) is much lower than their thermal energy at room temperature, are extensively used for scientific purposes. Intense molecular beams [$\geq 10^{20}$ molecules/(sr s)] with kinetic energies of one to several tens millielectronvolts are used to study chemical reactions, elastic and inelastic collisions, and interactions of molecules with the surface [1–3]. In recent years, their use in experiments with trapping molecules has been initiated [4, 5].

The most widely used technique for obtaining intense molecular beams is their isolation with skimmers from gas-dynamically cooled jets produced by pulsed nozzles [6]. The kinetic energy of molecules in beams is determined by gas temperature T_0 prior to expansion through a nozzle,

$$\frac{1}{2}mv^2 = \frac{\gamma}{\gamma - 1}k(T_0 - T), \qquad (1)$$

where v is the stationary flow velocity, m is the mass of one molecule, $\gamma = c_p/c_v$ is the ratio between the specific heat capacities, k is the Boltzmann constant, and T is the stationary temperature. If the gas is at room temperature in the source, the kinetic energy of molecules in the beam amounts to $E_{\text{kin}} \approx 50$–60 meV (for a monoatomic gas) or $E_{\text{kin}} \geq 150$–200 meV (for a gas of polyatomic molecules) depending on index γ.

It follows from (1) that gases should be cooled in gas sources to obtain low-energy molecular beams. Cooling pulsed sources of molecular beams to low temperatures is, however, impracticable, because the materials used in such sources (in particular, elastomers and plastics) lose their elastic and plastic properties and become brittle. In addition, gas pressure decreases as temperature lowers, which makes it difficult to generate gas-dynamically cooled jets.

Low-kinetic-energy molecular beams (≤50 meV) are usually obtained with the use of effusion sources, which can operate both at room temperature and with cooling gases within them to liquid nitrogen or, to produce He beams, liquid helium temperatures [2, 7]. The kinetic energy of molecules in an effusion beam is determined by the temperature of the gas in the source ($E_{\text{kin}} \approx kT_0$). The intensities of molecular beams obtained in such a way are, however, comparatively low [$\leq 10^{16}$–10^{17} molecules/(sr s)] [2]. In addition, the spread of the velocities of molecules in these beams is large, which additionally decreases their intensities.

At the same time, precisely beam intensities are the determining factor in many experiments with molecular beams [1, 2, 6]. The development of methods for obtaining intense beams of low-energy molecules is therefore a very topical area. Recently [8], we suggested a comparatively simple method for generating such beams and obtained preliminary results. In this work, we describe the suggested method in detail and report the results of a thorough study of the generation of intense low-energy molecular beams with controlled kinetic energy in the range of one to several tens millielectronvolts.

2. EXPERIMENTAL SETUP AND METHOD

The suggested method for generating low-energy pulsed molecular beams uses a pressure shock [9–11] formed in the interaction between an intense pulsed gas-dynamically cooled molecular beam (or flow) with

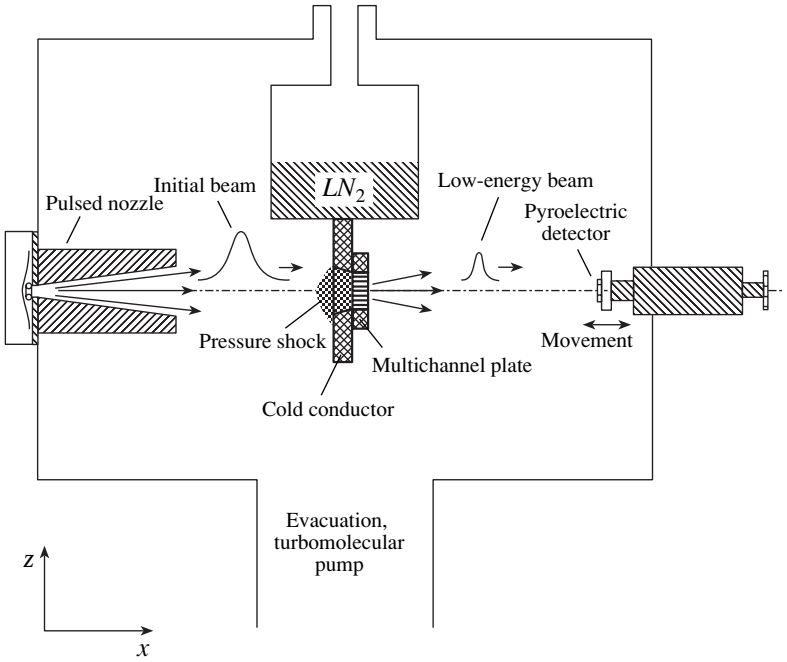


Fig. 1. Scheme of experiments.

a solid surface. The essence of the method is as follows (see Fig. 1). An intense [$\geq 10^{21}$ molecules/(sr s)] wide-aperture (divergence $\omega \approx 0.05$ sr) pulsed molecular beam fell on a copper cold conductor and a multichannel plate attached to it, both cooled by liquid nitrogen. The plate was made of duralumin, its thickness was L = 4 mm, and the diameter of channels in it was d_0 = 0.5 mm. The channels were arranged according to the close packing scheme. The distance between orifice centers was 0.75 mm. The cold conductor contained a hole in the form of a convergent cone. The diameters of the cone of the entrance and exit were 11 and 9 mm, respectively. The cold conductor was 8 mm thick. The distance between the nozzle throat and the surface of the multichannel plate was 70 mm. To prevent plate operation in the "transparent" mode, it was tilted at a small angle $\alpha \geq d_0/L \approx 7°$ with respect to the incident beam axis.

When an intense supersonic molecular beam falls onto the cooled multichannel plate, a cold pressure shock is formed in front of the plate and in its channels [12, 13]. The characteristic pressure shock dimensions are on the order of the free path of molecules Λ [9, 10]. Such a pressure shock is a convenient source for generating intense secondary pulsed molecular beams [14–16]. If $\Lambda_{\text{eff}} \geq d_0$ (Λ_{eff} is the effective mean free path of molecules in the channels, $\Lambda_{\text{eff}} > \Lambda$ [2]), the gas has sufficient time to cool to the temperature of channel walls while passing through the channels (approximately to liquid nitrogen temperatures). As a result, a beam of low-energy molecules is formed from the pressure shock. The specified condition was almost always met in our experiments. As a result, we were able to obtain intense gas-dynamically cooled molecular beams with the kinetic energy of molecules [Eq. (1)] determined by the gas temperature in the pressure shock (≈77 K). If, however, gas pressure in the pressure shock was low and gas dynamic flow could not form, the transition to the following off of effusion occurred, and the mean velocity of molecules in the beam was close to the mean velocity of molecules at nitrogen temperatures. Naturally, the spread of molecular velocities in the beam then increased.

We used a pulsed nozzle of the current loop type [17]. The diameter of the orifice was 0.75 mm. The time of opening was 70–100 μs at half-height depending on the composition and pressure of the gas above the nozzle. Gas pressure above the nozzle varied in the range 0.1–7 atm. The nozzle throat had the form of a cone with a total aperture angle of 15°. The length of the cone was 35 mm. The vacuum chamber in which molecular beams were formed was evacuated to approximately 1×10^{-6} torr by a turbomolecular pump. The number of molecules flowing off from the nozzle per pulse depended on the gas pressure above the nozzle and varied in the range of approximately 8×10^{15} to 2×10^{18} molecules/pulse in our experiments. The method for measuring the number of molecules per pulse was described in detail in [18, 19].

We studied the dependences of the intensity and velocity of low-energy molecular beams on the intensity and velocity of the initial beam and the gas temperature in the pressure shock. We also measured the spread of molecular velocities (the degree of gas cool-

ing) in both beams. Measurements were performed by the time-of-flight method based on using a pyroelectric detector with a time resolution of about 3–5 μs as a detector of molecular beams [20, 21]. The size of the active pyroelectric detector element was 4×4 mm^2. It detected molecules propagating within the solid angle determined by the size of its active element and the distance from the nozzle to the detector. The design of the detector was described in detail in [22]. It could be moved along the beam axis with the use of a mechanical device. This allowed us to measure the time-of-flight spectra of molecules at various distances from the nozzle. These spectra were used to determine beam velocities and the spreads of molecular velocities in the beams.

3. RESULTS AND DISCUSSION

Experiments were performed with H_2, He, CH_4, N_2, and Kr molecular beams. In all of them, we observed the generation of molecular beams with low velocities. The dependence of the kinetic energy of a low-energy He molecular beam on the gas pressure above the nozzle is shown in Fig. 2 (curve *1*). For comparison, similar dependences for the secondary beam generated when the gas was not cooled in the pressure shock ($T_0 \approx$ 300 K) (curve *2*) and for the initial beam (curve *3*) are also plotted. According to Fig. 2, the kinetic energy of molecules in the low-energy beam noticeably decreased and approached the mean energy of He atoms at $T \approx 77$ K, equal to approximately 6.6 meV as the intensity of the initial beam (gas pressure above the nozzle) lowered.

The dependences of the kinetic energies of He and CH_4 beams on the gas temperature in the pressure shock obtained at a gas pressure of 2 atm above the nozzle are shown in Fig. 3. Note that, in the studied temperature range, the kinetic energy of molecules decreased somewhat faster rather than linearly as temperature lowered, because, under our experimental conditions, the gas pressure in the pressure shock (in the source of the low-energy beam) also decreased as temperature lowered, which caused an additional decrease in the velocity of beams.

The table contains the results of measurements of the mean velocities and energies of low-energy H_2, CH_4, N_2, and Kr molecular beams. The spreads of molecular velocities in the beams are also given. For comparison, similar data on the initial beams are included. We obtained H_2 molecular beams with a kinetic energy of $E_{kin} \le 6.9$ meV and CH_4, N_2, and Kr molecular beams with kinetic energies of $E_{kin} \le 11.5$, ≤9.9, and ≤7.1 meV, respectively. Note that the mean velocity of Kr atoms in the beam was $V \approx 130$ m/s.

We also performed experiments with H_2/Kr and He/Kr molecular beams (the pressure ratio was 1 : 5 in both cases). At a gas pressure of 0.8 atm above the nozzle and a gas temperature of 77 K in the pressure shock,

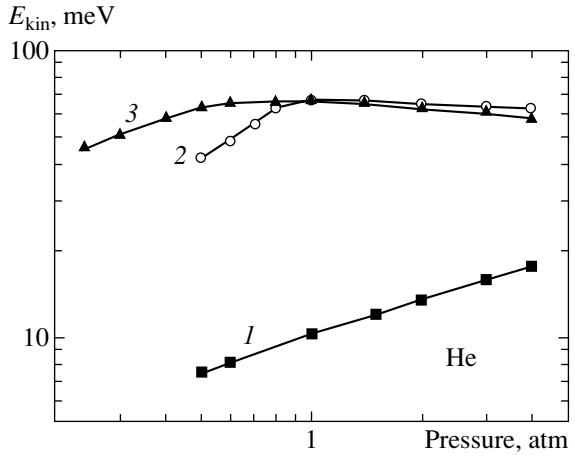


Fig. 2. Dependence of the kinetic energy of the low-energy He molecular beam on gas pressure above the nozzle (curve *1*). Curves *2* and *3* are similar dependences for the secondary beam when gas in the pressure shock is not cooled and for the initial molecular beam, respectively.

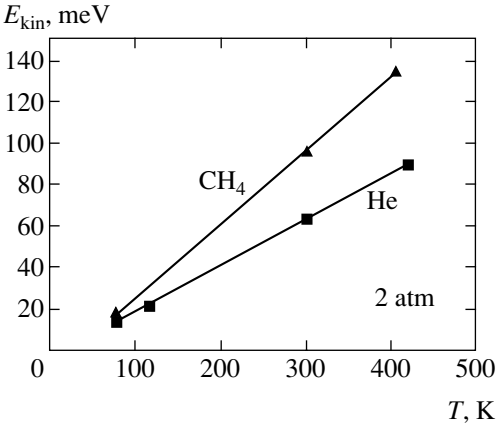


Fig. 3. Dependences of the kinetic energy of He and CH_4 molecular beams on gas temperature in the pressure shock. Gas pressure above the nozzle is 2.0 atm.

the velocities of molecular beams were approximately 165 m/s, which corresponded to the kinetic energies of H_2 and He molecules in the beams $E_{kin} \le 0.3$ and ≤0.6 meV, respectively.

We also studied the feasibility of generating low-energy molecular beams with the use of a convergent–divergent cone of the type of the Laval nozzle cooled to $T \approx 77$ K. The cone was made of duralumin. The diameters of the entrance orifice, waist, and exit orifice were $d_{in} = 13$ mm, $d_0 = 2.5$ mm, and $d_{out} = 6$ mm. The total cone length was 30 mm, and the length of the convergent part was 24 mm. Note that precisely with the use of the convergent–divergent cone of the Laval nozzle type, intense secondary beams formed from the pressure shock were obtained in [14–16]. To prevent the transparent mode of cone operation, the cone was insignificantly (by 1.5–2 mm) displaced with respect to the

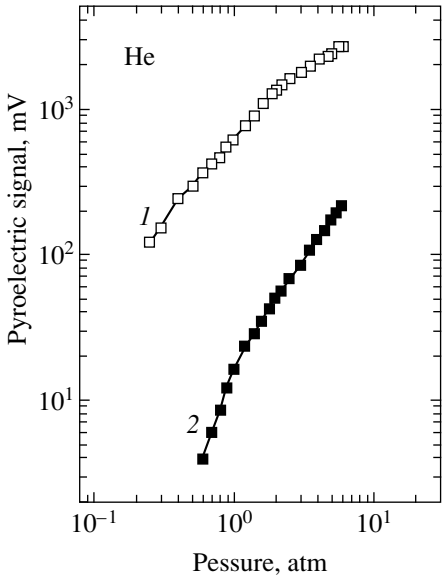


Fig. 4. Dependences of pyroelectric detector signals for (*1*) the initial He molecular beam and (*2*) the secondary molecular beam when the gas is not cooled in the pressure shock on gas pressure above the nozzle. The distances from the nozzle to the detector and the multichannel plate are 143 and 73 mm, respectively.

initial beam axis. In these experiments, we obtained He and H_2 molecular beams with mean velocities of molecules equal to $v \approx 640$ and ≈1300 m/s, which corresponded to the kinetic energies of He and H_2 molecules $E_{kin} \approx 8.6$ and ≈17.8 meV, respectively. The intensities of the beams were comparable with the intensities of the beams obtained with the use of the multichannel plate. With the cone, low-energy beams were, however, generated at much lower initial beam intensities (at lower gas pressures above the nozzle). For instance, the He, H_2, and CH_4 low-energy beams were generated at gas pressures above the nozzle of $P \le 0.6$, ≤0.4, and ≤0.3 atm, respectively. Indeed, the condition for gas cooling given above ($\Lambda_{eff} \ge d_0$) was satisfied in experiments with the cone at lower gas concentrations in the pressure shock, because the diameter of the cone waist was much larger than the diameter of channels in the plate. At higher gas pressures above the nozzle, this condition was not met. A gas within the cone then had no time to cool to the temperature of cone walls. As a result, high-kinetic-energy molecular beams were generated.

To compare the intensities of low-energy molecular beams with the initial beam intensities, we studied the dependences of the signals induced on the detector by primary and secondary beams on the gas pressure above the nozzle. Such dependences for the He molecular beams are shown in Fig. 4. The distance between the nozzle and the detector was 143 mm, and the distance between the multichannel plate and the detector, about 73 mm. The figure shows that the signal generated by the secondary beam was 20–30 times weaker than the signal generated by the primary beam at a gas pressure above the nozzle $P \ge 1$ atm. Taking into account the difference of the distances between the sources of primary and secondary beams and the detector, we find that the intensity of the secondary He molecular beam was approximately two orders of magnitude lower than the intensity of the primary beam. The intensity of the low-energy He beam was still lower, because a decrease in gas temperature (therefore,

Molecular beam velocity and energy measurements

Gas composition and pressure (atm) above the nozzle	Initial beam			Low-energy beam		
	v, m/s	E_{kin}, meV	$v/\Delta v$	v, m/s	E_{kin}, meV	$v/\Delta v$
H_2						
1.9	2950	91.7	8.5	1050	11.6	4.3
0.6	2620	72.4	4.2	810	6.9	1.3
CH_4						
2.0	1330	149.2	10.3	450	17.1	4.7
0.6	1250	131.8	4.5	370	11.5	2.1
N_2						
2.2	910	122.2	9.3	355	18.6	5.1
1.0	860	109.2	8.7	260	9.9	2.7
Kr						
2.0	385	62.5	6.5	168	11.9	1.8
1.0	360	54.7	5.4	130	7.1	

in the gas pressure) in the pressure shock also caused a decrease in the intensity of the beam.

The intensities of beams were determined by measuring gas consumption in n nozzle pulses and the geometric dimensions (extent and cross section) of the beams (e.g., see [18, 19]). These measurements allowed us to fairly accurately determine the intensities of the beams [18, 19]. For instance, let us determine the intensity of the He molecular beam from the measurement results. It was found from the gas flow rate that the total number of He atoms flowing off from the nozzle in one pulse at a pressure of 2 atm above the nozzle was approximately 7×10^{17}. As the pulse width of the initial molecular beam was about 100 μs and the beam divergence was $\omega \approx 0.05$ sr, the intensity of the beam was $I_1 \approx 1.4 \times 10^{23}$ molecules/(sr s). The intensity of the low-energy He molecular beam under the specified conditions was approximately 100–150 times lower; that is, it equaled $I_2 \approx 10^{21}$ molecules/(sr s). It follows that the intensities of the low-energy molecular beams obtained in our experiments were four to five orders of magnitude higher than the intensities characteristic of "standard" effusion beams.

The concentration of helium atoms in the initial beam and in the pressure shock was estimated based on measurements of the total number of molecules in the beam and the beam volume. This method was described in detail in [18, 19]. For instance, at a pulse width of about 100 μs and a mean beam velocity of approximately 1750 m/s, the spatial extent of the He molecular beam was approximately 17.5 cm. The cross section of the beam on the surface of the multichannel plate was about 2.4 cm^2 (it was determined by the cone aperture angle and the distance from the nozzle to the multichannel plate). It follows that the volume of the beam near the plate surface was $V_b \approx 42$ cm^3. At a gas pressure of 2 atm above the nozzle, the concentration of helium atoms in the initial beam was therefore $N_1 \approx 1.7 \times 10^{16}$ cm^{-3}. The concentration in the pressure shock (N_2) was estimated from the relation for the limiting concentration value in the normal pressure shock [9–11],

$$N_2/N_1 \approx \frac{\gamma + 1}{\gamma - 1}, \quad \gamma = 1.66 \qquad (2)$$

(see [23]). This gave $N_2 \approx 7 \times 10^{16}$ cm^{-3}.

4. CONCLUSIONS

To summarize, the method described in this work can be used to generate intense pulsed molecular beams with kinetic energies of one to several tens millielectron-volts. We showed that the energy of molecules in beams can be controlled by changing the intensity of primary beams or gas temperature in the pressure shock. The intensities of low-energy molecular beams obtained in our experiments were four to five orders of magnitude higher than the intensities of standard effusion beams, from which low-energy molecules are usually obtained. In generating low-energy He and H_2 molecular beams, liquid helium can be used to attain deeper gas cooling in the pressure shock and obtain still slower beams. Note in conclusion that using Kr or Xe as carrier gases in the method described above allows CO, N_2, NO, and O_2 molecular beams with kinetic energies of $E_{kin} \leq 1$–2 meV to be generated.

Lastly, note that slow pulsed beams have low spatial extents. This is an advantage in certain experiments on the excitation of molecules in beams by laser radiation. At a small beam extent, virtually all molecules can be excited or dissociated by high-intensity laser pulses, which considerably increases the effectiveness of laser radiation action on molecular beams [24, 25].

ACKNOWLEDGMENTS

The author thanks V.M. Apatin and A.N. Petin for technical assistance. This work was financially supported by the Russian Foundation for Basic Research (project no. 00-03-33003).

REFERENCES

1. *Atomic and Molecular Beam Methods*, Ed. by G. Scoles (Oxford Univ. Press, New York, 1988).
2. H. Pauly, in *Atomic and Molecular Beam Methods*, Ed. by G. Scoles (Oxford Univ. Press, New York, 1988).
3. R. Vetter, Ch. Naulin, and M. Costes, in *Book of Abstracts: XVIII International Symposium on Molecular Beams* (Ameland, Netherlands, 1999), p. 189.
4. S. K. Sekatskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **62**, 900 (1995) [JETP Lett. **62**, 916 (1995)].
5. R. T. Jongma, G. von Helden, G. Berden, and G. Meijer, Chem. Phys. Lett. **270**, 304 (1997).
6. W. R. Gentry, in *Atomic and Molecular Beam Methods*, Ed. by G. Scoles (Oxford Univ. Press, New York, 1988).
7. R. Feltgen, H. Kirst, K. A. Kohler, and H. Pauly, J. Chem. Phys. **76**, 2360 (1982).
8. G. N. Makarov, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 341 (2002) [JETP Lett. **76**, 283 (2002)].
9. Ya. B. Zel'dovich and Yu. P. Raĭzer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, 2nd ed. (Nauka, Moscow, 1966; Academic, New York, 1966 and 1967), Vols. 1 and 2.
10. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics*, 3rd ed. (Nauka, Moscow, 1986; Pergamon, New York, 1987).
11. G. N. Abramovich, *Applied Gas Dynamics*, 5th ed. (Nauka, Moscow, 1991), Part 1.
12. G. N. Makarov and A. N. Petin, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 583 (2000) [JETP Lett. **71**, 399 (2000)].
13. G. N. Makarov and A. N. Petin, Chem. Phys. Lett. **323**, 345 (2000).

14. V. M. Apatin, G. N. Makarov, and V. V. Nesterov, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 735 (2001) [JETP Lett. **73**, 651 (2001)].

15. V. M. Apatin, G. N. Makarov, and V. V. Nesterov, Chem. Phys. Lett. **347**, 101 (2001).

16. G. N. Makarov, Zh. Éksp. Teor. Fiz. **120**, 1411 (2001) [JETP **93**, 1222 (2001)].

17. W. R. Gentry and C. F. Giese, Rev. Sci. Instrum. **49**, 595 (1978).

18. G. N. Makarov and A. N. Petin, Zh. Éksp. Teor. Fiz. **119**, 5 (2001) [JETP **92**, 1 (2001)].

19. G. N. Makarov and A. N. Petin, Chem. Phys. **266**, 125 (2001).

20. V. M. Apatin, L. M. Dorozhkin, G. N. Makarov, and L. M. Pleshkov, Appl. Phys. B **29**, 273 (1982).

21. V. M. Apatin and G. N. Makarov, Zh. Éksp. Teor. Fiz. **84**, 15 (1983) [Sov. Phys. JETP **57**, 8 (1983)].

22. R. V. Ambartzumian, L. M. Dorozhkin, G. N. Makarov, *et al.*, Appl. Phys. **22**, 409 (1980).

23. *CRC Handbook of Chemistry and Physics*, Ed. by D. R. Lide (CRC Press, Boca Raton, 1993–1994).

24. G. N. Makarov, D. E. Malinovskiĭ, and D. D. Ogurok, Laser Chem. **17**, 205 (1998).

25. G. N. Makarov, D. E. Malinovskiĭ, and D. D. Ogurok, Zh. Tekh. Fiz. **69**, 35 (1999) [Tech. Phys. **44**, 31 (1999)].

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 247–256.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 283–292.

PLASMA, GASES

Bound States of the Josephson Degrees of Freedom and Trap Oscillations¶

L. A. Manakova

Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia
e-mail: manakova@kurm.polyn.kiae.su

Received September 11, 2002

Abstract—It is shown that the interaction of the Josephson degrees of freedom with states of condensate motion can produce their equilibrium bound states. As a result of the appearance of these states, first, the tunneling splitting is significantly increased in double-well trapped condensates. Second, the bound states can realize an absolute minimum of the thermodynamic energy for a sufficiently strong interaction. Transition to the new ground state is a second-order phase transition. The existence of the bound state leads to an equilibrium distortion of the condensate shape. This implies that the Josephson states can be detected by observing the change in the condensate shape. *© 2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Since the discovery of two-component condensates [1] and condensates in a double-well potential [2], the phenomena caused by phase coherence of two condensate modes attract considerable attention, both experimental and theoretical (see, e.g., [3] and references therein). In [2], spatial quantum coherence was observed by means of an interference pattern in two overlapping condensates. This interference pattern was confirmed in [4] by numerical simulation of the Gross–Pitaevskii equation. In [5], coherent oscillations of the relative populations were observed in driven two-component condensates with different internal states. As is well known, a clear manifestation of phase coherence is the Josephson effect. In numerous studies devoted to the Josephson effect in systems of two condensates in different internal states [5, 6] or in a double-well potential [7, 8], coherent Josephson oscillations are considered for various dynamical regimes caused by the competition between tunneling and intracondensate interaction (nonlinearity). In [7], the Josephson coupling energy is calculated for small-amplitude oscillations in a double-well potential. Damping effects due to the normal currents at a finite temperature are estimated there. In [8], it is shown that for a relatively weak interaction, the particle number oscillations between the condensates are complete. They are suppressed when the total number of atoms in the condensates exceeds a critical value and the behavior of the system is governed by nonlinearity. Nonlinear Josephson-type oscillations in the relative oscillations of driven two-component condensates have been studied in [6]. Decoherence effects and quantum corrections to mean-field solutions have been considered in [11, 12]. In [13], the damping effects of the Josephson current (even at zero temperature) are derived within the functional integral approach. A detailed treatment of the nonlinear classical dynamics of the condensates in a double-well potential was given in [9, 10]. In [14], the quantum and thermal fluctuations of the phase have been studied for condensates in the double-well potential.

We emphasize that experimental observation of the Josephson effect is difficult because the small energy splitting associated with the Josephson coupling implies that thermal and quantum fluctuations destroy the phase coherence between two condensates even at the lowest achievable temperatures [13, 14]. While the energy splitting can be increased, e.g., by lowering the barrier height, it then becomes comparable with that of the motion states of the condensates.

However, the problem of the interaction between the Josephson degrees of freedom and states of motion (oscillations) of the trapped condensate has yet to be analyzed. The present paper focuses on mechanisms of increasing the tunneling splitting in a double-well potential and of formation of the bound states of the Josephson degrees of freedom with trap oscillations. The mechanisms are generated by a sufficiently strong interaction between the Josephson and oscillation states. These mechanisms may be important for experimental detection of the Josephson states. The consideration proposed in the present paper is suitable for the double-well trapped condensates and two-component condensates in the same trap.

The results obtained in this paper are as follows.

¶This article was submitted by the author in English.

1. As is well known [3, 7, 14], the Gross–Pitaevskii equations for two condensates with a weak Josephson coupling have stationary solutions corresponding to the lowest states with the eigenenergies $\pm E_J$, where E_J is the Josephson coupling energy. This implies that the double-well condensates form a macroscopic two-level subsystem with the tunneling energy splitting $2E_J$ if the dynamical Josephson oscillations of the particle number are disregarded. In what follows, it is shown that a sufficiently strong interaction between the macroscopic two-level subsystem and the condensate oscillation results in their equilibrium bound state. The appearance of the bound state generates an essential increase in the tunneling splitting of the macroscopic two-level subsystem.

2. In Sections 3–5, we consider the interaction between anharmonic trap oscillations and the Josephson degrees of freedom generated by the particle number transfer between two condensates. Nonlinear dynamic oscillations of the particle number between two condensates with the Josephson coupling are considered in [9, 10]. In the present paper, we derive the quantized spectrum of the particle number generated by the Josephson coupling in order to formulate the problem of the interaction between the Josephson and oscillation degrees of freedom in an adequate manner. The states of this spectrum represent a quantum analogue of the nonlinear coherent Josephson oscillations considered in [10]. In what follows, the states of the quantized spectrum are called the Josephson states. The spectrum is highly nonequidistant and has a logarithmic singularity in the density of states at an energy of $2E_J$. We show that any Josephson state can be realized by means of a given initial disbalance of the particle number in two condensates.

3. We consider the interaction between trap oscillations and the excited Josephson states corresponding to a sufficiently large initial disbalance of the particle number. We show that this interaction is responsible for the formation of a bound state of $\bar{n}_m \gg 1$ oscillation quanta with the Josephson state corresponding to the initial disbalance of the particle number. In the Thomas–Fermi approximation at $\mu \gg \omega_0$, where μ is the chemical potential of the condensate and ω_0 is the characteristic frequency of the trap, $h = 1$ and the bound state arises in the region of a sufficiently dense oscillation spectrum. In this region, the level separations are small compared with the harmonic oscillation frequency ω_0. The equilibrium values of the oscillation quanta and the initial particle number disbalance are coupled self-consistently and can realize an absolute minimum of the thermodynamic energy at a sufficiently strong interaction. The thermodynamic average $\bar{n}_m \neq 0$ generates an equilibrium distortion of the condensate shape. This allows detection of the Josephson states by observing a change in the condensate shape.

2. BOUND STATE OF THE MACROSCOPIC TWO-LEVEL SYSTEM AND TRAP OSCILLATIONS

The Josephson coupling is realized for condensates in a symmetric–asymmetric double-well potential formed by two different traps with a barrier between them [2]. The barrier is created by laser light, and its height is directly proportional to the laser power and can therefore be varied easily. The proposed mechanism is also suitable for condensates in different internal states in the same trap. Experimentally, this may be a superposition of two Rb^{87} condensates in the states [5, 6]

$$|F = 1, m_F = -1\rangle, |F = 2, m_F = 1\rangle.$$

For a weak Josephson coupling, the basis states are the self-consistent ground states in the two condensates, separately. The wave function of a condensate with the Josephson coupling is given by a superposition of these states, namely,

$$\Psi(\mathbf{r}, t) = \psi_1(\mathbf{r})a_1(t) + \psi_2(\mathbf{r})a_2(t),$$

where $\psi_i(\mathbf{r})$ are normalized solutions of the Gross–Pitaevskii equation,

$$a_i(t) = N_i^{1/2}(t)e^{i\theta_i(t)}, \quad i = 1, 2,$$

with N_i and $\theta_i(t)$ being the particle numbers and phases of each condensate.

As is well known [3, 7, 14], the Hamiltonian of two condensates with a weak Josephson coupling is given by

$$H_J - E_0 = E_C(\Delta N)^2 - 2E_J\cos\phi, \qquad (1)$$

where

$$\Delta N = N_1 - N_2, \quad E_C = \partial\mu/\partial N,$$

$\mu \equiv \mu_1 = \mu_2$ are the chemical potentials, N is the total particle number, and $\phi = \theta_1 - \theta_2$ is the relative phase of the condensates. The quantities E_C and E_J depend on the total particle number N. In (1), the energy origin is the mean-field total energy of the condensates, namely, $E_0 \equiv \mu N$. The variables ΔN and ϕ are canonical. The equations of motion can be written in the Hamiltonian form,

$$\frac{\partial\phi}{\partial t} = \frac{\partial H_J}{\partial(\Delta N)}, \quad \frac{\partial(\Delta N)}{\partial t} = -\frac{\partial H_J}{\partial\phi}.$$

The lowest stationary solutions of these equations have the eigenenergies $E_{s,a} = \pm E_J$ corresponding to symmetric ($\phi_s = 2\pi n$, $(\Delta N)_s = 0$) and antisymmetric ($\phi_a = \pi(2n + 1)$, $(\Delta N)_a = 0$) eigenfunctions.

Thus, the double-well trapped condensates form a macroscopic two-level system with the tunneling splitting $2E_J$, if we disregard the dynamical Josephson oscillations of the particle number.

In what follows, we showed that a sufficiently strong interaction between the macroscopic two-level system and the condensate oscillation is responsible for the formation of their equilibrium bound state. In this state, a definite equilibrium number of the oscillation quanta is coupled to the two-level system.

The interaction can be realized by the following mechanisms. First, the interaction can be implemented if we allow the Josephson coupling energy to depend on the atom displacement. The latter is generated by the condensate oscillation. Second, the interaction can be realized by applying a pair of traveling-wave laser beams with the same Rabi frequency Ω and the wave-vector difference q, for instance, in the x direction. The one-dimensional condensate is considered for simplicity. The pulse frequencies are chosen to be resonant with the transition energy between the two stationary Josephson states. A similar mechanism is used for the detection of the motion states of a single, trapped, two-level atom (see [15] and references therein) and N two-level atoms in a trap [16].

In what follows, we assume that interaction generates the transition between the states $|e, n'\rangle$ and $|g, n\rangle$. Here, $|e\rangle$ and $|g\rangle$ are two Josephson states, and n and n' are the numbers of oscillation quanta. The transition matrix element can be written as

$$\begin{aligned} &G\langle e, n'|\sigma^+ \exp(i\eta(\hat{a}^+ + \hat{a})) \\ &+ \sigma^- \exp(-i\eta(\hat{a}^+ + \hat{a}))|g, n\rangle. \end{aligned} \tag{2}$$

The operators $\hat{a}^+$ and $\hat{a}$ are the creation and annihilation operators associated with the oscillation state and $\sigma^+ = |e\rangle\langle g|$. Each of two values $N\hat{\sigma}^z = \pm N$, where $\sigma^z = |e\rangle\langle e| - |g\rangle\langle g|$, corresponds to one of the two stationary Josephson states. The quantity G is determined by the specific mechanism inducing the interaction. If the interaction is realized by applying laser beams, it follows that in the rotating frame, $G = \Omega$ and $\eta = q(2MN\omega_0)^{-1/2}$ is the Lamb–Dicke parameter caused, for example, by the center-of-mass motion of N atoms in a trap with the characteristic frequency ω_0.

We consider the classical states of motion of the condensate. These states can be described in terms of the complex amplitudes a^*, $a = n^{1/2}e^{\pm i\varphi_1}$, where

$$n = \langle a|\hat{a}^+\hat{a}|a\rangle = |a|^2$$

is the average number of quanta in the coherent state $|a\rangle$. The variables n and φ_1 are canonical. By the classical state of motion, we mean that its number of quanta is very large, $n \gg 1$. It is convenient to specify the relation between the amplitudes a, a^* and operators $\hat{a}$, $\hat{a}^+$ as $a = N^{-1/2}\hat{a}$. The commutator of a and a^* is then equal to zero with macroscopic accuracy,

$$[a, a^*] = 1/N \longrightarrow 0.$$

The Hamiltonian of the motion states can be written as $N\epsilon(n)$.

The Hamiltonian of the interacting stationary Josephson and oscillation degrees of freedom therefore becomes

$$\hat{H} = N\epsilon(n) + N\frac{\varepsilon}{2}\sigma^z + H_{\text{int}}, \quad \varepsilon \equiv \frac{2E_J}{N}. \tag{3}$$

We are interested in the situation where the resonance condition

$$k\frac{d\epsilon}{dn} = \varepsilon \tag{4}$$

is satisfied with an integer k. In this case, the Hamiltonian H_{int} can be represented as

$$\begin{aligned} H_{\text{int}} &= N[g_k\sigma^+ a^k + g_k^* a^{*k}\sigma^-], \\ g_k &= G\left(\frac{i^k}{k!}\right)\left(\frac{q^2}{2M\omega_0}\right)^{k/2}. \end{aligned} \tag{5}$$

Dynamical solutions generated by the Hamiltonian of type (3), (5) have been studied in numerous papers both for a single trapped atom, $N = 1$ (see, e.g., [15, 17], and references therein), and for N atoms in a trap [16]. In the present paper, we focus on the equilibrium properties of the system described by Hamiltonian (3), (5).

We note that in order to determine the partition function $Z(N; T)$ in what follows, we must have a macroscopic two-level system. The phase transitions within the Dicke model, which corresponds to the case where $k = 1$ in Eqs. (3) and (5), have been discussed in the context of superradiance [18] and, recently, for exciton condensation [19].

Substituting Hamiltonian (3), (5) in the expression $Z(N, T) = \text{Sp}e^{-\beta\hat{H}}$ for the partition function, we obtain in the case of the classical oscillation degrees of freedom

$$\begin{aligned} Z(N; T) &= \int_0^\infty dn e^{-\beta N\epsilon(n)}(\text{Sp}e^{-\beta\hat{H}_\sigma})^N, \\ \hat{H}_\sigma &= \frac{\varepsilon}{2}\sigma^z + [g_k\sigma^+ a^k + g_k^* a^{+k}\sigma^-], \end{aligned} \tag{6}$$

where $\beta = 1/T$; T is the temperature. Using the eigenenergies

$$E_{1,2} = \pm\frac{\varepsilon}{2}\left(1 + \frac{4|g_k|^2 n^k}{\varepsilon^2}\right)^{1/2}$$

of Hamiltonian $\hat{H}_\sigma$, we arrive at the following expres-

sion for the partition function:

$$Z(N;\,T) = \int_0^\infty dn e^{-\beta N\epsilon(n)}(e^{-\beta E_1} + e^{-\beta E_2})^N$$
$$= \int_0^\infty dn \exp\left\{N\left(-\beta\varepsilon(n) + \ln\left[2\cosh\frac{\beta\varepsilon}{2}\left(1 + \frac{4|g_k|^2 n^k}{\varepsilon^2}\right)^{1/2}\right]\right)\right\} \equiv \int_0^\infty dn e^{-NF(n)}. \quad (7)$$

The partition function in Eq. (7) has a maximum at the value $|a|_m^2 \equiv \bar{n}_m$, realizing a minimum of the function $F(n)$. That is, the value $\bar{n}_m(T)$ is determined by the equation

$$\left(\frac{d\epsilon}{dn}\right)_{n=\bar{n}_m} = \frac{|g_k|^2 k\bar{n}_m^{k-1}}{\varepsilon}\left(1 + \frac{4|g_k|^2\bar{n}_m^k}{\varepsilon^2}\right)^{-1/2} \times \tanh\left[\beta\frac{\varepsilon}{2}\left(1 + \frac{4|g_k|^2\bar{n}_m^k}{\varepsilon^2}\right)^{1/2}\right] \quad (8)$$

and is the number of oscillation quanta per atom of the two-level system. This quantity plays the part of the order parameter of the system for $T < T_c$, where T_c is the transition temperature to the state with $\bar{n}_m \neq 0$.

The new splitting $E_1 - E_2 = \Delta\varepsilon$ of the Josephson levels depends on the temperature and is defined by the expression

$$\Delta\varepsilon(T) = \varepsilon\left(1 + \frac{4|g_k|^2\bar{n}_m^k}{\varepsilon^2}\right)^{1/2}. \quad (9)$$

At $T = 0$, Eq. (8) becomes

$$\left(1 + \frac{4|g_k|^2\bar{n}_m^k}{\varepsilon^2}\right) = \left(\frac{k^2|g_k|^2}{\varepsilon^2}\right)^2 \bar{n}_m^{2(k-1)}, \quad (10)$$

where we used

$$\frac{d\epsilon}{dn} = \frac{\varepsilon}{k}.$$

For a simple resonance with $k = 1$, the solution for $\bar{n}_m(T = 0)$ is given by

$$\bar{n}_m = \left(\frac{\varepsilon}{2G(ql_0)}\right)^2\left[\left(\frac{G^2(ql_0)^2}{\varepsilon^2}\right)^2 - 1\right], \quad l_0^2 = \frac{1}{2M\omega_0}. \quad (11)$$

We used $g_1 = G(ql_0)$ in accordance with Eq. (5). We note that

$$\left(\frac{d^2F(n)}{dn^2}\right)_{n=\bar{n}_m} > 0.$$

A solution for $\bar{n}_m$ exists when the interaction is sufficiently strong, that is,

$$\left(\frac{G}{\varepsilon}\right)^2(ql_0)^2 > 1. \quad (12)$$

This relation implies that the system in question undergoes a quantum phase transition at the critical value $G_c = \varepsilon(ql_0)$ of the interaction strength.

At $G(ql_0) \gg \varepsilon$, the tunneling splitting $\Delta\varepsilon(T = 0)$ and thermodynamic energy E_m of the system described by Hamiltonian (3), (5) are determined by the expressions

$$\Delta\varepsilon \approx \frac{G^2(ql_0)^2}{\varepsilon} \gg \varepsilon, \quad E_m \approx -N\frac{G^2(ql_0)^2}{4\varepsilon}. \quad (13)$$

Therefore, first, the splitting caused by the interaction is much greater than the "unperturbed" splitting ε. Second, the bound state realizes the absolute minimum of the thermodynamic energy within the range $g_1 \gg \varepsilon$.

It is worth noting that the consideration proposed above applies in the case when $qr_0 \ll 1$, where r_0 is the (typical) mean interatom spacing. The l_0 value is the characteristic size of the condensate. As a result, we arrive at the relation

$$ql_0 \gg 1.$$

Transition to the state with $\bar{n}_m \neq 0$ is second-order at $k = 1$. Imposing $\bar{n}_m \longrightarrow 0$ at the transition temperature T_c, we obtain from (8) at $k = 1$ and $|g_1| \gg \varepsilon$

$$T_c = \frac{|g_1|^2}{2\varepsilon} = \frac{G^2(ql_0)^2}{2\varepsilon} = \frac{\Delta\varepsilon(T = 0)}{2}. \quad (14)$$

It is interesting to discuss the solutions of Eq. (10) for multiple resonances with $k > 1$. As can be easily seen, the character of the ground state changes drastically at $k \neq 1$. Already at $k = 2$, solutions of Eq. (10) with $\bar{n}_m \gg 1$ are absent.

At $k \gg 1$, Eq. (10) implies

$$\bar{n}_m^k \approx \frac{2\varepsilon^2}{\pi k^2|g_k|^2}, \quad \Delta\varepsilon \approx \varepsilon\left(1 + \frac{4}{\pi k^2}\right). \quad (15)$$

We see that $\bar{n}_m \sim 1$ and $\Delta\varepsilon \approx \varepsilon$ at $k \gg 1$, and therefore, the effect of the interaction is negligible, unlike the solution for a simple resonance with $k = 1$ obtained above.

For a one-dimensional condensate, the appearance of a state with $\bar{n}_m \neq 0$ corresponds to the center-of-mass oscillation of N atoms with a displaced zero point. The

"capture" of oscillation quanta by the macroscopic two-level system can therefore be considered a mechanism of the formation of the equilibrium coherent state of motion.

In the general case, the nonzero value for the existence of $\bar{n}_m$ leads to an equilibrium distortion of the shape of the condensate. The distortion is self-consistently coupled to the tunneling splitting. In particular, the stationary Josephson states can be detected by observing a change in the condensate shape, and vice versa.

A remark is in order. The quantities $d\epsilon/dn \equiv \omega(n)$ specify the level separations of the oscillation spectrum. They are independent of n for harmonic oscillations, where $\omega(n) = \omega_0$ and for the states in the region of a "dense" (semiclassical) oscillation spectrum. In this region, the level separations $\omega(n)$ are small compared with ω_0 and, in addition, are a slowly varying function of n, $\omega(n) \approx \omega \ll \omega_0$.

3. QUANTUM SPECTRUM OF THE PARTICLE NUMBER

The quantization of Hamiltonian (1) produces the particle number spectrum in the Josephson potential $E_J\cos\phi$. As shown in what follows, any Josephson state can be realized by means of a given initial disbalance of the particle number. It is therefore interesting to obtain the complete spectrum generated by Hamiltonian (1) and to represent it as a function of the initial disbalance.

The Schrödinger equation for Hamiltonian (1) is derived by the quantization rule

$$(\Delta N) \longrightarrow -i\partial/\partial\phi.$$

As a result, we obtain the Mathieu equation

$$\left[-E_C\frac{d^2}{d\phi^2} - 2E_J\cos\phi_r\right]\Psi = \varepsilon\Psi, \qquad \varepsilon \equiv H_J - E_0. \tag{16}$$

For $\varepsilon > 2E_J$, this equation has a continuum spectrum. The states of this spectrum correspond to classical states with an unlimited phase change, $-\infty < \phi < +\infty$; they are called the self-trapping states in [9, 10]. In the region $-2E_J < \varepsilon < 2E_J$, Eq. (16) has a discrete spectrum. It corresponds to the finite-motion region of Hamiltonian (1), where the relative phase changes within

$$-\arccos(\varepsilon/2E_J) < \phi < \arccos(\varepsilon/2E_J)$$

for each ε. In the Josephson regime [3] at $E_J \gg E_C$, the number of levels in a well is large and the discrete spectrum is determined by the Bohr–Sommerfeld formula

$$\begin{aligned}\nu(\varepsilon_\nu) &= \oint\frac{d\phi_r}{\pi}\Delta N(\phi_r;\,\varepsilon_\nu)\\ &= \oint\frac{d\phi_r}{\pi}\left[\frac{1}{E_C}(\varepsilon_\nu + 2E_J\cos\phi)\right]^{1/2}\\ &= \nu_c[E(\kappa) - (1-\kappa^2)K(\kappa)],\end{aligned} \tag{17}$$

where

$$\nu_c = \frac{8}{\pi}\left(\frac{E_J}{E_C}\right)^{1/2}, \quad \kappa^2 = \frac{\varepsilon_\nu + 2E_J}{4E_J}, \quad \nu_c \gg 1, \tag{18}$$

and $K(\kappa)$ and $E(\kappa)$ are the complete elliptical integrals of the first and second kind. The density of states $\rho_d(\varepsilon_\nu)$ follows from Eq. (17) and is equal to

$$\rho_d(\varepsilon_\nu) = \frac{1}{2\pi}\frac{d\nu}{d\varepsilon_\nu} = \frac{K(\kappa)}{\pi^2\omega_m}. \tag{19}$$

The level separations in (17) are given by

$$\omega(\varepsilon_\nu) = \frac{d\varepsilon}{d\nu} = \frac{\pi\omega_m}{2K(\kappa)}, \quad \omega_m = 2(E_CE_J)^{1/2}. \tag{20}$$

At $(E_J/E_C)^{1/2} \gg 1$, we have the relation $\omega_m \ll 2E_J$. The quantity ω_m determines the maximum splitting of the levels in the Josephson well. In what follows, the states with $\nu < \nu_c$ are called libration states.

In the region $\nu \geq \nu_c$, the $\nu(\varepsilon)$ dependence and density of states are determined by

$$\begin{aligned}\nu(\varepsilon) &= \frac{4}{\pi^2}\left(\frac{E_J}{E_C}\right)^{1/2}\kappa E(\kappa^{-1}),\\ \rho_c(\varepsilon) &= \frac{1}{2\pi^2}\frac{\kappa^{-1}K(\kappa^{-1})}{\omega_m}.\end{aligned} \tag{21}$$

States (21) with $\nu > \nu_c$ are called self-trapping states.

Equations (17) and (21) imply that

$$\begin{aligned}\varepsilon(\nu) &\approx -2E_J + \omega_m\nu, \quad 1 \ll \nu \ll \nu_c,\\ \varepsilon(\nu) &\approx 4\pi^2E_C\nu^2, \quad \nu \gg \nu_c.\end{aligned} \tag{22}$$

At the same time, it is easy to show that

$$\frac{d^2\varepsilon(\nu)}{d\nu^2} < 0 \quad \text{for} \quad \nu < \nu_c$$

and

$$\frac{d^2\varepsilon(\nu)}{d\nu^2} > 0 \quad \text{for} \quad \nu > \nu_c.$$

At $\nu = \nu_c$, curve $\varepsilon(\nu)$ has an inflection point.

Because the energy is conserved, the state with a given value of ν can be realized by defining the initial values of $(\Delta N)_0$ and $\phi(0)$ as

$$\varepsilon(\nu) = E_C(\Delta N)_0^2 - 2E_J\cos\phi(0).$$

Supposing that $\phi(0) = 0$ we obtain the following relation between ν and $(\Delta N)_0$:

$$\varepsilon(\nu) = -2E_J + E_C(\Delta N)_0^2. \quad (23)$$

Using Eqs. (22) and (23), we arrive at the expressions

$$\nu = \frac{1}{2\pi}|(\Delta N)_0|$$

for the self-trapping (sf) states and

$$\nu = \left(\frac{E_C}{\omega_m}\right)(\Delta N)_0^2 \quad (24)$$

for the "libration" (l) states.

Combining (19) and (21), we obtain the dependence

$$\rho_{d,c}(\varepsilon) \propto \omega_m^{-1}\ln\left|1 - \frac{\varepsilon}{2E_J}\right|^{-1}, \quad \varepsilon \longrightarrow 2E_J^{\pm}. \quad (25)$$

Therefore, a new logarithmic singularity appears at the boundary separating the libration and self-trapping spectra.

4. INTERACTION OF THE EXCITED JOSEPHSON STATES AND ANHARMONIC CONDENSATE OSCILLATIONS

In this and the next sections, we show that the spectrum of the system can change drastically due to the interaction between excited Josephson states (22)–(24) with sufficiently large values of ν and oscillations of the condensate.

As in Section 2, two mechanisms can be proposed for the realization of the interaction. First, the interaction can be produced by a dependence of E_C in Eq. (23) on the atom displacements. The latter are generated by the condensate oscillation. Second, the interaction can be realized by applying a two-photon traveling-wave laser pulse with a Rabi frequency of Ω. The pulse creates condensates with different particle numbers and induces interaction of atom displacements with the excited Josephson states corresponding to the particle number disbalance created by the pulse. The general description proposed in what follows is independent of the specific mechanism producing the interaction.

The states of motion of the condensate are classical. These states are defined in Section 2. For semiclassical Josephson states with $\nu \gg 1$, the c_ν, c_ν^* amplitudes can be written as

$$c_\nu = \nu^{1/2}e^{i\varphi_2}.$$

It is convenient to rewrite $\varepsilon(\nu)$ and c_ν in terms of the variable

$$x = \frac{|(\Delta N)_0|}{N^{1/2}} \gg 1.$$

Combining this inequality with the requirement that

$$x = \frac{|(\Delta N)_0|}{N} \ll 1,$$

we arrive at the conditions for the x values,

$$1 \ll x \ll N^{1/2}. \quad (26)$$

Using Eqs. (22) and (23), we find that

$$\nu = \nu(x),$$
$$\varepsilon(\nu) = N(-E_J/N + E_C x^2).$$

In the general case, the $\nu(x)$ dependence is implicit. It is determined by Eqs. (17), (21), and (23). However, in the particular cases of the libration ($\varepsilon(\nu) \ll E_J$) and self-trapping ($\varepsilon(\nu) \gg E_J$) states, the relationships between ν and $(\Delta N)_0$ can be represented in a simple form, as can be seen from (24). Using Eqs. (23) and (24), we arrive at the following expressions:

$$H_0 \equiv N\varepsilon_0(n, x) = N\left[-\frac{E_J}{N} + \epsilon(n) + E_C x^2\right], \quad (27)$$

$$c_\nu^{(sf)} = N^{1/4}x^{1/2}e^{i\varphi_2} \equiv N^{1/4}c_x^{(sf)},$$
$$c_\nu^{(l)} = N\left(\frac{E_C}{\omega_m}\right)^{1/2}xe^{i\varphi_2} \equiv N\left(\frac{E_C}{\omega_m}\right)^{1/2}c_x^{(l)}. \quad (28)$$

For any mechanism producing interaction between two subsystems, the interaction Hamiltonian can be written in the form of a multiple Fourier series in φ_1, φ_2,

$$H_{\text{int}} = N\sum_{k_1k_2}[g_{k_1k_2}(N)a^{*k_1}c_\nu^{k_2} + \text{c.c.}]$$
$$= N\sum_{k_1,k_2}[g_{k_1k_2}^{(sf,l)}(N)n^{k_1/2}x^{\alpha k_2}\exp(i(k_1\varphi_1 - k_2\varphi_2)) + \text{c.c.}], \quad (29)$$

where $\alpha_{sf} = 1/2$, $\alpha_l = 1$, and k_1 and k_2 are integers. For simplicity, we disregard the phase-independent interaction. Using Eq. (28), we obtain

$$g^{(sf)}_{k_1k_2}(N) = gN^{-1+k_2/4},$$
$$g^{(l)}_{k_1k_2}(N) = gN^{-1+k_2}\left(\frac{E_C}{\omega_m}\right)^{k_2/2}. \quad (30)$$

The constant g is determined by the specific mechanism producing the interaction.

We now assume that the term with the phase

$$\phi^r_k = k_{1r}\varphi_1 - k_{2r}\varphi_2,$$

which varies anomalously slowly with time, can be dropped in sum (29). This can be done under two conditions. The first condition is

$$k_{1r}\left(\frac{d\epsilon(n)}{dn}\right) = k_{2r}\left(\frac{d\varepsilon_J(x)}{dx}\right), \quad (31)$$

or equivalently,

$$x_m = \frac{k_{1r}\epsilon'_n}{2k_{2r}E_C}, \quad \epsilon'_n \equiv \left(\frac{d\epsilon(n)}{dn}\right).$$

The second condition is

$$\left(\frac{d^2H_0}{dx^2}\right)_{x=x_m}(\Delta x)_{\max} \gg \left(\frac{dH_{\rm int}}{dx}\right)_{x=x_m}. \quad (32)$$

In writing this equation, we took into account that H_0 and $H_{\rm int}$ are functions of a single dynamic variable, e.g., x. The quantity ϵ'_n defines the separation of levels of the oscillation spectrum.

As is shown in what follows, condition (31) is equivalent to the condition of the minimum of the function $\varepsilon_0(n, x)$ with respect to x. When the minimum exists, it can provide the leading contribution to the thermodynamic functions.

In addition, condition (31) implies that the phase ϕ^r_k is an approximate integral of motion if the dependence on x near x_m is ignored,

$$\frac{d\phi^r_k}{dt} \approx k_{1r}\frac{\partial H_0}{\partial n} - k_{2r}\frac{\partial H_0}{\partial x} \approx 0.$$

Inequality (32) implies that the width of the near-minimum region is large at the characteristic interaction variation scale. From Eqs. (31) and (32), we can see that time changing the ϕ^r_k phase is proportional to $(d^2H_0/dx^2)_m\Delta x$, where Δx is the variation of x near the x_m value. The maximum value $(\Delta x)_{\max}$ specifies the width of the near-minimum region such that

$$\frac{d\phi^r_k}{dt} \sim \Delta x.$$

The estimate for $(\Delta x)_{\max}$ is given in what follows. Thus, the leading term in sum (29) is given by

$$H^{(r)}_{\rm int} = Ng^{(sf,\,l)}_k(N; n, x)\cos\phi_k,$$
$$\phi_k = k_{1r}\varphi_1 - k_{2r}\varphi_2. \quad (33)$$

All the remaining terms in this sum are rapidly oscillating perturbations and are disregarded in this work. Here and below, the index k in g_k and ϕ_k denotes the set k_{1r}, k_{2r}.

It can be easily shown that in addition to the energy

$$H = N[\varepsilon_0(n, x) + g_k(N; n, x)\cos\phi_k],$$

the system in question has the integral of motion

$$n_0 = \frac{n}{k_{1r}} + \frac{x}{k_{2r}}, \quad \frac{dn_0}{dt} = 0.$$

Owing to this, condition (31) is equivalent to that of the minimum of $\varepsilon_0(n_0, x)$ at x_m for a given value of n_0, as mentioned above.

Using Eqs. (31)–(33), it is straightforward to write the Hamiltonian

$$H_m = H_0 + H^{(r)}_{\rm int}$$

near the minimum to the first nonvanishing order in Δx,

$$H_m = N\left[\varepsilon_0(n_m; x_m) + \left(\frac{d^2\varepsilon_0}{dx^2}\right)_m(\Delta x)^2 - g_{km}\cos\phi_k\right], \quad (34)$$

where

$$n_m = n_0 - \frac{k_{1r}x_m}{k_{2r}},$$
$$\left(\frac{d^2\varepsilon_0}{dx^2}\right)_m = 2E_c, \quad g_{km} = g^{(sf,\,l)}_k(N; n_m, x_m).$$

The terms with the derivatives of $H_{\rm int}$ are absent from Eq. (34) because of condition (32).

Making use of the fact that

$$E_C \sim \omega_0(a/a_0)^{2/5}N^{-3/5}$$

in the Thomas–Fermi approximation [20] (where a and a_0 are the scattering and oscillator lengths, respectively), we can represent the range $1 \ll x_m \ll N^{1/2}$ as

$$\frac{1}{N}\left(\frac{Na}{a_0}\right)^{2/5} \ll \frac{k_{1r}}{k_{2r}}\left(\frac{\epsilon'_n}{\omega_0}\right) \ll \frac{1}{N^{1/2}}\left(\frac{Na}{a_0}\right)^{2/5}. \quad (35)$$

As is known [20], the relation $(Na/a_0) \gg 1$ is valid in the Thomas–Fermi approximation, but $N^{-1/2}(Na/a_0)^{2/5} \ll 1$, and therefore, condition (35) (or equivalently, condi-

tion (31)) specifies the region of the dense oscillation spectrum, where $\epsilon_n' \ll \omega_0$. Here and in what follows, we suppose that $k_{1r} = k_{2r} = 1$ for simplicity.

From Eq. (34), the value of $(\Delta x)_{\max}$ can be estimated as

$$(\Delta x)_{\max} \sim \left(\frac{g_{km}}{E_C}\right)^{1/2}.$$

Hence, condition (32) becomes

$$\frac{|x_m - n_m|}{x_m n_m} \ll \left(\frac{E_C}{g_{km}(N)}\right)^{1/2}. \quad (36)$$

In what follows (see Eq. (49)), we show that the relation (36) is satisfied with macroscopic accuracy.

5. THE GROUND STATE

At a fixed value of n_0, the leading contribution to the partition function comes from the neighborhood of the minimum at $x = x_m$. The expression for $Z(n_0; x_m; T)$ is

$$\begin{aligned} Z(n_0; x_m; T) &= \text{const}\int_{-\infty}^{\infty} d\Delta x \int_{-\pi}^{\pi} d\phi_k \\ &\times \exp(-\beta H_m(n_0, \Delta x, \phi_k)) = \frac{\text{const}}{(\beta N E_C)^{1/2}} \\ &\times \exp[-\beta N \varepsilon_0(N; n_m; x_m) + \ln I_0(\beta N g_{km})], \end{aligned} \quad (37)$$

where $I_0(x)$ is the modified Bessel function. Equation (37) implies that the free energy of the system is given by

$$\begin{aligned} F &= N\varepsilon_0(n_m; x_m) + \frac{1}{2}T\ln(\beta N E_C) \\ &- T\ln I_0(\beta N g_{km}). \end{aligned} \quad (38)$$

Using Eq. (38), we obtain the equation for the $\bar{n}_m$ value realizing the minimum of the free energy,

$$\left(\frac{d\varepsilon_{0m}}{dn_m}\right)_{n_m = \bar{n}_m} = \left(\frac{dg_{km}}{dn_m}\right)_{n_m = \bar{n}_m} \frac{I_1(\beta N g_{km})}{I_0(\beta N g_{km})}, \quad (39)$$

where $I_1(x) = I_0'(x)$. In addition to $\bar{n}_m$, the thermodynamic average of $\cos\phi_k$ can be determined from Eqs. (37) or (38). This average is equal to

$$\begin{aligned} \langle \cos\phi_k \rangle_T &= -\frac{\partial \ln Z}{\partial(\beta N g_{km})} \\ &= \frac{\partial F}{T\partial(\beta N g_{km})} = \frac{I_1(\beta N g_{km})}{I_0(\beta N g_{km})}. \end{aligned} \quad (40)$$

The order parameters $\bar{n}_m$ and $\langle \cos\phi_k \rangle_T$ describe the new coherent state. There is a bound state of the $\bar{n}_m$ oscillation quanta and the Josephson state generated by the initial disbalance of the particle number that corresponds to the x_m value. In addition, this state has an equilibrium phase coherence factor of $\langle \cos\phi_k \rangle_T$. The $\bar{n}_m \neq 0$ value provides the equilibrium distortion of the condensate shape. The above equations imply that the shape distortion is self-consistently coupled to the x_m value defining the equilibrium initial disbalance of the particle number.

At $T = 0$, the $\bar{n}_m$ value realizes the minimum of the thermodynamic energy

$$E_m = N[\varepsilon_0(N; n_m, x_m) - g_{km}(N; n_m, x_m)]. \quad (41)$$

To determine $\bar{n}_m(T = 0)$, it is suitable to use the following consideration. It is well known that the level separations ϵ_n' are slowly varying functions of n within the a dense (semiclassical) spectrum. We can therefore suppose that $\epsilon_n' \approx \text{const} \equiv \omega_b \ll \omega_0$. Under this assumption, the x_m value is independent of n_m and the $\bar{n}_m$ is equal to

$$\bar{n}_m^{1/2} = \frac{g^{(sf,\,l)}(N)x_m^{\alpha}}{2\omega_b}, \quad \alpha_{sf,\,l} = \frac{1}{2}, 1. \quad (42)$$

(It is worth noting that $(\partial^2 E_m/\partial n_m^2) > 0$.) Here and in what follows, the notation $g_{k=1}^{(sf,\,l)}(N) \equiv g^{(sf,\,l)}(N)$ is used. Taking into account Eq. (42), we readily obtain

$$\begin{aligned} \bar{g}_{1m}^{(sf)} &= \frac{g^{(sf)2}(N)x_m}{2\omega_b}, \quad \bar{g}_{1m}^{(l)} = \frac{g^{(l)2}(N)x_m^2}{2\omega_b}, \\ \frac{E_m^{(sf)}}{N} &= -\frac{E_J}{N} - \frac{g^{(sf)2}(N)}{8E_C}\left(1 - \frac{8E_C^2 x_m^2}{g^{(sf)2}(N)}\right), \\ \frac{E_m^{(l)}}{N} &= -\frac{E_J}{N} - \frac{g^{(l)2}(N)x_m}{16E_C}\left(1 - \frac{16E_C^2 x_m}{g^{(l)2}(N)}\right). \end{aligned} \quad (43)$$

The expressions for the energies imply, first, that we obtain the minimum in the region of a sufficiently dense oscillation spectrum that satisfies condition (35). The minimum corresponds to the formation of a bound state for the $\bar{n}_m$, x_m values. Second, as can be seen from Eq. (43), the absolute minimum of E_m can be realized within the ranges

$$1 \ll x_m < \frac{g^{(sf)}(N)}{8E_C}, \quad 1 \ll x_m < \frac{g^{(l)2}(N)}{16E_C^2}. \quad (44)$$

These conditions are satisfied when the interaction matrix elements $g^{(sf,\,l)}(N)$ are sufficient. We estimate the condensate parameters required for the existence of the absolute minimum. In the Thomas–Fermi approxima-

tion, inequality $g^{(sf)}(N) \gg E_C$ is valid if the total particle number is not very large, namely,

$$N \ll \frac{g}{\omega_0}\left(\frac{a_0}{a}\right)^{8/3}. \quad (45)$$

In turn, the relation $g^{(l)2}(N) \gg E_C^2$ is valid within the range

$$N^{0.1}\frac{g^2}{\Omega^{1/2}\omega_0^{3/2}}\left(\frac{a_0}{a}\right)^{0.6} \gg 1, \quad (46)$$

where we use $E_J = \Omega N$. Condition (46) is satisfied for all admissible parameters if

$$\frac{g^2}{\Omega^{1/2}\omega_0^{3/2}} \sim 1.$$

We emphasize that the inequalities in the right-hand side of Eq. (44) are much stronger than the condition $x_m \ll N^{1/2}$.

The transition to the state with $\bar{n}_m \neq 0$ and $\langle \cos\phi_k \rangle_T \neq 0$ is second-order. Requiring that $\bar{n}_m \longrightarrow 0$ at the transition temperature, we obtain from Eq. (39)

$$T_c^{(sf,l)} = \frac{g^{(sf,l)2}(N)x_m^{\alpha}}{\omega_b}N. \quad (47)$$

The dependences of transition temperatures (47) on the total particle number are given by

$$T_c^{(sf)} \sim N^{-0.2}, \quad T_c^{(l)} \sim N^{0.8}. \quad (48)$$

The transition temperature $T_c^{(sf)}$ therefore has macroscopic smallness in comparison with the $T_c^{(l)}$ temperature. Along with conditions (45) and (46), this fact implies that the libration Josephson state forms a bound state with condensate oscillation rather than a self-trapping state.

6. CONCLUSIONS

We have found that the interaction between the Josephson and oscillation states results in a new coherent ground state of the double-well trapped condensate. There is a bound state of the $\bar{n}_m$ oscillation quanta and the Josephson states. The latter are either two stationary states forming the macroscopic two-level subsystem or the excited Josephson states generated by the definite initial disbalance of the particle number. Both the x_m value defining the disbalance and the tunneling splitting $N\varepsilon$ of the stationary Josephson states are self-consistently coupled to the number of the oscillation quanta entering the bound state.

We emphasize that the bound states arise near the extremum points of the "unperturbed" spectrum of the system. For a two-level Josephson subsystem, there is resonance between the "initial" tunneling splitting ε and the oscillation frequency.

For the excited Josephson states in Sections 3 and 4, a new coherent state forms in the neighborhood of the minimum of the energy $\varepsilon_0(n_0; x)$ if conditions (31), (32) are satisfied.

The resonance and the minimum conditions specify the type of the interaction between the Josephson and oscillation states.

In addition, minimum condition (31) imposes substantial restrictions on the spectrum of the oscillation states that can effectively interact with the Josephson degrees of freedom. This condition implies that the excited Josephson states can interact only with the oscillation states having a sufficiently large density of states (such that inequalities (35) are satisfied). For instance, in an asymmetric double-well potential, the energy $\varepsilon(n)$ of its classical oscillation states has three branches. Two branches have energies of $\varepsilon_{1,2}(n) \leq V_b$, where V_b is the barrier height. There is a maximum at $\varepsilon_{1,2}(n_{\max}) = V_b$. The third branch has an energy of $\varepsilon_3(n) \geq V_b$ and a minimum at an energy of $\varepsilon_3(n_{\min}) = V_b$. Therefore, the region of the dense oscillation spectrum exists in the neighborhood of the top of the barrier.

A similar situation can occur for the stationary Josephson states in Section 2. In the Thomas–Fermi approximation, resonance condition (4) can be satisfied within the range of the dense spectrum close to the top of the barrier.

We emphasize that in both cases, the bound states of the highly excited oscillation state and Josephson degrees of freedom are formed. These states can realize the absolute minimum of the thermodynamic energy, as shown in Eqs. (13) and (43).

As shown in Section 4, condition (32) is realized within range (36). Substituting the expressions for $\bar{n}_m$ and $\bar{g}_{1m}$ obtained above in (36), we find that this equation becomes

$$|x_m - n_m|^{(sf,l)} \ll \left(\frac{g_{1m}^{(sf,l)}(N)}{E_C}\right)^{1/2}, \quad (49)$$

and is therefore satisfied with macroscopic accuracy.

The bound states exist at a sufficiently strong interaction determined by conditions (12) and (44). These conditions are experimentally controlled by means of either the interaction matrix elements G, g or the particle number N (see Eqs. (12), (45), and (46)).

For the stationary Josephson states in Section 2, the mechanism proposed above provides a substantial increase in the tunneling splitting that turns out to be dependent on temperature. It is defined by (11). For the excited Josephson states interacting with the anharmonic oscillation, two order parameters describe the new ground state. These are the number $\bar{n}_m$ of the oscil-

lation quanta entering the bound state and the phase coherence factor in Eq. (40).

The second-order phase transitions to the states with $\bar{n}_m \neq 0$ can be observed at $T = T_c$, where the transition temperatures T_c are defined by Eqs. (14), (47), and (48).

We finally note that the appearance of bound states generates an distortion in the equilibrium of the condensate shape specified by the $\bar{n}_m$ values in Eqs. (11) and (42). This mechanism can provide experimental detection of the Josephson states. The latter can be observed by changing the condensate shape.

ACKNOWLEDGMENTS

I am grateful to Yu.M. Kagan and L.A. Maksimov for helpful discussions.

This work was supported by the Russian Foundation for Basic Research (project no. 01-02-16513).

REFERENCES

1. C. J. Myatt, E. A. Burt, R. W. Christ, *et al.*, Phys. Rev. Lett. **78**, 586 (1997).
2. M. R. Andrews, C. G. Townsend, H.-J. Meisner, *et al.*, Science **275**, 637 (1997).
3. A. J. Leggett, Rev. Mod. Phys. **73**, 307 (2001).
4. A. Rohrl, M. Naraschewski, A. Schenzle, and H. Wallis, Phys. Rev. Lett. **78**, 4143 (1997).
5. D. S. Hall, M. R. Mattews, C. E. Wiemann, and E. A. Cornell, Phys. Rev. Lett. **81**, 1543 (1998).
6. J. Williams, R. Walser, J. Cooper, *et al.*, Phys. Rev. A **59**, R31 (1999).
7. I. Zapata, F. Sols, and A. J. Leggett, Phys. Rev. A **57**, R28 (1998).
8. G. J. Millburn, J. Corney, E. M. Wright, and D. F. Walls, Phys. Rev. A **55**, 4318 (1997).
9. A. Smerzi and S. Fantoni, Phys. Rev. Lett. **78**, 3589 (1997).
10. S. Raghavan, A. Smerzi, S. Fantoni, and V. B. Shenoy, Phys. Rev. A **59**, 620 (1999).
11. M. Lewenstein and L. You, Phys. Rev. Lett. **77**, 3489 (1996).
12. A. Imamoglu, M. Lewenstein, and L. You, Phys. Rev. Lett. **78**, 2511 (1997).
13. F. Meir and W. Zweger, Phys. Rev. A **64**, 033610 (2001).
14. L. P. Pitaevskii and S. Stringari, Phys. Rev. Lett. **87**, 180402 (2001).
15. D. M. Meechof, C. Monroe, B. E. King, *et al.*, Phys. Rev. Lett. **76**, 1796 (1996).
16. E. Solano, R. L. de Matos Filho, and N. Zagury, Phys. Rev. Lett. **87**, 060402 (2001).
17. W. Vogel and R. L. de Matos Filho, Phys. Rev. A **52**, 4214 (1995).
18. K. Hepp and E. H. Lieb, Ann. Phys. (N.Y.) **76**, 36 (1973).
19. P. R. Eastham and P. B. Littwood, Phys. Rev. B **64**, 235101 (2001).
20. F. Dalfovo, S. Giorgini, L. P. Pitaevskii, and S. Stringari, Rev. Mod. Phys. **71**, 463 (1999).

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 257–267.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 293–304.

SOLIDS
Electronic Properties

Electronic Structure and Phase Separation in La_xMnO_3 Films ($0.83 \le x \le 1.10$): Optical and Magnetooptical Data

Yu. P. Sukhorukov[a,*], N. N. Loshkareva[a], E. A. Gan'shina[b], E. V. Mostovshchikova[a], I. K. Rodin[b], A. R. Kaul[b], O. Yu. Gorbenko[b], A. A. Bosak[b], A. S. Moskvin[c], and E. V. Zenkov[c]

[a]*Institute of Metal Physics, Ural Division, Russian Academy of Sciences, ul. S. Kovalevskoĭ 18, Yekaterinburg, 620219 Russia*

[b]*Moscow State University, Vorob'evy gory, Moscow, 119899 Russia*

[c]*Ural State University, pr. Lenina 51, Yekaterinburg, 620083 Russia*

**e-mail: suhorukov@imp.uran.ru*

Received June 7, 2002

Abstract—The variations in the electronic structure and the evolution of phase separation as a result of creation of vacancies or excess of lanthanum in lanthanum manganites are studied on the basis of analysis of optical and magnetic properties of La_xMnO_3 epitaxial films ($0.83 \le x \le 1.10$) in the fundamental absorption range. The Kerr effect, the temperature dependences of resistivity, optical density, and magnetoabsorption of light indicate the charge and magnetic phase separation in the films. The fine structure observed in the spectrum is attributed to spectral overlapping of electron transitions with charge transfer and geometrical resonances reflecting an inhomogeneous nanoscopic structure of the films, which strongly depends on stoichiometry and stresses emerging during film deposition. It is shown that, in contrast to bulk polycrystals, the gradient of stresses over the film thickness significantly affects the phase separation in the films. © *2003 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

In studying lanthanum manganites with a colossal magnetoresistance, the main attention is paid to manganites obtained by doping antiferromagnetic insulator $LaMnO_3$ (T_N = 141 K) with divalent ions (Ba, Sr, Ca). The creation of La deficit in $LaMnO_3$, as well as doping with Ca [1], gives rise to spontaneous magnetization and elevates the Curie temperature (T_C) up to 260 K. An analysis of optical properties of polycrystalline lanthanum manganites La_xMnO_3 with a small La deficit in the IR range reveals the existence of an insulator–metal transition in the transmittance of light near T_C despite the semiconductor-type behavior of resistance $\rho(T)$ in the entire temperature range [2]. This proves the existence of inhomogeneous charge states leading to the formation of highly conducting regions (metallic drops) in the insulating matrix, i.e., charge separation of phases. The Kerr effect observed in ferromagnetic regions [3] indicates the magnetic separation of phases in lanthanum manganite. Simultaneous analysis of optical and magnetooptical properties of manganites, in particular, $(La_{0.5}Pr_{0.5})_{0.7}Ca_{0.3}MnO_3$ epitaxial films [4], shows that ferromagnetic regions are associated with magnetic drops, which speaks in favor of the phase separation model [5].

In view of their high absorptance, polycrystalline samples are not suitable for studying optical spectra in the visible range, where optical transitions providing information on the electronic structure are manifested. The most appropriate objects for this purpose are films. An analysis of optical and magnetooptical properties of lanthanum manganite films provides information on the behavior of charge carriers above and below the magnetic ordering temperature and makes it possible to reveal changes in the electronic structure of $LaMnO_3$ upon a deviation from the stoichiometric composition and due to stresses emerging during film deposition in view of epitaxy of the film on the substrate. The presence of charge and magnetic inhomogeneities in the form of individual regions was visualized in $La_{0.67}Ca_{0.33}MnO_3$ films on $LaAlO_3$ (LAO) substrates [6]. The formation of inhomogeneous regions due to structural distortions and/or deviations from the stoichiometric composition must noticeably affect magnetic, electrical, and optical properties of the films. We use here optical methods to study the changes in the electronic structure and the evolution in phase separation as a result of creation of vacancies or excess of La in La_xMnO_3 epitaxial films. The experimental results are discussed qualitatively on the basis of a theoretical model assuming that the main contribution to the formation of the optical response of manganites comes

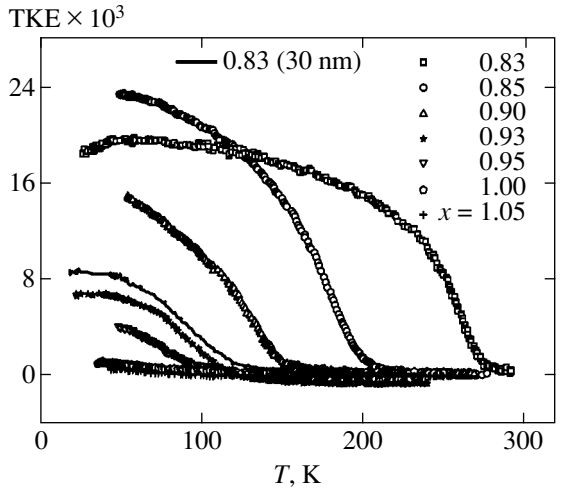


Fig. 1. Temperature dependences of TKE for $H = 0.35$ T and $E = 2.9$ eV for La_xMnO_3 films with different La deficits, obtained during heating.

from transitions with charge transfer. The model takes into account the inhomogeneous nanoscopic electronic structure of the films, which is a consequence of phase separation.

2. SAMPLES AND EXPERIMENTAL CONDITIONS

Epitaxial films of La_xMnO_3 (LXMO) with different La/Mn concentration ratios ($0.83 \le x \le 1.10$) were grown on $LaAlO_3$ (001) substrates. The film thickness was 200 nm; the thickness of a film with $x = 0.83$ was 30 nm. The methods of preparing the films, as well as the conditions of their quality and thickness control, are described in [7]. The optical density spectra of the films in the temperature range 80–295 K were measured in magnetic fields up to 0.8 T in the energy range 0.1–4.0 eV on IKS-21 and KSVU-12 automated spectrometers. The equatorial Kerr effect (TKE) was measured in the energy range 1–4.5 eV in the temperature interval 20–300 K in magnetic fields up to 0.35 T. We measured experimentally the relative change in the intensity of p polarized light reflected from the film, $\delta = [I(H) - I(0)]/I(0)$, where $I(H)$ and $I(0)$ are the intensities of reflected light in the presence and absence of a magnetic field. The TKE measuring technique is described in greater detail in [8]. The resistivity and magnetoresistance of the films were measured by the two-probe method in the temperature range 75–300 K in magnetic fields up to 0.8 T. Silver contacts were fixed by In solder with the help of an ultrasonic solderer.

3. EXPERIMENTAL RESULTS

3.1. Kerr Effect

Figure 1 shows the temperature dependences of the TKE of LXMO films, reflecting the temperature dependence of magnetization. An increase in lanthanum deficit elevates the phase-transition temperature T_C, which attains the value $T_C = 264$ K for $x = 0.83$ (200 nm). In addition, a considerable increase in the TKE is observed for $x < 0.95$. An increase in the La concentration ($x > 1$) is accompanied by a shift in T_C towards lower temperatures and a decrease in the TKE. Small values of TKE for the nominally stoichiometric composition $LaMnO_3$ (LMO) and for films with $x > 1.0$ correspond to the antiferromagnetic ground state of undoped $LaMnO_3$ [9].

The field dependence of the TKE for LMO is linear, the slope decreasing with an excess of La ($x = 1.05$) (see insets to Fig. 2). Such a form of the field dependence of TKE also indicates that the magnetic ordering is predominantly antiferromagnetic, especially for the compound with $x = 1.05$. The relative changes in the slopes of the field dependences of TKE for compositions with $x = 1.0$ and 1.05 may indicate the presence of a canted structure or the presence of a small number of ferromagnetic regions in the antiferromagnetic matrix, which is determined by stoichiometry. In the case of samples with a considerable La deficit ($x < 0.9$), the curves describing the field dependence attain saturation even in weak field (e.g., at $x = 0.85$ for $H \sim 1$ kOe), which is typical of ferromagnets. It should be noted that, in contrast to samples doped with Ca and Pr [3], we did not observe a hysteresis in the temperature dependences of TKE for films with $x > 0.95$, measured under cooling–heating conditions (Fig. 2). For samples with $x \le 0.95$, temperature curves acquire a hysteresis. Such a behavior indicates a considerable magnetic inhomogeneity of these films as well as the presence of ferromagnetic and antiferromagnetic components. A comparison of these temperature dependences of TKE for LXMO films with the temperature dependence of the magnetic susceptibility of bulk polycrystals of the same system [2] suggests that the more gently sloping temperature dependences of TKE of the films are due to peculiarities of the film state and are caused by the stresses emerging during the film growth.

The spectral dependences of the TKE were measured for all the samples at temperatures for which the TKE attains its peak value. The magnetooptical spectra for films with various compositions (Fig. 3), as well as for $LaMnO_{3.11}$ polycrystals [10] (i.e., with cation deficit), are characterized by two negative peaks and one positive peak. The amplitudes of the peaks for the LMO film are an order of magnitude smaller than the amplitudes of bands in $LaMnO_{3.11}$. With increasing La deficit in the LXMO films, the amplitudes of the TKE peaks increase and attain values typical of heavily doped polycrystalline lanthanum manganites for $x = 0.85$. It was established earlier [10] that the magnetooptical activity of manganites with a perovskite structure in the spectra range under investigation is associated with transitions in octahedral complexes of Mn^{3+} and Mn^{4+}

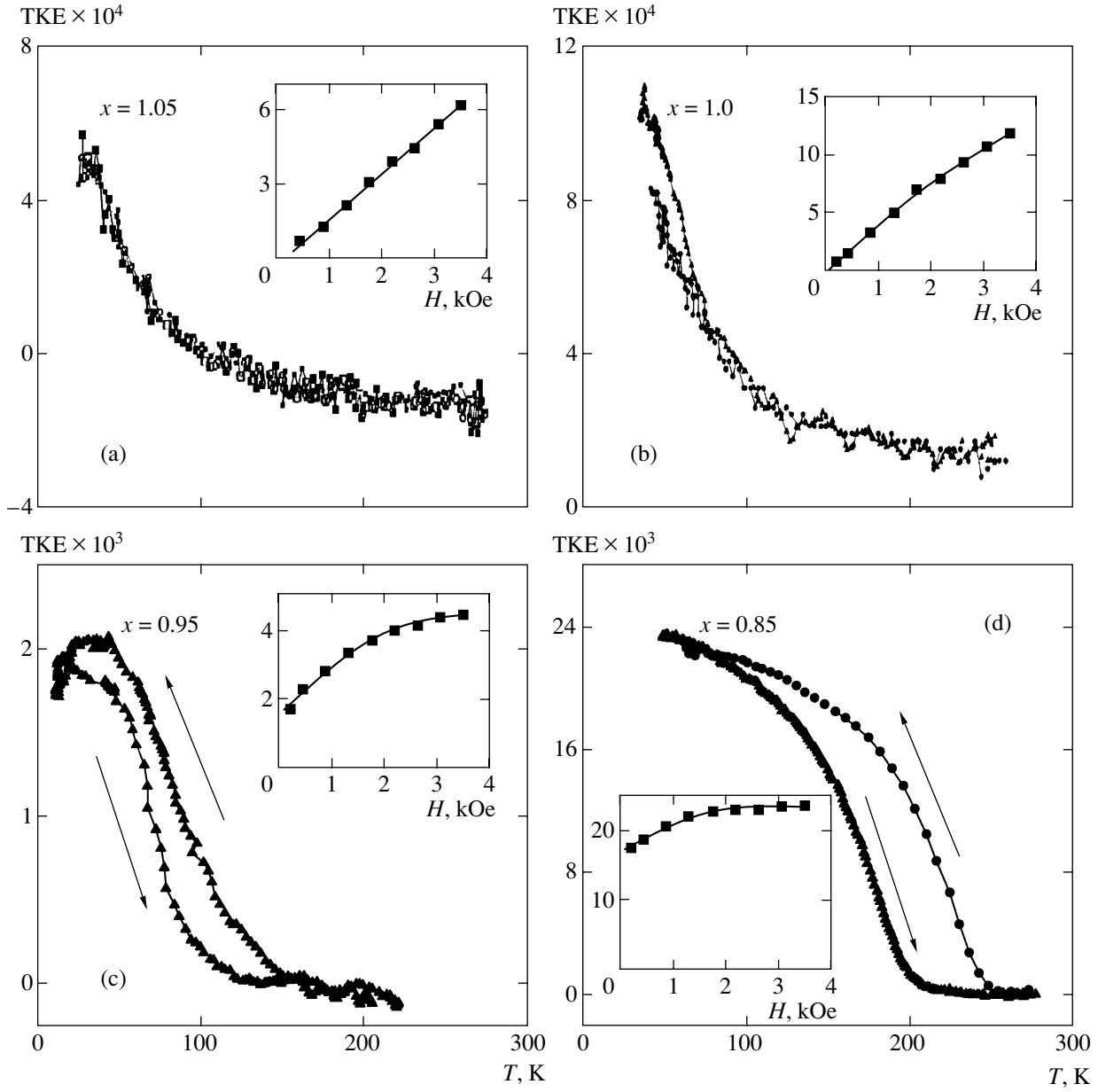


Fig. 2. Temperature dependences of TKE for La_xMnO_3 films with different values of x, obtained in cooling–heating regimes for H = 0.35 T and E = 2.9 eV. The direction of temperature variation is indicated by arrows. The insets show the field dependences of TKE for corresponding values of x.

ions. A shift in the spectral peak near 3 eV indicates that the ratio of the magnetically active Mn^{3+} and Mn^{4+} ions changes upon an increase in the La deficit. The increase in the TKE amplitude for compounds with $x \le 0.95$ indicates that the volume of the ferromagnetic phase increases. We have not detected a shift in the spectral peak for an excess of La (e.g., for x = 1.05), but observed a decrease in the amplitudes of the TKE peaks as compared to the amplitudes for the LMO film. This may be due to a decrease in the fraction of the ferromagnetic phase due to a decrease in the total concentration of Mn ions because an excess of La in a perovskite lattice is structurally realized owing to the formation of Mn vacancies.

3.2. Optical Absorption

The optical density spectrum for a nominally pure $LaMnO_3$ film (Fig. 4) clearly displays a band at 1.7 eV, which is also observed in the spectrum of a $LaMnO_3$ single crystal (see inset to Fig. 4). For energies above 2.5 eV, all LXMO films exhibit an increase in the optical density, which corresponds to the edge of the absorption band in manganites at ~5 eV (see, for exam-

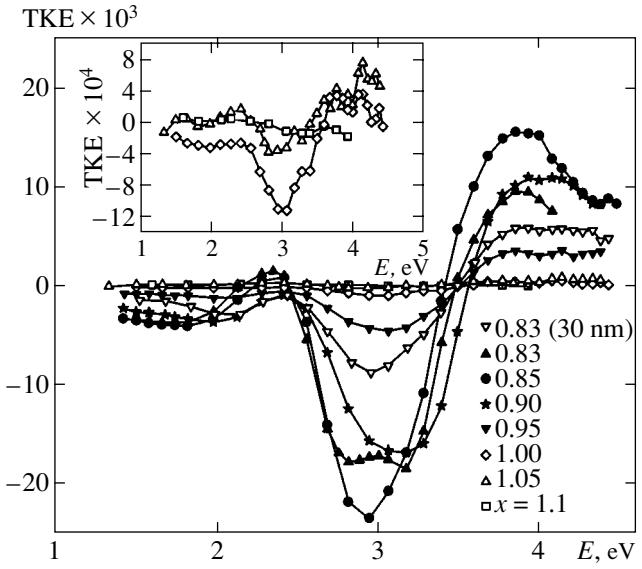


Fig. 3. Spectral dependences of TKE for fields with different La deficits, obtained at $T = 50$ K, $H = 0.35$ T. The inset shows the curves for concentrations $x = 1.00$, 1.05, and 1.10 at $T = 35$ K.

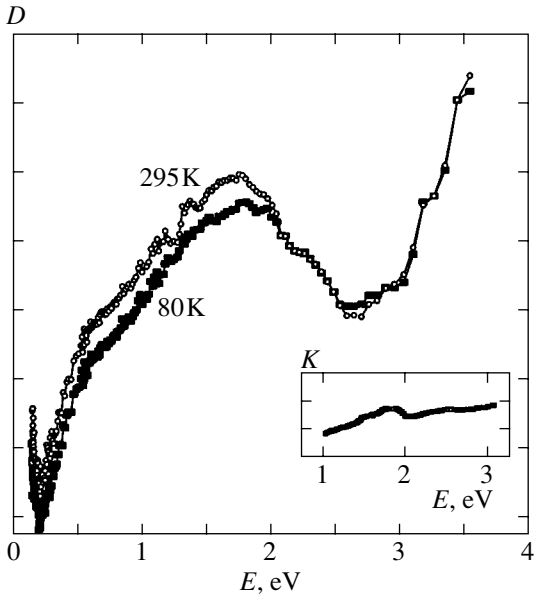


Fig. 4. Optical density spectrum for a $LaMnO_3$ film at two temperatures. The inset shows the absorption spectrum for a $LaMnO_3$ single crystal.

ple, [11]). A wide spectral band at 1.7 eV has a fine structure. Against the background of this band, we can single out a number of bands centered at 1.12, 1.60, 2.00, 2.35, and 2.80 eV (Fig. 5).

With an excess of lanthanum (i.e., after the introduction of Mn vacancies), the fine structure is smoothed. The introduction of La vacancies leads to a clearer manifestation of fine structure bands, a change in the intensity ratio of these bands, and a displacement of the center of gravity of the wide band at 1.7 eV towards lower energies. A film with a composition of $La_{0.83}MnO_3$ and a thickness of 30 nm displays a considerable improvement in resolution of these bands at 295 K for the same values of energy as for films with a thickness of 200 nm (inset to Fig. 6). The emergence of these bands is not associated with interference of light since the film has a high absorptance in the visible range and the interference condition for thin films is not satisfied for the normal incidence of light.

Cooling of the films from 295 to 80 K changes the spectrum, the change being different for LMO and for films with $0.83 \le x \le 0.90$. The LMO film displays a shift of the edge of the band at 1.7 eV towards higher energies (see Fig. 4), which is typical of semiconductors. LXMO films ($0.83 \le x \le 0.90$) exhibit an increase in the optical density in the range of low energies $E <$ 2.5 eV (Fig. 6). For films with $x = 0.83$, cooling leads to a considerable increase in the optical density and smoothens the fine structure (see Fig. 6). The difference in the optical densities (absorptances) for a film with $x = 0.83$ (200 nm) at two temperatures (80 and 269 K) in the ferromagnetic region is proportional to the square of the wavelength (in the spectral range from 2000 to 8000 nm). This indicates that the increase in absorptance as a result of film cooling may be due to the emergence of the contribution from free charge carriers. We observed similar temperature variations in the spectra of heavily doped $La_{0.7}Sr_{0.3}MnO_3$ films [12] and attributed it to a redistribution of the spectral weight from the region of fundamental absorption to the IR spectral region in which the contribution from free charge carriers shows up. The shape of the temperature dependence of sample transmittance t in the IR spectral range gives an idea of the behavior of charge carriers upon a transi-

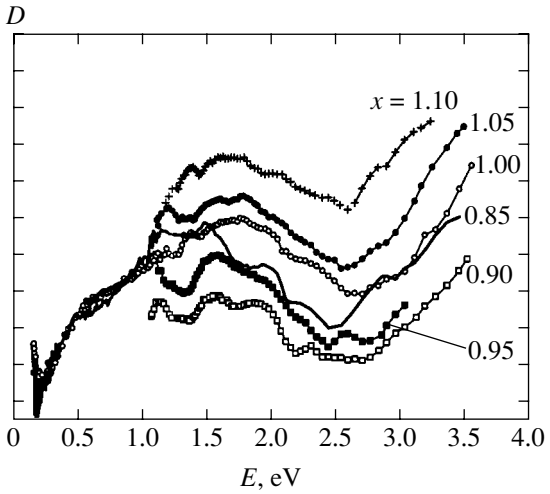


Fig. 5. Optical density spectrum for La_xMnO_3 films with an excess and deficit of lanthanum at $T = 295$ K. The curves corresponding to compositions with $x = 1.10$, 1.05, 0.95, and 0.90 are displaced along the ordinate axis for better visualization.

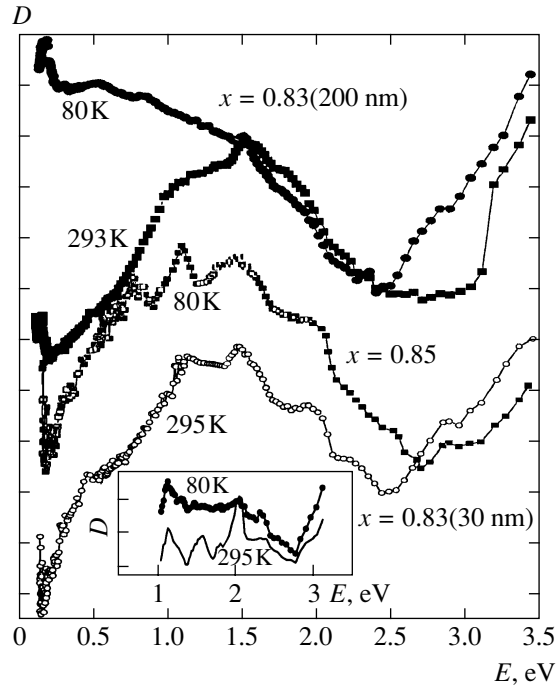


Fig. 6. Optical density spectra for La_xMnO_3 films with x = 0.83 and 0.85 at two temperatures. The curves corresponding to a composition with x = 0.83 are shifted along the ordinate axis through a constant interval for convenience of visualization. The inset shows the optical density spectra for a thin film with x = 0.83.

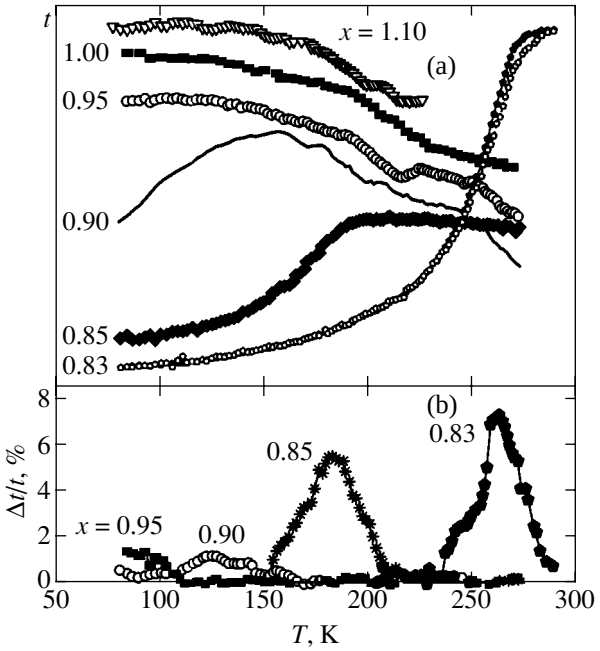


Fig. 7. (a) Temperature dependences of the optical transmittance of La_xMnO_3 films ($0.83 < x < 1.10$) for E = 0.14 eV. The curves are displaced along the ordinate axis for better visualization. The temperature dependence of transmittance of a film with x = 0.83 and thickness 200 nm is denoted by dark circles for H = 0 and by light circles for H = 0.8 T. (b) Temperature dependences of magnetotransmittance of the films for H = 0.8 T.

tion from the paramagnetic to the ferromagnetic region. The energy of 0.14 eV corresponding to the optical density minimum in front of the phonon spectrum (Figs. 5, 6) for $T > T_C$ is a convenient point for investigating the temperature behavior of the contribution from charge carriers.

Let us compare the temperature dependences of transmittance (Fig. 7a) and resistivity ρ (Fig. 8) of the films. As a result of cooling of films with $0.95 \le x \le 1.10$, their transmittance increases monotonically as in the case of an undoped LMO single crystal [13], which is typical of semiconductors. However, an increase of the lanthanum deficiency leads to the inverse pattern even for x = 0.90: after the attainment of its peak value near T_C, the transmittance of a film decreases in spite of continuous increase in resistivity in the temperature range under investigation. A further increase in the lanthanum deficit ($x \le 0.85$) is accompanied by an insulator–metal transition near T_C, which can be seen from the $t(T)$ and $\rho(T)$ curves (see Figs. 7a and 8). A similar temperature dependence of transmittance (insulator–metal transition below T_C) was observed for LXMO polycrystals with $x \le 1$ [2]. According to the $\rho(T)$ (Fig. 8) and $t(T)$ (Fig. 7a) curves, films with $x \ge 0.95$ exhibit no insulator–metal transition in the temperature range $T > 80$ K under investigation. A higher value of ρ for a film with x = 0.83 is due to difficulties involved when the shape of the sample is taken into consideration. An increase in the value of ρ and activation energy for films with $x > 1$ near room temperature (see table) indicates the presence of lanthanum and manganese vacancies even in the nominally stoichiometric LMO due to deviation from oxygen stoichiometry. For $x > 1$, the number of manganese vacancies in the structure in fact increases. Both doping of lanthanum manganite with bivalent ions and an increase in the concentration of La vacancies (increase in the ion concentration ratio Mn^{4+}/Mn^{3+}) lead to an increase in the

Table

Composition, x	T_C^*, K	T_{MR}, K	T_{MT}, K	E_a, eV
0.83	264	267	268	–
0.85	180	190	184	0.112
0.90	126	125	126	0.125
0.95	–	–	<80	0.125
1.00	–	–	–	0.142
1.05	–	–	–	0.145
1.10	–	–	–	0.162
0.83 (30 nm)	100	–	–	–

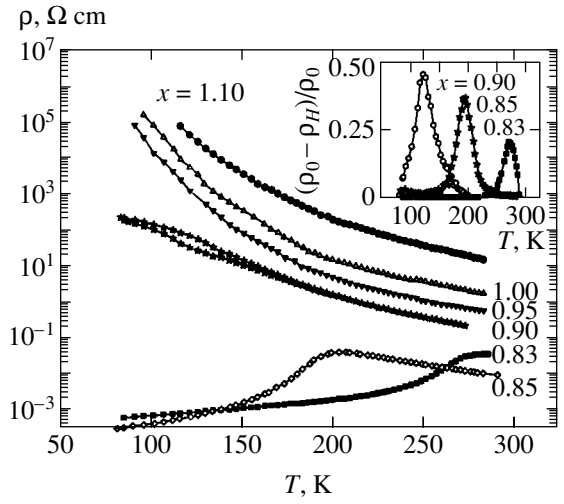


Fig. 8. Temperature dependences of resistivity of La_xMnO_3 films. Dark stars correspond to the temperature dependence of the film with $x = 0.90$ for $H = 0$; light stars correspond to $H = 0.8$ T. The inset shows the temperature dependences of the magnetoresistance of films with $x = 0.83$ and 0.90 for $H = 0.8$ T.

magnetic ordering temperature and the temperature corresponding to the maximum film transmittance, which is associated with it.

The application of a magnetic field reduces the transmittance and shifts the $t(T)$ curve towards higher temperatures in analogy to the behavior of the $\rho(T)$ curve in a magnetic field. The magnetotransmittance (MT) effect (relative change in the intensity of light transmitted by the sample under the action of the magnetic field, $\Delta t/t = |(t_H - t)/t|$) emerging in this case is similar to the magnetoresistance (MR) effect (inset to Fig. 8). Figure 7b shows the temperature dependences of the magnetotransmittance for films in a magnetic field $H = 0.8$ T applied along the normal to the film surface. The magnetotransmittance of the films attains its peak value in the vicinity of the magnetic ordering temperature. The maximum value of MT ~ 8% was obtained for a film with $x = 0.83$ (200 nm). The magnetotransmittance associated with free charge carriers is practically absent in the region of fundamental absorption; for instance, the value of MT for a composition with $x = 0.85$ amounts to less than 0.5% for an energy of 3.0 eV.

For most compositions of LXMO films, the gently sloping temperature dependence of the TKE indicates that the films are magnetically inhomogeneous and apparently consist of a set of ferromagnetic regions with different Curie temperatures. In this case, it is expedient to assume that the temperature T_C^* corresponding to the minimum of the first derivative dM/dT of magnetization with respect to temperature is the effective temperature of magnetic ordering for an inhomogeneous film in accordance with the Landau theory. The temperature T_C^* determined in this way for $[d(\mathrm{TKE})/dT]_{\min} = [dM/dT]_{\min}$ is close to temperatures corresponding to the maxima of magnetoresistance and magnetotransmittance (see table). The Curie temperature obtained by differentiating the curve $M(T)$ measured on a SQUID magnetometer for a film with $x = 0.90$ coincided with T_C^* measured from the TKE.

4. DISCUSSION

Even for oxides with stoichiometric compositions of transition 3*d* metals, the origin of low-energy electron-hole excitations is one of the most important and disputable questions in the physics of these strongly correlated systems. It is generally accepted at present that the most intense low-energy electronic excitations in dielectric 3*d* oxides are associated with the electron transfer from an oxygen anion to a cation of a 3*d* metal; consequently, these materials are insulators with charge transfer (CT). Unfortunately, a convincing interpretation of various absorption bands in the spectra of 3*d* oxides has not been worked out as yet. For example, wide bands in the optical conductivity spectra of $LaMnO_3$ with centers near 2.0 and 5.0 eV are attributed by some authors [14, 16] to allowed electric-dipole CT transitions of the types $t_{2g}^3 e_g^1 - t_{2g}^3 e_g^2 \underline{L}$ and $t_{2g}^3 e_g^1 - t_{2g}^4 e_g^1 \underline{L}$ (L is an electron of the ligand) to the e_g or t_{2g} state of the 3*d* cation, respectively. Other authors attribute the low-energy band to various *d*–*d* transitions [11, 17] or CT transitions involving Mn^{4+} ions [18].

In view of the ambiguity in the interpretation of manganite optics, the well-known quantum-mechanical cluster approach acquires special importance. This approach enables us to include consistently all correlation effects with the help of the well-developed apparatus of the theory of the crystal field and the field of ligands [19]. Such an approach was used in [20] for the classification and analysis of various CT states and O2*p*–Mn3*d* CT transitions in $[MnO_6]^{-9}$ octahedra and for a semiquantitative description of CT bands in $LaMnO_3$. According to [20], single-electron CT transitions from odd purely oxygen $t_{1u}(\sigma)$, $t_{1u}(\pi)$, and $t_{2u}(\pi)$ orbitals to the e_g or t_{2g} state of a 3*d* cation form 60 (!) allowed electric-dipole transitions between the corresponding multielectron configurations with energies in a wide spectral range from 2.5 to 11 eV and a set of dipole-forbidden transition from oxygen states $a_{1g}(\sigma)$ and $t_{1g}(\pi)$ with a minimal energy of 1.7 eV $t_{1g}(\pi)$–e_g transition). Thus, the low-energy spectral range of the anomalously wide CT band in $LaMnO_3$ is formed by a series of CT transitions in a $[MnO_6]^{-9}$ octahedron with increasing intensity, starting from the lower dipole-forbidden transition $t_{2u}(\pi)$–e_g with a peak at 1.7 eV, followed by a weak dipole-allowed transition $t_{1g}(\pi)$–e_g with a peak at 2.5 eV and a relatively more intense dipole-allowed $t_{1u}(\pi)$–e_g transition with a peak at 3.5 eV.

It should be noted that the theoretical model developed in [20] corresponds to nominally stoichiometric $LaMnO_3$. However, manganites, as well as many other 3*d* oxides, are characterized by deviations from stoichiometry for all sublattices, including oxygen nonstoichiometry with charge compensation by holes of the Mn^{4+} and/or O^- type. The model calculations made in a recent publication [18] show that a charge transfer with the participation of such holes may lead to a noticeable contribution to the low-energy optical response in a range near 1.5 eV.

The interpretation of the optical response of nonstoichiometric La_xMnO_3 or doped manganites of the $La_{1-x}Sr_xMnO_3$ type is considerably complicated due to the emergence of strong charge fluctuations and phase separation presuming the presence of significant intrinsic electronic inhomogeneities. These nanoscopic inhomogeneities in the simplest approximation can be presented as a system of metal-like drops generated by doping or vacancies, which are distributed in the insulating matrix. The inhomogeneous texture of bulk crystals may differ significantly from the texture of films, for which the existence of a stress gradient over the film thickness must be taken into account. It should be emphasized that the optical properties of nonstoichiometric or doped manganites should be described taking into account these inhomogeneities as one of the most important factors of formation of the optical response.

The properties of nanoscopically inhomogeneous media differ significantly in many respects from the properties of homogeneous media. Their peculiarities are manifested in optical and magnetooptical spectra in the form of so-called geometrical resonances, viz., absorption peaks, which have no analog in homogeneous systems and reflect the discrete structure of the excitation spectrum of intrinsic electromagnetic modes of a small particle. Semiquantitative description of these features is usually based on the theory of an effective medium [21, 22], which is an analog of the coherent potential approximation widely used in the physics of disordered systems and "operates" well when the particle size is smaller than the wavelength of light. In this case, intrinsic absorption of particles predominates and scattering of light by particles is disregarded as a rule. The effective permittivity ε_{eff} can be determined from an expression of the type

$$\left\langle \sum_i p_i \frac{\varepsilon_i - \varepsilon_{\text{eff}}}{\varepsilon_{\text{eff}} + L_i(\varepsilon_i - \varepsilon_{\text{eff}})} \right\rangle = 0, \qquad (1)$$

where ε_i, p_i, and L_i are the permittivity, volume fraction, and depolarization factors of particles of the *i*th species, whose shape is assumed to be ellipsoidal; angle brackets indicate averaging over orientations of particles, their size, shape, etc.

Even in its simplest version, the theory of an effective medium successfully describes [23] the known spectral dependences of the optical conductivity of doped manganite $La_{1-x}Sr_xMnO_3$ for various extents of doping [16]. Nevertheless, the model of "homogeneous metallic drops" (1) in an insulator is very simple and is intended only for a semiquantitative description of basic properties of an optical response of an inhomogeneous medium. In the general case, we can expect a considerable spread in the shape, size, and internal structure of inclusions (in particular, their Drude parameters). Obviously, metal-like inclusions in an insulator always have a certain transient boundary layer with parameters that are not typical of the insulator or the metal. All these factors may lead to a more complex structure of geometrical resonance (its nonuniform broadening or even "splitting"). One of the familiar models of an effective medium that enables us to take into account the inhomogeneous structure of inclusions is the Sheng model of "two-layer" inclusions [24]. In this model, instead of the permittivity of the *i*th component of the mixture, the modified expression

$$\hat{\varepsilon}_i = \varepsilon_{\text{out}} \frac{(\varepsilon_{\text{in}} - \varepsilon_{\text{out}})(fL_i^{\text{out}} - L_i^{\text{in}} - f) - \varepsilon_{\text{out}}}{(\varepsilon_{\text{in}} - \varepsilon_{\text{out}})(fL_i^{\text{out}} - L_i^{\text{in}}) - \varepsilon_{\text{out}}} \qquad (2)$$

is substituted into Eq. (1), where the indices "in" and "out" corresponds to the core and shell of compound particles, *f* being the ratio of the volumes of the inner and outer shells.

In order to illustrate the role of electron inhomogeneity effects, we use the effective medium formalism for describing the spectral and temperature dependences of the optical density of nonstoichiometric manganates, obtained by us for a thick $La_{0.83}MnO_3$ film. Figure 9 shows the calculated spectral dependences of optical conductivity $\sigma = (\omega/4\pi)\text{Im}\varepsilon$ and optical density $D = \ln(1/t)$ of this film, which is regarded as a system of the dielectric manganite type with various concentrations *p* of nanoscopic metal-like spherical inclusions whose permittivity is described by the Drude formula:

$$\varepsilon(\omega) = \varepsilon_\infty - \frac{\omega_p^2}{\omega(\omega + i\gamma)}.$$

The optical conductivity of the stoichiometric dielectric manganite $LaMnO_3$ (Fig. 9a) was simulated by a superposition of Lorentzians with centers at $\omega_1 = 2.3$ eV, $\omega_2 = 5.0$ eV, and $\omega_3 = 6.5$ eV, which successfully reproduces the experimental data obtained in [13–15]. The calculated spectrum of the optical density of a $La_{0.83}MnO_3$ film at 295 K (curve *1c* in Fig. 9b) corresponds to the simple model of homogeneous metallic drops (1) with concentration $p = 0.2$, plasma frequency $\omega_p = 3.8$ eV, relaxation rate $\gamma = 1.3$ eV, and $\varepsilon_\infty = 1$. Curves *1a* and *1b* correspond to a more realistic model of inhomogeneous two-layer inclusions (2) with the same total concentration, but consisting of a core with

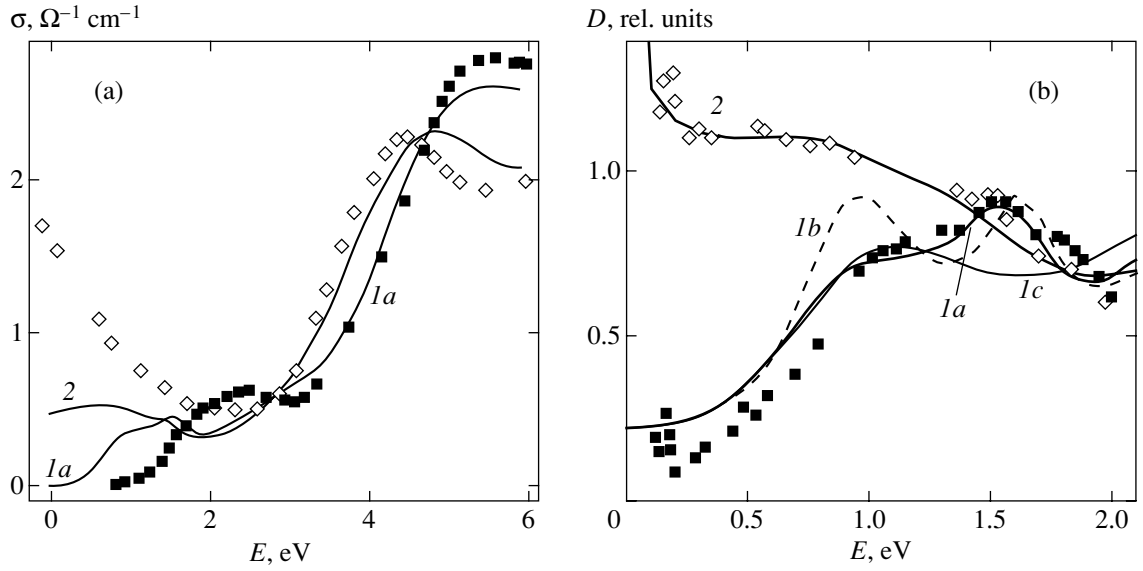


Fig. 9. Results of simulation of the high-temperature (T = 295 K, curves *1a* and *1b*) and low-temperature (T = 80 K, curve *2*) optical responses of a $La_{0.83}MnO_3$ film in the effective medium approximation with two-layer inclusions. (a) Calculated optical conductivity spectra with the quasimetallic phase volume p = 0.2 (curve *1a*) and 0.7 (curve *2*). The experimental results obtained for an allied system $La_{1-x}Sr_xMnO_3$ at 295 K [16] are shown for comparison (squares correspond to x = 0 and rhombi to x = 0.3). (b) Calculated optical density spectra with a wide (*1a*) and narrow (*1b*) spectrum of inhomogeneity fluctuations. Curve *1c* is the spectrum calculated in the model of homogeneous inclusions. Squares and rhombi correspond to experimental results obtained for a $La_{0.83}MnO_3$ film at 295 and 80 K.

volume fraction f = 0.3 and with relatively "good" metallic properties ($\varepsilon_{1\infty}$ = 1, ω_{1p} = 4.7 eV, and γ_1 = 0.05 and 0.01 eV for curves *1a* and *1b*, respectively) and a shell of a "poor" metal ($\varepsilon_{2\infty}$ = 2.5, ω_{2p} = 3.0 eV, and γ_2 = 1.1and 0.7 eV for curves *1a* and *1b*, respectively). In the case of an inhomogeneous two-layer inclusion, a geometrical resonance existing for a "homogeneous" inclusion "splits" into two. This effect is manifested most clearly for relatively small values of effective relaxation parameter γ, which also describes the structural distribution of the inclusion, i.e., the specific texture of the material. Curve *1a* describes fairly well the experimental data on the optical density of the film in the high-temperature region, which enables us to use the corresponding parameters as the initial parameters for simulating the temperature and concentration effects. In the general case, all parameters of a model inhomogeneous electronic structure may obviously exhibit a noticeable temperature dependence. If we assume the existence of double exchange for metallic inclusions, the ferromagnetic ordering below T_C must be accompanied by "metallization" with increasing plasma frequency and volume fraction of inclusions at the expense of replacement of the dielectric phase, which is disadvantageous from the energy viewpoint. The separation of these two effects ("electronic" and "volume" effects in the effective medium model) encounters certain difficulties and requires more detailed experimental studies. For the sake of simplicity, we confine our subsequent analysis to the volume effect only. Curve *2* (Fig. 9b) was calculated for the parameters p = 0.7, f = 0.65, ω_{1p} = 3.2 eV, γ_1 = 0.8 eV, $\varepsilon_{1\infty}$ = 1, and ω_{2p} = 1.9 eV, γ_2 = 2.0 eV, $\varepsilon_{2\infty}$ = 1. In other words, we assume a sharp increase in the volume fraction of metallic drops; as a result, metallic properties for curve *2* are manifested more clearly than for curves *1a* and *1b*. It can easily be seen that curve *2* successfully simulates the experimental optical density spectrum of the film for T = 80 K, which enables us to consider the volume effect as one of the most important factors determining the temperature behavior of the low-energy spectral response of doped manganites. The spectral singularity in the vicinity of 0.2 eV, which is clearly seen at room as well as nitrogen temperature, may be due to allowed electric-dipole transitions with charge transfer in the hole centers MnO_6^{8-}, where, according to estimates obtained in [18], the CT energy is a small quantity on the order of 0.2 eV. It should be noted that the optical response observed for a real inhomogeneous system might also be a complex effect of overlapping of electronic transitions with geometrical resonances. It is well known [22] that the interference between geometrical and electronic resonances may lead to a complex shape of the resultant absorption spectrum and the TKE spectrum, which is quite sensitive even to small changes both in the type of inhomogeneity and in the electronic transition. In particular, the initiation of forbidden or weak allowed transitions could be observed. The effects of interference of geometrical and electron resonances in manganites can be

manifested in the range 1.5–2.0 eV owing to the forbidden transition $t_{1g}(\pi)$–e_g with CT.

Thus, the model of an effective medium makes it possible by and large to describe the main spectral features of nonstoichiometric manganites as well as their dependence on the vacancy concentration, temperature, and the structure of electronic inhomogeneity as a whole. A comparison of the optical response of nonstoichiometric manganites L_xMnO_3 with the results of various experimental studies of the spectral response of doped $La_{1-x}Sr_xMnO_3$-type manganites based the simple model of an effective medium leads to very important conclusions concerning the universal nature of their electronically inhomogeneous state and, in addition, concerning the instability of the $LaMnO_3$ system in the formation of electronic inhomogeneity. Indeed, the optical response of the entire wide set of LXMO films under investigation in the mean IR range in the region of interaction between light and charge carriers directly indicates the existence of an electronic inhomogeneity of the type of metallic drops in an insulating matrix. For example, a $La_{0.9}MnO_3$ film exhibits an optical response typical of the insulator–metal transition (see Fig. 7a) in spite of the fact that the resistivity of this film is high and is of the semiconductor type (see Fig. 8). An increase in magnetoabsorptance for the $La_{0.95}MnO_3$ film at $T < 100$ K (see Fig. 7b) indicates that this film contains metallic drops. The size of the drops, the volume occupied by them, and the spatial pattern of inhomogeneity depend not only on the number of acceptors (La and Mn vacancies), but also on the number of donors (oxygen vacancies) whose presence leads to a partly compensated state [25].

It is noteworthy that the quantities of MT and MR exhibit different dependences on the Curie temperature of the films. A decrease in T_C^* leads to an increase in the magnetoresistance of the films, which attains its peak value of about 45% for a film with $x = 0.90$ ($T_C^* = 125$ K). This composition is apparently close to the percolation threshold. On the contrary, the magnetotransmittance decreases with T_C^*. Such a difference in the behavior of MT and MR is due to the fact that the resistivity data are averaged over the sample volume. Magnetoresistance is low for small drops. When magnetotransmittance is being measured, a change in the transmittance in individual regions (drops) is observed. Above the percolation threshold, the emergence of a simply connected metallic region results in a strong magnetotransmittance effect accompanying the insulator-metal transition near T_C^* (see Fig. 7b). At the same time, the MT effect is quite sensitive for small-size drops. For example, for a composition with $x = 0.95$, the MR effect is absent even at $T > 80$ K, while the magnetotransmittance for this composition is appreciable. However, the MT peak is not attained for this composition, which means that an insulator–metal transition is possible at $T < 80$ K. The temperature range in which metallic drops exist coincides with the existence of TKE, i.e., with the ferromagnetic ordering; consequently, metallic drops are ferromagnetic.

It can be seen from the temperature dependence of TKE that the Curie temperature for a film with $x = 0.83$ having a smaller thickness (30 nm) is 100 K, which is approximately one-third the value of T_C for a thick (200 nm) film of the same composition ($T_C^* = 264$ K). This is probably due to a decrease in the volume and the total fraction of ferromagnetic regions in a thin film owing to high compressive stresses emerging in the film–substrate system. In view of the small thickness (30 nm) of the film with $x = 0.83$, magnetotransmittance is not observed.

While comparing the optical properties of La_xMnO_3 epitaxial films in the visible and IR regions with the behavior of manganites of the $La_{1-x}Sr_xMnO_3$ type, we must emphasize once again the universal nature of the phase separation and the properties of the inhomogeneous electronic structure in systems with nonstoichiometric compositions and with heterovalent doping. It appears that the $LaMnO_3$ system is unstable to the emergence of fluctuations of the metallic ferromagnetic phase, which is stabilized upon the formation of vacancies in the La sublattice or upon heterovalent doping.

It was noted above that the formation of fine structure in the spectra for nonstoichiometric LXMO in a fairly wide range with the center at 1.7 eV and the peculiarities in its spectral, concentration, and temperature dependences may be associates with the splitting of geometrical resonances as well as with the interference of geometrical and electron resonances. The fine structure of the wide band changes with the vacancy concentration, temperature, and film thickness. The formation of Mn vacancies for $x > 1$, when the system composition is apparently close to the stoichiometric composition, leads to smoothing of the fine structure (see Fig. 5). On the contrary, an increase in the La vacancy concentration is accompanied by a noticeable intensification of the fine structure, accompanied by a gradual displacement of the spectral center of gravity towards lower energies in analogy to the effect of heterovalent doping in the $La_{1-x}Sr_xMnO_3$ system [16]. We believe that the optical data obtained in the visible and near IR ranges reflects both the effects of nonuniform broadening of a geometrical resonance up to its splitting into several resonances, and the effects of spectral overlapping of geometrical and electron resonances in fluctuating low-symmetry crystal fields, removing the forbiddenness from the electronic CT transition $t_{1g}(\pi)$–e_g and leading to splitting of degenerate initial and final states of the MnO_6 center. The low-symmetry crystal field in LXMO films may also be due to inhomogeneity associated with deviations from the stoichiometric composition as well as the stresses emerging in the film because

of the difference in the parameters of the film and the substrate. The most significant enhancement and resolution of the fine structure are observed in a thin $La_{0.83}MnO_3$ film (see inset to Fig. 6) whose thickness is approximately equal to one-seventh of the thickness of a thick film with a more smoothened spectrum. This experimental fact appears quite natural if we take into account a wider spectrum of fluctuations of electronic inhomogeneities in the thick film as compared to the thin film for which the stress gradient from the substrate side is a selective factor narrowing the spectrum of inhomogeneity fluctuations. In the simple model of two-layer inclusions considered above, a decrease in the effective relaxation parameters γ directly connected with fluctuations in the insulator–metal system makes it possible to simulate the transition from the thick to the thin film (curves *1a* and *1b* in Fig. 9), which is accompanied by splitting of a nonuniformly broadened geometrical resonance.

Thus, an analysis of the optical response of lanthanum manganite films makes it possible to study not only deviations from stoichiometry and the insulator–metal volumetric balance, but also finer features of the electronically inhomogeneous texture as well as the effect of various external factors. An increase in the La vacancy concentration leads not only to the emergence of a fine structure against the background of a wide absorption band at 1.7 eV, but also to a considerable increase and displacement of the position of the main TKE peak, indicating a change in the ratio of the number of magnetically active Mn^{3+} and Mn^{4+} ions.

It should be noted that the emergence of geometrical resonances in an inhomogeneous system also leads to a considerable modification of magnetooptical spectra [26]. A detailed analysis of the TKE spectra for films of nonstoichiometric manganites as nanoscopically inhomogeneous systems may become a subject of special investigation.

5. CONCLUSIONS

Analysis of optical and magnetooptical properties of La_xMnO_3 epitaxial films with a deficit and excess of lanthanum ($0.83 \le x \le 1.10$) in the visible and IR ranges and their comparison with the behavior of the $La_{1-x}Sr_xMnO_3$ systems indicate the universal nature of phase separation and properties of an inhomogeneous electronic structure in a system with a nonstoichiometric composition and with heterovalent doping. It is shown that the optical and magnetooptical responses of the La_xMnO_3 system are determined not only by the nominal concentration of vacancies, but also depend significantly on the internal nanoscopic metal–insulator texture which, in turn, depends on the composition, temperature, and thickness of the film. The spectral, concentration, and temperature dependences of optical density are interpreted quantitatively in the framework of the theory of an effective medium. The fine structure of the fundamental absorption band of La_xMnO_3 has been discovered and analyzed. The origin of the fine structure is associated with the splitting of a geometrical resonance and with the interference of geometrical and electron resonances in an inhomogeneous system with phase separation. A comparison of the optical properties of La_xMnO_3 films with different degrees of nonstoichiometry ($0.83 \le x \le 1.10$) suggests that even a compound with a nominally stoichiometric composition ($x = 1.0$) is a system unstable to the formation of electronic inhomogeneities.

ACKNOWLEDGMENTS

The authors are grateful to A.V. Korolev, who performed magnetic measurements, and to N.G. Bebenin for fruitful discussions.

This study was supported financially by the Federal Center for Science and Technology Projects (contract no. 40.012.1.1.1153-14/02), Russian Foundation for Basic Research (project nos. 01-02-96404, 02-02-16429, and 02-03-33258), Ministry of Education of the Russian Federation (grant no. E00-3.4-280), and CRDF (grant no. REC-005).

REFERENCES

1. V. Ferris, L. Brohan, M. Ganne, *et al.*, Eur. J. Solid State Inorg. Chem. **32**, 131 (1995).
2. N. N. Loshkareva, N. I. Solin, Yu. P. Sukhorukov, *et al.*, Physica B (Amsterdam) **293**, 390 (2001).
3. E. A. Gan'shina, O. Yu. Gorbenko, A. R. Kaul, *et al.*, J. Phys.: Condens. Matter **12**, 2857 (2000).
4. N. N. Loshkareva, Yu. P. Sukhorukov, E. A. Gan'shina, *et al.*, Zh. Éksp. Teor. Fiz. **119**, 533 (2001) [JETP **92**, 462 (2001)].
5. É. L. Nagaev, Usp. Fiz. Nauk **166**, 833 (1996) [Phys.–Usp. **39**, 781 (1996)].
6. Amlan Biswas, M. Rajeswari, R. C. Srivastava, *et al.*, Phys. Rev. B **63**, 184424 (2001).
7. A. A. Bosak, O. Yu. Gorbenko, A. R. Kaul, *et al.*, J. Magn. Magn. Mater. **211**, 61 (2000).
8. E. A. Balykina, E. A. Gan'shina, and G. S. Krinchik, Zh. Éksp. Teor. Fiz. **93**, 1879 (1987) [Sov. Phys. JETP **66**, 1073 (1987)].
9. E. O. Wollan and W. C. Koehler, Phys. Rev. **100**, 545 (1955).
10. E. A. Balykina, E. A. Gan'shina, G. S. Krinchik, and A. Yu. Trifonov, J. Magn. Magn. Mater. **117**, 259 (1992).
11. J. F. Lawler, J. G. Lunney, and J. M. D. Coey, Appl. Phys. Lett. **65**, 3017 (1994).
12. N. N. Loshkareva, Yu. P. Sukhorukov, V. E. Arkhipov, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **41**, 475 (1999) [Phys. Solid State **41**, 426 (1999)].
13. N. N. Loshkareva, Yu. P. Sukhorukov, E. V. Mostovshchikova, *et al.*, Zh. Éksp. Teor. Fiz. **121**, 412 (2002) [JETP **94**, 350 (2002)].

14. T. Arima and Y. Tokura, J. Phys. Soc. Jpn. **64**, 2488 (1995).
15. Y. Okimoto, T. Katsufui, T. Ishikawa, *et al.*, Phys. Rev. Lett. **75**, 109 (1995); Y. Okimoto, T. Katsufui, T. Ishikawa, *et al.*, Phys. Rev. B **55**, 4206 (1997).
16. K. Takenaka, K. Iida, Y. Sawaki, and Y. Morimoto, J. Phys. Soc. Jpn. **68**, 1828 (1999).
17. J. H. Jung, K. H. Kim, T. W. Noh, *et al.*, Phys. Rev. B **57**, R11043 (1998).
18. N. N. Kovaleva, J. L. Gavartin, A. L. Shluger, *et al.*, E-print archives, cond-mat/0108207.
19. I. B. Bersuker, *Electronic Structure and Properties of Transition Metal Compounds (Introduction to the Theory)* (Wiley, New York, 1996).
20. A. S. Moskvin, E-print archives, cond-mat/0111198.
21. D. J. Bergman and D. Stroud, Solid State Phys. **46**, 148 (1992).
22. W. A. de Heer, Rev. Mod. Phys. **65**, 611 (1993).
23. A. S. Moskvin, E. V. Zenkov, Yu. D. Panov, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **44**, 1452 (2002) [Phys. Solid State **44**, 1519 (2002)].
24. Ping Sheng, Phys. Rev. Lett. **45**, 60 (1980).
25. B. I. Shklovskiĭ and A. L. Éfros, *Electronic Properties of Doped Semiconductors* (Nauka, Moscow, 1979; Springer, New York, 1984).
26. T. K. Xia, P. M. Hiu, and D. Stroud, J. Appl. Phys. **67**, 2736 (1990).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 268–285.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 305–324.

SOLIDS
Electronic Properties

Quantum Theory of Conductivity of Spatially Inhomogeneous Systems

I. G. Lang[a], L. I. Korovin[a,*], J. A. de la Cruz-Alcaz[b], and S. T. Pavlov[c,]**

[a]*Ioffe Physicotechnical Institute, Russian Academy of Sciences, Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia*
[b]*Facultad de Fisica de la UAZ, Apartado Postal C-580, 98060 Zacatecas, Zac., Mexico*
[c]*Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 119991 Russia*
**e-mail: korovin@mail.ioffe.ru*
***e-mail: pavlov@sci.lpi.ac.ru; pavlov@ahabon.reduaz.mx*
Received March 28, 2002

Abstract—A quantum theory of conductivity is constructed for semiconductor objects such as quantum wells, wires, and dots. The mean values of current and charge densities induced by a weak electromagnetic field are calculated. It is shown that the mean values of current and charge densities consist of two parts, the first of which is expressed in terms of the electric field and the second is expressed in terms of derivatives of the electric field with respect to spatial coordinates. Appropriate expressions are derived for the conductivity tensor that depends on coordinates; these expressions can be applied to any spatially inhomogeneous systems. The results obtained can be used in the theory of secondary radiation from objects of reduced dimension in the cases of monochromatic or pulsed irradiation. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The growing interest in the experiments and theory on the reflection and absorption of light by semiconductor objects of reduced dimension such as quantum wells, wires, and dots under pulsed light excitation (see, for example, [1, 2]) revives the question of what form of interaction between electrons and an electromagnetic wave is more convenient—the one containing vector potential $\mathbf{A}(\mathbf{r}, t)$ or that with electric field $\mathbf{E}(\mathbf{r}, t)$.

Paper [3] is devoted to the consideration of the same question as applied to calculating the differential cross section of inelastic scattering of light by infinite crystals. In [3], it is shown that exact expressions for scattering cross sections (that involve sets of exact wave functions of electrons in a crystal) obtained with the use of two different forms of interaction between electrons and an electromagnetic wave coincide. However, the set of exact wave functions of electrons in a crystal (that accurately take into account, for example, the electron–phonon interaction) is unknown; therefore, approximate methods are used for calculating cross-sections—namely, one takes into account the lowest order electron–phonon interaction and does not take into account all intermediate states of the electron system. In this approximation, various forms of electron–light interaction (those involving **A** or **E**) yield different results. The authors of [3] assert that, in the case of nonresonant scattering, the interaction containing **E** gives better results (i.e., results closer to exact ones).

We will construct a general theory of secondary radiation of light by low-dimensional semiconductor objects such as quantum wells, quantum wires, and quantum dots. First of all, this theory should describe the reflection and absorption of light by such objects, as well as various types of light scattering (Raman and Rayleigh scatterings). The theory must be applicable in the cases of monochromatic as well as pulsed radiation. We restrict the analysis to the linear approximation of the intensity of exciting light. It follows from the aforesaid that the problems solved here are different from those treated in [3], where the authors consider the light scattering in bulk crystals. Therefore, the problem of choosing the form of interaction (either in terms of **A** or **E**) should be solved anew.

In the present paper, we calculate the mean densities of current and charge induced by electromagnetic field in the case of an inhomogeneous medium. This case includes semiconductor objects of reduced dimension. Having calculated the linear (in electric and magnetic fields) contributions to the mean densities of current and charge, we can determine these fields inside and outside the semiconductor objects by solving the Maxwell equations. In this way, we can obtain expressions for the fields that correspond to the light reflected from and transmitted through an object. Such a procedure, which takes into account all orders of interaction of electrons with light, was carried out in [4], where the intensities of reflected and absorbed light were calculated under monochromatic irradiation of a quantum well of finite thickness. In the present paper, we derive

an expression for the mean density of induced current that was used in [4].

The operator of interaction of charged particles with electric and magnetic fields is expressed in terms of the vector $\mathbf{A}(\mathbf{r}, t)$ and scalar $\varphi(\mathbf{r}, t)$ potentials. Therefore, the mean densities of induced current and charge are also expressed in terms of these potentials. However, the application of these expressions is hampered due to the contribution

$$-\frac{e}{mc}\langle 0|\rho(\mathbf{r})|0\rangle \mathbf{A}_\alpha(\mathbf{r}, t)$$

to the mean density of current, where e and m are the charge and mass of a particle, respectively, and $\rho(\mathbf{r})$ is the charge-density operator (see Section 4 below). Therefore, we express the mean values of current and charge densities in terms of electric and magnetic fields. This is possible because all observable quantities are expressed via the electric $\mathbf{E}(\mathbf{r}, t)$ and magnetic $\mathbf{H}(\mathbf{r}, t)$ fields. Our task is to pass from expressions for physical quantities containing potentials to expressions containing the fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{H}(\mathbf{r}, t)$. In the present paper, we consider the case of zero temperature $T = 0$.

The paper is organized as follows. In Sections 2–4, we give an account of the statement of the problem and introduce the operators of current and charge densities and their averaged values over the ground state of the system. In Sections 5–9, we solve the problem of expressing these mean values in terms of electric field and its spatial derivatives. Section 10 is devoted to the elimination of diagonal matrix elements of operators $\mathbf{r}_i$ from the expressions of the mean values. In Section 11, we present a general expression for the conductivity tensor of an inhomogeneous system. In Sections 12 and 13, we consider the case of zero electric and constant magnetic fields.

2. STATEMENT OF THE PROBLEM

Consider a system of N particles of mass m and charge e in an arbitrary weak electromagnetic field characterized by intensities $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{H}(\mathbf{r}, t)$. Let us introduce vector $\mathbf{A}(\mathbf{r}, t)$ and scalar $\varphi(\mathbf{r}, t)$ potentials in terms of which the fields are expressed:

$$\begin{aligned}\mathbf{E}(\mathbf{r}, t) &= -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} - \nabla\varphi,\\ \mathbf{H}(\mathbf{r}, t) &= \mathrm{curl}\mathbf{A}.\end{aligned} \tag{1}$$

The fields are assumed to be classical. The gage of the potentials $\mathbf{A}(\mathbf{r}, t)$ and $\varphi(\mathbf{r}, t)$ is arbitrary. For the sake of completeness, we assume that the system of particles is placed in a constant magnetic field $\mathbf{H}_c$, which may be strong. This field corresponds to a vector potential $\mathcal{A}(\mathbf{r})$, so that

$$\mathbf{H}_c = \mathrm{curl}\mathcal{A}(\mathbf{r}).$$

The total Hamiltonian $\mathcal{H}_{\mathrm{tot}}$ is expressed as

$$\begin{aligned}\mathcal{H}_{\mathrm{tot}} &= \frac{1}{2m}\sum_i\left(\mathcal{P}_i - \frac{e}{c}\mathcal{A}(\mathbf{r}_i) - \frac{e}{c}\mathbf{A}(\mathbf{r}_i, t)\right)^2\\ &+ V(\mathbf{r}_1, \ldots, \mathbf{r}_N) + e\sum_i \varphi(\mathbf{r}_i, t),\end{aligned} \tag{2}$$

where $\mathcal{P}_i = -i\hbar(\partial/\partial\mathbf{r}_i)$ is a generalized-momentum operator [5, 6] and $V(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is the potential energy that includes the interaction between particles and an external potential. We have to take into account the noncommutativity of $\mathcal{P}_i$ with $\mathcal{A}(\mathbf{r}_i)$ and $\mathbf{A}(\mathbf{r}_i, t)$ in (2). Let us single out the energy U of interaction between particles and electromagnetic field from (2) by including the interaction with a strong magnetic field in the total Hamiltonian:

$$\mathcal{H}_{\mathrm{tot}} = \mathcal{H} + U, \tag{3}$$

$$\begin{aligned}\mathcal{H} &= \frac{1}{2m}\sum_i \mathbf{p}_i^2 + V(\mathbf{r}_1, \ldots, \mathbf{r}_N),\\ \mathbf{p}_i &= \mathcal{P}_i - \frac{e}{c}\mathcal{A}(\mathbf{r}_i),\end{aligned} \tag{4}$$

where

$$\begin{aligned}U &= U_1 + U_2,\\ U_1 &= -\frac{1}{c}\int d(^3r\,\mathbf{j}(\mathbf{r})\cdot\mathbf{A}(\mathbf{r}, t)) + \int d^3r\rho(\mathbf{r})\varphi(\mathbf{r}, t),\\ U_2 &= \frac{e}{2mc}\int d^3r\rho(\mathbf{r})\mathbf{A}^2(\mathbf{r}, t)\end{aligned} \tag{5}$$

and the following current- and charge-density operators are introduced:

$$\begin{aligned}\mathbf{j}(\mathbf{r}, t) &= \sum_i \mathbf{j}_i(\mathbf{r}, t),\\ \mathbf{j}_i(\mathbf{r}, t) &= \frac{e}{2}\{\delta(\mathbf{r} - \mathbf{r}_i)\mathbf{v}_i + \mathbf{v}_i\delta(\mathbf{r} - \mathbf{r}_i)\}, \quad \mathbf{v}_i = \frac{\mathbf{p}_i}{m},\\ \rho(\mathbf{r}) &= \sum_i \rho_i(\mathbf{r}), \quad \rho_i(\mathbf{r}) = e\delta(\mathbf{r} - \mathbf{r}_i).\end{aligned} \tag{6}$$

Our aim is to calculate the averaged (over the ground state of the system) densities of induced current and charge in a linear approximation with respect to the external fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{H}(\mathbf{r}, t)$.

3. DEFINITION OF OPERATORS

In the Schrödinger representation, the charge-density operator $\rho(\mathbf{r})$ does not contain components propor-

tional to fields, while the current-density operator under switched-on fields is expressed as

$$\mathbf{j}(\mathbf{r}, t) + \Delta\mathbf{j}(\mathbf{r}),$$

where

$$\Delta\mathbf{j}(\mathbf{r}) = \frac{e}{2}\sum_i \{\Delta\mathbf{v}_i\delta(\mathbf{r}-\mathbf{r}_i) + \delta(\mathbf{r}-\mathbf{r}_i)\Delta\mathbf{v}_i\}, \quad (7)$$

$$\Delta\mathbf{v}_i = \frac{i}{\hbar}[U, \mathbf{r}_i] = -\frac{e}{mc}\mathbf{A}(\mathbf{r}_i, t), \quad (8)$$

$[F, Q]$ is the commutator of operators F and Q, and hence,

$$\Delta\mathbf{j}(\mathbf{r}) = -\frac{e}{mc}\rho(\mathbf{r})\mathbf{A}(\mathbf{r}, t). \quad (9)$$

In the interaction representation, we have

$$\begin{aligned}\rho(\mathbf{r}, t) &= e^{i\mathcal{H}t/\hbar}\rho(\mathbf{r})e^{-i\mathcal{H}t/\hbar},\\ \mathbf{j}(\mathbf{r}, t) &= e^{i\mathcal{H}t/\hbar}\mathbf{j}(\mathbf{r})e^{-i\mathcal{H}t/\hbar},\\ \Delta\mathbf{j}(\mathbf{r}, t) &= -\frac{e}{mc}\rho(\mathbf{r}, t)\mathbf{A}(\mathbf{r}, t).\end{aligned} \quad (10)$$

Now, we define the current- and charge-density operators in the Heisenberg representation. In [7, p. 82], it is shown that the relation between the operator $F(t)$ in the interaction representation and the operator $F_G(t)$ in the Heisenberg representation is expressed as

$$F_G(t) = S^{-1}(t)F(t)S(t), \quad (11)$$

where the S matrix is defined as

$$\begin{aligned}S(t) &= S(t, -\infty) = 1 - \frac{i}{\hbar}\int_{-\infty}^{t} dt_1 U(t_1)\\ &+ \left(-\frac{i}{\hbar}\right)^2 \int_{-\infty}^{t} dt_2 U(t_2)\int_{-\infty}^{t_2} dt_1 U(t_1) + \ldots,\\ U(t) &= e^{i\mathcal{H}t/\hbar}Ue^{-i\mathcal{H}t/\hbar}.\end{aligned} \quad (12)$$

Using (12), we find that additions, linear in the potentials $\mathbf{A}(\mathbf{r}, t)$ and $\varphi(\mathbf{r}, t)$, of current and charge densities in the Heisenberg representation are equal to

$$\begin{aligned}j_{1\alpha}(\mathbf{r}, t) &= \Delta j_\alpha(\mathbf{r}, t) - \frac{i}{\hbar}\int_{-\infty}^{t} dt'[j_\alpha(\mathbf{r}, t), U_1(t')],\\ \rho_1(\mathbf{r}, t) &= -\frac{i}{\hbar}\int_{-\infty}^{t} dt'[\rho(\mathbf{r}, t), U_1(t')].\end{aligned} \quad (13)$$

The subscript "1" indicates the first order with respect to the potentials $\mathbf{A}(\mathbf{r}, t)$ and $\varphi(\mathbf{r}, t)$. Substituting (9) and (5) into the last expressions, we obtain

$$\begin{aligned}j_{1\alpha}(\mathbf{r}, t) &= -\frac{e}{mc}\rho(\mathbf{r}, t)A_\alpha(\mathbf{r}, t)\\ &+ \frac{i}{\hbar c}\int d^3r' \int_{-\infty}^{t} dt'[j_\alpha(\mathbf{r}, t), j_\beta(\mathbf{r}', t')]A_\beta(\mathbf{r}', t')\\ &- \frac{i}{\hbar}\int d^3r' \int_{-\infty}^{t} dt'[j_\alpha(\mathbf{r}, t), \rho(\mathbf{r}', t')]\varphi(\mathbf{r}', t'),\end{aligned} \quad (14)$$

$$\begin{aligned}\rho_1(\mathbf{r}, t) &= \frac{i}{\hbar c}\int d^3r' \int_{-\infty}^{t} dt'[\rho(\mathbf{r}, t), j_\beta(\mathbf{r}', t')]A_\beta(\mathbf{r}', t')\\ &- \frac{i}{\hbar}\int d^3r' \int_{-\infty}^{t} dt'[\rho(\mathbf{r}, t), \rho(\mathbf{r}', t')]\varphi(\mathbf{r}', t').\end{aligned} \quad (15)$$

4. AVERAGING OVER THE GROUND STATE OF THE SYSTEM

Consider the case of zero temperature and average operators (14) and (15) over the ground state of the system. In all calculations below, we will assume that there are no charges and currents at infinity and that the fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{H}(\mathbf{r}, t)$ vanish as $t \longrightarrow -\infty$, which corresponds to the adiabatic switching on of these fields. In [7, p. 84], it is shown that, while averaging, one should use the wave functions $|0\rangle$ of the ground state without taking into consideration the interaction U. For the averaged values of the current and charge densities, we introduce the notations $\langle 0|\mathbf{j}_1(\mathbf{r}, t)|0\rangle$ and $\langle 0|\rho_1(\mathbf{r}, t)|0\rangle$. In (14) and (15), we change the integration variable t' to $t'' = t' - t$. When averaging $\langle 0|\ldots|0\rangle$, we take into consideration that

$$\langle 0|\hat{F}(t)|0\rangle = \langle 0|e^{i\mathcal{H}t/\hbar}\hat{F}e^{-i\mathcal{H}t/\hbar}|0\rangle = \langle 0|\hat{F}|0\rangle, \quad (16)$$

where $\hat{F}$ is an arbitrary operator. We obtain

$$\begin{aligned}\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle &= -\frac{e}{mc}\langle 0|\rho(\mathbf{r})|0\rangle A_\alpha(\mathbf{r}, t) + \frac{i}{\hbar c}\\ &\times \int d^3r' \int_{-\infty}^{0} dt' \langle 0|[j_\alpha(\mathbf{r}), j_\beta(\mathbf{r}', t')]|0\rangle A_\beta(\mathbf{r}', t+t')\\ &- \frac{i}{\hbar}\int d^3r' \int_{-\infty}^{0} dt' \langle 0|[j_\alpha(\mathbf{r}), \rho(\mathbf{r}', t')]|0\rangle \varphi(\mathbf{r}', t+t'),\end{aligned} \quad (17)$$

$$\langle 0|\rho_1(\mathbf{r}, t)|0\rangle = \frac{i}{\hbar c}\int d^3r' \times \int_{-\infty}^{0} dt' \langle 0|[\rho(\mathbf{r}, t), j_\beta(\mathbf{r}', t')]|0\rangle A_\beta(\mathbf{r}', t+t') - \frac{i}{\hbar}\int d^3r' \int_{-\infty}^{0} dt' \langle 0|[\rho(\mathbf{r}), \rho(\mathbf{r}', t')]|0\rangle \varphi(\mathbf{r}', t+t'). \quad (18)$$

Thus, we have obtained expressions for the averaged, over the ground state, current and charge densities in terms of the vector and scalar potentials. However, the averaged values of any quantities must be expressed through measurable quantities—the fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{H}(\mathbf{r}, t)$ and their derivatives. Below, we express the quantities $\langle 0|\mathbf{j}_1(\mathbf{r}, t)|0\rangle$ and $\langle 0|\rho_1(\mathbf{r}, t)|0\rangle$ in terms of fields.

5. TIME DERIVATIVES OF MEAN DENSITIES OF CURRENT AND CHARGE

Let us calculate the time derivatives of (17) and (18):

$$\frac{\partial}{\partial t}\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle = \frac{\partial}{\partial t}\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_A + \frac{\partial}{\partial t}\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_\varphi,$$
$$\frac{\partial}{\partial t}\langle 0|\rho_1(\mathbf{r}, t)|0\rangle = \frac{\partial}{\partial t}\langle 0|\rho_1(\mathbf{r}, t)|0\rangle_A + \frac{\partial}{\partial t}\langle 0|\rho_1(\mathbf{r}, t)|0\rangle_\varphi, \quad (19)$$

where the subscripts A and φ denote the contributions of the vector and scalar potentials, respectively, which are given by

$$\frac{\partial}{\partial t}\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_A = -\frac{e}{mc}\langle 0|\rho(\mathbf{r})|0\rangle\frac{\partial A_\alpha(\mathbf{r}, t)}{\partial t} + \frac{i}{\hbar c}\int d^3r' \int_{-\infty}^{0} dt' \langle 0|[j_\alpha(\mathbf{r}), j_\beta(\mathbf{r}', t')]|0\rangle \frac{\partial A_\beta(\mathbf{r}', t+t')}{\partial t}, \quad (20)$$

$$\frac{\partial}{\partial t}\langle 0|\rho_1(\mathbf{r}, t)|0\rangle_A = \frac{i}{\hbar c}\int d^3r' \times \int_{-\infty}^{0} dt' \langle 0|[\rho(\mathbf{r}), j_\beta(\mathbf{r}', t')]|0\rangle \frac{\partial A_\beta(\mathbf{r}', t+t')}{\partial t}, \quad (21)$$

$$\frac{\partial}{\partial t}\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_\varphi = -\frac{i}{\hbar}\int d^3r' \times \int_{-\infty}^{0} dt' \langle 0|[j_\alpha(\mathbf{r}, t), \rho(\mathbf{r}', t')]|0\rangle \frac{\partial \varphi(\mathbf{r}', t+t')}{\partial t}, \quad (22)$$

$$\frac{\partial}{\partial t}\langle 0|\rho_1(\mathbf{r}, t)|0\rangle_\varphi = -\frac{i}{\hbar}\int d^3r' \times \int_{-\infty}^{0} dt' \langle 0|[\rho(\mathbf{r}), \rho(\mathbf{r}', t')]|0\rangle \frac{\partial \varphi(\mathbf{r}', t+t')}{\partial t}. \quad (23)$$

Let us transform expressions (22) and (23), which contain the scalar potential φ. We apply the identity

$$\frac{\partial \varphi(\mathbf{r}, t+t')}{\partial t} = \frac{\partial \varphi(\mathbf{r}, t+t')}{\partial t'},$$

and then integrate by parts with respect to t'. We obtain

$$\frac{\partial}{\partial t}\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_\varphi = -\frac{i}{\hbar}\int d^3r' \langle 0|[j_\alpha(\mathbf{r}), \rho(\mathbf{r}')]|0\rangle \varphi(\mathbf{r}', t) + \frac{i}{\hbar}\int d^3r' \int_{-\infty}^{0} dt' \langle 0|\left[j_\alpha(\mathbf{r}), \frac{\partial \rho(\mathbf{r}', t')}{\partial t'}\right]|0\rangle \varphi(\mathbf{r}', t+t'). \quad (24)$$

In the first term in the right-hand side of (24), we integrate with respect to $\mathbf{r}'$ and use the explicit expression (6) for the operators $j(\mathbf{r})$ and $\rho(\mathbf{r})$. To calculate the second term, we apply the continuity equation

$$\nabla \mathbf{j}(\mathbf{r}, t) + \frac{\partial \rho(\mathbf{r}, t)}{\partial t} = 0, \quad (25)$$

which is valid for the operators defined in (10) with regard to the constant magnetic field $\mathbf{H}_c$. Next, we integrate by parts with respect to r'_β in this term, carrying over the differentiation operation to the scalar potential $\varphi(\mathbf{r}', t+t')$. As a result, we obtain

$$\frac{\partial}{\partial t}\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_\varphi = -\frac{e}{m}\langle 0|\rho(\mathbf{r})|0\rangle\frac{\partial \varphi(\mathbf{r}, t)}{\partial r_\alpha} + \frac{i}{\hbar}\int d^3r' \int_{-\infty}^{0} dt' \langle 0|[j_\alpha(\mathbf{r}), j_\beta(\mathbf{r}', t')]|0\rangle \frac{\partial \varphi(\mathbf{r}', t+t')}{\partial r'_\beta}. \quad (26)$$

Adding up (20) and (26) and using (1), we obtain

$$\frac{\partial}{\partial t}\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle = \frac{e}{m}\langle 0|\rho(\mathbf{r})|0\rangle E_\alpha(\mathbf{r}, t) - \frac{i}{\hbar}\int d^3r' \times \int_{-\infty}^{0} dt' \langle 0|[j_\alpha(\mathbf{r}), j_\beta(\mathbf{r}', t')]|0\rangle E_\alpha(\mathbf{r}', t+t'). \quad (27)$$

Completely analogously, we obtain

$$\frac{\partial \langle 0|\rho_1(\mathbf{r}, t)|0\rangle}{\partial t} = -\frac{i}{\hbar}\int d^3r' \times \int_{-\infty}^{0} dt' \langle 0|[\rho(\mathbf{r}), j_\beta(\mathbf{r}', t')]|0\rangle E_\beta(\mathbf{r}', t+t'). \quad (28)$$

Thus, we have expressed the time derivatives of the mean densities of current and charge through electric

fields, getting rid of vector and scalar potentials.

6. MEAN VALUES OF CURRENT AND CHARGE DENSITIES EXPRESSED IN TERMS OF ELECTRIC FIELD

Integrating (27) and (28) with respect to time, we obtain the following expressions for the mean densities of current and charge:

$$\begin{aligned}\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle &= \int_{-\infty}^{t} dt' \frac{\partial \langle 0|[j_{1,\alpha}(\mathbf{r}, t')]|0\rangle}{\partial t'} + C_\alpha,\\ \langle 0|\rho_1(\mathbf{r}, t)|0\rangle &= \int_{-\infty}^{t} dt' \frac{\partial \langle 0|\rho_1(\mathbf{r}, t')|0\rangle}{\partial t'} + C';\end{aligned} \tag{29}$$

here, we assume that $C_\alpha = C' = 0$, which corresponds to the absence of induced currents and charges in the infinite past.

Introduce the notation

$$\mathbf{a}(\mathbf{r}, t) = -c\int_{-\infty}^{t} dt' \mathbf{E}(\mathbf{r}, t'). \tag{30}$$

Then, using (27)–(29), we obtain

$$\begin{aligned}\langle 0|[j_{1\alpha}(\mathbf{r}, t)]|0\rangle &= -\frac{e}{mc}\langle 0|\rho(\mathbf{r}, t)|0\rangle a_\alpha(\mathbf{r}, t)\\ &+ \frac{i}{\hbar c}\int d^3r' \int_{-\infty}^{t} dt' \langle 0|[j_\alpha(\mathbf{r}, t), j_\beta(\mathbf{r}', t')]|0\rangle a_\beta(\mathbf{r}', t'),\end{aligned} \tag{31}$$

$$\begin{aligned}\langle 0|\rho_1(\mathbf{r}, t)|0\rangle &= \frac{i}{\hbar c}\int d^3r'\\ &\times \int_{-\infty}^{t} dt' \langle 0|[\rho(\mathbf{r}, t), j_\beta(\mathbf{r}', t')]|0\rangle a_\beta(\mathbf{r}', t').\end{aligned} \tag{32}$$

Comparing (31) and (32) with (17) and (18), we can see that the latter pair of equations differs from the former by the absence of scalar potential and the replacement of the vector potential $\mathbf{A}(\mathbf{r}, t)$ by the vector $\mathbf{a}(\mathbf{r}, t)$ defined in (30). Thus, we have expressed the mean values of induced current and charge densities through the electric field alone. However, expressions (31) and (32) cannot be applied without further transformations when passing to a time-independent electric field since the integration with respect to t in (30) leads to uncertainty: the frequency ω in the denominator vanishes. This also applies to the case of $\mathbf{E} = 0$, $\mathbf{H} \neq 0$. We will return to this case in Section 12.

7. TRANSFORMATION OF EXPRESSIONS FOR THE MEAN DENSITIES OF CURRENT AND CHARGE

Now, we pass to the expressions that explicitly correspond to the interaction of electrons with a field of the form

$$\tilde{U}_1 = -e\sum_i r_{i\beta}E_\beta(t), \tag{33}$$

which, for example, was used by Kubo [8] in the case of a coordinate-independent electric field $\mathbf{E}(t)$. Let us transform expressions (31) and (32). Introduce fictitious operators $\mathbf{j}_1^f(\mathbf{r}, t)$ and $\rho_1^f(\mathbf{r}, t)$ that satisfy the relations

$$\begin{aligned}\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle &= \left\langle 0\middle|j_{1\alpha}^{\mathrm{f}}(\mathbf{r}, t)\middle|0\right\rangle,\\ \langle 0|\rho_1(\mathbf{r}, t)|0\rangle &= \left\langle 0\middle|\rho_1^{\mathrm{f}}(\mathbf{r}, t)\middle|0\right\rangle.\end{aligned} \tag{34}$$

From (31) and (32), we obtain

$$\begin{aligned}j_{1\alpha}^{\mathrm{f}}(\mathbf{r}, t) &= -\frac{e}{mc}\rho(\mathbf{r}, t)a_\alpha(\mathbf{r}, t)\\ &+ \frac{i}{\hbar c}\int d^3r' \int_{-\infty}^{t} dt' [j_\alpha(\mathbf{r}, t), j_\beta(\mathbf{r}', t')]a_\beta(\mathbf{r}', t'),\end{aligned} \tag{35}$$

$$\begin{aligned}\rho_1^{\mathrm{f}}(\mathbf{r}, t) &= \frac{i}{\hbar c}\int d^3r'\\ &\times \int_{-\infty}^{t} dt' [\rho(\mathbf{r}, t), j_\beta(\mathbf{r}', t')]a_\beta(\mathbf{r}', t').\end{aligned} \tag{36}$$

Comparing fictitious operators (35) and (36) with real operators (14) and (15), we conclude that, to pass from the real to the fictitious operators, one has to set $\varphi(\mathbf{r}, t) = 0$ and replace the vector potential $\mathbf{A}(\mathbf{r}, t)$ by $\mathbf{a}(\mathbf{r}, t)$, defined in (30).

We also introduce a fictitious operator of interaction of particles with a field,

$$U_1^{\mathrm{f}} = -\frac{1}{c}\int d^3r\ \mathbf{j}(\mathbf{r})\mathbf{a}(\mathbf{r}, t), \tag{37}$$

which differs from the operator U_1 defined in (5) by the condition $\varphi(\mathbf{r}, t) = 0$ and the replacement of $\mathbf{A}(\mathbf{r}, t)$ by $\mathbf{a}(\mathbf{r}, t)$. Interaction (37) corresponds to the linear (in the field) addition

$$\Delta j_\alpha^{\mathrm{f}} = -\frac{e}{mc}\rho(\mathbf{r})a_\alpha(\mathbf{r}, t) \tag{38}$$

to the current-density operator. It is obvious that (35)

and (36) can be rewritten as

$$j^{\mathrm{f}}_{1\alpha}(\mathbf{r}, t) = \Delta j^{\mathrm{f}}_{\alpha}(\mathbf{r}, t) - \frac{i}{\hbar}\int\limits_{-\infty}^{t} dt'[j_{\alpha}(\mathbf{r}, t), U^{\mathrm{f}}_{1}(t')], \quad (39)$$

$$\rho^{\mathrm{f}}_{1}(\mathbf{r}, t) = -\frac{i}{\hbar}\int\limits_{-\infty}^{t} dt'[\rho(\mathbf{r}, t), U^{\mathrm{f}}_{1}(t')], \quad (40)$$

which is analogous to (13), in which real operators are replaced by fictitious ones. Let us transform (39) and (40) so that the first term in the right-hand side of (39) vanishes. The integrals of the form

$$\int\limits_{-\infty}^{t} dt'[F(\mathbf{r}, t), U^{\mathrm{f}}_{1}(t')]$$

in the right-hand sides of (39) and (40) can be rewritten as

$$\int\limits_{-\infty}^{t} dt'[F(\mathbf{r}, t), \tilde{U}_1(t')] - [F(\mathbf{r}, t), R(t)] \quad (41)$$

if

$$\tilde{U}_1 = U^{f}_{1}(t) + \frac{dR(t)}{dt}, \quad (42)$$

where $F(\mathbf{r}, t)$ is an operator equal to $j_{\alpha}(\mathbf{r}, t)$ for (39) and $\rho(\mathbf{r}, t)$ for (40); $R(t)$ is an arbitrary operator in the interaction representation,

$$R(t) = e^{i\mathcal{H}t/\hbar}R_{\mathrm{Sch}}(\mathbf{r})e^{-i\mathcal{H}t/\hbar}; \quad (43)$$

and R_{Sch} is the operator in the Schrödinger representation. One can show that, if

$$R_{\mathrm{Sch}} = R_{\mathrm{Sch}}(\mathbf{r}_1, \ldots, \mathbf{r}_N, t), \quad (44)$$

i.e., if the operator R_{Sch} does not contain momenta, then the relation

$$\frac{i}{\hbar}[j_{\alpha}(\mathbf{r}, t), R(t)] = \Delta j_{R\alpha}(\mathbf{r}, t), \quad (45)$$

holds, where

$$\Delta j_{R\alpha}(\mathbf{r}, t) = \frac{e}{2}e^{i\mathcal{H}t/\hbar} \times \sum_i [\Delta v_{iR\alpha}\delta(\mathbf{r} - \mathbf{r}_i) + \delta(\mathbf{r} - \mathbf{r}_i)\Delta v_{iR\alpha}]e^{-i\mathcal{H}t/\hbar}, \quad (46)$$

$$\Delta v_{iR\alpha} = \frac{i}{\hbar}[\dot{R}_{\mathrm{Sch}}, r_{i\alpha}], \quad (47)$$

$$\dot{R}_{\mathrm{Sch}} = \frac{i}{\hbar}[\mathcal{H}, R_{\mathrm{Sch}}] + \frac{\partial R_{\mathrm{Sch}}}{\partial t}. \quad (48)$$

It is also obvious that, under condition (44),

$$[\rho(\mathbf{r}, t), R(t)] = e^{i\mathcal{H}t/\hbar}[\rho(\mathbf{r}), R_{\mathrm{Sch}}]e^{-i\mathcal{H}t/\hbar} = 0. \quad (49)$$

Thus, it is proved that, instead of $U^{\mathrm{f}}_{1}(t)$, one can choose any operator defined in (42) if R_{Sch} does not contain momenta and, instead of (39) and (40), one can write

$$j^{\mathrm{f}}_{1\alpha}(\mathbf{r}, t) = \Delta\tilde{j}_{\alpha}(\mathbf{r}, t) - \frac{i}{\hbar}\int\limits_{-\infty}^{t} dt'[j_{\alpha}(\mathbf{r}, t), \tilde{U}_1(t')], \quad (50)$$

$$\rho^{\mathrm{f}}_{1}(\mathbf{r}, t) = -\frac{i}{\hbar}\int\limits_{-\infty}^{t} dt'[\rho(\mathbf{r}, t), \tilde{U}_1(t')], \quad (51)$$

where

$$\Delta\tilde{j}_{\alpha}(\mathbf{r}) = \frac{e}{2}\sum_i [\Delta\tilde{v}_{i\alpha}\delta(\mathbf{r} - \mathbf{r}_i) + \delta(\mathbf{r} - \mathbf{r}_i)\Delta\tilde{v}_{i\alpha}], \quad (52)$$

$$\Delta\tilde{v}_{i\alpha} = \frac{i}{\hbar}[\tilde{U}_1, r_{i\alpha}]. \quad (53)$$

Substituting (42) and (52) into (50) and (51), we obtain

$$j^{f}_{1\alpha}(\mathbf{r}, t) = -\frac{e}{mc}\rho(\mathbf{r}, t)a_{\alpha}(\mathbf{r}, t) + \frac{i}{\hbar}[j_{\alpha}(\mathbf{r}, t), R(t)] + \frac{i}{\hbar c}\int d^3r' \int\limits_{-\infty}^{t} dt'[j_{\alpha}(\mathbf{r}, t), j_{\beta}(\mathbf{r}', t')]a_{\beta}(\mathbf{r}', t') - \frac{i}{\hbar}\int\limits_{-\infty}^{t} dt'\left[j_{\alpha}(\mathbf{r}, t), \frac{dR(t')}{dt'}\right], \quad (54)$$

$$\rho^{\mathrm{f}}_{1\alpha}(\mathbf{r}, t) = \frac{i}{\hbar c}\int d^3r' \int\limits_{-\infty}^{t} dt'[\rho(\mathbf{r}, t), j_{\beta}(\mathbf{r}', t')]a_{\beta}(\mathbf{r}', t') - \frac{i}{\hbar}\int\limits_{-\infty}^{t} dt'\left[\rho(\mathbf{r}, t), \frac{dR(t')}{dt'}\right]. \quad (55)$$

8. CHOICE OF THE OPERATOR R_{Sch}

Choose the operator R_{Sch} in the form

$$R_{\mathrm{Sch}} = \frac{e}{c}\sum_i r_{i\beta}a_{\beta}(\mathbf{r}_i, t). \quad (56)$$

Then,

$$R(t) = e^{i\mathcal{H}t/\hbar}R_{\mathrm{Sch}}(\mathbf{r})e^{-i\mathcal{H}t/\hbar} = \frac{1}{c}\int d^3r\, d_{\beta}(\mathbf{r}, t)a_{\beta}(\mathbf{r}, t), \quad (57)$$

where

$$\mathbf{d}(\mathbf{r}) = e\sum_i \mathbf{r}_i\delta(\mathbf{r}-\mathbf{r}_i) = e\mathbf{r}\sum_i \delta(\mathbf{r}-\mathbf{r}_i) = \mathbf{r}\rho(\mathbf{r}). \quad (58)$$

Let us calculate the terms in the right-hand sides of (54) and (55) that contain $R(t)$ and $dR(t)/dt$. We can show that

$$\frac{i}{\hbar}[j_\alpha(\mathbf{r}, t), R(t)] = \frac{e}{mc}\rho(\mathbf{r}, t)a_\alpha(\mathbf{r}, t) + \frac{e}{mc}d_\beta(\mathbf{r}, t)\frac{\partial a_\beta(\mathbf{r}, t)}{\partial r_\alpha}. \quad (59)$$

In addition, we can obtain

$$\frac{dR(t)}{dt} = \frac{1}{c}\int d^3r\, d_\beta(\mathbf{r}, t)\frac{\partial a_\beta(\mathbf{r}, t)}{\partial t} + \frac{1}{c}\int d^3r\, j_\beta(\mathbf{r}, t)a_\beta(\mathbf{r}, t) + \frac{1}{c}\int d^3r Y_{\beta\gamma}(\mathbf{r}, t)\frac{\partial a_\beta(\mathbf{r}, t)}{\partial r_\gamma}, \quad (60)$$

where

$$Y_{\beta\gamma}(\mathbf{r}) = r_\beta j_\gamma(\mathbf{r}). \quad (61)$$

Substituting (59) and (60) into the right-hand side of Eqs. (54) and (55), we find that the terms that do not contain the derivatives $\partial a_\beta(\mathbf{r}, t)/\partial t$ or $\partial a_\beta(\mathbf{r}, t)/\partial r_\alpha$ cancel out. As a result, we have

$$j^{\mathrm{f}}_{1\alpha} = \frac{e}{mc}d_\beta(\mathbf{r}, t)\frac{\partial a_\beta(\mathbf{r}, t)}{\partial r_\alpha} - \frac{i}{\hbar c}\int d^3r'\int_{-\infty}^{t} dt'[j_\alpha(\mathbf{r}, t), d_\beta(\mathbf{r}', t')]\frac{\partial a_\beta(\mathbf{r}', t')}{\partial t'} - \frac{i}{\hbar c}\int d^3r'\int_{-\infty}^{t} dt'[j_\alpha(\mathbf{r}, t), Y_{\beta\gamma}(\mathbf{r}', t')]\frac{\partial a_\beta(\mathbf{r}', t')}{\partial r'_\gamma}, \quad (62)$$

$$\rho^{\mathrm{f}}_1(\mathbf{r}, t) = -\frac{i}{\hbar c}\int d^3r'\int_{-\infty}^{t} dt'[\rho(\mathbf{r}, t), d_\beta(\mathbf{r}', t')]\frac{\partial a_\beta(\mathbf{r}', t')}{\partial t'} - \frac{i}{\hbar c}\int d^3r'\int_{-\infty}^{t} dt'[\rho(\mathbf{r}, t), Y_{\beta\gamma}(\mathbf{r}', t')]\frac{\partial a_\beta(\mathbf{r}', t')}{\partial r'_\gamma}. \quad (63)$$

Averaging the fictitious operators (62) and (63) over the ground state, we obtain the required expressions for the mean densities of induced current and charge.

Taking into account definition (30), we find that the terms in the right-hand sides of (62) and (63) fall into two categories; the first includes the terms that contain electric field, while the second includes the terms that contain the derivatives of this field with respect to coordinates. Therefore, it is convenient to represent the mean values of (62) and (63) as

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle = \langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_E + \langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_{\partial E/\partial r'}, \quad (64)$$

$$\langle 0|\rho_1(\mathbf{r}, t)|0\rangle = \langle 0|\rho_1(\mathbf{r}, t)|0\rangle_E + \langle 0|\rho_1(\mathbf{r}, t)|0\rangle_{\partial E/\partial r}, \quad (65)$$

where

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_E = \frac{i}{\hbar}\int d^3r' \times \int_{-\infty}^{t} dt'\langle 0|[j_\alpha(\mathbf{r}, t), d_\beta(\mathbf{r}', t')]|0\rangle E_\beta(\mathbf{r}', t'), \quad (66)$$

$$\langle 0|\rho_1(\mathbf{r}, t)|0\rangle_E = \frac{i}{\hbar}\int d^3r' \times \int_{-\infty}^{t} dt'\langle 0|[\rho(\mathbf{r}, t), d_\beta(\mathbf{r}', t')]|0\rangle E_\beta(\mathbf{r}', t'), \quad (67)$$

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_{\partial E/\partial r} = -\frac{e}{m}\langle 0|d_\beta(\mathbf{r})|0\rangle\int_{-\infty}^{t} dt'\frac{\partial E_\beta(\mathbf{r}, t')}{\partial r_\alpha} + \frac{i}{\hbar}\int d^3r'\int_{-\infty}^{t} dt'\langle 0|[j_\alpha(\mathbf{r}, t), Y_{\beta\gamma}(\mathbf{r}', t')]|0\rangle \times \int_{-\infty}^{t'} dt''\frac{\partial E_\beta(\mathbf{r}', t'')}{\partial r'_\gamma}, \quad (68)$$

$$\langle 0|\rho_1(\mathbf{r}, t)|0\rangle_{\partial E/\partial r} = \frac{i}{\hbar}\int d^3r'\int_{-\infty}^{t} dt'\langle 0|[\rho(\mathbf{r}, t), Y_{\alpha\beta}(\mathbf{r}', t')]|0\rangle\int_{-\infty}^{t'} dt''\frac{\partial E_\beta(\mathbf{r}', t'')}{\partial r'_\gamma}. \quad (69)$$

Thus, the problem of expressing the current and charge densities, averaged over the ground state of the system, in the linear approximation with respect to electric field and its coordinate derivatives is solved.

9. TRANSFORMED FICTITIOUS OPERATOR OF INTERACTION

To complete the picture, we define the form of the fictitious operator of interaction $\tilde{U}_1 = U_1^f + \dot{R}_{\mathrm{Sch}}$. Using (37) and (60), we obtain the following relations in the Schrödinger representation:

$$\tilde{U}_1 = \tilde{U}_{1E} + \tilde{U}_{1\partial E/\partial r}, \quad (70)$$

$$\tilde{U}_{1E} = -\int d^3r\, d_\beta(\mathbf{r})E_\beta(\mathbf{r}, t), \quad (71)$$

$$\tilde{U}_{1\partial E/\partial r} = -\int d^3r Y_{\beta\gamma}(\mathbf{r}) \int_{-\infty}^{t} dt' \frac{\partial E_\beta(\mathbf{r}, t')}{\partial r_\gamma}. \quad (72)$$

Integrating with respect to $\mathbf{r}$ in the right-hand sides of (71) and (72), we obtain

$$\tilde{U}_{1E} = -e\sum_i r_{i\beta} E_\beta(\mathbf{r}_i, t), \quad (73)$$

$$\tilde{U}_{1\partial E/\partial r} = -\frac{e}{4}\int_{-\infty}^{t} dt' \sum_i \left\{ \left(v_{i\gamma} \frac{\partial E_\beta(\mathbf{r}_i, t')}{\partial r_{i\gamma}} + \frac{\partial E_\beta(\mathbf{r}_i, t')}{\partial r_{i\gamma}} v_{i\gamma} \right) r_{i\beta} \right.$$
$$\left. + r_{i\beta} \left(v_{i\gamma} \frac{\partial E_\beta(\mathbf{r}_i, t')}{\partial r_{i\gamma}} + \frac{\partial E_\beta(\mathbf{r}_i, t')}{\partial r_{i\gamma}} v_{i\gamma} \right) \right\} \quad (74)$$
$$= -\frac{e}{2}\int_{-\infty}^{t} dt' \sum_i \left\{ v_{i\gamma} \frac{\partial E_\beta(\mathbf{r}_i, t')}{\partial r_{i\gamma}} r_{i\beta} + r_{i\beta} \frac{\partial E_\beta(\mathbf{r}_i, t')}{\partial r_{i\gamma}} v_{i\gamma} \right\}.$$

Thus, the term $\tilde{U}_{1E}$ contains electric field, while $\tilde{U}_{1\partial E/\partial r}$ contains the derivatives of electric field with respect to coordinate and the integral with respect to time. When the contribution of the terms containing the coordinate derivatives of the electric field is small for one or another reason,[1] one can use expression (73). This expression coincides with the formula for the interaction of charged particles with a field [8] (see also [9, 10]). Note that expression (70) for the fictitious interaction $\tilde{U}_1$ can be rewritten in the following compact form:

$$\tilde{U}_1 = \frac{e}{2c}\sum_i \left\{ \frac{da_\beta(\mathbf{r}_i, t)}{dt} r_{i\beta} + r_{i\beta} \frac{da_\beta(\mathbf{r}_i, t)}{dt} \right\}, \quad (75)$$

where $da_\beta(\mathbf{r}_i, t)/dt$ is the total time derivative:

$$\frac{da_\beta(\mathbf{r}_i, t)}{dt} = \frac{\partial a_\beta(\mathbf{r}_i, t)}{\partial t} + \frac{i}{\hbar}[\mathcal{H}, a_\beta(\mathbf{r}_i, t)], \quad (76)$$

and

$$\frac{i}{\hbar}[\mathcal{H}, a_\beta(\mathbf{r}_i, t)] = \frac{1}{2}\left\{ v_{i\gamma} \frac{\partial a_\beta(\mathbf{r}_i, t)}{\partial r_{i\gamma}} + \frac{\partial a_\beta(\mathbf{r}_i, t)}{\partial r_{i\gamma}} v_{i\gamma} \right\}. \quad (77)$$

The interaction (70) corresponds to the following linear (in the field) addition to the velocity:

$$\Delta\tilde{v}_{i\alpha} = \frac{e}{mc} r_{i\beta} \frac{\partial a_\beta(\mathbf{r}_i, t)}{\partial r_{i\alpha}}, \quad (78)$$

which, according to (52), corresponds to the addition to the current density

$$\Delta\tilde{j}_\alpha(\mathbf{r}) = \frac{e}{mc} d_\beta(\mathbf{r}) \frac{\partial a_\beta(\mathbf{r}, t)}{\partial r_\alpha}. \quad (79)$$

Passing to the interaction representation, we obtain the first term in the right-hand side of (62).

10. ELIMINATION OF THE DIAGONAL ELEMENTS OF THE OPERATORS OF PARTICLE COORDINATES

Let us return to expressions (64)–(69) for the mean densities of induced current and charge for $T = 0$ that were obtained in Section 8. Taking into account definition (58) of the operator $\mathbf{d}(\mathbf{r})$, as well as definition (61) of the operator $Y_{\beta\gamma}(\mathbf{r})$, which can be rewritten as

$$Y_{\beta\gamma}(\mathbf{r}) = \frac{1}{2}\sum_i (j_{i\gamma} r_{i\beta} + r_{i\beta} j_{i\gamma}), \quad (80)$$

we can see that these operators contain the particle coordinates $\mathbf{r}_i$. However, the mean values $\langle 0|\mathbf{j}_1(\mathbf{r}, t)|0\rangle$ and $\langle 0|\rho_1(\mathbf{r}, t)|0\rangle$ must be independent of the coordinates $\mathbf{r}_i$. Let us transform expressions (64)–(69) so that the latter property becomes obvious. Divide vector $\mathbf{r}_i$ into two parts:

$$\mathbf{r}_i = \bar{\mathbf{r}}_i + \langle 0|\mathbf{r}_i|0\rangle. \quad (81)$$

It is obvious that the matrix elements of operator $\bar{\mathbf{r}}_i$, both diagonal and off-diagonal, are independent of change point readout of $\mathbf{r}_i$. Let us show that operators $\mathbf{r}_i$ in the expressions for the mean values $\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle$ and $\langle 0|\rho_1(\mathbf{r}, t)|0\rangle$ can be replaced by $\bar{\mathbf{r}}_i$. Rewrite the mean value $\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle$ as

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle = \overline{\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle} + \sum_i \langle 0|r_{i\beta}|0\rangle x_{i\alpha\beta}(\mathbf{r}, t), \quad (82)$$

where $\overline{\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle}$ represents the contribution of operators $\bar{\mathbf{r}}_i$ and

$$x_{i\alpha\beta}(\mathbf{r}, t) = \frac{i}{\hbar}\int d^3r' \int_{-\infty}^{t} dt' \langle 0|[j_\alpha(\mathbf{r}, t), \rho_i(\mathbf{r}', t')]|0\rangle E_\beta(\mathbf{r}', t')$$
$$+ \frac{e}{mc}\langle 0|\rho_i(\mathbf{r})|0\rangle \frac{\partial a_\beta(\mathbf{r}, t)}{\partial r_\alpha} \quad (83)$$
$$- \frac{i}{\hbar c}\int d^3r' \int_{-\infty}^{t} dt' \langle 0|[j_\alpha(\mathbf{r}, t), j_{i\gamma}(\mathbf{r}', t')]|0\rangle \frac{\partial a_\beta(\mathbf{r}', t')}{\partial r'_\gamma}.$$

Integrating with respect to $\mathbf{r}'$ in the first and third

[1] Strictly speaking, the Maxwell equations imply that, if the electric field depends on time, then it also depends on coordinates; i.e., the derivatives of the field components with respect to coordinates are different from zero.

terms on the right-hand side of (83), we obtain

$$x_{i\alpha\beta}(\mathbf{r}, t) = -\frac{ie}{\hbar c}\int_{-\infty}^{t} dt' \langle 0|\left[j_\alpha(\mathbf{r}, t), \frac{da_\beta(\mathbf{r}_i(t'), t')}{dt'}\right]|0\rangle + \frac{e}{mc}\langle 0|\rho_i(\mathbf{r})|0\rangle \frac{\partial a_\beta(\mathbf{r}, t)}{\partial r_\alpha}, \tag{84}$$

where we used the following expression for the total time derivative:

$$\frac{da_\beta(\mathbf{r}_i(t), t)}{dt} = \frac{d}{dt}(e^{i\mathcal{H}t/\hbar} a_\beta(\mathbf{r}_i, t) e^{-i\mathcal{H}t/\hbar}) = e^{i\mathcal{H}t/\hbar}\frac{da_\beta(\mathbf{r}_i, t)}{\partial t} e^{-i\mathcal{H}t/\hbar}, \tag{85}$$

while the derivative $da_\beta(\mathbf{r}_i, t)/dt$ is defined in (76). Integrating with respect to t' in the first term in the right-hand side of (84) and calculating the commutator $[j_\alpha(\mathbf{r}), a_\beta(\mathbf{r}_i, t)]$, we obtain

$$x_{i\alpha\beta}(\mathbf{r}, t) = 0, \tag{86}$$

which implies

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle = \overline{\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle}. \tag{87}$$

Similarly, we can obtain

$$\langle 0|\rho_1(\mathbf{r}, t)|0\rangle = \overline{\langle 0|\rho_1(\mathbf{r}, t)|0\rangle}. \tag{88}$$

Let us divide formulas (87) and (88) into two parts,

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle = \langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_{\mathrm{I}} + \langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_{\mathrm{II}}, \tag{89}$$

$$\langle 0|\rho_1(\mathbf{r}, t)|0\rangle = \langle 0|\rho_1(\mathbf{r}, t)|0\rangle_{\mathrm{I}} + \langle 0|\rho_1(\mathbf{r}, t)|0\rangle_{\mathrm{II}}, \tag{90}$$

where

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_{\mathrm{I}} = \frac{i}{\hbar}\int d^3r' \times \int_{-\infty}^{t} dt' \left\langle 0\left|[j_\alpha(\mathbf{r}, t), \bar{d}_\beta(\mathbf{r}', t')]\right|0\right\rangle E_\beta(\mathbf{r}', t'), \tag{91}$$

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_{\mathrm{II}} = \frac{e}{mc}\left\langle 0\left|\bar{d}_\beta(\mathbf{r})\right|0\right\rangle \frac{da_\beta(\mathbf{r}, t)}{\partial r_\alpha} - \frac{i}{c\hbar}\int d^3r' \times \int_{-\infty}^{t} dt' \left\langle 0\left|[j_\alpha(\mathbf{r}, t)\bar{Y}_{\beta\gamma}(\mathbf{r}', t')]\right|0\right\rangle \frac{\partial a_\beta(\mathbf{r}', t')}{\partial r'_\gamma}, \tag{92}$$

$$\langle 0|\rho_1(\mathbf{r}, t)|0\rangle_{\mathrm{I}} = \frac{i}{\hbar}\int d^3r' \times \int_{-\infty}^{t} dt' \left\langle 0\left|[\rho(\mathbf{r}, t), \bar{d}_\beta(\mathbf{r}', t')]\right|0\right\rangle E_\beta(\mathbf{r}', t'), \tag{93}$$

$$\langle 0|\rho_1(\mathbf{r}, t)|0\rangle_{\mathrm{II}} = -\frac{i}{\hbar c}\int d^3r' \times \int_{-\infty}^{t} dt' \left\langle 0\left|[\rho(\mathbf{r}, t), \bar{Y}_{\beta\gamma}(\mathbf{r}', t')]\right|0\right\rangle \frac{\partial a_\beta(\mathbf{r}', t')}{\partial r'_\gamma}, \tag{94}$$

where

$$\bar{\mathbf{d}}(\mathbf{r}) = e\sum_i \bar{\mathbf{r}}_i \rho(\mathbf{r}), \tag{95}$$

$$\bar{Y}_{\beta\gamma}(\mathbf{r}) = \frac{1}{2}\sum_i (j_{i\gamma}\bar{r}_{i\beta} + \bar{r}_{i\beta} j_{i\gamma}). \tag{96}$$

Formulas (89)–(94) represent the main result of this paper. We emphasize that the partitions (89) and (90) of the mean values into two parts do not coincide with partitions (64) and (65).

The terms with subscript I will be called basic terms since they do not vanish when the electric field is independent of coordinates **r**. The terms with subscript II contain the coordinate derivatives of the electric field.

11. CONDUCTIVITY TENSOR THAT DEPENDS ON COORDINATES

First, consider the main part of the density of induced current, which is denoted by subscript I in (89). Rewrite (91) as

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_{\mathrm{I}} = \int d^3r \int_{-\infty}^{\infty} dt' l_{\alpha\beta}(\mathbf{r}, t, \mathbf{r}', t') E_\beta(\mathbf{r}', t'), \tag{97}$$

where

$$l_{\alpha\beta}(\mathbf{r}, t, \mathbf{r}', t') = \frac{i}{\hbar}\theta(t - t')\left\langle 0\left|[j_\alpha(\mathbf{r}, t), \bar{d}_\beta(\mathbf{r}', t')]\right|0\right\rangle \tag{98}$$

and θ is a step function. Introduce the tensor

$$\sigma_{\mathrm{I}\alpha\beta}(\mathbf{r}', t'|\mathbf{r}, t) = l_{\alpha\beta}(\mathbf{r}, t, \mathbf{r} - \mathbf{r}', t - t'). \tag{99}$$

Here, the notation of vertical line is borrowed from [11]. Then, (97) can be rewritten as

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_{\mathrm{I}} = \int d^3r' \int_{-\infty}^{\infty} dt' \times \sigma_{\mathrm{I}\alpha\beta}(\mathbf{r}', t'|\mathbf{r}, t) E_\beta(\mathbf{r} - \mathbf{r}', t - t'), \tag{100}$$

where

$$\sigma_{\mathrm{I}\alpha\beta}(\mathbf{r}', t'|\mathbf{r}, t) = \frac{i}{\hbar}\theta(t')\left\langle 0\left|[j_\alpha(\mathbf{r}, t), \bar{d}_\beta(\mathbf{r} - \mathbf{r}', t - t')]\right|0\right\rangle. \tag{101}$$

It is clear from (101) that the tensor $\sigma_{I\alpha\beta}(\mathbf{r}', t'|\mathbf{r}, t)$ is independent of t.

Now, we apply the Fourier transform. Let us write the electric field as

$$E_\alpha(\mathbf{r}, t) = E_\alpha^{(+)}(\mathbf{r}, t) + E_\alpha^{(-)}(\mathbf{r}, t), \tag{102}$$

where

$$E_\alpha^{(+)}(\mathbf{r}, t) = \frac{1}{(2\pi)^4}\int d^3k \int_0^\infty d\omega E_\alpha(\mathbf{k}, \omega) e^{i\mathbf{k}\cdot\mathbf{r} - i\omega t'}, \tag{103}$$

$$E_\alpha^{(-)}(\mathbf{r}, t) = (E_\alpha^{(+)}(\mathbf{r}, t))^*, \tag{104}$$

$$E_\alpha(\mathbf{k}, \omega) = \int d^3r \int_{-\infty}^\infty dt E_\alpha(\mathbf{r}, t) e^{-i\mathbf{k}\cdot\mathbf{r} + i\omega t}. \tag{105}$$

Let us introduce the Fourier image of the tensor $\sigma_{I\alpha\beta}(\mathbf{r}', t'|\mathbf{r}, 0)$ with respect to the variables $\mathbf{r}'$ and t':

$$\sigma_{I\alpha\beta}(\mathbf{k}, \omega|\mathbf{r}) = \int d^3r' \int_{-\infty}^\infty dt' \sigma_{I\alpha\beta}(\mathbf{r}', t'|\mathbf{r}, 0) e^{-i\mathbf{k}\cdot\mathbf{r}' + i\omega t'}. \tag{106}$$

Then,

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_I = \langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_I^{(+)} + \langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_I^{(-)}, \tag{107}$$

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_I^{(+)} = \frac{1}{(2\pi)^4}\int d^3k \int_0^\infty d\omega \times \sigma_{I\alpha\beta}(\mathbf{k}, \omega|\mathbf{r}) E_\beta(\mathbf{k}, \omega) e^{i\mathbf{k}\cdot\mathbf{r} - i\omega t}, \tag{108}$$

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_I^{(-)} = (\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_I^{(+)})^*. \tag{109}$$

Substituting (101) into (106), we obtain

$$\sigma_{I\alpha\beta}(\mathbf{k}, \omega|\mathbf{r}) = \frac{i}{\hbar}\int d^3r' \int_{-\infty}^\infty dt' \theta(t') e^{-i\mathbf{k}\cdot\mathbf{r}' + i\omega t'} \times \langle 0|[j_\alpha(\mathbf{r}), \bar{d}_\beta(\mathbf{r} - \mathbf{r}', -t')]|0\rangle. \tag{110}$$

In a similar way we can obtain the contribution of the term with subscript II to the conductivity. Finally, we have

$$\langle 0|j_\alpha(\mathbf{r}, t)|0\rangle^{(+)} = \frac{1}{(2\pi)^4}\int d^3k \int_0^\infty d\omega \times \sigma_{\alpha\beta}(\mathbf{k}, \omega|\mathbf{r}) E_\beta(\mathbf{k}, \omega) e^{i\mathbf{k}\cdot\mathbf{r} - i\omega t}, \tag{111}$$

$$\sigma_{\alpha\beta}(\mathbf{k}, \omega|\mathbf{r}) = \sigma_{I\alpha\beta}(\mathbf{k}, \omega|\mathbf{r}) + \sigma_{II\alpha\beta}(\mathbf{k}, \omega|\mathbf{r}). \tag{112}$$

The quantity $\sigma_{I\alpha\beta}(\mathbf{k}, \omega|\mathbf{r})$ is defined in (110),

$$\sigma_{II\alpha\beta}(\mathbf{k}, \omega|\mathbf{r}) = \frac{ek_\alpha}{m\omega}\langle 0|\bar{d}_\beta(\mathbf{r})|0\rangle - \frac{ik_\gamma}{m\omega}\int d^3r' \int_{-\infty}^\infty dt' \theta(t') e^{-i\mathbf{k}\cdot\mathbf{r}' + i\omega t'} \times \langle 0|[j_\alpha(\mathbf{r}), \bar{Y}_{\beta\gamma}(\mathbf{r} - \mathbf{r}', -t')]|0\rangle. \tag{113}$$

The conductivity tensor $\sigma_{\alpha\beta}(\mathbf{k}, \omega|\mathbf{r})$ is independent only of coordinates $\mathbf{r}$ when the system is spatially homogeneous. When analyzing semiconductor objects of reduced dimension, the dependence of the conductivity tensor on $\mathbf{r}$ is very substantial.

In our previous works [4, 12–14], we used formula (91) for calculating the density of induced current.

12. TRANSITION TO EXPRESSIONS CONTAINING MAGNETIC FIELD

Up to now, we have not considered the case when the electric field $\mathbf{E}$ is independent of time, in particular, the case $\mathbf{E} = 0$, $\mathbf{H}(\mathbf{r}, t) = \text{const}$. To consider the latter case, we transform the expressions for the mean densities of induced current and charge, obtained at the end of Section 8, by introducing into them magnetic field $\mathbf{H}(\mathbf{r}, t)$. To this end, we divide each of the quantities $\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_{\partial E/\partial r}$ and $\langle 0|\rho_1(\mathbf{r}, t)|0\rangle_{\partial E/\partial r}$ defined in (68) and (69), respectively, into two parts as follows:[2]

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_{\partial E/\partial r} = \langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle^{(+)} + \langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle^{(-)}, \tag{114}$$

$$\langle 0|\rho_1(\mathbf{r}, t)|0\rangle_{\partial E/\partial r} = \langle 0|\rho_1(\mathbf{r}, t)|0\rangle^{(+)} + \langle 0|\rho_1(\mathbf{r}, t)|0\rangle^{(-)}, \tag{115}$$

where

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle^{(\pm)} = \frac{e}{2mc}\langle 0|d_\beta(\mathbf{r})|0\rangle \times\left(\frac{\partial a_\beta(\mathbf{r}, t)}{\partial \mathbf{r}_\alpha} \pm \frac{\partial a_\alpha(\mathbf{r}, t)}{\partial \mathbf{r}_\beta}\right) - \frac{i}{2\hbar c}\int d\mathbf{r}' \int_{-\infty}^t dt' \langle 0|[j_\alpha(\mathbf{r}, t), Y_{\beta\gamma}(\mathbf{r}', t')]|0\rangle \times\left(\frac{\partial a_\beta(\mathbf{r}', t')}{\partial \mathbf{r}'_\gamma} \pm \frac{\partial a_\gamma(\mathbf{r}', t')}{\partial \mathbf{r}'_\beta}\right), \tag{116}$$

[2] Here, the superscripts "plus" and "minus" have nothing to do with analogous superscripts in Section 11.

$$\langle 0|\rho_1(\mathbf{r}, t)|0\rangle^{(\pm)} = -\frac{i}{2\hbar c}\int d\mathbf{r}' \int_{-\infty}^{t} dt' \times \langle 0|[\rho(\mathbf{r}, t), Y_{\beta\gamma}(\mathbf{r}', t')]|0\rangle \times\left(\frac{\partial a_\beta(\mathbf{r}', t')}{\partial \mathbf{r}'_\gamma} \pm \frac{\partial a_\gamma(\mathbf{r}', t')}{\partial \mathbf{r}'_\beta}\right). \quad (117)$$

First, we consider the contributions with the minus sign. Let us return to the vector $\mathbf{A}(\mathbf{r}, t)$ and scalar $\varphi(\mathbf{r}, t)$ potentials. Taking into account definition (30) of vector $\mathbf{a}(\mathbf{r}, t)$ and the first formula in (1), we obtain

$$\mathbf{a}(\mathbf{r}, t) = \mathbf{A}(\mathbf{r}, t) + c\int_{-\infty}^{t} dt' \frac{\partial \varphi(\mathbf{r}, t')}{\partial \mathbf{r}}. \quad (118)$$

Substituting (118) into the right-hand sides of the expressions for $\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle^{(-)}$ and $\langle 0|\rho_1(\mathbf{r}, t)|0\rangle^{(-)}$, we find that the contributions of the scalar potential φ vanish and that

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle^{(-)} = \frac{e}{2mc}\langle 0|d_\beta(\mathbf{r})|0\rangle \times\left(\frac{\partial A_\beta(\mathbf{r}, t)}{\partial \mathbf{r}_\alpha} - \frac{\partial A_\alpha(\mathbf{r}, t)}{\partial \mathbf{r}_\beta}\right) - \frac{i}{2\hbar c}\int d\mathbf{r}' \int_{-\infty}^{t} dt' \langle 0|[j_\alpha(\mathbf{r}, t), Y_{\beta\gamma}(\mathbf{r}', t')]|0\rangle \times\left(\frac{\partial A_\beta(\mathbf{r}', t')}{\partial \mathbf{r}'_\gamma} - \frac{\partial A_\gamma(\mathbf{r}', t')}{\partial \mathbf{r}'_\beta}\right), \quad (119)$$

$$\langle 0|\rho_1(\mathbf{r}, t)|0\rangle^{(-)} = -\frac{i}{2\hbar c}\int d\mathbf{r}' \int_{-\infty}^{t} dt' \times \langle 0|[\rho(\mathbf{r}, t), Y_{\beta\gamma}(\mathbf{r}', t')]|0\rangle \times\left(\frac{\partial A_\beta(\mathbf{r}', t')}{\partial \mathbf{r}'_\gamma} - \frac{\partial A_\gamma(\mathbf{r}', t')}{\partial \mathbf{r}'_\beta}\right). \quad (120)$$

Taking into account that $\mathbf{H}(\mathbf{r}, t) = \mathrm{curl}\,\mathbf{A}(\mathbf{r}, t)$ (the second equality in (1)), we easily obtain

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle^{(-)} = -\frac{e}{2mc}[\mathbf{H}(\mathbf{r}, t)\times\mathbf{r}]_\alpha\langle 0|\rho(\mathbf{r})|0\rangle + \frac{i}{2\hbar c}\int d\mathbf{r}' \int_{-\infty}^{t} dt' [\mathbf{H}(\mathbf{r}', t')\times\mathbf{r}']_\beta \times \langle 0|[j_\alpha(\mathbf{r}, t), j_\beta(\mathbf{r}', t')]|0\rangle, \quad (121)$$

$$\langle 0|\rho_1(\mathbf{r}, t)|0\rangle^{(-)} = \frac{i}{2\hbar c}\int d\mathbf{r}' \int_{-\infty}^{t} dt' \times [\mathbf{H}(\mathbf{r}', t')\times\mathbf{r}']_\beta\langle 0|[\rho(\mathbf{r}, t), j_\beta(\mathbf{r}', t')]|0\rangle. \quad (122)$$

Thus, we have expressed the quantities $\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle^{(-)}$ and $\langle 0|\rho_1(\mathbf{r}, t)|0\rangle^{(-)}$ in terms of a magnetic field.

Now, we transform expressions (116) and (117) for the terms with the superscript "plus." We will proceed as follows. Consider, for example, (116). The right-hand side contains

$$Y_{\beta\gamma}(\mathbf{r}') = r'_\beta j_\gamma(\mathbf{r}'). \quad (123)$$

For $j_\gamma(\mathbf{r})$, we can easily derive the relation

$$j_\gamma(\mathbf{r}) = \dot{d}_\gamma(\mathbf{r}) + \partial Y_{\gamma\delta}(\mathbf{r})/\partial r_\delta. \quad (124)$$

Let us substitute (123) and (124) into the second term in the right-hand side of (116); as a result, the latter equation is decomposed into two parts that come from $\partial Y_{\gamma\delta}(\mathbf{r}')/\partial r'_\delta$ and $\dot{d}_\gamma(\mathbf{r})$. In the first term, we integrate by parts with respect to r'_δ, while in the second, we integrate, also by parts, with respect to t'. As a result, we obtain

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle^{(+)} = -\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle^{(+)} - \frac{e}{mc}\langle 0|\rho(\mathbf{r})|0\rangle r_\beta r_\gamma \frac{\partial^2 a_\beta(\mathbf{r}, t)}{\partial r_\alpha \partial r_\gamma} + \frac{i}{\hbar c}\int d\mathbf{r}' r'_\beta r'_\gamma \int_{-\infty}^{t} dt' \langle 0|[j_\alpha(\mathbf{r}, t), j_\delta(\mathbf{r}', t')]|0\rangle \times \frac{\partial^2 a_\beta(\mathbf{r}', t')}{\partial r'_\gamma \partial r'_\delta} + \frac{i}{\hbar c}\int d\mathbf{r}' r'_\beta \int_{-\infty}^{t} dt' \times \langle 0|[j_\alpha(\mathbf{r}, t), d_\gamma(\mathbf{r}', t')]|0\rangle \frac{\partial^2 a_\beta(\mathbf{r}', t')}{\partial r'_\gamma \partial t'}. \quad (125)$$

Using the relation

$$-c^{-1}\partial a_\beta(\mathbf{r}, t)/\partial t = E_\beta(\mathbf{r}, t),$$

from (125) we finally derive

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle^{(+)} = -\frac{e}{2mc}\langle 0|\rho(\mathbf{r})|0\rangle r_\beta r_\gamma \frac{\partial^2 a_\beta(\mathbf{r}, t)}{\partial r_\alpha \partial r_\gamma} + \frac{i}{2\hbar c}\int d\mathbf{r}' r'_\beta r'_\gamma \times \int_{-\infty}^{t} dt' \langle 0|[j_\alpha(\mathbf{r}, t), j_\delta(\mathbf{r}', t')]|0\rangle \frac{\partial^2 a_\beta(\mathbf{r}', t')}{\partial r'_\gamma \partial r'_\delta} - \frac{i}{2\hbar}\int d\mathbf{r}' r'_\beta r'_\gamma \int_{-\infty}^{t} dt' \langle 0|[j_\alpha(\mathbf{r}, t), \rho(\mathbf{r}', t')]|0\rangle \frac{\partial E_\beta(\mathbf{r}', t')}{\partial r'_\gamma} \quad (126)$$

and, similarly,

$$\langle 0|\rho(\mathbf{r}, t)|0\rangle^{(+)} = \frac{i}{2\hbar c}\int d\mathbf{r}' r'_\beta r'_\gamma \times \int_{-\infty}^{t} dt' \langle 0[\rho(\mathbf{r}, t), j_\delta(\mathbf{r}', t')]0\rangle \frac{\partial^2 a_\beta(\mathbf{r}', t')}{\partial r'_\gamma \partial r'_\delta} \quad (127)$$

$$-\frac{i}{2\hbar}\int d\mathbf{r}' r'_\beta r'_\gamma \int_{-\infty}^{t} dt' \langle 0|[\rho(\mathbf{r}, t), \rho(\mathbf{r}', t')]|0\rangle \frac{\partial E_\beta(\mathbf{r}', t')}{\partial r'_\gamma}.$$

Note that, in contrast to expressions (68) and (69), formulas (126) and (127) contain only the second derivatives of the vector $\mathbf{a}(\mathbf{r}, t)$. The results obtained can be expressed in the following symmetric form:

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle^{(+)} = -\frac{e}{4mc}\langle 0|\rho(\mathbf{r})|0\rangle r_\beta r_\gamma \times \frac{\partial}{\partial r_\alpha}\left(\frac{\partial a_\beta(\mathbf{r}, t)}{\partial r_\gamma} + \frac{\partial a_\gamma(\mathbf{r}, t)}{\partial r_\beta}\right) + \frac{i}{4\hbar c}\int d\mathbf{r}' r'_\beta r'_\gamma \times \int_{-\infty}^{t} dt' \langle 0|[j_\alpha(\mathbf{r}, t), \Omega_{\beta\gamma}(\mathbf{r}', t')]|0\rangle, \quad (128)$$

$$\langle 0|\rho(\mathbf{r}, t)|0\rangle^{(+)} = \frac{i}{4\hbar c}\int d\mathbf{r}' r'_\beta r'_\gamma \times \int_{-\infty}^{t} dt' \langle 0|[\rho(\mathbf{r}, t), \Omega_{\beta\gamma}(\mathbf{r}', t')]|0\rangle, \quad (129)$$

where

$$\Omega_{\beta\gamma}(\mathbf{r}, t) = j_\delta(\mathbf{r}, t)\frac{\partial}{\partial r_\delta}\left(\frac{\partial a_\beta(\mathbf{r}, t)}{\partial r_\gamma} + \frac{\partial a_\gamma(\mathbf{r}, t)}{\partial r_\beta}\right) + \rho(\mathbf{r}, t)\frac{\partial}{\partial t}\left(\frac{\partial a_\beta(\mathbf{r}, t)}{\partial r_\gamma} + \frac{\partial a_\gamma(\mathbf{r}, t)}{\partial r_\beta}\right) \quad (130)$$

is a symmetric tensor.

13. THE CASE OF A CONSTANT MAGNETIC FIELD

Consider the case when the magnetic field is constant in time and space, $\mathbf{H}$ = const, and $\mathbf{E} = 0$. Recall that we have included a constant magnetic field in the unperturbed Hamiltonian $\mathcal{H}$ (see Section 2). However, here we do not include the field in the total Hamiltonian but assume that it is so weak that one can restrict the analysis to linear (in the field) contributions to the induced current and charge densities. Since $\mathbf{E} = 0$, the contributions with the subscript E to the mean densities vanish (see (66) and (67)). The contributions with the subscripts $\partial E/\partial r$ are divided into two parts, which we denoted by superscripts "plus" and "minus." The parts with the superscripts "plus" are equal to zero. This fact can easily be verified by choosing the vector and scalar potentials, for example, in the calibration [5],

$$\mathbf{A}(\mathbf{r}) = \frac{1}{2}[\mathbf{r}\times\mathbf{H}], \quad \varphi = 0, \quad (131)$$

and by applying (118), (126), and (127).

Thus, the remaining contributions are those with the superscripts "minus" defined in (121) and (122). Setting $\mathbf{H}(\mathbf{r}, t) = \mathbf{H}$ and changing the variable t' to $t'' = t' - t$, we obtain the following time-independent results:

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_H = -\frac{e}{2mc}[\mathbf{H}\times\mathbf{r}]_\alpha\langle 0|\rho(\mathbf{r})|0\rangle + \frac{i}{2\hbar c}\int d\mathbf{r}' \int_{-\infty}^{0} dt''[\mathbf{H}\times\mathbf{r}']_\beta\langle 0|[j_\alpha(\mathbf{r}), j_\beta(\mathbf{r}', t'')]|0\rangle, \quad (132)$$

$$\langle 0|\rho_1(\mathbf{r}, t)|0\rangle_H = \frac{i}{2\hbar c}\int d\mathbf{r}' \times \int_{-\infty}^{0} dt''[\mathbf{H}\times\mathbf{r}']_\beta\langle 0|[\rho(\mathbf{r}), j_\beta(\mathbf{r}', t'')]|0\rangle. \quad (133)$$

The subscripts H in the left-hand sides indicate that $\mathbf{H}$ = const and $\mathbf{E} = 0$. In these expressions, we pass from operators $\mathbf{r}_i$ to operators $\bar{\mathbf{r}}_i$, defined in (81). Applying a method similar to that described in Section 10, we obtain

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_H = -\frac{e}{2mc}[\mathbf{H}\times\langle 0|\bar{\mathbf{d}}(\mathbf{r})|0\rangle]_\alpha + \frac{ie}{2\hbar c}\int_{-\infty}^{0} dt' \sum_i \{\langle 0|[j_\alpha(\mathbf{r}, -t'), [\mathbf{H}\times\bar{\mathbf{r}}_i]_\beta v_{i\beta}]|0\rangle\}, \quad (134)$$

$$\langle 0|\rho_1(\mathbf{r}, t)|0\rangle_H = \frac{ie}{2\hbar c}\int_{-\infty}^{0} dt' \times \sum_i \{\langle 0|[\rho(\mathbf{r}, -t'), [\mathbf{H}\times\bar{\mathbf{r}}_i]_\beta v_{i\beta}]|0\rangle\}. \quad (135)$$

Note that the following relations hold:

$$\mathrm{div}\langle 0|\mathbf{j}_1(\mathbf{r}, t)|0\rangle_H = 0,$$

$$\int d\mathbf{r}\langle 0|\mathbf{j}_1(\mathbf{r}, t)|0\rangle_H = 0.$$

14. CONCLUSIONS

Let us summarize the main results. Since the operator of interaction between charged particles and an electromagnetic field is expressed in terms of potentials $\mathbf{A}(\mathbf{r}, t)$ and $\varphi(\mathbf{r}, t)$ rather than in terms of electric $\mathbf{E}(\mathbf{r}, t)$

and magnetic $\mathbf{H}(\mathbf{r}, t)$ fields, the initial expressions for the mean densities of induced current and charge are also expressed in terms of potentials (see (17) and (18)). However, expression (17) for the current density is inconvenient because it contains a contribution proportional to the mean value $\langle 0|\rho(\mathbf{r}, t)|0\rangle$ of the charge density and this contribution is not small.

Therefore, our task has been to express the induced densities in terms of the fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{H}(\mathbf{r}, t)$, which can certainly be accomplished because the mean densities of current and charge are observable quantities.

As a result, we obtained formulas (31) and (32), which contain the electric field alone. However, formula (31) has the same shortcoming as (17): it contains a contribution proportional to $\langle 0|\rho(\mathbf{r}, t)|0\rangle$.

If we use an approximation in which the electric field $\mathbf{E}(\mathbf{r}, t)$ is independent of coordinates but depends only on time, then the contribution proportional to $\langle 0|\rho(\mathbf{r}, t)|0\rangle$ can be removed if we introduce into the expression for the mean current density the operators $\mathbf{r}_i$ of the coordinates of the ith particle using the relation $\mathbf{v}_i = (i/\hbar)[\mathcal{H}, \mathbf{r}_i]$. This method has actually been used in [3], where the light scattering in bulk crystals is discussed. In Appendix A, we show that the same method allows one to eliminate the concentration of charged particles from the formula for the conductivity of bulk intrinsic semiconductors at $T = 0$.

The same result is obtained if we express the interaction between particles and a field as $-e\sum_i r_{i\beta}E_\beta(t)$ in the approximation $\mathbf{E}(\mathbf{r}, t) \approx \mathbf{E}(t)$, as was done by Kubo in [8].

However, the problem is complicated if we take into consideration the dependence of electric field on coordinates. Our aim has been to obtain such an expression for the density of induced current that would reduce to the Kubo formula when passing from $\mathbf{E}(\mathbf{r}, t)$ to $\mathbf{E}(t)$, as well as similar expressions for the density of induced charge. For the case $T = 0$, this problem is solved in Sections 7–9.

The essence of the method used consists in the following. The operator of interaction between particles and a field, expressed in terms of electric and magnetic fields, does not exist. However, we introduce a fictitious interaction operator that yields correct results for the densities of induced current and charge averaged over the ground state of the system. This fictitious operator U_1^f (37) is expressed in terms of the electric field $\mathbf{E}(\mathbf{r}, t)$ alone under the sign of time integral. Then, we apply the transformation of interaction $U_1^f \longrightarrow \tilde{U}_1$ that does not change the values of the mean densities; however, this transformation eliminates the contribution containing $\langle 0|\rho(\mathbf{r}, t)|0\rangle$ from the expression for the mean induced density of current. As a result, we obtained expressions (64)–(69) for $\langle 0|\mathbf{j}_1(\mathbf{r}, t)|0\rangle$ and $\langle 0|\rho_1(\mathbf{r}, t)|0\rangle$ in which the principal terms contain the electric field, while the additional terms contain the derivatives of $\mathbf{E}(\mathbf{r}, t)$ with respect to coordinates. The transformed fictitious interaction $\tilde{U}_1$ is also divided into principal and additional parts. The former is equal to $-e\sum_i r_{i\beta}E_\beta(\mathbf{r}_i, t)$, while the latter contains the derivatives $\partial E_\alpha(\mathbf{r}, t)/\partial \mathbf{r}_\beta$ under the sign of time integral.

In Section 10, expressions (64)–(69) for the mean densities of induced currents and charges have been transformed so that they do not explicitly depend on the particle coordinates $\mathbf{r}_i$. This is the main result of the present paper. In Section 11, we have obtained expressions for the coordinate-dependent conductivity tensor $\sigma(\mathbf{k}, \omega|\mathbf{r})$ for spatially inhomogeneous systems.

In Section 12, we have passed from the expressions for $\langle 0|\mathbf{j}_1(\mathbf{r}, t)|0\rangle$ and $\langle 0|\rho_1(\mathbf{r}, t)|0\rangle$ that contain only electric field and its derivatives with respect to coordinates to the expressions that contain the magnetic field $\mathbf{H}(\mathbf{r}, t)$ as well. This is necessary for analyzing the case $\mathbf{H}$ = const, $\mathbf{E} = 0$, which was considered in Section 13. In Appendix B, we derive an expression for the operator of total acceleration of a system of charged particles. The acceleration is attributed to a weak external electromagnetic field. We have shown that the total acceleration averaged over the ground state of the system can be expressed in terms of electric and magnetic fields. In the case of free particles, the result obtained reduces, in the limit, to the correct result that contains a force due to the electric field and the Lorentz force.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research (project no. 00-02-16904), the ISTC program Physics of Semiconductor Nanostructures, and the federal program Integration.

APPENDIX A

Three Expressions for the Conductivity Tensor in the Case of a Spatially Homogeneous Medium and a Coordinate-Independent Electric Field

Using formula (31) for the mean density of induced current and taking into account the following relation between Fourier components,

$$\mathbf{a}(\mathbf{k}, \omega) = (ic/\omega)\mathbf{E}(\mathbf{k}, \omega), \tag{A.1}$$

which follows from definition (30), we obtain

$$\sigma_{\alpha\beta}(\mathbf{k}, \omega|\mathbf{r}) = \frac{ie}{m\omega}\langle 0|\rho(\mathbf{r})|0\rangle\delta_{\alpha\beta} + \frac{1}{\hbar\omega}\int d\mathbf{r}'\int_{-\infty}^{\infty} dt' \times e^{-i\mathbf{k}\cdot\mathbf{r}' + i\omega t'}\theta(t')\langle 0|j_\alpha(\mathbf{r}), j_\beta(\mathbf{r}-\mathbf{r}', -t')|0\rangle. \tag{A.2}$$

In the case of a spatially homogeneous system, the tensor $\sigma_{\alpha\beta}(\mathbf{k}, \omega|\mathbf{r})$ is independent of $\mathbf{r}$. In the first term,

we use the relation

$$\langle 0|\rho(\mathbf{r})|0\rangle = en, \tag{A.3}$$

where n is the concentration of charged particles; in the second term, we integrate with respect to $\mathbf{r}$ and $\mathbf{r}'$ and divide the result obtained by the normalized volume V_0. We have

$$\sigma_{\alpha\beta}(\mathbf{k}, \omega) = \frac{ie^2n}{m\omega}\delta_{\alpha\beta} + \frac{e^2}{4m^2\hbar\omega V_0}\sum_{i,j}\int_{-\infty}^{\infty} dt\,\theta(t)e^{i\omega t} \times\langle 0|[\{e^{i\mathcal{H}t/\hbar}(e^{-i\mathbf{k}\cdot\mathbf{r}_i}p_{i\alpha} + p_{i\alpha}e^{-i\mathbf{k}\cdot\mathbf{r}_i})e^{-i\mathcal{H}t/\hbar}\}, (e^{i\mathbf{k}\cdot\mathbf{r}_j}p_{j\beta} + p_{i\beta}e^{i\mathbf{k}\cdot\mathbf{r}_j})]|0\rangle, \tag{A.4}$$

where $\mathbf{p}_i$ is the momentum operator defined in (4) with regard to a strong constant magnetic field.

When the electric field $\mathbf{E}(t)$ is independent of coordinates, we introduce the frequency representation

$$E_\alpha(\omega) = \int_{-\infty}^{\infty} dt E_\alpha(t)e^{i\omega t}, \tag{A.5}$$

where

$$E_\alpha(\mathbf{k}, \omega) = (2\pi)^3\delta(\mathbf{k})E_\alpha(\omega). \tag{A.6}$$

We also introduce the conductivity $\sigma_{\alpha\beta}(\omega|\mathbf{r})$ that enters in the definition of the mean density of current

$$\langle 0|j_{1\alpha}(\mathbf{r}, t)|0\rangle_{\mathrm{h}}^{(+)} = \frac{1}{2\pi}\int_0^{\infty} d\omega\sigma_{\alpha\beta}(\omega|\mathbf{r})E_\alpha(\omega)e^{-i\omega t}, \tag{A.7}$$

where the subscript "h" indicates that the electric field is spatially homogeneous.

Using the results obtained in Section 11, we can easily show that

$$\sigma_{\alpha\beta}(\omega|\mathbf{r}) = \sigma_{\alpha\beta}(\mathbf{k} = 0, \omega|\mathbf{r}). \tag{A.8}$$

In the case of a spatially homogeneous medium and a field $\mathbf{E}(t)$, from (A.4) and (A.8) we obtain

$$\sigma^{\mathrm{I}}_{\alpha\beta}(\omega) = \frac{ie^2n}{m\omega}\delta_{\alpha\beta} + \frac{e^2}{\hbar\omega V_0}\int_{-\infty}^{\infty} dt\,\theta(t)e^{i\omega t}\langle 0|[V_\alpha(t), V_\beta]|0\rangle, \tag{A.9}$$

where $V_\alpha = (1/m)\sum_i p_{i\alpha}$ is the operator of the total velocity of charged particles. (A.9) is the first formula for the conductivity tensor. To derive the two other formulas, we apply the relation

$$V_\alpha = (i/\hbar)[\mathcal{H}, R_\alpha], \tag{A.10}$$

where $R_\alpha = \sum_i r_{i\alpha}$. Substituting (A.10) into (A.8), we obtain

$$\sigma_{\alpha\beta}(\omega) = \frac{ie^2n}{m\omega}\delta_{\alpha\beta} + \frac{e^2}{\hbar\omega V_0}\int_{-\infty}^{\infty} dt\,\theta(t) \times e^{i\omega t}\frac{d}{dt}\langle 0|[R_\alpha(t), V_\beta]|0\rangle. \tag{A.11}$$

Let us integrate by parts with respect to t. As $t \longrightarrow \infty$,

$$\theta(t)e^{i\omega t}\frac{d}{dt}\langle 0|R_\alpha(t), V_\beta|0\rangle \longrightarrow 0$$

under the substitution of $\omega + i\delta$ for ω with $\delta \longrightarrow 0$. We also take into account that $d\theta(t)/dt = \delta(t)$ and

$$(1/V_0)[R_\alpha, V_\beta] = (i\hbar n/m)\delta_{\alpha\beta}. \tag{A.12}$$

Then, we obtain

$$\sigma^{\mathrm{II}}_{\alpha\beta}(\omega) = -\frac{ie}{\hbar V_0} \times\int_{-\infty}^{\infty} dt\,\theta(t)e^{i\omega t}\langle 0|[D_\alpha(t), V_\beta]|0\rangle, \tag{A.13}$$

where we used the notation

$$\mathbf{D} = e\mathbf{R}. \tag{A.14}$$

Formula (A.13) is the second expression for the conductivity tensor. We stress that, when passing from (A.9) to (A.13), the first term in the right-hand side of (A.9) containing concentration n is canceled out.

When passing to the third expression, we use (A.10) for the operator V_β in the right-hand side of (A.13). Further, we proceed in the same way as when passing from (A.9) to (A.13). Since the commutator $[D_\alpha, D_\beta] = 0$, we obtain the third formula,

$$\sigma^{\mathrm{III}}_{\alpha\beta}(\omega) = \frac{\omega}{\hbar V_0}\int_{-\infty}^{\infty} dt\,\theta(t)e^{i\omega t}\langle 0|[D_\alpha(t), D_\beta]|0\rangle. \tag{A.15}$$

For various systems, it is convenient to apply one of the formulas (A.9), (A.13), or (A.15).

For free particles, when $V(\mathbf{r}_1, \ldots, \mathbf{r}_N) = 0$ and $\mathbf{H}_c = 0$, the velocity operator $\mathbf{V}$ commutes with the Hamiltonian $\mathcal{H}_{\mathrm{free}} = m\sum_i v_i^2$. Using (A.9) and taking into account that $V_\alpha(t) = V_\alpha$ and $[V_\alpha, V_\beta] = 0$, we obtain the following well-known result:

$$\sigma_{\alpha\beta\ \mathrm{free}}(\omega) = \frac{ie^2n}{m\omega}\delta_{\alpha\beta}. \tag{A.16}$$

When the excited states are separated from the ground state by a gap, i.e., when the energy of these states is $E_n = \hbar\omega_n \neq 0$, it is more convenient to apply (A.13) or (A.15). Using the exact wave functions $|n\rangle$ of the

excited states and calculating the time integral, we express (A.15) in terms of the matrix elements of the operator **D**:

$$\sigma^{\mathrm{III}}_{\alpha\beta}(\omega) = \frac{i\omega}{\hbar V_0} \times \sum_n \left\{ \frac{\langle 0|D_\alpha|n\rangle\langle n|D_\beta|0\rangle}{\omega - \omega_n} - \frac{\langle 0|D_\beta|n\rangle\langle n|D_\alpha|0\rangle}{\omega + \omega_n} \right\}, \quad \text{(A.17)}$$

whence it follows that, for systems with a gap, $\sigma_{\alpha\beta}(\omega) \longrightarrow 0$ as $\omega \longrightarrow 0$ (we do not consider the case of superconductivity). Systems with a gap in the energy spectrum include bulk semiconductors without impurities and defects.

The conductivity tensor $\sigma_{\alpha\beta}(\omega)$ is related to the permittivity tensor $\varepsilon_{\alpha\beta}(\omega)$ by the well-known formula

$$\varepsilon_{\alpha\beta}(\omega) = \delta_{\alpha\beta} + 4\pi\chi_{\alpha\beta}(\omega),$$

$$\chi_{\alpha\beta}(\omega) = (i/\omega)\sigma_{\alpha\beta}(\omega),$$

so that from (A.17) we obtain

$$\chi_{\alpha\beta}(\omega) = \frac{1}{\hbar V_0} \times \sum_n \left\{ \frac{\langle 0|D_\alpha|n\rangle\langle n|D_\beta|0\rangle}{\omega_n - \omega} + \frac{\langle 0|D_\beta|n\rangle\langle n|D_\alpha|0\rangle}{\omega_n + \omega} \right\}. \quad \text{(A.18)}$$

Using the relations

$$\langle 0|D_\alpha|n\rangle = (ie/\omega_n)\langle 0|V_\alpha|n\rangle,$$

$$\langle n|D_\alpha|0\rangle = -(ie/\omega_n)\langle n|V_\alpha|0\rangle$$

between the matrix elements, from (A.18) we derive

$$\chi_{\alpha\beta}(\omega) = \frac{e^2}{\hbar V_0} \times \sum_n \frac{1}{\omega_n^2}\left\{ \frac{\langle 0|V_\alpha|n\rangle\langle n|V_\beta|0\rangle}{\omega_n - \omega} + \frac{\langle 0|V_\beta|n\rangle\langle n|V_\alpha|0\rangle}{\omega_n + \omega} \right\}. \quad \text{(A.19)}$$

If $\omega \ll \omega_n$, then the quantity $\chi_{\alpha\beta}$ is real and does not depend on frequency ω; however, when $\omega \approx \omega_n$, this dependence becomes strong and a nonzero imaginary part of the tensor $\chi_{\alpha\beta}$ arises that determines the resonant absorption of light at frequencies $\omega \approx \omega_n$. To calculate this imaginary part, one should replace frequency ω by $\omega + i\delta$, $\delta \longrightarrow 0$, assuming that the switching on of the field is adiabatic or taking into account that the lifetime of the system in state n is finite.

APPENDIX B

Acceleration of a System of Particles

To obtain the operator $\mathbf{J}_1(t)$ of the induced current of a system of particles, we apply formula (14), which contains vector and scalar potentials. Integrating with respect to **r** and **r'**, we obtain

$$J_{1\alpha}(t) = -\frac{e^2}{mc}\sum_i A_\alpha(\mathbf{r}_i(t), t) + \frac{ie}{\hbar}\int_{-\infty}^{t} dt' \left[U_1(t'), \sum_i v_{i\alpha}(t) \right], \quad \text{(B.1)}$$

where the notation

$$\mathbf{A}(\mathbf{r}_i(t), t) = e^{i\mathcal{H}t/\hbar}\mathbf{A}(\mathbf{r}_i, t)e^{-i\mathcal{H}t/\hbar}$$

is used. Differentiating (B.1) with respect to time and dividing by e, we obtain the operator $\mathbf{W}_1(t)$ of the induced total acceleration,

$$W_{1\alpha}(t) = -\frac{e}{mc}\frac{d}{dt}\sum_i A_\alpha(\mathbf{r}_i(t), t) + \frac{i}{\hbar}\left[U_1(t), \sum_i v_{i\alpha}(t) \right] + \frac{i}{\hbar}\int_{-\infty}^{t} dt'[U_1(t'), W_\alpha(t)], \quad \text{(B.2)}$$

where $W_\alpha(t) = \sum_i w_{i\alpha}(t)$ and $w_{i\alpha}(t) = dv_{i\alpha}(t)/dt$. The interpretation of the first two terms yields

$$W_{1\alpha}(t) = -\frac{e}{mc}\sum_i \frac{\partial A_\alpha(\mathbf{r}_i(t), t)}{\partial t} - \frac{e}{m}\sum_i \frac{\partial \varphi(\mathbf{r}_i(t), t)}{\partial \mathbf{r}_{i\alpha}(t)} + \frac{e}{2mc}\sum_i \left\{ v_{i\alpha}(t)\left(\frac{\partial A_\beta(\mathbf{r}_i(t), t)}{\partial \mathbf{r}_{i\alpha}(t)} - \frac{\partial A_\alpha(\mathbf{r}_i(t), t)}{\partial \mathbf{r}_{i\beta}(t)} \right) + \left(\frac{\partial A_\beta(\mathbf{r}_i(t), t)}{\partial \mathbf{r}_{i\alpha}(t)} - \frac{\partial A_\alpha(\mathbf{r}_i(t), t)}{\partial \mathbf{r}_{i\beta}(t)} \right) v_{i\alpha}(t) \right\} - \left(\frac{e}{mc}\right)^2 \sum_i [\mathbf{A}(\mathbf{r}_i(t), t) \times \mathbf{H}_c]_\alpha + \frac{i}{\hbar}\int_{-\infty}^{t} dt'[U_1(t'), W_\alpha(t)], \quad \text{(B.3)}$$

where the notations

$$\frac{\partial A_\alpha(\mathbf{r}_i(t), t)}{\partial t} = e^{i\mathcal{H}t/\hbar}\frac{\partial A_\alpha(\mathbf{r}_i, t)}{\partial t}e^{-i\mathcal{H}t/\hbar},$$

$$\frac{\partial \varphi(\mathbf{r}_i(t), t)}{\partial \mathbf{r}_{i\alpha}(t)} = e^{i\mathcal{H}t/\hbar}\frac{\partial \varphi(\mathbf{r}_i, t)}{\partial \mathbf{r}_{i\alpha}}e^{-i\mathcal{H}t/\hbar}$$

are used and $\mathbf{H}_c$ is a strong constant magnetic field included in the basic Hamiltonian $\mathcal{H}$. When passing

from (B.2) to (B.3), we used the relation $\mathbf{v}_i \times \mathbf{v}_i = (i\hbar e/m^2 c)\mathbf{H}_c$, where the left-hand side is nonzero because of the noncommutativity of various projections of the velocity operator, for example,

$$[\mathbf{v}_i \times \mathbf{v}_i]_z = [v_{ix}, v_{iy}] = (i\hbar e/m^2 c)H_{cz}. \quad \text{(B.4)}$$

The acceleration operator $\mathbf{w}_i(t)$ of the ith particle is equal to

$$w_{i\alpha} = \frac{i}{\hbar} e^{i\mathcal{H}t/\hbar}[\mathcal{H}, v_{i\alpha}]e^{-i\mathcal{H}t/\hbar}$$
$$= -\frac{1}{m} e^{i\mathcal{H}t/\hbar}\frac{\partial V(\mathbf{r}_1, \ldots, \mathbf{r}_N)}{\partial r_{i\alpha}} e^{-i\mathcal{H}t/\hbar} + \frac{e[\mathbf{v}_i(t) \times \mathbf{H}_c]_\alpha}{mc}. \quad \text{(B.5)}$$

Using (1), we can easily see that the expression in curly brackets in (B.3) is equal to

$$e^{i\mathcal{H}t/\hbar}\{[\mathbf{v}_i \times \mathbf{H}(\mathbf{r}_i, t)]_\alpha - [\mathbf{H}(\mathbf{r}_i, t) \times \mathbf{v}_i]_\alpha\}e^{-i\mathcal{H}t/\hbar}.$$

Therefore, expression (B.3) is transformed into

$$\mathbf{W}_1(t) = \frac{e}{m}\sum_i \mathbf{E}(\mathbf{r}_i(t), t) + \frac{e}{2mc}$$
$$\times \sum_i \{\mathbf{v}_i(t) \times \mathbf{H}(\mathbf{r}_i(t), t) - \mathbf{H}_i(\mathbf{r}_i(t), t) \times \mathbf{v}_i(t)\} \quad \text{(B.6)}$$
$$-\left(\frac{e}{mc}\right)^2 \sum_i \mathbf{A}(\mathbf{r}_i(t), t) \times \mathbf{H}_c + \frac{i}{\hbar}\int_{-\infty}^{t} dt'[U_1(t'), \mathbf{W}(t)].$$

It is obvious that the second term corresponds to the Lorentz force, which is written with regard to the noncommutativity of the operators $\mathbf{v}_i$ and $\mathbf{H}(\mathbf{r}_i, t)$. The third term, which is attributed to $\mathbf{H}_c$, contains the addition $\Delta\mathbf{v}_i$ to velocity that is defined in (8) and is induced by a weak electromagnetic field.

In the case of free particles,

$$\mathbf{H}_c = 0, \quad V(\mathbf{r}_1, \ldots, \mathbf{r}_N) = 0, \quad \mathbf{W} = 0, \quad \text{(B.7)}$$

and only the first two terms containing weak electric and magnetic fields are preserved in the right-hand side of expression (B.6).

However, if particles are not free, then operator (B.6) cannot be expressed in terms of fields alone because the last two terms contain vector and scalar potentials. The mean value $\langle 0|\mathbf{W}_1(t)|0\rangle$ of the induced acceleration must be expressed in terms of the fields alone, as we will now prove. To this end, we calculate the mean value $\langle 0|\mathbf{W}_1(t)|0\rangle$ by a different method and then check if the two methods yield the same result. Using expression (27), we integrate both its parts with respect to $\mathbf{r}$. We obtain

$$\langle 0|W_{1\alpha}(t)|0\rangle = \frac{e}{m}\sum_i \langle 0|E_\alpha(\mathbf{r}_i, t)|0\rangle - \frac{ie}{2\hbar}$$
$$\times \int_{-\infty}^{0} dt' \langle 0|\left[\sum_j v_{j\beta}, \sum_i \{v_{i\beta}(t')E_\beta(\mathbf{r}_i(t'), t + t') \right. \quad \text{(B.8)}$$
$$\left. + E_\beta(\mathbf{r}_i(t'), t + t')v_{j\beta}(t')\}\right]|0\rangle.$$

This expression contains electric fields alone. On the other hand, averaging operator (B.6), we obtain

$$\langle 0|W_{1\alpha}(t)|0\rangle = \frac{e}{m}\sum_i \langle 0|E_\alpha(\mathbf{r}_i, t)|0\rangle$$
$$+ \frac{e}{2mc}\sum_i \langle 0|[\mathbf{v}_i \times \mathbf{H}(\mathbf{r}_i, t)]_\alpha - [\mathbf{H}(\mathbf{r}_i, t) \times \mathbf{v}_i]_\alpha|0\rangle \quad \text{(B.9)}$$
$$-\left(\frac{e}{mc}\right)^2 \sum_i \langle 0|[\mathbf{A}(\mathbf{r}_i, t) \times \mathbf{H}_c]_\alpha|0\rangle + C_\alpha(t),$$

where

$$C_\alpha(t) = \frac{i}{\hbar}\int_{-\infty}^{t} dt' \langle 0|[U_1(t'), W_\alpha(t)]|0\rangle. \quad \text{(B.10)}$$

Let us transform (B.10). Integrating by parts, we obtain

$$C_\alpha(t) = C_\alpha^1(t) + C_\alpha^2(t), \quad \text{(B.11)}$$

$$C_\alpha^1(t) = -i/\hbar\langle 0|\left[U_1, \sum_i v_{i\alpha}\right]|0\rangle, \quad \text{(B.12)}$$

$$C_\alpha^2(t) = \frac{i}{\hbar}\frac{d}{dt}\int_{-\infty}^{t} dt'\langle 0|\left[U_1(t'), \sum_i v_{i\alpha}(t)\right]|0\rangle. \quad \text{(B.13)}$$

Substitute the expression (5) for U_1 into the right-hand side of (B.12) and calculate the commutator. We obtain

$$C_\alpha^1(t) = \frac{e}{m}\sum_i \langle 0|\frac{\partial\varphi(\mathbf{r}_i, t)}{\partial r_{i\alpha}}|0\rangle$$
$$+\left(\frac{e}{mc}\right)^2 \sum_i \langle 0|[\mathbf{A}(\mathbf{r}_i, t) \times \mathbf{H}_c]_\alpha|0\rangle \quad \text{(B.14)}$$
$$-\frac{e}{2mc}\sum_i \langle 0|v_{i\beta}\frac{\partial A_\beta(\mathbf{r}_i, t)}{\partial r_{i\alpha}} + \frac{\partial A_\beta(\mathbf{r}_i, t)}{\partial r_{i\alpha}}v_{i\beta}|0\rangle.$$

In (B.13), we pass to $t'' = t - t'$ in the integrand and partition the integral into two parts, the first of which

contains a vector potential and the second, a scalar potential:

$$C_\alpha^2(t) = C_{A\alpha}^2(t) + C_{\varphi a}^2(t), \tag{B.15}$$

$$C_{A\alpha}^2(t) = \frac{ie}{2c\hbar}\int_{-\infty}^{0} dt'$$

$$\times\langle 0|\Bigg[\sum_j V_{j\alpha}, \sum_i \Bigg\{ V_{i\beta}(t')\frac{\partial A_\beta(\mathbf{r}_i(t'), t+t')}{\partial t} \tag{B.16}$$

$$+ \frac{\partial A_\beta(\mathbf{r}_i(t'), t+t')}{\partial t} V_{i\beta}(t') \Bigg\}\Bigg]|0\rangle,$$

$$C_{\varphi\alpha}^2(t) = -\frac{ie}{\hbar}$$

$$\times\int_{-\infty}^{0} dt'\langle 0|\Bigg[\sum_j V_{j\alpha}, \sum_i \frac{\partial\varphi(\mathbf{r}_i(t'), t+t')}{\partial t}\Bigg]|0\rangle. \tag{B.17}$$

We leave (B.16) unchanged for the present and apply the relation

$$\partial\varphi(\mathbf{r}_i(t'), t+t')/\partial t = e^{i\mathcal{H}t'/\hbar}\frac{\partial\varphi(\mathbf{r}_i, t+t')}{\partial t}e^{-i\mathcal{H}t'/\hbar}$$

$$= e^{i\mathcal{H}t'/\hbar}\frac{\partial\varphi(\mathbf{r}_i, t+t')}{\partial t'}e^{-i\mathcal{H}t'/\hbar}$$

to (B.17). Then, we integrate by parts with respect to t' and obtain

$$C_{\varphi\alpha}^2(t) = -\frac{ie}{\hbar}\sum_i\langle 0|[V_{i\alpha}, \varphi(\mathbf{r}_i, t)]|0\rangle - \frac{e}{\hbar^2}\int_{-\infty}^{0} dt' \tag{B.18}$$

$$\times\langle 0|\Bigg[\sum_j V_{j\alpha}, e^{i\mathcal{H}t'/\hbar}\sum_i[\mathcal{H}, \varphi(\mathbf{r}_i, t+t')]e^{-i\mathcal{H}t'/\hbar}\Bigg]|0\rangle.$$

We calculate the commutator in the first term. To transform the second term, we note that

$$\frac{i}{\hbar}[\mathcal{H}, \varphi(\mathbf{r}_i, t)] = \frac{1}{2}\Bigg(V_{i\beta}\frac{\partial\varphi(\mathbf{r}_i, t)}{\partial r_{i\beta}} + \frac{\partial\varphi(\mathbf{r}_i, t)}{\partial r_{i\beta}}V_{i\beta}\Bigg). \tag{B.19}$$

Substituting (B1.9) into (B.18), we obtain

$$C_{\varphi\alpha}^2(t) = -\frac{e}{m}\sum_i\langle 0|\frac{\partial\varphi(\mathbf{r}_i, t)}{\partial r_{i\alpha}}|0\rangle + \frac{ie}{2\hbar}\int_{-\infty}^{0} dt'$$

$$\times\langle 0|\Bigg[\sum_j V_{j\alpha}, \sum_i \Bigg\{ V_{i\beta}(t')\frac{\partial\varphi(\mathbf{r}_i(t'), t+t')}{\partial r_{i\beta}(t')} \tag{B.20}$$

$$+ \frac{\partial\varphi(\mathbf{r}_i(t'), t+t')}{\partial r_{i\beta}(t')}V_{i\beta}(t') \Bigg\}\Bigg]|0\rangle.$$

Adding up (B.14), (B.16), and (B.20), we finally obtain

$$C_\alpha(t) = -\frac{e}{2mc}\langle 0|V_{i\beta}\frac{\partial A_\beta(\mathbf{r}_i, t)}{\partial r_{i\alpha}} + \frac{\partial A_\beta(\mathbf{r}_i, t)}{\partial r_{i\alpha}}V_{i\beta}|0\rangle$$

$$+\Bigg(\frac{e}{mc}\Bigg)^2\sum_i\langle 0|[\mathbf{A}(\mathbf{r}_i, t)\times\mathbf{H}_c]_\alpha|0\rangle - \frac{ie}{2\hbar}\int_{-\infty}^{0} dt'$$

$$\times\langle 0|\Bigg[\sum_j V_{j\alpha}, \sum_i \{ V_{i\beta}(t')E_\beta(\mathbf{r}_i(t'), t+t') \tag{B.21}$$

$$+ E_\beta(\mathbf{r}_i(t'), t+t')V_{i\beta}(t') \}\Bigg]|0\rangle.$$

Substituting (B.21) into (B.9), we obtain

$$\langle 0|W_{1\alpha}|0\rangle = \frac{e}{m}\sum_i\langle 0|E_\alpha(\mathbf{r}_i, t)|0\rangle$$

$$-\frac{ie}{2\hbar}\int_{-\infty}^{0} dt'\langle 0|\Bigg[\sum_j V_{j\alpha}, \sum_i \{ V_{i\beta}(t')E_\beta(\mathbf{r}_i(t'), t+t') \tag{B.22}$$

$$+ E_\beta(\mathbf{r}_i(t'), t+t')V_{i\beta}(t') \}\Bigg]|0\rangle - \frac{e}{2mc}$$

$$\times\sum_i\langle 0|V_{i\beta}\frac{\partial A_\alpha(\mathbf{r}_i, t)}{\partial r_{i\beta}} + \frac{\partial A_\alpha(\mathbf{r}_i, t)}{\partial r_{i\beta}}V_{i\beta}|0\rangle.$$

This expression coincides with formula (B.8) except for the last term in the right-hand side of (B.22). However, we can show that this term is equal to zero. Indeed,

$$\frac{1}{2}\langle 0|V_{i\beta}\frac{\partial A_\alpha(\mathbf{r}_i, t)}{\partial r_{i\beta}} + \frac{\partial A_\alpha(\mathbf{r}_i, t)}{\partial r_{i\beta}}V_{i\beta}|0\rangle$$

$$= \frac{i}{\hbar}\langle 0|[\mathcal{H}, A_\alpha(\mathbf{r}_i, t)]|0\rangle = 0, \tag{B.23}$$

because the operator $\mathcal{H}$ has only diagonal matrix elements $\langle 0|\mathcal{H}|0\rangle$. Thus, we have checked that the results (B.8) and (B.9), obtained by different methods, coincide.

Consider the case of free particles when conditions (B.7) are satisfied. Then, $C_\alpha(t) = 0$ and the penultimate term in the right-hand side of (B.9) also vanishes. Comparing (B.8) and (B.4), we find that the mean Lorentz force must satisfy the relation

$$\frac{e}{2c}\langle 0|[\mathbf{v}_i \times \mathbf{H}(\mathbf{r}_i, t)]_\alpha - [\mathbf{H}(\mathbf{r}_i, t) \times \mathbf{v}_i]_\alpha|0\rangle_{\text{free}}$$

$$= -\frac{iem}{2\hbar}\int_{-\infty}^{0} dt' \langle 0| \Bigg[\sum_j v_{j\alpha}, \sum_i \{ v_{i\beta}(t') E_\beta(\mathbf{r}_i(t'), t + t') + E_\beta(\mathbf{r}_i(t'), t + t') v_{i\beta}(t') \} \Bigg] |0\rangle_{\text{free}}, \quad \text{(B.24)}$$

where the subscript "free" indicates that conditions (B.7) are satisfied. Let us verify (B.24) by direct calculation, transforming the right-hand side. Since

$$\mathcal{H}_{\text{free}} = \frac{m}{2}\sum_i v_i^2$$

and $v_{j\alpha}$ commutes with $\mathcal{H}_{\text{free}}$, we can write

$$v_{j\alpha} = e^{i\mathcal{H}_{\text{free}}t'/\hbar} v_{j\alpha} e^{-i\mathcal{H}_{\text{free}}t'/\hbar} = v_{j\alpha}(t'),$$

then, the "bordering"

$$\exp(i\mathcal{H}_{\text{free}}t'/\hbar)\ldots - \exp(-i\mathcal{H}_{\text{free}}t'/\hbar)$$

is removed from the right-hand side of (B.24). We denote the right-hand side of (B.24) by $\Psi_\alpha(t)$. After calculating the commutator, it proves to be equal to

$$\Psi_\alpha(t) = -\frac{e}{2}\int_{-\infty}^{0} dt' \sum_i \langle 0| v_{i\beta} \frac{\partial E_\beta(\mathbf{r}_i, t + t')}{\partial r_{i\alpha}} + \frac{\partial E_\beta(\mathbf{r}_i, t + t')}{\partial r_{i\alpha}} v_{i\beta}|0\rangle. \quad \text{(B.25)}$$

Note that the equality

$$\frac{1}{2}\langle 0| v_{i\beta}\frac{\partial E_\alpha(\mathbf{r}_i, t)}{\partial r_{i\beta}} + \frac{\partial E_\alpha(\mathbf{r}_i, t)}{\partial r_{i\beta}} v_{i\beta}|0\rangle = (i/\hbar)\langle 0|[\mathcal{H}, E_\alpha(\mathbf{r}_i, t)]|0\rangle = 0$$

holds, which is analogous to (B.23); therefore, $\Psi_\alpha(t)$ can be rewritten as

$$\Psi_\alpha(t) = -\frac{e}{2}\int_{-\infty}^{0} dt' \times \sum_i \langle 0| v_{i\beta}\left(\frac{\partial E_\beta(\mathbf{r}_i, t + t')}{\partial r_{i\alpha}} - \frac{\partial E_\alpha(\mathbf{r}_i, t + t')}{\partial r_{i\beta}}\right) + \left(\frac{\partial E_\beta(\mathbf{r}_i, t + t')}{\partial r_{i\alpha}} - \frac{\partial E_\alpha(\mathbf{r}_i, t + t')}{\partial r_{i\beta}}\right) v_{i\beta}|0\rangle_{\text{free}}, \quad \text{(B.26)}$$

which is equal to

$$\Psi_\alpha(t) = -\frac{e}{2}\int_{-\infty}^{0} dt' \sum_i \langle 0|[\mathbf{v}_i \times \text{curl}\mathbf{E}(\mathbf{r}_i, t + t')]_\alpha - [\text{curl}\mathbf{E}(\mathbf{r}_i, t + t') \times \mathbf{v}_i]_\alpha|0\rangle. \quad \text{(B.27)}$$

Using the Maxwell equation $\text{curl}\mathbf{E}(\mathbf{r}, t) = -(1/c)(\partial\mathbf{H}(\mathbf{r}, t)/\partial t)$ and calculating the integral with respect to t', we find that $\Psi_\alpha(t)$ is equal to the left-hand side of (B.24), which was to be proved.

Note that, when $\mathbf{E} = 0$ and $\mathbf{H} = \text{const}$, both sides of Eq. (B.24) vanish. For the right-hand side, this result is obvious; the left-hand side contains the matrix elements $\langle 0|v_{i\beta}|0\rangle = (i/\hbar)\langle 0|[\mathcal{H}, r_{i\beta}]|0\rangle = 0$ because the operator $\mathcal{H}$ is diagonal.

REFERENCES

1. H. Stolz, *Time Resolved Light Scattering from Excitons* (Springer, Berlin, 1994).
2. J. Shah, *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures* (Springer, Berlin, 1996).
3. R. Zeyher, H. Bilz, and M. Cardona, Solid State Commun. **19**, 57 (1976).
4. L. I. Korovin, I. G. Lang, D. A. Kontreras-Solorio, and S. T. Pavlov, Fiz. Tverd. Tela (St. Petersburg) **43**, 2091 (2001) [Phys. Solid State **43**, 2182 (2001)]; E-print archives, cond-mat/0104262.
5. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 6th ed. (Nauka, Moscow, 1973; Pergamon, Oxford, 1975).
6. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed. (Nauka, Moscow, 1974; Pergamon, New York, 1977).
7. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics*, 2nd ed. (Prentice Hall, Englewood Cliffs, N.J., 1963; Dobrosvet, Moscow, 1998).
8. R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957); R. Kubo, *Problems of Quantum Theory of Irreversible Processes* (Inostrannaya Literatura, Moscow, 1961), p. 59.
9. S. Nakajima, Proc. Phys. Soc. London **69**, 441 (1956).
10. O. V. Konstantinov and V. I. Perel', Zh. Éksp. Teor. Fiz. **37**, 786 (1959) [Sov. Phys. JETP **10**, 560 (1959)].
11. R. Enderlein, K. Peuker, and F. Bechstedt, Phys. Status Solidi B **92**, 149 (1979).
12. D. A. Contreras-Solorio, S. T. Pavlov, L. I. Korovin, and I. G. Lang, Phys. Rev. B **62**, 16815 (2000); E-print archives, cond-mat/0002229.
13. L. I. Korovin, I. G. Lang, V. A. Kontreras-Solorio, and S. T. Pavlov, Fiz. Tverd. Tela (St. Petersburg) **44**, 1681 (2002) [Phys. Solid State **44**, 1759 (2002)]; E-print archives, cond-mat/0203390.
14. L. I. Korovin, I. G. Lang, D. A. Kontreras-Solorio, and S. T. Pavlov, Fiz. Tverd. Tela (St. Petersburg) **44**, 2084 (2002) [Phys. Solid State **44**, 171 (2002)]; E-print archives, cond-mat/0104262.

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 286–300.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 325–340.

SOLIDS
Electronic Properties

On the Theory of Ferromagnetism in Transition Metals with Cubic Symmetry

R. O. Zaĭtsev

Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia
e-mail: zaitsev@mbslab.kiae.ru

Received June 4, 2002

Abstract—The idea of strong interaction within the same unit cell is used to establish the possibility of the existence of ferromagnetic instability in a system with jumps between cations of transition elements. The phase diagram of the existence of the ferromagnetic ordering as a function of the average number of holes (h_d) in the $3d^{10}$ shell of transition elements is constructed. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The exchange interaction of free electrons considered as a perturbation always leads to a negative correction, which corresponds to the tendency of electrons to ferromagnetism.

When localized *s*-electrons are excited, strong Hubbard repulsion under a small density of electrons also leads to a considerable increase of the spin part of paramagnetic susceptibility; however, no ferromagnetism phenomenon occurs in this case [1].

In this paper, we show that the strong interaction amplitude is negative only for a limited interval of energies, which corresponds to a finite interval of *d*-hole concentration.

The number of 4*s*-electrons per unit cell n_s is actually used as an adjustable parameter. According to band calculations, which are qualitatively confirmed by experiments, this number does not exceed unity.

Using the electrical neutrality equation, we find that the average number of 3*d*-holes h_d per cell is less than unity for Ni, lies in the interval $1 < h_d < 2$ for Co, and in the interval $2 < h_d < 3$ for Fe.

For each of those intervals, we construct a magnetic phase diagram and determine the temperature dependence of the magnetic susceptibility in the paramagnetic phase. We also explain the absence of ferromagnetism in manganese and chromium, and in palladium and platinum, for which $3 < h_d < 5$ and $h_d < 1$, respectively.

The computation of magnetic susceptibility is based on the one-loop approximation using the seed semielliptic density of states. The corresponding half-width is the only energy parameter, since the Hubbard energy is assumed infinite in all cases.

2. GENERAL RELATIONS

Upon the diagonalization of the zero Hamiltonian, which corresponds to nonoverlapping atomic states, the creation and destruction operators are represented as an expansion in terms of all possible transitions between *N*- and (*N* + 1)-hole states:

$$\hat{a}^{+}_{(m,\sigma)}(\mathbf{r}) = \sum_{\alpha} g_{\alpha}^{m,\sigma}\hat{X}^{\alpha}_{\mathbf{r}},$$
$$\hat{a}_{(n,\sigma)}(\mathbf{r}) = \sum_{\beta} g_{\beta}^{n,\sigma}\hat{X}^{\beta}_{\mathbf{r}}. \tag{1}$$

Here, the indices α and β correspond to reverse transitions $s \longrightarrow m$, i.e., $\beta(m, s) = -\alpha(s, m)$. The quantities $g_{\alpha}^{m,\sigma}$ are called fractional parentage coefficients and are calculated below.

Equations for the average occupation numbers n_m are found from the definition of the temperature Green function calculated for every pair of adjoint *X*-operators (see [2]):

$$D^{\alpha,\beta}(\mathbf{r}, \tau; \mathbf{r}, \tau') = -\Theta(\tau - \tau')\langle X^{\alpha}_{\mathbf{r}}(\tau)X^{\beta}_{\mathbf{r}}(\tau')\rangle + \Theta(\tau' - \tau)\langle X^{\beta}_{\mathbf{r}}(\tau')X^{\alpha}_{\mathbf{r}}(\tau)\rangle. \tag{2}$$

To calculate the single-particle Green function, we use the simplest one-loop approximation of the self-consistent field. In this approximation, the Fourier components of the single-particle Green function $D_{\omega}^{\alpha,\beta}(\mathbf{p})$ differ from the so-called virtual Green function $G_{\omega}^{\alpha,\beta}(\mathbf{p})$ only in the factors f_{β}. In turn, this virtual function satisfies the Dyson equation

$$D_{\omega}^{\alpha,\beta}(\mathbf{p}) = G_{\omega}^{\alpha,\beta}(\mathbf{p})f_{\beta},$$
$$\{\hat{G}^{-1}_{\omega}(\mathbf{p})\}^{\alpha}_{\beta} = \{i\omega - \epsilon_m + \epsilon_s\}\delta(\alpha + \beta) - \Sigma_{\omega}^{\alpha,\beta}(\mathbf{p}). \tag{3}$$

Here, $(\epsilon_m - \epsilon_s)$ is the transition energy corresponding to the transition number α and $\omega = T(2n + 1)\pi$.

When the numbers of the single-particle transition $\beta(m, s)$ are given, every end factor f_β is equal (by definition) to the sum of average occupation numbers of the initial and terminal states [3]. On the other hand, the self-energy part in our approximation is a sum of products of the end factor by the generalized jump matrix and the one-loop correction, which is independent of the frequency and momentum:

$$f_{\alpha(s,m)} = n_s + n_m, \quad \Sigma^{\alpha,\beta}(\mathbf{p}) = f_\alpha t_\beta^\alpha(\mathbf{p}) + \Sigma^{\alpha,\beta}, \qquad t_\beta^\alpha(\mathbf{p}) = g_\alpha^{k,\sigma} t_s^k(\mathbf{p}) g_\beta^{s,\sigma}. \tag{4}$$

The average values of the occupation numbers n_{N+1}^m of the terminal states are found from the diagonal component at $\beta = -\alpha$:

$$\lim_{\delta \to +0} D^{\alpha,\beta}(\mathbf{r}, \tau; \mathbf{r}, \tau + \delta) = \lim_{\delta \to +0} T \sum_{\omega, \mathbf{p}} D_\omega^{\alpha,\beta}(\mathbf{p}) \exp(i\omega\delta) \tag{5}$$

$$= \langle X_{\mathbf{r}}^{\beta(m,s)} X_{\mathbf{r}}^{\alpha(s,m)} \rangle = \langle X_{\mathbf{r}}^{m,s} X_{\mathbf{r}}^{s,m} \rangle = \langle X_{\mathbf{r}}^{m,m} \rangle = n_{N+1}^m.$$

Equations (4) determine all end factors $f(\alpha(s, m)) = n_s + n_m$ that occur in the definition of the diagonal components of the single-particle Green function, which, in turn, is expressed in terms of various end factors.

In the one-loop approximation, the self-energy parts are independent of the momentum and frequency. In the simplified calculations conducted below, they prove to be diagonal with respect to the transition number; i.e., $\Sigma^{\alpha,\beta} = \delta_{\alpha,\beta}\Sigma_\alpha$. If we consider only transitions between high-spin N- and $(N + 1)$-hole states, then it is convenient to use the projection of the spin of the $(N + 1)$-hole state with the spin $S = (N + 1)/2$ instead of the indices α and β (see, e.g., [4]). In this case, the fractional parentage coefficients squared are determined in terms of the spin magnitude S and its projection M:

$$|g_S(M)| = \sqrt{\frac{S + M}{2S}}, \qquad M = S, S - 1, S - 2, \ldots, -S. \tag{6}$$

The average occupation numbers appearing in the expression for the Green function (2) are also determined in terms of the magnitude of the total spin and its projection. When a magnetic field is turned on, the energy levels are split, which determines corrections to the average occupation numbers.

The relationship between the variations of N- and $(N \pm 1)$-particle states can be expressed in terms of the mean value of the diagonal component of the Green function at the zero value of the magnetic field:

$$K = \frac{T}{g^2} \lim_{\delta \to +0} \sum_{\omega, \mathbf{p}} \sum_{\alpha,\beta} g_\alpha G_\omega^{\alpha,\beta}(\mathbf{p}) g_\beta f_\beta \exp(i\omega\delta), \qquad g^2 = \sum_\alpha g_\alpha^2. \tag{7}$$

In order to obtain equations for the end factors, we average the T-product of the destruction operator (1) by a linear combination of adjoint X-operators with arbitrary coefficients $\gamma_\alpha^{m,\sigma}$:

$$\hat{Y}_{(m,\sigma)} = \sum_\beta \gamma_\beta^{m,\sigma} \hat{X}_{(m,\sigma)}^\beta \times \langle \hat{T}(\hat{a}_{(n,\sigma)}(\mathbf{r}, \tau) \hat{Y}_{(m,\sigma)}(\mathbf{r}', \tau')) \rangle \tag{8}$$

$$= \sum_{\alpha,\beta} g_\alpha^{n,\sigma} \gamma_\beta^{m,\sigma} \langle \hat{T}(\hat{X}_{\mathbf{r},\tau}^\alpha, \hat{X}_{\mathbf{r},\tau'}^\beta) \rangle.$$

Here, we use the expansion of the annihilation operator (1) on the basis of the known fractional parentage coefficients $g_\beta^{n,\sigma}$. Passing to the limit $\tau' \longrightarrow \tau$ $(\tau' > \tau)$, we obtain equations for finding all $(N + 1)$-particle occupation numbers n_{N+1}^s:

$$\sum_\alpha g_{\alpha(m,s)}^{(k,\sigma)} n_{N+1}^s \gamma_{\alpha(s,m)}^{(k,\sigma)} = T \lim_{\delta \to +0} \sum_{\omega, \mathbf{p}} \sum_{\alpha,\beta} g_{\alpha(m,s)}^{(k,\sigma)} \times D_\omega^{\alpha,\beta}(\mathbf{p}) \gamma_{\beta(s,m)}^{(k,\sigma)} \exp(i\omega\delta). \tag{9}$$

Substitute expression (3) for the Green function into this equation to obtain

$$\sum_\alpha g_{\alpha(m,s)}^{(k,\sigma)} n_{N+1}^s \gamma_{\alpha(s,m)}^{(k,\sigma)} = T \sum_{\alpha,\beta} g_{\alpha(m,s)}^{(k,\sigma)} \times \lim_{\delta \to +0} \sum_{\omega, \mathbf{p}} G_\omega^{\alpha,\beta}(\mathbf{p}) \exp(i\omega\delta) f_\beta \gamma_{\beta(s,m)}^{(k,\sigma)}. \tag{10}$$

Now, we explicitly calculate the sum in the right-hand side over α:

$$\sum_\alpha g_{\alpha(m,s)}^{(k,\sigma)} G_\omega^{(\alpha,\beta)}(\mathbf{p}) = W_\omega^\beta(\mathbf{p}). \tag{11}$$

To this end, we use the simplified expression for the inverse Green function

$$[\hat{G}_\omega^{-1}(\mathbf{p})]_\beta^\alpha = (E_\omega^{\alpha,\sigma})\delta_{\alpha,\beta} - g_\alpha f_\alpha t_{\mathbf{p}} g_\beta. \tag{12}$$

For simplicity, we assume that the Fourier components of the matrix of jump integrals $t_{\mathbf{p}}$ are independent of the numbers of the terminal and initial states for the selected group of transitions. This assumption leads to

the elimination of the off-diagonal (with respect to the indices α and β) components of the self-energy Σ_α, which are independent of the frequency and quasi-momentum in the one-loop approximation. The single-particle energy E_α, which depends on the difference of N- and $(N+1)$-hole states, is assumed to be equal to the sum of the chemical potential and the magnitude of the spin splitting in the given magnetic field H:

$$E_\omega^{\alpha,\sigma} = i\omega + \mu + \sigma H - \Sigma_\alpha, \qquad \omega = (2k+1)\pi T, \quad \sigma = \pm 1. \tag{13}$$

Thus, we are interested only in the spin dependence on the magnitude of the applied weak magnetic field, which corresponds to the further possibility to determine the spin magnetic susceptibility and investigate the possible occurrence of the ferromagnetic instability.

An explicit expression for the inverse Green function (12) makes it possible to find the desired sum over the indices α in Eq. (11):

$$W_\omega^\beta(\mathbf{p}) = \frac{g_\beta}{E_\omega^{\beta,\sigma}}\left[1 - \sum_\alpha \frac{g_\alpha^2 t_\mathbf{p} f_\alpha}{E_\omega^{\alpha,\sigma}}\right]^{-1}. \tag{14}$$

As a result, we find the explicit form of Eqs. (10):

$$\sum_\beta g_\beta n_{N+1}^\beta \gamma_\beta = T\lim_{\delta\to+0}\sum_{\omega,\mathbf{p}}\exp(i\omega\delta) \times \sum_\beta \frac{g_\beta f_\beta \gamma_\beta}{E_\omega^{\beta,\sigma}}\left[1 - \sum_\alpha \frac{g_\alpha^2 t_\mathbf{p} f_\alpha}{E_\omega^{\alpha,\sigma}}\right]^{-1}. \tag{15}$$

Since the numerical coefficients γ_β are arbitrary, the number of independent equations in (15) (under the given values of the fractional parentage coefficients g_β) is equal to the number of possible values of β, which equals the number of different transitions determined by the initial expansion (1).

In the limit $H \longrightarrow 0$, the end factors and one-loop self-energy parts become independent of the transition number β; i.e.,

$$\lim_{H\to 0} f_\beta = f_0, \quad \lim_{H\to 0} E_\omega^{\beta,\sigma} = i\omega + \mu_0.$$

Setting $\gamma_\beta = g_\beta$ in Eqs. (15) and passing to the limit as $H \longrightarrow 0$, we find an equation for the average $(N+1)$-hole states in the form

$$n_{N+1} = R_{1+N} f_0 K,$$
$$\text{where } K = T\lim_{\delta\to+0}\sum_{\omega,\mathbf{p}}\frac{\exp(i\omega\delta)}{[i\omega - \xi_\mathbf{p}]} = \sum_\mathbf{p} n_F(\xi_\mathbf{p}), \tag{16}$$

where R_{N+1} is the multiplicity of the $(N+1)$-hole state in the absence of the magnetic field, $n_F(\epsilon)$ is the Fermi distribution, $\xi_\mathbf{p} = (g^2 f_0 t_\mathbf{p} - \mu_0)$ is the energy of the single-particle excitations, and g^2 is the complete sum of the fractional parentage coefficients with account for the degeneracy order of the selected group of transitions: $g^2 = \sum_\alpha g_\alpha^2$.

In the absence of field, the end factor equals the sum of occupation numbers of N- and $(N+1)$-hole states; therefore, the total number of holes is expressed in terms of the integral function K by the equation

$$n = [n] + R_{1+N} f K, \quad f_0 = \frac{1+[n]-n}{R_N} + \frac{n-[n]}{R_{N+1}}, \qquad [n] \le n < [n]+1. \tag{17}$$

Here, the square brackets denote the integer part, and the second relation can be written with regard for the fact that in our approximation the end factor is a linear function of the occupation numbers.

The system of equations (16), (17) determines the relationship between the chemical potential, temperature, and density of holes n. For convenience of computations, we define the renormalized chemical potential and the renormalized temperature:

$$\mu_0 = \tilde{\mu} f_0 g^2, \quad T = \tilde{T} f_0 g^2. \tag{18}$$

Then, the integral function K depends on $\tilde{T}$ and $\tilde{\mu}$ as if these variables were the temperature and chemical potential of the ideal gas. Furthermore, using Eqs. (17), we find the dependence of these parameters on the end factor f_0 and the average number of hole states n; then, we use Eqs. (18) to obtain the true temperature and chemical potential.

To find out how the spin magnetic susceptibility depends on temperature and concentration, we set $\gamma_\beta = g_\beta$ in the equations for the transitions that differ in the value of the total spin projection at the terminal $(N+1)$-hole state. For other transitions with the same terminal spin projection, we set $\gamma_\beta = 0$ and then vary the right-hand side of Eq. (15) with respect to the magnitude of the magnetic field, end factors, and one-loop self-energy parts:

$$\sum_{k=-S+1}^{S} g_k^2 \delta n_{N+1}^{(k)} = (K + g^2 f_0 D_1)\sum_{k=-S+1}^{S} g_k^2 \delta f_k + D_0 f_0 \sum_{k=-S+1}^{S} g_k^2 \delta\Sigma_k - b^2 f_0 D_0 \delta H. \tag{19}$$

The coefficients in this equation are calculated at the zero external magnetic field:

$$D_s = T\lim_{\delta\to+0}\sum_{\omega,\mathbf{p}} t_\mathbf{p}^s \frac{\exp(i\omega\delta)}{[i\omega - \xi_\mathbf{p}]^2} = \sum_\mathbf{p} t_\mathbf{p}^s n_F'(\xi_\mathbf{p}). \tag{20}$$

The quantity $b^2 = \sum_k g_k^2$ is the sum of the fractional parentage coefficients squared that correspond to the states with the given projection of the total spin at the

terminal state and the fixed collection of single-particle orbital states. Using definition (6), we find

$$b^2 = \sum_{M=-S}^{S} \frac{S+M}{2S} = S + \frac{1}{2}. \tag{21}$$

Considering all possible one-loop transitions from high-spin $(N+1)$-hole states to the states with a given total spin S, we obtain several groups of transitions satisfying definition (6). In this case, the sum of the squared fractional parentage coefficients g^2, for the given total spin S of high-spin $(N+1)$-hole states and the given number κ of different orbital states, is found in terms of the combinatorial factor

$$g^2 = b^2 C_{\kappa-1}^{2S-1} = b^2 \frac{(\kappa-1)!}{(2S-1)!(\kappa-2S)!}. \tag{22}$$

If the total number of orbital components κ is given, one must generally consider various values of the total spin from $S = 1/2$ to $S = \kappa/2$.

The other equations for the variation of $(N+1)$-hole states are conveniently written with the use of auxiliary coefficients γ_k satisfying the condition of orthogonality to the set of different fractional parentage coefficients

$$\sum_{k=-S+1}^{S} g_k \gamma_k = 0. \tag{23}$$

Varying Eq. (15) with regard for these conditions, we obtain $2S-1$ possible relations, which are independent of the variation of the magnetic field:

$$\sum_{k=-S+1}^{S} g_k \gamma_k \delta n_{N+1}^{(k)} = K \sum_{k=-S+1}^{S} g_k \gamma_k \delta f_k + f_0 A \sum_{k=-S+1}^{S} g_k \gamma_k \delta \Sigma_k. \tag{24}$$

Here, there is a new coefficient A, which has a singularity at the zero energy of excitations:

$$A = \sum_{\mathbf{p}} \frac{n_F(\xi_{\mathbf{p}}) - n_F(-\mu)}{\xi_{\mathbf{p}} + \mu}. \tag{25}$$

The right-hand side of the system (20), (24) includes variations of the average occupation numbers only implicitly through the variations of the end factors. Therefore, we must complement these equations with relations that express variations of the average occupation numbers δn_{N+1}^{k} in terms of the variations of the end factors δf_k.

These relations are easily obtained if we observe that, in an external magnetic field, the variations of the occupation numbers δn^k corresponding to the given projection of the spin $k = S^z \neq 0$ differ from the variation of the occupation numbers corresponding to the same state with the opposite spin projection only in sign.

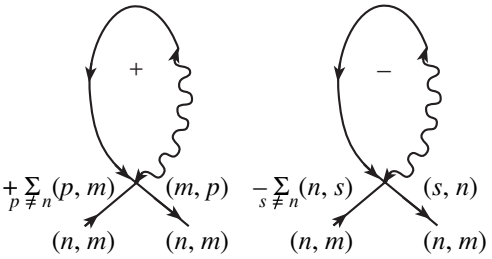


Fig. 1. One-loop self-energy parts.

Taking into account the relationship between the number of the end factor $k = S^z$ and the values of the spin projection for N- and $(N+1)$-hole states $f_k = n_{N+1}^{(k)} + n_N^{(k-1/2)}$, we immediately find the desired relations

$$\sum_{k=1-S+p}^{S-p} \delta f_k = \sum_{k=1-S+p}^{S-p} \delta n_{N+1}^{(k)} + \sum_{k=1-S+p}^{S-p} \delta n_N^{(k-1/2)} = \sum_{k=1-S+p}^{S-p-1} \delta n_{N+1}^{(k)} + \delta n_{N+1}^{(S-p)} = \delta n_{N+1}^{(S-p)}. \tag{26}$$

Here, p can take all integer values in the range from zero to $[S]$. Therefore, relations (26) give the complete solution to the problem of elimination of the variation of all $(N+1)$-hole occupation numbers from the left-hand side of Eqs. (19) and (24).

In order to obtain equations for the variation of the self-energy parts Σ_α, we represent this part as a sum of two terms each of which contains the single-particle Green function summed over momenta and frequencies (see Fig. 1):

$$\Sigma_\alpha = A_{\alpha,\beta} Y_\beta^\sigma + B_{\alpha,\beta} Y_\beta^{-\sigma}. \tag{27}$$

Here, $A_{\alpha,\beta}$ and $B_{\alpha,\beta}$ are given numerical matrices determined by the matrix elements of the kinematic interaction characteristic of the given group of spin states. The functions Y_β^σ are expressed in terms of the same sum $W_\omega^\alpha(\mathbf{p})$ as that involved in the calculation of the occupation numbers:

$$Y_\alpha^\sigma = g_\alpha T \sum_{\omega,\mathbf{p}} t_{\mathbf{p}} \sum_\beta g_\beta [G_\omega(\mathbf{p}, \sigma)]_\alpha^\beta = g_\alpha T \sum_{\omega,\mathbf{p}} t_{\mathbf{p}} W_\omega^\alpha(\mathbf{p}). \tag{28}$$

Substitute expression (14) into this formula to obtain

$$Y_\alpha^\sigma = g_\alpha^2 T \sum_{\omega, \mathbf{p}} t_\mathbf{p} \frac{1}{E_\omega^{\alpha,\sigma}} \left[1 - \sum_\nu \frac{g_\nu^2 t_\mathbf{p} f_\nu^\sigma}{E_\omega^{\nu,\sigma}} \right]^{-1}. \quad (29)$$

Under a zero external magnetic field, the one-loop corrections are independent of the transition number, which yields an insignificant correction to the chemical potential.

For the calculation of the variation Σ with respect to the external magnetic field, we use the obvious equation

$$\delta\Sigma_\alpha = A_{\alpha,\beta}\delta Y_\beta^\sigma - B_{\alpha,\beta}\delta Y_\beta^\sigma. \quad (30)$$

Furthermore, it is natural to define the numerical matrix

$$S_{\alpha,\beta} = A_{\alpha,\beta} g_\beta^2 - B_{\alpha,\beta} g_\beta^2, \quad (31)$$

which is involved in the expressions of the variation of the right-hand side of Eqs. (29) and (24):

$$\begin{aligned} \delta\Sigma_k &= g^2 \sum_{p=-S+1}^{S} D_{k,p}^{(2)} \delta f_p \\ &+ \sum_{p=-S+1}^{S} [D_{k,p}^{(1)} - F_{k,p}] \delta\Sigma_p - R_k D_1 \delta H. \end{aligned} \quad (32)$$

Here, we introduced the components of the numerical vector $R_k = \sum_{p=-S+1}^{S} S_{k,p}$ that can be used to express all other matrices:

$$\begin{aligned} D_{k,p}^{(n)} &= D_n U_{k,p}, \\ F_{k,p} &= Q W_{k,p}, \quad Q = \frac{[K - n(-\mu)]}{f_0 g^2}. \end{aligned} \quad (33)$$

Here, $g^2 = \sum_\alpha g_\alpha^2$ differs from $b^2 = \sum_k b_k^2$ only in the factor $C_{\kappa-1}^{2S-1}$,

$$\begin{aligned} U_{k,p} &= \frac{R_k g_p^2}{b^2}, \quad b^2 = S + \frac{1}{2}, \\ W_{k,p} &= U_{k,p} - S_{k,p}, \end{aligned} \quad (34)$$

where S is magnitude of the spin of $(N+1)$-hole states and κ is the total number of distinguished orbital states.

Thus, given the numbers $S_{k,p}$ and the coefficients D_n, A, K, and Q, the system of equations (19), (24), (26), (32) makes it possible to determine the variations of all end factors and one-loop self-energy parts (see Fig. 1).

To determine the spin susceptibility χ, one should find corrections to the spin magnetic moment:

$$\begin{aligned} \chi &= 2\kappa \frac{\mu_B^2}{v_0} \frac{\delta M}{\delta H}, \\ \delta M &= \sum_{k=-S}^{S} k \delta n_{N+1}^{(k)} + \sum_{m=-S+1/2}^{S-1/2} m \delta n_N^{(m)}, \end{aligned} \quad (35)$$

where v_0 is the volume of the unit cell.

The linear combination of the variations of occupation numbers involved in the definition of the spin moment must be expressed in terms of the variations of end factors, which are determined from the system of equations (19), (24), (26), (32).

In order to obtain the corresponding relation, we first calculate an auxiliary sum using the definition of end factors (6):

$$\begin{aligned} \sum_{k=-S+1}^{S} k\delta f_k &= \sum_{k=-S+1}^{S} k\delta n_{N+1}^{(k)} + \sum_{k=-S+1}^{S} k\delta n_N^{(k-1/2)} \\ &= \sum_{k=-S}^{S} k\delta n_{N+1}^{(k)} + S\delta n_{N+1}^{(-S)} \\ &+ \sum_{k=-S+1}^{S} \left(k - \frac{1}{2}\right)\delta n_N^{(k-1/2)} + \frac{1}{2}\sum_{k=-S+1}^{S} \delta n_N^{(k-1/2)}. \end{aligned} \quad (36)$$

The last sum vanishes, because it is taken over all projections of the total spin pertaining to N-hole states. The first and the third sums provide the correction to the spin moment, and the second term can be written as the sum of variations of the end factors by formula (26) at $p = 0$:

$$S\delta n_N^{(-S)} = -S\delta n_N^{(S)} = -S \sum_{k=1-S}^{S} \delta f_k. \quad (37)$$

As a result, we obtain the variation of the spin moment in terms of the variations of the end factors:

$$\begin{aligned} \delta M &= \sum_{k=1-S}^{S} (S+k)\delta f_k \\ &= \sum_{p=1}^{2S} p\delta \tilde{f}_{2S+1-p} = \sum_{n=1}^{2S} (2S+1-n)\delta \tilde{f}_n. \end{aligned} \quad (38)$$

Here and in what follows, we use another indexing of the end factors: $\tilde{f}_n = f_{S+1-n}$ $(n = 1, 2, \ldots, 2S)$.

Equations (26), (35), and (38) enable us to use only end factors and their variations as variables and then to determine the susceptibility as a function of the renormalized temperature and chemical potential (see (18)).

To find the conditions for ferromagnetic instability, it is sufficient to write solvability conditions for the

homogeneous system of equations obtained from Eqs. (19), (24), (26), and (32) at $\delta H = 0$.

As a result, we find a relation between the renormalized temperature and chemical potential, which can be used to obtain the magnetic phase diagram in the variables *concentration–temperature.*

3. FERROMAGNETISM IN NICKEL

In the case of Ni, Pd, and Pt, the *d*-system resonates between empty (ten-electron) and one-hole (nine-electron) states.

Consider the limiting case when the influence of the crystalline field can be neglected compared to the width of the hole conduction band. At the zero magnetic field and an infinite Hubbard energy, we obtain by (17) the state equation

$$h_d = 2\kappa f \sum_{\mathbf{p}} n_F(\xi_{\mathbf{p}}), \quad \xi_{\mathbf{p}} = f t_{\mathbf{p}} - \mu, \qquad f = 1 - \frac{h_d(2\kappa - 1)}{2\kappa}. \tag{39}$$

Here, $\kappa = 5$ is the total number of the orbital one-hole states and f is the end factor which is a linear function of the average number of holes per the cell n_h in the absence of the magnetic field. The chemical potential μ includes the sum of all one-loop self-energy parts, which are independent of spin and orbital indices in the absence of the magnetic field.

If a finite magnetic field is applied, it is convenient to separate the spin and orbital parts and write the equation for the average number of holes that have a given projection of the spin σ and of the single-particle orbital moment:

$$n_m^\sigma = f_m^\sigma \sum_{\mathbf{p}} n_F(\xi_m^\sigma(\mathbf{p})), \qquad \xi_m^\sigma(\mathbf{p}) = f_m^\sigma t_{\mathbf{p}} + \Sigma_m^\sigma - \mu - \sigma H - mH, \qquad f_m^\sigma = n_0 + n_m^\sigma. \tag{40}$$

Here, the end factors are equal to the sum of the average number of empty, n_0, and one-hole, n_m^σ, states, so that their variation, which is connected with the variation of the magnetic field, coincides with the variation of the one-hole states: $\delta n_m^\sigma = \delta f_m^\sigma$.

Equations for the one-loop self-energy parts have a very simple form, since we assume, for simplicity, that the spectrum of energy excitations is diagonal with respect to both the spin and orbital indices:

$$\Sigma_m^\sigma = -T \sum_{\mathbf{p}, k, \omega} t_{\mathbf{p}} \{ G_\omega^{\sigma, k}(\mathbf{p}) + G_\omega^{-\sigma, k}(\mathbf{p}) \} + T \sum_{\mathbf{p}, \omega} t_{\mathbf{p}} G_\omega^{\sigma, m}(\mathbf{p}). \tag{41}$$

Here, the last term and the first term in the first sum cancel out, which corresponds to the zero scattering amplitude of excitations with identical spin and orbital indices. In our representation, the single-particle Green function is diagonal and the summation over the frequency $\omega_n = \pi T(2n + 1)$ yields the well-known Fermi function

$$\Sigma_m^\sigma = -\sum_{\mathbf{p}, k} t_{\mathbf{p}} \{ n_F(\xi_k^\sigma(\mathbf{p})) + n_F(\xi_k^{-\sigma}(\mathbf{p})) \} + \sum_{\mathbf{p}} t_{\mathbf{p}} n_F(\xi_m^\sigma(\mathbf{p})). \tag{42}$$

To find the conditions for ferromagnetic instability, it is sufficient to determine the conditions for the appearance of nonzero variations of the average occupation numbers δn_m^σ at the zero variation of the magnetic field magnitude H. Assume that the variations are independent of the orbital variable m, but change sign when the sign of the spin projection is changed:

$$\delta n_m^\sigma = -\delta n_m^{-\sigma} = \delta f^\sigma = -\delta f^{-\sigma}. \tag{43}$$

The fact that the variation is independent of the spin indices corresponds to the possibility of separating the susceptibility into the spin and orbital parts and the further separation of the singular part of the spin susceptibility. It can be proved that the orbital instability does not occur in the cubic lattice; therefore, below we do not consider the dependence on the orbital indices.

The variation of Eqs. (40) and (41) yields

$$\begin{pmatrix} 1 - K - f D_1 & -f D_0 \\ -D_2 & 1 - D_1 \end{pmatrix} \begin{pmatrix} \delta f^\sigma \\ \delta \Sigma^\sigma \end{pmatrix} = \begin{pmatrix} -f D_0 \sigma \delta H \\ -D_1 \sigma \delta H \end{pmatrix}. \tag{44}$$

All the coefficients are determined in terms of the seed density of states $\rho_0(\epsilon)$:

$$K = \int \rho(\epsilon) n_F(f\epsilon - \mu) d\epsilon, \qquad D_k = \int \rho(\epsilon) \epsilon^k n_F'(f\epsilon - \mu) d\epsilon. \tag{45}$$

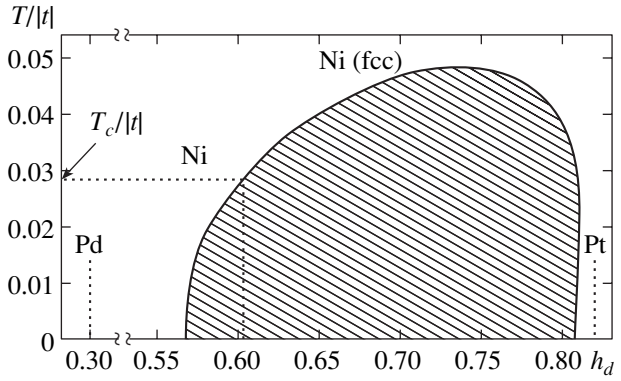


Fig. 2. The phase diagram of nickel, palladium, and platinum. The ferromagnetic region is shaded. The arrow marks the theoretical value of the transition temperature: $T_c/|t| = (627.4 \times 2)/(3.78 \times 11606) = 0.0286$.

Solving system (44), we find the susceptibility

$$\chi(T) = \frac{2\kappa\mu_B^2}{v_s}\frac{\delta f^\sigma}{\delta H} = 2\kappa\frac{\mu_B^2}{v_s} \times \frac{-fD_0}{1-K-fD_1-(1-K)D_1-f(D_0D_2-D_1^2)}, \quad (46)$$

where v_s is the volume of the unit cell and μ_B is the Bohr magneton. The condition for ferromagnetic ordering is found from the solvability condition for the homogeneous system (44) at $\delta H = 0$:

$$1-K = (f+1-K)D_1 + f(D_0D_2 - D_1^2), \quad (47)$$

$$1-K = \frac{fD_1}{1-D_1} \text{ at } T = 0. \quad (48)$$

The equation of state that relates the chemical potential to the temperature is also expressed in terms of the function K and a parameter κ that is interpreted as the number of orbital components:

$$n = 2\kappa fK, \quad f = 1 - n\left(1 - \frac{1}{2\kappa}\right), \quad (49)$$

where κ is the degree of orbital degeneration.

In the limit $T \longrightarrow 0$, these equations become much simpler:

$$K = \int_{-\infty}^{\nu} \rho(\epsilon)d\epsilon, \quad fD_1 = -\nu\rho(\nu), \quad f = \frac{1}{1+(2\kappa-1)K}, \quad n = \frac{2\kappa K}{1+(2\kappa-1)K}. \quad (50)$$

Since all the quantities here are expressed in terms of a single parameter ν, we can substitute them into the condition for ferromagnetic instability, one can calculate the critical values corresponding to the domain of existence of ferromagnetism at $T = 0$.

In the model of semielliptic density of states, when $\rho(\epsilon) = 2\sqrt{1-\epsilon^2}/\pi$, we have the condition

$$K = \frac{\alpha - \sin\alpha}{2\pi}, \quad fD_1 = \frac{\sin\alpha}{\pi}, \quad \frac{1-K}{-(2\kappa-1)K^2 + (2\kappa-2)K + 2} = \frac{\sin\alpha}{\pi}. \quad (51)$$

For the given degeneration multiplicity κ, the last equation determines the critical value K_c corresponding to the critical concentration $n_c = 2\kappa K_c/[1 + (2k-1)K_c]$.

At $\kappa = 1$ or 2, the equations have no solutions (see [5]). If $\kappa = 5$, we obtain two critical points $K_{c1} = 0.1161$, i.e., $n_{c1} = 0.5671$, and $K_{c2} = 0.295$, i.e., $n_{c2} = 0.8071$.

Thus, for the model of a semielliptic band, ferromagnetism occurs in the finite domain of concentrations from $n_{c1} = 0.56710$ to $n_{c2} = 0.8071$ (see Fig. 2).

The theory proposed above gives a qualitative explanation of the magnetic properties of Ni, Pd, and Pt, which belong to subgroup 8. All of them have a cubic unit cell of the fcc type. The total number of conduction electrons is ten, and the number of electrons on the unfilled *s*-shell does not exceed unity. According to band theory calculations, the number of *s*-electrons on the unfilled shell is 0.81, 0.59, and 0.94 for Ni, Pd, and Pt, respectively. Hence, we conclude that the number of hole *d*-states (h_d) is equal to 0.81, 0.59, and 0.94, respectively. Therefore, the number of hole states of nickel lies between h_d for palladium and platinum. It is natural to assume that the concentration of holes for Pd and Pt is outside the domain in which there is the ferromagnetic instability. At the same time, the intermediate value of hole concentration for Ni is within the domain in which there is the ferromagnetic instability (see Fig. 2, which illustrates the quantitative agreement between the experimental and theoretical values of the magnetic saturation moment).

4. FERROMAGNETISM IN COBALT

Consider the case when the system resonates between the one-hole and two-hole high-spin states with spin 1. Represent the variation of the two-hole occupation numbers in terms of the variation of two end factors. In accordance with (26), we have

$$\delta n_{II}^\sigma = \sum_{k=1}^{2} \delta f_k^\sigma, \quad \delta n_{II}^0 = 0. \quad (52)$$

These relations must be used for obtaining equations

for the variation of end factors δf_k^σ:

$$b_1^2\delta n_{II}^\sigma = \sum_{k=1}^{2}\delta f_k^\sigma = (K + g^2 f D_1)\sum_{k=1}^{2} b_k^2\delta f_k^\sigma + f D_0\sum_{k=1}^{2} b_k^2\delta\Sigma_k^\delta - f D_0 b^2\delta H. \quad (53)$$

Here, we must use the values of the fractional parentage coefficients (6) at $S = 1$ and notation (22):

$$b_1^2 = 1, \quad b_2^2 = \frac{1}{2}, \quad \sum_{k=1}^{2} b_k^2 = \frac{3}{2}, \quad g^2 = b^2(\kappa - 1), \quad (54)$$

which corresponds to the maximum possible number of degenerate *d*-orbitals ($\kappa = 2$, 3, or 5).

Here and in what follows, we have

$$K = \sum_{\mathbf{p}} n_F(\xi_{\mathbf{p}}), \quad D_k = \sum_{\mathbf{p}} (t_{\mathbf{p}})^k n_F'(\xi_{\mathbf{p}}), \quad \xi_{\mathbf{p}} = g^2 f t_{\mathbf{p}} - \mu. \quad (55)$$

Two other equations are obtained from (24) with the use of the additional orthogonality condition $\gamma_1 b_1 = -\gamma_2 b_2$:

$$b_1\gamma_1\delta n_{II}^\sigma = K\sum_{k=1}^{2} b_k\gamma_k\delta f_k^\sigma + A\sum_{k=1}^{2} b_k\gamma_k\delta\Sigma_k^\sigma. \quad (56)$$

Relations (52) make it possible to eliminate the variations of two-particle occupation numbers from the equations, so that Eqs. (53) and (56) are written in the form

$$\delta f_1 + \delta f_2 = (K + g^2 f D_1)\sum_{k=1}^{2} b_k^2\delta f_k^\sigma + f D_0\sum_{k=1}^{2} b_k^2\delta\Sigma_k^\sigma - f D_0 b^2\delta H, \quad (57)$$

$$\delta f_1 + \delta f_2 = (\delta f_1 - \delta f_2)K + fA(\delta\Sigma_1^\sigma - \delta\Sigma_2^\sigma). \quad (58)$$

Here, we used the condition $\gamma_1 b_1 = -\gamma_2 b_2$ and the resonance coefficient *A* defined in (25).

The resulting equations at the zero corrections to the self-energy parts $\delta\Sigma$ determine the susceptibility and the boundary of the ferromagnetic instability in the zero-loop approximation:

$$K(1 - K) - fg^2(1/3 + K)D_1. \quad (59)$$

Now, the equation that connects the variations of the occupation numbers also includes variations of the self-energy parts (see Fig. 1).

Considering one-loop self-energy diagrams, one can notice that the off-diagonal (with respect to the transition numbers (k, p = 1, 2)) self-energy parts vanish, while all the diagonal elements are determined in terms of integrals of Green functions with the given projection of the spin σ:

$$\Sigma_1^{(a,\sigma)} = -(\kappa - 3)J_1^{(\sigma)} - (\kappa - 1)J_2^{(-\sigma)}, \quad \Sigma_2^{(a,\sigma)} = -(\kappa - 1)J_1^{(-\sigma)} + 2J_2^{(-\sigma)} - (\kappa - 3)J_2^{(\sigma)}. \quad (60)$$

Here, κ is the total number of orbital components; in our case, $\kappa = 2$, 3, or 5.

The quantities on the right-hand side are independent of the number of the atomic state and, for the given σ, differ in a factor proportional to the corresponding fractional parentage coefficient squared: $g_1^2 = 1$ or $g_2^2 = 1/2$. We have

$$J_k^{(\sigma)} = T\sum_{n,\omega,\mathbf{p}} t_{\mathbf{p}}^{(k,n)} G_\omega^{(n,k)} \propto g_k^2 C(\sigma). \quad (61)$$

In the absence of the magnetic field, the functions $C(\sigma)$ are independent of the spin projection. Therefore, substituting them into (61), we find that in this limit, both self-energy parts are reduced to a constant, which leads to the correction to the chemical potential and is not taken into account below.

Equations for $\delta\Sigma$ are obtained in agreement with the general relations (31)–(34). The numerical values of the matrix $\hat{S}$ constructed in accordance with the definition of the self-energy matrix using (60) and Fig. 1 has the following form in the case under consideration:

$$\hat{S} = \begin{pmatrix} 3 - \kappa & (\kappa - 1)/2 \\ \kappa - 1 & (1 - \kappa)/2 \end{pmatrix}, \quad \hat{R} = \begin{pmatrix} (5 - \kappa)/2 \\ (\kappa - 1)/2 \end{pmatrix}. \quad (62)$$

The matrix $\hat{U}$ is represented in the form of the products $U_{k,n} = R_k b_n / b^2$:

$$\hat{U} = \begin{pmatrix} (5 - \kappa)/3 & (5 - \kappa)/6 \\ (\kappa - 1)/3 & (\kappa - 1)/6 \end{pmatrix}. \quad (63)$$

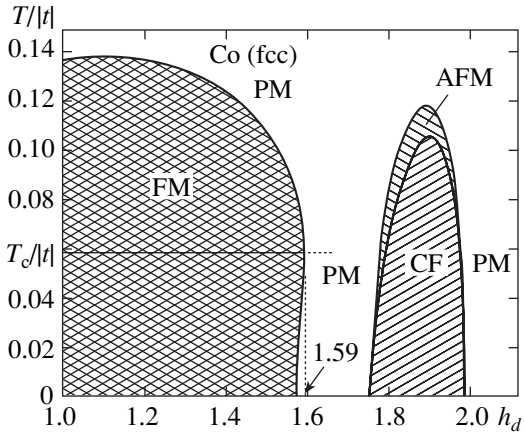


Fig. 3. The phase diagram of cobalt. FM denotes the ferromagnetic regions, AFM denotes the antiferromagnetic ones, CF is the mixed phase, and PM denotes the paramagnetic regions; $T_c/|t| = (1388 \times 2)/(4.35 \times 11606) = 0.05984$.

The operator $\hat{F}^{(0)} = Q(\mu)\hat{W}$, where Q is defined in (33), and the matrix $\hat{W}$ has a zero sum of elements in each row:

$$\hat{W} = \hat{U} - \hat{S} = \begin{pmatrix} 2(\kappa-2)/3 & 2(2-\kappa)/3 \\ 2(-\kappa+1)/3 & 2(\kappa-1)/3 \end{pmatrix}. \quad (64)$$

The inhomogeneous term is proportional to the two-dimensional vector $\hat{R}$.

Let us write equations for the corrections to the end factors for high-spin 3*d*-hole states when $\kappa = 5$. We have

$$\begin{pmatrix} 1-b_1^2P & 1-b_2^2P & -fD_0b_1^2 & -fD_0b_2^2 \\ 1-K & 1+K & -fA & fA \\ -g^2D_2K & -g^2D_2K & 1-D_1b_1^2 & 3/4-D_1b_2^2 \\ 0 & 0 & 1+2Q & -2Q \end{pmatrix} \times \begin{pmatrix} \delta f_1 \\ \delta f_2 \\ \delta\Sigma_1 \\ \delta\Sigma_2 \end{pmatrix} = \begin{pmatrix} -fD_0b^2\delta H \\ 0 \\ -D_1b^2\delta H \\ 0 \end{pmatrix}, \quad (65)$$

where

$$P = K + g^2 f D_1, \quad f = \frac{1}{10(2K+1)}, \quad g^2 = 4b^2 = 6, \quad n_d = \frac{1+5K}{1+2K}. \quad (66)$$

Equate the corresponding determinant to zero to obtain an equation for finding the ferromagnetic instability condition. At $T = 0$, when the condition $D_2D_0 - D_1^2 = 0$ is satisfied, we have the equation

$$-\frac{4}{3}g^2 fAD_2 + K(1-K)(3+14Q) = \frac{1}{3}fg^2D_1(3K+1)(3+14Q) + 2D_1(1+6Q)K(1-K). \quad (67)$$

All the coefficients here depend on the position of the chemical potential.

For the semielliptic band model, all quantities in the equations are conveniently expressed in terms of the parameter $\alpha = \pi + 2\arcsin(\mu/g^2 f)$:

$$K = \frac{\alpha - \sin\alpha}{2\pi}, \quad D_0 = -\frac{2\sin(\alpha/2)}{f\pi g^2}, \quad D_1 = \frac{\sin\alpha}{f\pi g^2}, \quad D_2 = \frac{D_1^2}{D_0} = -\frac{\cos(\alpha/2)\sin\alpha}{f\pi g^2}, \quad A(\mu) = \frac{2}{\pi f g^2}\left\{\ln\left|\tan\left[\frac{\pi-\alpha}{4}\right]\right| + \sin\frac{\alpha}{2}\right\}, \quad Q(\mu) = \frac{K_0 - \theta(\alpha-\pi)}{fg^2}. \quad (68)$$

The logarithmic divergence in the coefficient $A(\mu)$ at $\mu \longrightarrow 0$ in (67) is compensated for by the factor D_2.

In accordance with the general formula (38), the correction to the magnetic moment is expressed in terms of a linear combination of corrections to the end factors:

$$\delta M = 2\delta n_{\mathrm{II}}^1 + \delta n_{\mathrm{I}}^{1/2} = 2\delta f_1 + \delta f_2. \quad (69)$$

Calculating the last sum, we obtain the susceptibility

$$\chi = \frac{2\kappa\mu_b^2}{v_0}\frac{[-fD_0(1+3K)(3+14Q) - 4AD_1]}{D}, \quad (70)$$

where the denominator has the form

$$D = -\frac{4}{3}g^2AD_2 + K(1-K)(3+14Q) - fg^2D_1\left(K+\frac{1}{3}\right)(14Q+3) - 2D_1(1+6Q)K(1-K) - 2fg^2\left(\frac{1}{3}+K\right)(1+6Q)(D_2D_0 - D_1^2). \quad (71)$$

The numerical solution of Eq. (67) determines the domain of ferromagnetic instability (see Fig. 3)

$$1 < n_d < 1.5732. \quad (72)$$

In the zero-loop approximation, we find a narrower domain of ferromagnetic instability, $1 < n_d < 1.3946$.

Special properties of the phase diagram, which manifest themselves in the region of positive values of the chemical potential, are caused by the function A, which (at $T = 0$) has a logarithmic singularity at the zero chemical potential (i.e., at $\alpha = \pi$). This singularity is compensated for by the factor D_2, which vanishes at the same point. However, outside this point and at $\alpha \neq 0, 2\pi$, the product AD_2 is positive. This is the cause of the fact that the system remains paramagnetic in a wide range of concentrations except for the two narrow domains adjacent to the points corresponding to the critical values $\alpha = 0$ and π.

Another special feature of the system under consideration is the possibility to change the sign of the magnetic susceptibility by setting to zero the numerator (N) of the magnetic susceptibility. Thus, the susceptibility is negative in a narrow domain. In the limit $T = 0$, the width of the left part is zero and the right part of this domain has a small finite width,

$$1.995 < h_d < 1.999. \tag{73}$$

In the region of high temperatures, the boundary of this domain is defined by the vanishing condition for the magnetic susceptibility. For low temperatures, the boundary is determined by the condition of infinity of the susceptibility. Hence, we conclude that the narrow domain of negative susceptibility corresponds to the antiferromagnetic phase. Inside this domain, the system seems to be ordered as well, since it is adjacent to the ferromagnetic phase and turns into infinity at the boundary with this phase.

5. FERROMAGNETISM IN IRON

In this case, the system resonates between two- and three-hole high-spin states. Accordingly, we write equations for the three-hole occupation numbers. We have from the general equations (19)

$$\begin{aligned} &b_1^2\delta n_{\mathrm{III}}^{(3\sigma/2)} + b_2^2\delta n_{\mathrm{III}}^{(\sigma/2)} + b_3^2\delta n_{\mathrm{III}}^{(-\sigma/2)} \\ &= K_0 \sum_{k=1,2,3} b_k^2\delta f_k^{\sigma} + f \sum_{k=1,2,3} b_k^2\delta\Sigma_k^{\sigma} D_0 \\ &+ b^2 f \sum_{k=1,2,3} b_k^2\delta f_k^{\sigma} D_1 - b^2 f\sigma\delta H D_0. \end{aligned} \tag{74}$$

Here, we must use the numerical values of the fractional parentage coefficients obtained from (6) for the transitions to the three-hole states with the spin $S = 3/2$:

$$\begin{aligned} b_1^2 &= 1, \quad b_2^2 = \frac{2}{3}, \quad b_3^2 = \frac{1}{3}, \\ b^2 &= \sum_k b_k^2 = 2, \\ g^2 &= b^2 C_{\kappa-4}^2 = (\kappa-4)(\kappa-3). \end{aligned} \tag{75}$$

Using the general relations (26), the variations of the occupation numbers are expressed in terms of the variations of the end factors

$$\begin{aligned} \delta n_{\mathrm{III}}^{(3/2)} &= \delta f_1 + \delta f_2 + \delta f_3, \quad \delta n_{\mathrm{III}}^{(1/2)} = \delta f_2, \\ \delta n_{\mathrm{II}} &= -\delta f_2 - \delta f_3. \end{aligned} \tag{76}$$

Two additional equations are found from the orthogonality conditions. For $b_3\beta_3 = b_1\beta_1$ and $b_2\beta_2 = -2b_1\beta_1$, we have, in accordance with (24),

$$\begin{aligned} &(1-K)(\delta n_{\mathrm{III}}^{(3/2)} - 3\delta n_{\mathrm{III}}^{(1/2)}) \\ &= fA(\mu)(\delta\Sigma_1 - 2\delta\Sigma_2 + \delta\Sigma_3). \end{aligned} \tag{77}$$

Setting $\beta_2 = 0$ and $b_3\beta_3 = -b_1\beta_1$, we obtain

$$\begin{aligned} &(1-K_0)(\delta n_{\mathrm{III}}^{(3/2)} + \delta n_{\mathrm{III}}^{(1/2)}) - 2K_0\delta n_{\mathrm{II}} \\ &= fA(\mu)(\delta\Sigma_1 - \delta\Sigma_3). \end{aligned} \tag{78}$$

The coefficients K, D_k, and A are determined by the same general formulas as in the preceding section.

Three equations for the diagonal self-energy parts are written, in accordance with Fig. 1, in the form

$$\begin{aligned} \Sigma_1^{(a,\sigma)} &= 2J_1^{(\sigma)} - (\kappa-3)J_1^{(\sigma)} - (\kappa-2)J_3^{(-\sigma)}, \\ \Sigma_2^{(a,\sigma)} &= 2J_2^{(\sigma)} - (\kappa-2)J_2^{(-\sigma)} \\ &\quad - (\kappa-3)J_2^{(\sigma)} + 3J_3^{(-\sigma)}, \\ \Sigma_3^{(a,\sigma)} &= -(\kappa-2)J_1^{(-\sigma)} + 3J_2^{(-\sigma)} + 2J_3^{(\sigma)} - (\kappa-3)J_3^{(\sigma)}. \end{aligned} \tag{79}$$

Here, $J_s^{(\sigma)} = T\sum_{n,\omega,\mathbf{p}} t_{\mathbf{p}}^{(s,n)} G_{\omega}^{(n,s)}(\mathbf{p})$ is the sum of products of the matrix elements of the transition matrix $\hat{t}(\mathbf{p})$ by the elements of the virtual Green function $\hat{G}_{\omega}(\mathbf{p})$ for the given spin projection and the given a-state ($a = xy$, yz, zx, $x^2 - y^2$, $3z^2 - r^2$).

Our next aim is to calculate the corrections $\delta\Sigma_k$ proportional to the magnetic field.

Three variations $\delta\Sigma$ satisfy the general equations (32) with the coefficients expressed in terms of the elements of the matrix $\hat{S}$.

In turn, the matrices $\hat{S}$ and $\hat{R}$ are expressed in terms of the coefficients of Eq. (79):

$$\begin{aligned} \hat{S} &= \begin{pmatrix} 5-\kappa & 0 & (\kappa-2)/3 \\ 0 & 2 & -1 \\ \kappa-2 & -2 & (5-\kappa)/3 \end{pmatrix}, \\ \hat{R} &= \begin{pmatrix} (13-2\kappa)/3 \\ 1 \\ (13-2\kappa)/3 \end{pmatrix}. \end{aligned} \tag{80}$$

For d-electrons, we have $\kappa = 3$, 4, or 5. The matrix $\hat{U}$ is expressed in terms of the products $U_{k,n} = R_k b_n^2/2$:

$$\hat{U} = \begin{pmatrix} (13-2\kappa)/6 & (13-2\kappa)/9 & (13-2\kappa)/18 \\ 1/2 & 1/3 & 1/6 \\ (13-2\kappa)/6 & (13-2\kappa)/9 & (13-2\kappa)/18 \end{pmatrix}. \quad (81)$$

We have $\hat{F}^{(0)} = Q(\mu)\hat{W}$, where Q is defined in (33) and the sum of elements in every row of the matrix $\hat{W} = \hat{U} - \hat{S}$ is zero:

$$\hat{W} = \begin{pmatrix} (4\kappa-17)/6 & (13-2\kappa)/9 & (25-8\kappa)/18 \\ 1/2 & -5/3 & 7/6 \\ (5-4\kappa)/6 & (2\kappa+11)/9 & (8\kappa-37)/18 \end{pmatrix}. \quad (82)$$

In the particular case $\kappa = 5$, we must solve the system of equations

$$\begin{pmatrix} 1-(K_0+fg^2D_1) & 4/3-2(K_0+fg^2D_1)/3 & 1-(K_0+fg^2D_1)/3 & -fD_0 & -2fD_0/3 & -fD_0/3 \\ 1-K_0 & -2(1-K_0) & 1-K_0 & -fA & 2fA & -fA \\ 1-K_0 & 2 & 1+K_0 & -fA & 0 & fA \\ -g^2D_2/2 & -g^2D_2/3 & -g^2D_2/6 & Q/2-D_1/2+1 & Q/3-D_1/3 & -5Q/6-D_1/6 \\ -g^2D_2/2 & -g^2D_2/3 & -g^2D_2/6 & Q/2-D_1/2 & -5Q/3-D_1/3+1 & 7Q/6-D_1/6 \\ -g^2D_2/2 & -g^2D_2/3 & -g^2D_2/6 & -5Q/6-D_1/2 & 7Q/3-D_1/3 & Q/6-D_1/6+1 \end{pmatrix} \quad (83)$$

$$\times \begin{pmatrix} \delta f_1 \\ \delta f_2 \\ \delta f_3 \\ \delta\Sigma_1 \\ \delta\Sigma_2 \\ \delta\Sigma_3 \end{pmatrix} = \begin{pmatrix} -b^2 f\sigma\delta H D_0 \\ 0 \\ 0 \\ -\sigma\delta H D_1 \\ -\sigma\delta H D_1 \\ -\sigma\delta H D_1 \end{pmatrix}.$$

The coefficients here are expressed in terms of the integrals of the Fermi distribution, and they depend on the renormalized temperature and chemical potential:

$$\tilde{T} = \frac{T}{fg^2}, \quad \tilde{\mu} = \frac{\mu}{fg^2}, \quad f = \frac{(6-h_d)}{120}, \quad g^2 = 12, \quad K = 3\frac{(h_d-2)}{(6-h_d)}. \quad (84)$$

Equate the determinant of this system to zero to find the condition for ferromagnetic instability

$$K(1-K) = g^2 f D_1\left(\frac{2}{3}+K\right) + K(1-K)D_1 + fg^2\left(\frac{2}{3}+K\right)(D_2D_0 - D_1^2). \quad (85)$$

To this equation, we add the equation of state in the form

$$fg^2 = \frac{6}{15+5K}, \quad h_d = 6\frac{1+K}{3+K}, \quad 2 < h_d < 3.$$

If $T = 0$, then $D_2D_0 = D_1^2$, and all quantities appearing in the equations in the model of the semielliptic band can be represented in the form of general relations (68), which

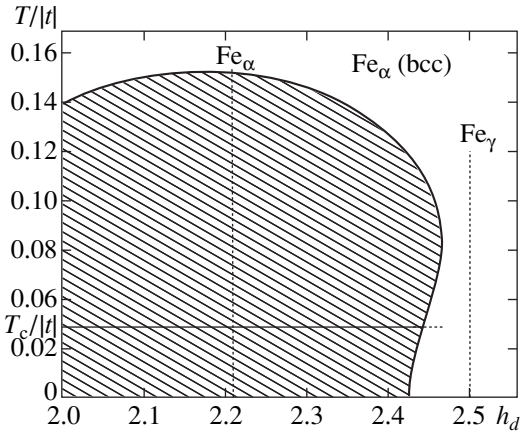


Fig. 4. The phase diagram of iron. The ferromagnetic region is shaded; $T_c/|t| = (1044 \times 2)/(4.82 \times 11606) = 0.0373$.

depend on the single parameter α; the total number of holes h_d can be expressed in terms of this parameter.

As a result, we obtain the phase diagram depicted in Fig. 4.

The numerical solution of Eq. (84) yields $\alpha_c = 2.6938$ and $K_c = 0.35826$. Therefore, the ferromagnetic instability is observed in a limited region of concentrations of $2 < h_d < h_c = 2.4298$. Outside this interval, the state is paramagnetic [6].

In the zero-loop approximation, one can also obtain a finite interval of concentrations in which there occurs ferromagnetic instability: $2 < h_d < h_{c0} = 2.339$.

Thus, ferromagnetism exists in the entire region from $h_d = 2$ to 2.428. However, beginning with $h_d = 2.26$ and up to $h_d = 2.428$, the instability due to one-loop corrections disappears and there remains only the instability caused mainly by zero-loop corrections.

This result concerns the ferromagnetism of volume-centered α-iron, which has a magnetic saturation moment equal to $2.2\mu_B$ (see, e.g., [7]).

As for the nonferromagnetic phase of γ-iron, measurements show that the number of electrons on the *d*-shell is about 7.5, so that the number of holes is 2.5.

Therefore, the existence of ferromagnetism in α-iron is explained by the fact that the number of hole states (≈ 2.2) is within the interval of the ferromagnetic instability. The absence of ferromagnetism in γ-iron is explained by the fact that the number of holes on the $3d$-shell, which is approximately equal to 2.5, exceeds the critical value obtained both in the zero-loop and one-loop approximation.

The qualitative conclusion about the possibility of ferromagnetic ordering only for a limited interval of concentrations is in agreement with both the zero-loop and one-loop approximations.

To find the magnetic permeability, it is sufficient to calculate the correction to the magnitude of the magnetic moment δm, which is a linear combination of the variations of end factors (in accordance with (38)):

$$\delta m = 2\left\{\frac{3}{2}\delta n_{\mathrm{III}}^{(3/2)} + \frac{1}{2}\delta n_{\mathrm{III}}^{(1/2)} + \delta n_{\mathrm{II}}^{(1)}\right\} = \delta f_3 + 2\delta f_2 + 3\delta f_1.$$

This linear combination is calculated using system (83), which yields the susceptibility

$$\chi = \frac{\mu_B^2}{v_0} 2\kappa \frac{\delta m}{\delta H} = 2\kappa\mu_B^2 \frac{-2fD_0(3K+2)}{v_0 D}, \qquad (86)$$

where v_0 is the volume of the unit cell, μ_B is the Bohr magneton, and

$$D = K(1-K) - g^2 f D_1\left(\frac{2}{3} + K\right) - K(1-K)D_1 - fg^2\left(\frac{2}{3} + K\right)(D_2D_0 - D_1^2). \qquad (87)$$

Thus, we determine the inverse magnetic susceptibility for the paramagnetic part of the phase diagram (Fig. 4).

In the case under consideration, the resonance coefficient A disappears from the final formula (86), and the numerator of the magnetic susceptibility is positive for all values of the temperature and chemical potential.

6. THE DOMAIN $3 < h_d < 4$

This domain corresponds to the cubic phase of the paramagnetic manganese, which goes to the antiferromagnetic phase with decreasing temperature. Nevertheless, this interval is of interest, because compounds containing manganese cations manifest high-spin ferromagnetism.

To analyze transitions between high-spin three- and four-hole states, we use the general relations (6) and (21):

$$b_1^2 = 1, \quad b_2^2 = \frac{3}{4}, \quad b_3^2 = \frac{1}{2}, \quad b_4^2 = \frac{1}{4}, \qquad b^2 = \sum_{k=1}^{4} b_k^2 = \frac{5}{2}, \quad g^2 = b^2 C_{\kappa-1}^3 = 10. \qquad (88)$$

This corresponds to the choice of the maximum possible number of degenerate *d*-orbitals ($\kappa = 5$).

The further consideration corresponds to the general relations given in Section 2. The formula for susceptibility is

$$\chi = 2\kappa \times 10 \frac{-fD_0(1+K) + AD_1}{v_0 D}, \qquad (89)$$

where the denominator is expressed in terms of the coefficients calculated for the paramagnetic phase:

$$D = K(1-K) + g^2 D_2 A - fg^2(K+1)D_1 - \frac{1}{2}(2Q+3)K(1-K)D_1 - \frac{1}{2}g^2 f(1+K)(2Q+3)(D_2D_0 - D_1^2). \qquad (90)$$

In the limit $T = 0$, when $D_1^2 - D_2D_0 = 0$, we find the con-

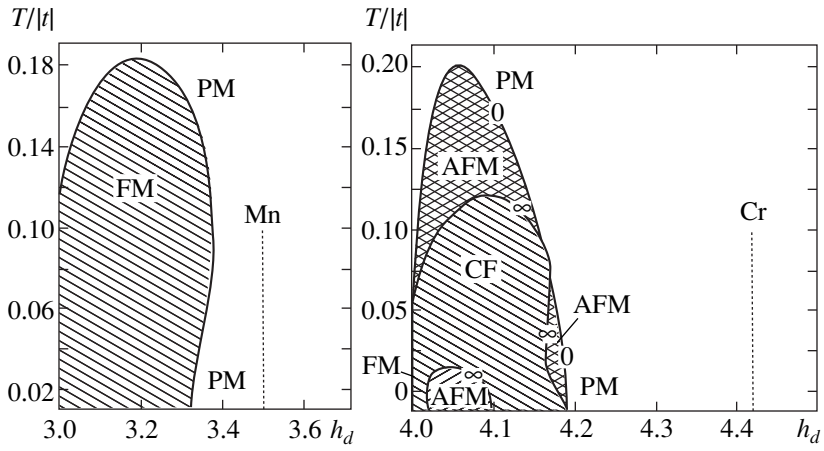


Fig. 5. The phase diagram of high-spin manganese and chromium. FM denotes the ferromagnetic regions, AFM denotes the antiferromagnetic ones, CF denotes the mixed phase, and PM denotes the paramagnetic regions.

dition for ferromagnetism

$$2K(1-K) + 2g^2 D_2 A = [2fg^2(K+1) + (3+2Q)K(1-K)]D_1. \quad (91)$$

These equations are complemented by the equation of state

$$\frac{1}{f} = 5(8-3n_d), \quad g^2 = 10, \quad n_d = 4\frac{6-K_0}{8-3K_0}. \quad (92)$$

The numerical solution of Eq. (91) yields a rather wide interval of the ferromagnetic instability of $3 < n_d < 3.318$. In the zero-loop approximation, we find a smaller interval of the ferromagnetic instability of $3 < n_d < 3.23497$.

The calculation of the coefficients at a finite temperature allows us to determine the Curie temperature in the entire interval of concentrations where ferromagnetism is possible (see Fig. 5).

7. THE DOMAIN $4 < h_d < 5$

The consideration follows the same scheme as that for manganese.

First, we represent the variations of the average occupation numbers $\delta n_V^{(5\sigma/2)}$, $\delta n_V^{(3\sigma/2)}$, and $\delta n_V^{(\sigma/2)}$ in terms of the variations of the five end factors:

$$\delta n_V^{(5\sigma/2)} = \sum_{k=1}^{5} \delta f_k^\sigma, \quad \delta n_V^{(3\sigma/2)} = \sum_{k=2}^{4} \delta f_k^\sigma, \quad \delta n_V^{(\sigma/2)} = \delta f_3^\sigma. \quad (93)$$

These expressions are substituted into the equations for the variations of the end factors δf_k^σ.

Furthermore, using (6) and (21), we calculate the fractional parentage coefficients

$$b_1^2 = 1, \quad b_2^2 = \frac{4}{5}, \quad b_3^2 = \frac{3}{5}, \quad b_4^2 = \frac{2}{5}, \quad b_5^2 = \frac{1}{5}, \quad b^2 = \sum_{k=1}^{5} b_k^2 = 3, \quad g^2 = b^2, \quad (94)$$

which corresponds to the choice of the maximum possible number of degenerate *d*-orbitals ($\kappa = 5$).

The correction to the magnetic moment is written, using (38), as a linear combination of the corrections to the end factors:

$$\delta M = 2\left(\frac{5}{2}\delta n_V^{(5/2)} + \frac{3}{2}\delta n_V^{(3/2)} + \frac{1}{2}\delta n_V^{(1/2)} + 2\delta n_{IV}^{(2)} + \delta n_{IV}^{(1)}\right) = 5\delta f_1 + 4\delta f_2 + 3\delta f_3 + 2\delta f_4 + \delta f_5. \quad (95)$$

Calculating this sum, we obtain the susceptibility

$$\chi = 2\kappa\frac{\mu_B^2}{v_0} \times \frac{[-5fD_0(3K+4)(15-26Q) + 80AD_1(7+2K)]}{D}, \quad (96)$$

where the denominator is written as

$$D = (15-26Q)K(1-K) + \frac{112}{3}g^2 D_2 A - (31+6Q)K(1-K)D_1 - \frac{1}{3}g^2 f(3K+4)(31+6Q)(D_2D_0 - D_1^2) + \frac{1}{3}fg^2(3K+4)(15-26Q)D_1. \quad (97)$$

To construct the phase diagram as a function of the hole concentration, these equations must be supplemented with the equation of state

$$f = \frac{1}{25-19K}, \quad n = 10\frac{10-7K}{25-19K}, \qquad K = 25\frac{(n-4)}{(19n-70)}. \tag{98}$$

In the limit $T = 0$, when $D_1^2 - D_2D_0 = 0$, the condition for ferromagnetism is

$$3(15-26Q)K(1-K) + 112g^2D_2A = [fg^2(3K+4)(15-26Q) + 3(31+6Q)K(1-K)]D_1. \tag{99}$$

The numerical solution of Eq. (99) yields a very narrow domain of ferromagnetic instability of $4 < n_d < 4.0008829$. In the zero-loop approximation, a wider domain of ferromagnetic instability of $4 < n_d < 4.23$ is obtained.

The considerable difference between the phase diagrams of chromium and manganese is caused by the function A. As has already been mentioned, this function has a logarithmic singularity at $T = 0$ and the zero value of the chemical potential (i.e., at $\alpha = \pi$). This singularity is compensated for by the factors D_1 and D_2, which vanish at this point. Outside this point, the product AD_2 is positive, but not very large; however, in the interval $4 < h_d < 5$, it is multiplied by a large factor of 112.

Another specific feature of the system under study is the possibility of sign reversal of the magnetic susceptibility due to the vanishing denominator (N) of the magnetic susceptibility.

At $T = 0$, there exists a rather wide domain of $4.00828829 < n_d < 4.1935$, in which the denominator is negative.

Therefore, at $T = 0$, in addition to the regions of ferromagnetic (100) and antiferromagnetic instability

$$4 < n_d < 4.0008829, \quad 4.01782 < n_d < 4.098442, \quad D < 0, \quad N > 0, \tag{100}$$

there exist regions with positive susceptibility (see Fig. 5) characterized by a negative numerator and denominator ($D < 0$ and $N < 0$):

$$4.008829 < n_d < 4.01782, \quad 4.098442 < n_d < 4.1935. \tag{101}$$

At finite temperatures, the latter domain is adjacent to the antiferromagnetic domain (see Fig. 5) and the susceptibility becomes infinite at the boundary. Hence, we intuitively conclude that there is an ordered mixed phase inside this domain.

It follows from Fig. 5 that as the temperature decreases, the system goes to the antiferromagnetic state with a negative magnetic susceptibility rather than to the ferromagnetic state.

As the temperature decreases further, the susceptibility increases in absolute value, becomes infinite, and than goes to a low-temperature mixed phase.

The cause of this phenomenon is a strong interaction of antiferromagnetic type, which manifests itself in the form of the resonance function A with a logarithmic singularity at $T = 0$. The intensity of this interaction depends on the magnitude of the total spin in the terminal and the initial state. If the initial spin is zero, as is the case for Ni, Pd, and Pt, there is no such interaction. In the case of cobalt, it appears with a negative sign and, therefore, manifests itself for large-energy excitations in the nonferromagnetic domain (see Fig. 3).

In the case of high-spin iron and manganese, the effect of resonance is insignificant. However, for chromium, the resonance influence domain overlaps with the domain of ferromagnetic instability, which actually leads to the elimination of significant domains of ferromagnetic ordering.

8. CONCLUSIONS

The theory developed in this paper gives a qualitative explanation of the magnetic properties of Ni, Pd, and Pt belonging to subgroup 8. Therefore, the number of hole states for nickel lies between h_d for palladium and platinum. One can naturally assume that the concentration of holes in Pd and Pt is outside the domain of ferromagnetic instability, while the intermediate hole concentration in Ni is inside this domain (see Fig. 2).

The examination of the phase diagram of cobalt, iron, and manganese reveals a qualitatively different situation. In this case, the system resonates between the high-spin (magnetic) states, so that the system is ferromagnetic when the number of excitations with negative amplitude of exchange scattering is small. As the Fermi energy increases, the exchange scattering amplitude changes sign so that the system ceases to be ferromagnetic for every integer-valued interval of the hole concentration beginning with a certain critical value.

For cobalt, the observed value of the magnetic saturation moment ($1.7\mu_B$) is somewhat greater than the critical value ($1.59\mu_B$) obtained in this paper.

On the contrary, the critical value obtained for bcc iron ($2.46\mu_B$) exceeds the magnetic saturation moment

($2.2\mu_B$). Therefore, the observed number of hole states belongs to the domain of ferromagnetic ordering.

For the nonferromagnetic phase of γ-iron and manganese, the observed (antiferromagnetic) values of the moment exceed the critical values obtained in this paper.

Therefore, the presented model yields a quantitative description of ferromagnetism in nickel and α-iron.

In addition, it provides a qualitative explanation of the absence of ferromagnetism in palladium, platinum, γ-iron, and in hypothetical (high-spin) phases of chromium and manganese.

As for cobalt, a further elaboration of the model is required aimed at its application to the hexagonal and fcc phases with account for the particular single-particle density of states.

REFERENCES

1. J. Hubbard and K. R. Jain, J. Phys. C **1**, 1650 (1968).
2. R. O. Zaĭtsev, Zh. Éksp. Teor. Fiz. **70**, 1100 (1976) [Sov. Phys. JETP **43**, 574 (1976)].
3. J. Hubbard, Proc. R. Soc. London, Ser. A **281**, 401 (1964).
4. D. T. Smirnov and Yu. F. Smirnov, *Theory of Optical Spectra of Transition-Metal Ions* (Nauka, Moscow, 1997).
5. R. O. Zaĭtsev, Zh. Éksp. Teor. Fiz. **112**, 2223 (1997) [JETP **85**, 1218 (1997)].
6. R. O. Zaĭtsev, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 275 (1998) [JETP Lett. **68**, 295 (1998)].
7. J. B. Goodenough, *Magnetism and the Chemical Bond* (Krieger, New York, 1976; Metallurgiya, Moscow, 1988).

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 301–309.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 341–350.

SOLIDS
Electronic Properties

Extremum of the Percolation Cluster Surface

P. S. Grinchuk* and O. S. Rabinovich

Lykov Institute of Mass and Heat Exchange, Belarussian Academy of Sciences, Minsk, 220072 Belarus
**e-mail: gps@hmti.ac.by*

Received April 22, 2002

Abstract—The internal and external surfaces of a percolation cluster, as well as the total surface of the entire percolation system, are investigated numerically and analytically. Numerical simulation is carried out using the Monte Carlo method for problems of percolation over lattice sites and bonds on square and simple cubic lattices. Analytic expressions derived by using the probabilistic approach describe the behavior of such surfaces to a high degree of accuracy. It is shown that both the external and total surface areas of a percolation cluster, as well as the total area of the surface of the entire percolation system, have a peak for a certain (different in the general case) fraction of occupied sites (in the site problem) or bonds (in the bond problem). Two examples of technological processes (current generation in a fuel cell and self-propagating high-temperature synthesis in heterogeneous condensed systems) in which the surface of a percolation cluster plays a significant role are discussed. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The interface between components participating in physical and chemical processes plays an important role in many cases. The areas of this surface are important for heat and mass transfer processes, for chemical reaction, etc. Components (phases) participating in such processes may have either deterministic or random structure. We will treat the latter case here. Percolation theory [1] is one of basic models providing a correct description of geometrical properties of systems with disordered structures. The main feature of such systems is the possibility of a geometrical phase transition leading to the formation of an infinitely large (percolation) cluster (PC).

We will consider two examples of systems in which the surface of the formed PC plays a significant role. The first system of this kind is a fuel cell. The main part of the fuel cell is a porous electrode. One of the basic problems in the theory of porous electrodes is to determine conditions for gathering all participants of an electrochemical process (viz., reagents, ions, electric current (electrons), and a catalyst) on the well-developed internal surface of the electrode and for removing useful products or byproducts of the electrochemical reaction. It is important that the electrochemical process extends over macroscopic distances (as compared to the microscopic characteristic size of the porous medium).

However, there exists a limitation associated with a small velocity of fuel (gas) transport to the surface of the electrode via an electrolyte. This limitation is due to the diffusion mechanism of transport of gas molecules in a liquid. At the same time, it is well known that the main part of the current is generated in an individual pore in the meniscus region, where the film of the electrolyte is thin enough for gas diffusion, while the electric resistance is still not high. Consequently, in order to obtain a stronger current, the number of such meniscuses (gas (hydrogen)–electrolyte–electrode "three-phase" boundaries) per unit volume must be as large as possible. However, we must take into account the following circumstance: the electric circuit must be closed for the ionic component of the current. This means that a connected system of pores filled with electrolyte (percolation cluster) must exist in the electrolyte. It is this cluster that plays the key role in the generation of current. Drops of electrolyte insulated from all sides (finite clusters) which may exist in the electrode make zero contribution to the overall generation of current by the fuel cell since the circuit is disconnected for the ionic component of the current. All "favorable" electrochemical reactions in this case occur most intensely near the above-mentioned three-phase boundary. This means that the electrochemical activity (generated current) is proportional to the area of a given interface. The useful interface between the three phases is, in turn, proportional to the outer surface of the PC formed by pores filled with electrolyte since it is only this surface that participates in the continuous process of current generation. If, for example, gas "bubbles" are formed in the PC, the surface of such bubbles, which is the internal surface of the PC in this case, drops out of the process of current generation after a certain time since the gas required for the electrochemical reaction is no longer supplied to these regions. Thus, the following optimization problem appears: we must find the degree of saturation of the porous electrode with electrolyte, for which the area of the external surface of the percolation cluster being formed has the maximal value. It should be noted that the problem of distribution of a liquid in a

porous body could be reduced to the problem of percolation over bonds [2, 3].

By way of another example, we consider a self-propagating high-temperature synthesis (SHS) in a heterogeneous condensed system (mixture of powders), which is one of the most effective and practicable methods for obtaining a wide range of inorganic compounds including intermetallides [4]. A number of phenomena and effects which have been observed while studying combustion processes in heterogeneous systems can be explained only taking into account the random structure of such systems [5–8]. We will briefly consider the effect of geometrical contingency on the interface and the processes in such systems connected with it.

In combustion reactions (except exothermal decomposition of some compounds), the usual participants are a fuel and an oxidizer, which must be mixed at the molecular level for the reaction to occur. If the fuel and the oxidizer are in the gaseous state, they can be mixed prior to the reaction in many cases. If, however, the components are in the condensed state, such a mixing is practically unattainable. In this case, the interface between the components becomes significant for the reaction. In mixtures consisting of particles that are infusible during combustion (e.g., Ta–C), the interaction between these particles occurs through solid-phase diffusion. Consequently, a mechanical contact of a powder particle with at least one particle of the other powder is required for the reaction to occur in such systems. The necessary condition for the propagation of the reaction over macroscopic distances as compared to the size of an individual particle is the presence of two PCs formed by particles of the first and second species. In this case, the maximum conversion of components can be reached when the area of the contact surface between these clusters has the maximum possible value. It should be noted that, by definition, this surface coincides with the external surface of a percolation cluster. If one of the components melts during the combustion of the components in the mixture (e.g., Ti–C or Zr–C system), the mutual arrangement of particles prior to the reaction is the major factor determining its occurrence since a fused particle may spread only in a small neighborhood of its initial position. If the fused component is sufficient for filling the entire or almost entire free space of a pore, the total surface of the percolated system, consisting of the PC surface and the surfaces of all finite clusters formed by particles of the infused component, will participate in the reaction. If, however, the amount of fused component in the system is small, the PC surface is the only factor determining the reaction.

For a monodisperse mixture of powders, the geometrical contingency can be modeled by the classical problem of percolation over lattice sites. In this case, the mixture of powders *A* and *B* is put into correspondence with a space lattice in which each site is occupied either by particles of species *A* with probability *p*, or by particles of species *B* with probability $(1 - p)$. Here, *p* is the numerical fraction of *A* particles in the mixture, which may be connected with the volume fraction of these particles through the Scher–Zallen construction [9]. Then the problem of SHS optimization in heterogeneous condensed systems can be formulated as follows: find the relation between the concentrations of initial components of the mixture for which the corresponding interface area and, hence, the degree of conversion have the highest values. The choice of the surface whose maximum area has to be determined in each specific case was discussed above.

It is important to note that, if we are speaking about the external surface of a PC, its maximum for one PC can be attained only in the region in which a PC formed by particles of the other species can exist. This condition is satisfied if $p_c \le p_m \le 1 - p_c$, where p_m is the point at which the surface area has the maximum value and p_c is the percolation threshold. This inequality leads to the condition imposed on the percolation threshold in the simulating lattice: $p_c \le 0.5$. In the 2D case, this condition is satisfied only for a triangular lattice (in the site problem) and for a square lattice (in the bond problem) and only at a single point, viz., percolation threshold ($p_c = 0.5$ [10]). However, in the 3D case, this condition is satisfied for all regular lattices.

Some other circumstances indirectly confirming the important role of interface in SHS are worth noting. It was pointed out in a number of experimental publications [11–13] that many heterogeneous condensed systems do not exhibit complete conversion of components in SHS processes. One possible reason is the statistical inhomogeneity of such mixtures. For example, not every particle has a particle of the other species as a neighbor, even for well-mixed initial components taken in stoichiometric proportion. It is impossible to obtain a system in which particles of different species would be arranged, for example, in staggered order. Particles of a powder in a mixture with other particles always have a certain tendency toward clusterization. For this reason, some of such particles have no particles of the other species among their neighbors; i.e., the interface (contact surface) area of the reagents is much smaller than the theoretically possible surface area of periodic (staggered) packing. An excess of one of the components in the reaction zone strongly decelerates the reaction upon an increase in the conversion depth, ultimately leading to incomplete conversion of the initial components in an SHS process [14]. Another experimental fact is also worth noting. It is well known that SHS can be realized only in certain concentration limits [11, 15]. If the concentration of one of the components is lower than the threshold value, SHS cannot be initiated at all. On the one hand, this is due to a decrease in the degree of isothermality of the reaction as a result of a decrease in the content of one of the powders. It was mentioned in [11], however, that another possible reason for this effect is a decrease in the contact area

between the reagents, which becomes critical for a low concentration of one of them. Moreover, the absence of a PC made of particles of a component in deficit (which may be required for the propagation of the reaction over the entire sample) can also be responsible for the effect in question.

Thus, this study aims at analysis of areas of interfaces in two-component heterogeneous systems on the basis of percolation models.

2. ANALYTICAL TREATMENT OF THE SURFACE OF A PERCOLATION CLUSTER

2.1. Site Problem

Let us consider a space lattice and assume that a certain liquid can flow from one site to another along tubes, which will be referred to as bonds. By introducing stochastic elements into this system in different ways, we obtain the problem of percolation either via lattice sites or via bonds. For example, in the site problem for a liquid percolating over the lattice, all bonds are assumed to be penetrable for the liquid, while sites are either overlapped or nonoverlapped (with a certain probability). In the bond problem, all sites are penetrable for the liquid, while part of the bonds are damaged. It should be noted above all that we assume in these problems that elements with different properties (occupied or empty sites and penetrable or impenetrable bonds) are distributed over the lattice absolutely at random and, hence, the probability that an arbitrarily chosen element possesses a given property is equal to the fraction of such elements in the lattice and is independent of the properties of neighboring elements. The assumption concerning the randomness of the distribution implies that each site (bond) in a lattice is occupied with the same probability p and is empty with probability $(1 - p)$, where p is the fraction of occupied sites (bonds) in the lattice.

As a measure of the surface of any cluster in a percolation system, one of two quantities (perimeter or "energy") is traditionally used [16]. For example, in the bond problem, the perimeter of a cluster is the number of empty lattice sites, which do not belong to the cluster, but are separated from the lattice sites belonging to it by the length of a bond. The "energy" is defined as the number of neighboring sites (pairs), one of which is occupied (belongs to the cluster), while the other is empty.

Let us prove that the PC surface indeed displays extremal behavior. We consider a regular d-dimensional lattice with coordination number z, in which the fraction of occupied sites is equal to p. We assume that a PC is formed on the lattice and the fraction of lattice sites belonging to it is $P_\infty^{(s)}(p)$. This quantity simultaneously indicates the probability that an arbitrary site belongs to the PC.

We consider a set of conditions under which an arbitrary lattice site can belong to the total perimeter of the PC. First, the site in question must be empty (with probability $(1 - p)$); second, at least one of its z neighbors must belong to the PC. We state that event Ω_j has occurred at an arbitrary site if any j from z sites neighboring the given site belong to the PC (with probability $[P_\infty^{(s)}(p)]^j z!/(j!(z-j)!)$, where $z!/(j!(z-j)!)$ is the number of combinations of z sites taken j at a time), while the remaining $(z - j)$ sites are either empty or belong to arbitrary finite clusters (with probability $[1 - P_\infty^{(s)}(p)]^{z-j}$). It should be noted that we could multiply the probabilities of elementary events corresponding to individual sites owing to the above assumption concerning absolute randomness of their distribution in the lattice. Then the probability of occurrence of event Ω_j is determined by the formula

$$P(\Omega_j) = \frac{z!}{j!(z-j)!}[P_\infty^{(s)}(p)]^j[1 - P_\infty^{(s)}(p)]^{z-j}. \quad (1)$$

A site belongs to the PC perimeter (event D) if it is empty and if any of events Ω_j (for j varying from 1 to z) has occurred at this site. Since events Ω_j are pairwise incompatible, the conditional probability is $P(D/\Omega_j) = 1$. Using the formula for total probability, we can write, taking into account the relations derived above, the following relation for the PC perimeter per lattice site:

$$\begin{aligned} S_\infty^{(s)}(p) &= (1-p)\sum_{j=1}^{z} P(\Omega_j)P(D/\Omega_j) \\ &= (1-p)\sum_{j=1}^{z}\frac{z!}{j!(z-j)!}[P_\infty^{(s)}(p)]^j[1 - P_\infty^{(s)}(p)]^{z-j} \quad (2) \\ &= (1-p)\{1 - [1 - P_\infty^{(s)}(p)]^z\}. \end{aligned}$$

It is important to note that the relation obtained is valid for any z. In the approach proposed here, this quantity determines only the number of terms that must be summed in Eq. (2). The application of the probabilistic approach is justified by the fact that the number of lattice sites over which averaging is carried out is macroscopically large. It should also be noted that the perimeter S of large clusters near the percolation threshold for $d = 2$ was investigated in [17, 18] using the Monte Carlo method. In accordance with theoretical predictions [1], it was found that $S = k(1 - p_c)/p_c$ for large clusters, where k is the number of sites in a cluster. It follows from formula (2) that, near the percolation

threshold, when $P_\infty^{(s)}(p)$ is small, the PC perimeter is defined as

$$S_\infty^{(s)}(p) \approx z(1-p)P_\infty^{(s)}(p).$$

Considering that $z \approx 1/p_c$, we see that formula (2) is in good agreement with the results obtained by other authors.

A similar expression can be obtained for the PC "energy" $E_\infty^{(s)}(p)$. The expression will differ from that for the perimeter in that the energy at a given site (at which event Ω_j occurs) is equal to j, which the contribution to the perimeter after the occurrence of any event Ω_j is equal to unity. In the site problem, we have the following expression for the PC energy:

$$E_\infty^{(s)}(p) = (1-p)\sum_{j=1}^{z} j\frac{z!}{j!(z-j)!}[P_\infty^{(s)}(p)]^j \times [1-P_\infty^{(s)}(p)]^{z-j} = z(1-p)P_\infty^{(s)}(p). \quad (3)$$

It should be noted that the total energy of a percolation cluster was simulated numerically both for $d = 2$ [19, 20] and for $d = 3$ [21]; it was found that $\langle E\rangle \sim k$ in both cases. It can be seen that formula (3) derived by us is in perfect agreement with this result.

In the limit of an infinitely large system, the PC density tends to zero at $p \longrightarrow p_c + 0$. Consequently, the measure of the PC surface (perimeter or energy) must also vanish at the percolation threshold. For $p = 1$, all lattice sites belong to the PC; consequently the measure of the PS surface must also be zero at this point. Since the PC density is a continuous function on segment $[p_c, 1]$ and a differentiable function on segment $(p_c, 1]$, it follows from Eqs. (2) and (3) that the PC surface measure (perimeter or energy) must also be continuous and differentiable on these segments.

Thus, the conditions of Rolle's theorem are satisfied for the PC perimeter (energy): it is continuous on $[p_c, 1]$, differentiable on $(p_c, 1)$, and $S_\infty^{(s)}(p_c) = S_\infty^{(s)}(1)$. Consequently, there exists a point on the given segment, at which the derivative of function $S_\infty^{(s)}(p)$ vanishes. Since this function is nonnegative, this point is the point of maximum.

The formulas for the perimeter and energy of the entire system can be obtained similarly. For this purpose, while deriving Eqs. (2) and (3), we must replace $P_\infty^{(s)}(p)$ by p; i.e., we require that, after the occurrence of event Ω_j, the corresponding sites may belong to the PC or to arbitrary finite clusters. In this case, we have

$$S_{\text{tot}}^{(s)}(p) = (1-p)[1-(1-p)^z], \quad (4)$$

$$E_{\text{tot}}^{(s)}(p) = z(1-p)p. \quad (5)$$

These equations show that the total perimeter and energy also exhibit extremal behavior. The maximum value of total energy for any lattice is attained at point $p_m = 0.5$. At the same time, the maximum of the total perimeter is attained at point $p_m = 1 - [1/(z+1)]^{1/z}$; i.e., it depends on the type of the lattice (we have $p_m = 0.37$ for $z = 3$, $p_m = 0.33$ for $z = 4$, and $p_m = 0.28$ for $z = 6$). It should be noted that the results of our numerical experiment confirm this conclusion.

The proposed approach can also be used for deriving formulas for the external perimeter and energy of a PC. An arbitrary site belongs to the external PC perimeter if it belongs to a PC formed by empty sites (with probability $P_\infty^{(s)}(1-p)$) and if any of events Ω_j defined above has occurred at this site. Analogously to Eq. (2), we obtain the following expressions for the external perimeter $S_\infty^{(s,\,\text{ext})}(p)$ and external energy $E_\infty^{(s,\,\text{ext})}(p)$ of the PC:

$$S_\infty^{(s,\,\text{ext})}(p) = P_\infty^{(s)}(1-p)\sum_{j=1}^{z}\frac{z!}{j!(z-j)!}P_\infty^{(s)}(p)^j \times [1-P_\infty^{(s)}(p)]^{z-j} = P_\infty^{(s)}(1-p)\{(1-[1-P_\infty^{(s)}(p)])^z\}, \quad (6)$$

$$E_\infty^{(s,\,\text{ext})}(p) = zP_\infty^{(s)}(1-p)P_\infty^{(s)}(p). \quad (7)$$

In the limit $p \longrightarrow 1$, we have $P_\infty^{(s)}(1-p) = 0$. Consequently, it follows from Eqs. (6) and (7) that the PC has only internal surface in this case. By virtue of relation $P_\infty^{(s)}(p) \approx p$, Eqs. (2) and (3) in this limit are transformed into Eqs. (4) and (5), respectively. This means that, for a large fraction of occupied sites, the surface of the entire system virtually coincides with the PC surface. These conclusions are also in perfect agreement with the results of our numerical experiment.

2.2. Bond Problem

In this problem, we analyze only the PC perimeter. It should be noted that an unfilled bond connecting two sites belonging to a cluster is counted twice in the perimeter of this cluster [2].

Let us consider a periodic lattice of bonds with coordination number z, in which the fraction of filled bonds is equal to p. In order to avoid the multiple inclusion of

the same unfilled lattice bond in the PC perimeter, we must consider an arbitrary lattice site (and not a bond) and find the probabilities of relevant elementary events for this site.

A bond adjoining an arbitrary lattice site belongs to the total PC perimeter if it is unfilled and if at least one of the bonds adjoining the same site belongs to the PC. It should be noted that all the bonds adjoining the given site and not belonging to the PC must be vacant by definition in the bond problem. Then we can write the following expression for the average PC perimeter per lattice bond, taking into account the fact that the lattice of N bonds contains $2N/z$ sites:

$$S_\infty^{(b)}(p) = \frac{1}{N}\frac{2N}{z}\sum_{j=1}^{z-1} j\frac{z!}{j!(z-j)!}[P_\infty^{(b)}(p)]^{z-j}(1-p)^j \quad (8)$$
$$= 2(1-p)([1-p+P_\infty^{(b)}(p)]^{z-1} - (1-p)^{z-1}).$$

Here, $P_\infty^{(b)}(p)$ is the connection function, or the fraction of bonds belonging to the PC [2].

The existence of the maximum of the PC perimeter on segment $(p_c, 1)$ can be proved in the same way as in the site problem.

In the bond problem, the total perimeter of the entire percolation system (including both the PC perimeter and the perimeters of all finite clusters) per bond can be determined in the same way as in the site problem by substituting p for $P_\infty^{(b)}(p)$:

$$S_{\text{tot}}^{(b)}(p) = 2(1-p)(1-[1-p]^{z-1}). \quad (9)$$

It should be noted that the expression for the total perimeter of a percolation system obtained in [2] for the bond problem coincides with Eq. (9).

The expression for the external PC perimeter can be derived in the same way as Eq. (6) in the site problem. It should be borne in mind that, in the case when a vacant bond adjoining this site belongs to the PC of vacant bonds, all other vacant bonds adjoining this site belong to the given PC by definition. In this case, for the external PC perimeter, we have

$$S_\infty^{(b,\,\text{ext})}(p) = 2P_\infty^{(b)}(1-p)$$
$$\times([P_\infty^{(b)}(1-p)+P_\infty^{(b)}(p)]^{z-1} - [P_\infty^{(b)}(1-p)]^{z-1}). \quad (10)$$

In the limit $p \longrightarrow 1$, the surfaces behave as in the site problem: expression (10) becomes equal to zero, while Eq. (8) is transformed into Eq. (9).

3. NUMERICAL EXPERIMENT

It is well known that a percolation cluster is a fractal object over lengths smaller than the correlation length ξ [1, 22]. In numerical simulation, this circumstance may play a significant role. In the classical work [23], the PC density $\rho(L)$, i.e., the ratio of the number of PC sites located in a rectangular region with a preset size L to the total number of sites in the given region, was analyzed in the site problem on a square lattice for various values of L. It was proved that $\rho(L)$ strongly depends on the relation between L and ξ. If $L < \xi$, the number of PC sites increases in proportion to L^D (D is the fractal dimension of the PC), while the area (volume) of the region obviously increases in proportion to L^d (d is the topological dimension of space). Since $D < d$, the PC density decreases with increasing L in this region. Over lengths larger than the correlation radius, the PC behaves as a homogeneous object, and its density in this region remains unchanged. It is important to note that it is this value of density that corresponds to an infinitely large system. Finally, if the boundaries of the region in question are at a distance smaller than the correlation length from the lattice edge, the value of $\rho(L)$ starts decreasing upon an increase in L (see [23] for details).

In our numerical experiment, we employed an analogous method for studying the PC surface. It was found that, in the region $L > \xi$, which is most interesting for us, the surface measure (perimeter or energy) of the PC is also independent of L (Fig. 1). It was the position of the given plateau that was taken for the corresponding measure of the surface in the given experiment. A different behavior was observed only for $L < \xi$: the densities of perimeter and energy in this region increased with L. We assume that a superposition of several effects takes place in this region. First, the fractal nature of the PC affects the density of the surface. Second, an empty site, which does not belong to the PC surface in the "old" region, may have a neighbor belonging to the PC in the "new" (increased) region. For small values of L, this effect can make a noticeable contribution to the increment of the PC perimeter and energy densities. It should be noted that this increase in the density should not be identified with any cluster and its fractal dimension since all empty sites within a hypercube with side L do not necessarily belong to the same cluster of vacant sites.

In order to calculate the measures (perimeter and energy) of the PC surface, we used the following algorithm. A lattice of appreciable size (501×501 for 2D problems and $151 \times 151 \times 151$ for 3D problems) with a preset fraction p of occupied sites (bonds) was filled with the help of the standard procedure [24]. For each site (bond), a pseudorandom number ζ distributed uniformly on interval [0, 1] was generated. If $\zeta \le p$, the corresponding site (bond) was assumed to be occupied; otherwise, it was regarded as free. Then a percolation cluster connecting the left and right faces of the lattice

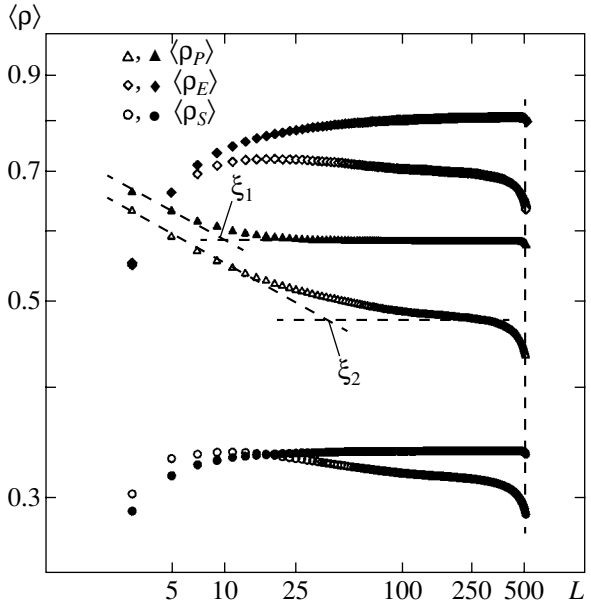


Fig. 1. Average density $\langle\rho_P\rangle$ of a percolation cluster, its average energy density $\langle\rho_E\rangle$, and average perimeter density $\langle\rho_S\rangle$ as functions of size L of the domain of averaging in the site problem on a 501×501 square lattice ($p_c \approx 0.593$, log–log scale); ξ_1 and ξ_2 are correlation lengths for the corresponding fractions of occupied sites: $p - p_c = 0.01$ (△, ◇, ○); $p - p_c = 0.04$ (▲, ◆, ●).

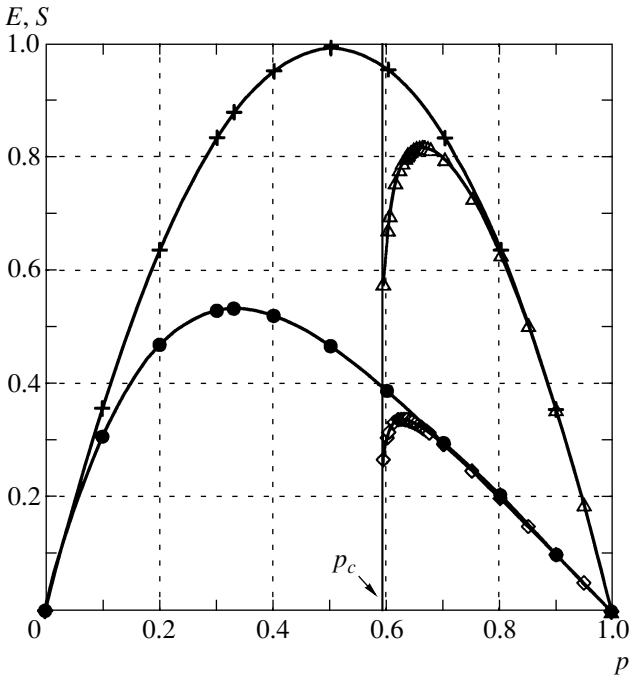


Fig. 2. Perimeter S and energy E of a percolation cluster as functions of the fraction p of occupied sites in the site problem on a 2D 501×501 square lattice; ◇ is the total perimeter of the percolation cluster, ● is the perimeter of the entire system, △ is the total energy of the percolation cluster, and + is the energy of the entire system.

was found and marked. In order to find the external surface of the PC formed by occupied sites, we also found and marked a PC formed by empty sites (if such a PC existed). Then, we counted the number of empty sites (bonds) adjoining the sites (bonds) belonging to the PC in the corresponding regions of the lattice. In order to improve the accuracy of numerical simulation, we took into account only the lattices for which the fraction of occupied sites (bonds) differed by less than 0.0001 of the preset value of p. In each of the four problems investigated by us here, we composed $(5–10) \times 10^3$ histories for each value of p, which ensured accuracy up to three decimal places. The computer time of the central processor using a dual-processor (2×800 MHz) PC Pentium III exceeded 2000 h.

4. DISCUSSION

The site problem was analyzed numerically on square and simple cubic lattices. The results of simulation are presented in Figs. 2 and 3 and in Table 1. In accordance with predictions made in Section 2, the PC surface, as well as the surface of the entire system, attains maximum values for a certain fraction of occupied sites. In this case, the position of the maximum and its absolute value depend to a considerable extent on the measure of the surface (perimeter or energy) used (see Table 1). Since the percolation threshold in the site

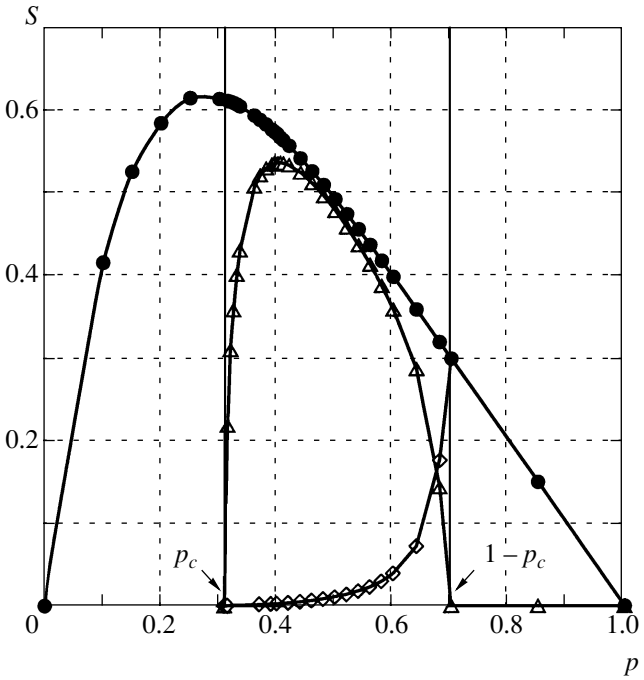


Fig. 3. Perimeter S of a percolation cluster as a function of the fraction p of occupied sites in the site problem on a 3D simple $151 \times 151 \times 151$ cubic lattice; △ is the external perimeter of the percolation cluster, ◇ is the internal perimeter of the percolation cluster, and ● is the perimeter of the entire system.

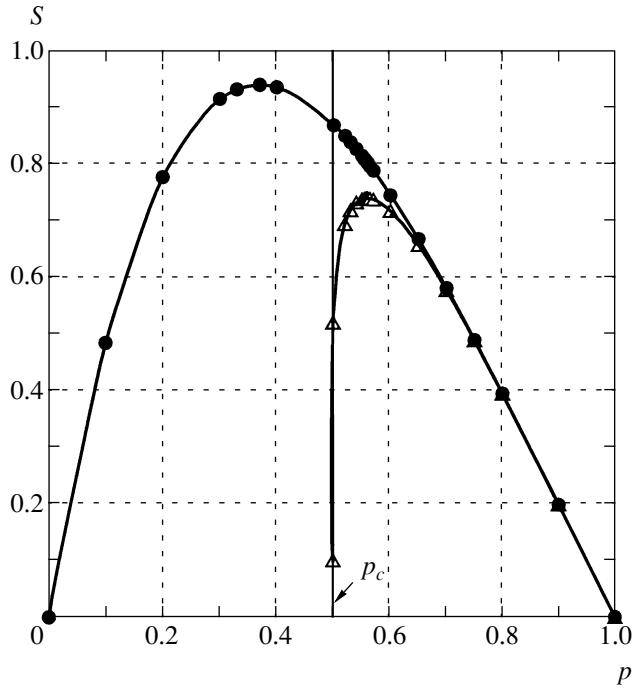


Fig. 4. Perimeter S of a percolation cluster as a functions of the fraction p of occupied bonds in the bond problem on a 2D 501×501 square lattice; △ is the total perimeter of the percolation cluster, ● is the perimeter of the entire system.

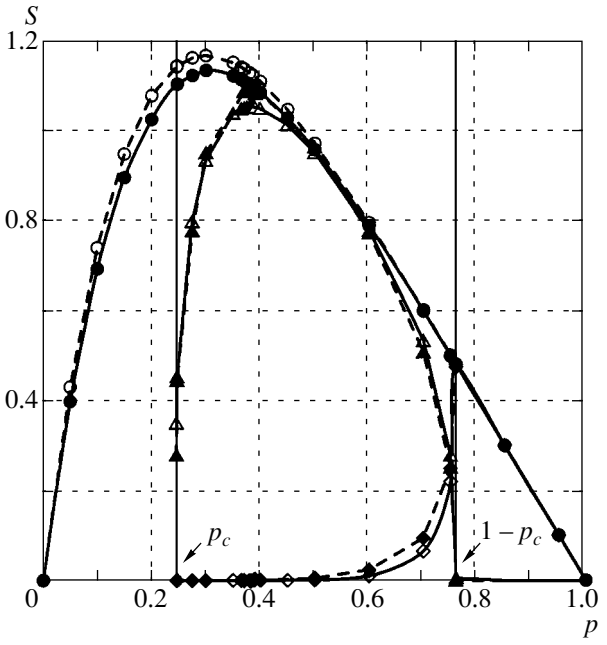


Fig. 5. Perimeter S of a percolation cluster as a function of the fraction p of filled bonds in the bond problem on a simple $151 \times 151 \times 151$ cubic lattice; △ is the external perimeter of the percolation cluster, ◇ is the internal perimeter of the percolation cluster, and ● is the perimeter of the entire system according to the results of numerical experiment. Dashed curves describe the same quantities calculated on the basis of Eqs. (8)–(10) using the data of simulation for the percolation cluster density $P_\infty^{(b)}(p)$.

problem on a square lattice is $p_c = 0.59264721 > 0.5$ [25], the entire PC surface is internal in this case (see Fig. 2). For a simple cubic lattice, $p_c = 0.3116080 < 0.5$ [26, 27]; consequently, the PC in this case has both internal and external surfaces (see Fig. 3). In the region in which the maximum value of the external surface of the PC is attained, the internal surface is negligibly small. It is interesting to note that, in the 3D case, the internal PC surface also has a maximum, which is always attained at point $1 - p_c$. In order to verify the accuracy of the calculations made here, we estimated the fractal dimension of the PC in the 2D problem from the slope of the curve describing the PC density for $L < \xi$. We obtained a value of $D = 1.889 \pm 0.004$, which is in good agreement with the exact result $D = 91/48 = 1.8958 \ldots$ [22].

The bond problem was also analyzed on square and simple cubic lattices. The results of simulation are presented in Figs. 4 and 5 and in Table 2. The behavior of surfaces in this case is qualitatively the same as in the

Table 1. Total ($S_\infty^{(s)}$) and external ($S_\infty^{(s,\,\mathrm{ext})}$) perimeters and total and external energies ($E_\infty^{(s)}$ and $E_\infty^{(s,\,\mathrm{ext})}$) of a percolation cluster in the site problem: points of maximum p_m and the values of these quantities at these points

	Square lattice ($d = 2$)		Simple cubic lattice ($d = 3$)	
	p_m	Value	p_m	Value
$S_\infty^{(s)}$	0.625 ± 0.001	0.3397 ± 0.0001	0.406 ± 0.001	0.538 ± 0.001
$S_\infty^{(s,\,\mathrm{ext})}$	–	–	0.402 ± 0.001	0.535 ± 0.001
$E_\infty^{(s)}$	0.660 ± 0.001	0.819 ± 0.003	0.538 ± 0.001	1.42 ± 0.01
$E_\infty^{(s,\,\mathrm{ext})}$	–	–	0.501 ± 0.001	1.35 ± 0.01

Table 2. Total ($S_\infty^{(b)}$) and external ($S_\infty^{(b,\,\mathrm{ext})}$) perimeters of a percolation cluster in the bond problem: points of maximum p_m and the values of perimeter at these points

	Square lattice ($d = 2$)		Simple cubic lattice ($d = 3$)	
	p_m	Value	p_m	Value
$S_\infty^{(b)}$	0.558 ± 0.001	0.737 ± 0.001	0.382 ± 0.001	1.050 ± 0.002
$S_\infty^{(b,\,\mathrm{ext})}$	–	–	0.382 ± 0.001	1.050 ± 0.002

site problem. A difference is observed only in specific numerical values.

The equations derived by us here were verified for all four problems. For this purpose, the PC density was calculated for sets $\{p_i\}$ and $\{1 - p_i\}$ and then numerical results were substituted into corresponding equations. The discrepancy between the surface measures (perimeters and energies) calculated in this way and the surface measures obtained directly in the numerical experiment did not exceed 10%. The perimeters obtained in the above two ways in the bond problem on a simple cubic lattice are compared in Fig. 5. It can be seen that the maximum discrepancy takes place in the vicinity of the percolation threshold, where size effects are manifested most strongly. This allows us to attribute the above discrepancy to the error of the numerical experiment rather than to approximate nature of the theory. We can expect that this discrepancy will be reduced upon an increase in the size of the system being simulated. For example, at large distances from the percolation threshold, where size effects can be neglected, the discrepancy between the surface measures calculated analytically and numerically is less than 3%.

5. CONCLUSIONS

Thus, the theoretical and numerical analyses of the surfaces of percolation systems revealed an interesting effect, viz., extremal behavior of the external and total surfaces of a percolation cluster. A similar behavior was also observed for the total surface a percolation system. The results of this study lead to the conclusion that the analytic approach proposed for describing various surfaces in percolation systems is correct. The analytic expressions obtained can be used for estimating the position of the extremum of the corresponding surface on lattices, which were not studied in this work. For this purpose, we must know only the density of a percolation cluster as a function of the fraction of occupied sites (bonds) on the corresponding lattice. The latter quantity was studied in detail in [3, 28–30]. We have also considered two examples of systems, viz., a fuel cell and heterogeneous condensed systems (mixtures of powders), in which the discovered effect can be used to optimize important technological processes such as generation of electric current and self-propagating high-temperature synthesis, respectively.

ACKNOWLEDGMENTS

This study was supported by the Belarussian Republic Foundation for Basic Research (project nos. T00-171 and T02P-031).

REFERENCES

1. D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor and Francis, London, 1995).
2. L. I. Kheĭfets and A. V. Neĭmark, *Multiphase Processes in Porous Media* (Khimiya, Moscow, 1982), p. 98.
3. A. V. Neĭmark, Zh. Tekh. Fiz. **56**, 2235 (1986) [Sov. Phys. Tech. Phys. **31**, 1338 (1986)].
4. V. I. Itin and Yu. S. Naĭborodenko, *High-Temperature Synthesis of Intermetallic Compounds* (Tomsk. Gos. Univ., Tomsk, 1989).
5. H. Tephany, J. Nahmias, and J. A. M. S. Duarte, Physica A (Amsterdam) **242**, 57 (1997).
6. N. Provatas, T. Ala-Nissita, M. Grant, *et al.*, Phys. Rev. E **51**, 4232 (1995).
7. A. S. Mukasyan, A. S. Rogachev, and A. Varma, Chem. Eng. Sci. **54**, 3357 (1999).
8. A. Varma, A. S. Mukasyan, and S. Hwang, Chem. Eng. Sci. **56**, 1459 (2001).
9. H. Scher and R. Zallen, J. Chem. Phys. **53**, 3759 (1970).
10. B. I. Shklovskiĭ and A. L. Éfros, Usp. Fiz. Nauk **117**, 401 (1975) [Sov. Phys. Usp. **18**, 845 (1975)].
11. T. F. Grigor'eva, M. A. Korchagin, A. P. Barinova, and N. Z. Lyakhov, Dokl. Akad. Nauk, Ser. Khim. **369**, 345 (1999).
12. S. W. Jo, G. W. Lee, J. T. Moon, and Y. S. Kim, Acta Mater. **44**, 4317 (1996).
13. T. El-Raghy and M. W. Barsoum, J. Am. Ceram. Soc. **82**, 2849 (1999).
14. V. I. Itin, Yu. S. Naĭborodenko, and A. D. Bratchikov, Fiz. Goreniya Vzryva **37** (4), 69 (2001).

15. G. A. Nersisyan, S. L. Kharatyan, L. S. Abovyan, and G. L. Khachatryan, Fiz. Goreniya Vzryva **37** (4), 63 (2001).
16. *Monte Carlo Methods in Statistical Physics*, Ed. by K. Binder (Springer, Berlin, 1979; Mir, Moscow, 1982).
17. P. L. Leath, Phys. Rev. B **14**, 5046 (1976).
18. P. L. Leath and G. R. Reich, J. Phys. C **11**, 4017 (1978).
19. C. Domb and E. Stoll, J. Phys. A **10**, 1141 (1976).
20. C. Domb, T. Schneider, and E. Stoll, J. Phys. A **8**, L90 (1975).
21. A. Sur *et al.*, J. Stat. Phys. **15**, 345 (1977).
22. J. Feder, *Fractals* (Plenum, New York, 1988; Mir, Moscow, 1991).
23. A. Kapitulnik, A. Aharony, G. Deutcher, and D. Stauffer, J. Phys. A **16**, L269 (1983).
24. A. L. Éfros, *Physics and Geometry of Disorder* (Nauka, Moscow, 1982), p. 54.
25. M. E. J. Newman and R. M. Ziff, Phys. Rev. Lett. **85**, 4104 (2000).
26. C. D. Lorenz and R. M. Ziff, J. Phys. A **31**, 8147 (1998).
27. H. G. Ballesteros *et al.*, J. Phys. A **32**, 1 (1999).
28. A. K. Sarychev, A. P. Vinogradoff, and A. M. Karimov, J. Phys. C **18**, L105 (1985).
29. H. L. Fish, J. M. Hammersley, and D. J. Welsh, Phys. Rev. **126**, 949 (1962).
30. D. S. Gaunt and M. F. Sykes, J. Phys. A **16**, 783 (1983).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 310–314.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 351–355.

SOLIDS
Electronic Properties

Coherent Electron–Hole BCS State: Study of Dynamics¶

P. P. Vasil'ev[a,*], H. Kan[b], H. Ohta[b], and T. Hiruma[b]

[a]*Lebedev Institute of Physics, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 119991 Russia*
[b]*Central Research Lab, Hamamatsu Photonics, K.K. 5000, Hirakuchi, Hamakita City, 434, Japan*
**e-mail: peter@mail1.lebedev.ru*

Received July 26, 2002

Abstract—The first experimental study of the evolution of a coherent electron–hole (*e–h*) BCS-like state in bulk GaAs at room temperature is presented. We explicitly demonstrate that the total spontaneous emission from *e–h* pairs located within the conduction and valence bands approaches zero when the radiative recombination of the *e–h* BCS state occurs. This confirms that a vast majority of electrons and holes available are condensed at the very bottoms of the bands and form the BCS state. The average lifetime of this state is measured to be around 300 fs. We also show that the coherence of electrons and holes of the BCS state is preserved for a much longer time compared to the intraband relaxation time T_2. *© 2003 MAIK "Nauka/Interperiodica".*

Electron–hole (*e–h*) systems in semiconductors and their interactions with resonant electromagnetic emission have been among the foremost topics in condensed matter physics for a long time. One of the most notable features of such systems is their ability to form macroscopic quantum states under appropriate conditions. For instance, depending on the *e–h* density, the macroscopic quantum state can be either an excitonic Bose–Einstein condensate or a cooperative *e–h* state that is similar to the BCS state of the Cooper pairs in a superconductor [1]. Semiconductors, including quantum-well structures and microcavity structures, which are highly excited under optical emission, are generally considered as the most promising candidates for observation of these phenomena [2–5]. The number of publications devoted to excitonic Bose condensates, their superfluidity, excitonic insulators, and crossover from Bose condensation to BCS states rapidly increased [6–9].

In our recent papers [10, 11], we presented the first experimental results from an investigation of the spectral characteristics of the radiative recombination of a transient BCS-like *e–h* state in bulk GaAs at room temperature. Instead of optical pumping, we used a strong current injection in a *p–i–n* semiconductor structure for achieving very high *e–h* densities. The *e–h* concentrations attained in our experiments were so high that the average interparticle distance (34–58 Å) was about or even smaller than the de Broglie wavelength of an *e–h* pair in GaAs at room temperature. The latter is about 107 and 60 Å for the electron–light hole and electron–heavy hole pairs, respectively. The collective pairing of electrons and holes and their condensation were responsible for the spectral features of the observed emission, including a large spectral shift in the emission line peak toward longer wavelengths. Fitting of the optical spectra against some theoretical curves allowed us to estimate order parameter Δ of the *e–h* quantum state (the *e–h* pairing gap), which proved to be about 2 meV [10]. We have also demonstrated that the order parameter decreases from 2.1 to 1.2 meV with increasing *e–h* density, whereas the Fermi energy of the quasiparticles lies in the range from 4 to 8 meV [11]. The latter values are much smaller than the Fermi energy of electrons in GaAs (60–170 meV) at room temperature and carrier concentrations of (2–6) $\times 10^{18}$ cm^{-3} achieved experimentally.

It is obvious that in contrast to the normal BCS state of the Cooper pairs, any *e–h* BCS-like state is essentially unstable due to recombination of electrons and holes. The radiative decay of the *e–h* BCS state should exhibit basic features of the cooperative recombination or superradiance of an ensemble of quantum oscillators. Some characteristics of the cooperative emission of electrons and holes were reported previously [12–14]. Because the *e–h* BCS-like state possesses a macroscopic polarization and occupies a sizable portion of the active region of a semiconductor structure, it should explicitly display the coherent interaction of the recombination emission field with *e–h* pairs of the BCS state. By observing this coherent interaction, we can prove that *e–h* pairs located at different parts of the structure (or of the BCS state) have a common phase, i.e., that they are coherent.

In this paper, we present additional experimental results for properties of the transient *e–h* BCS-like state. The main goal is to investigate the temporal behavior and evolution of the cooperative state. The observations are based on (a) direct measurements of time-resolved spectra of the *e–h* BCS state recombination emission utilizing a streak camera and a monochromator and (b) fringe-resolved second harmonic

¶This article was submitted by the authors in English.

generation (SHG) autocorrelation. The former method allows us to find which part of the *e–h* ensemble occupies which energy places in the bands and at which time intervals. The latter technique allows measuring coherent properties of the recombination emission.

We use the same semiconductor structures and the pumping technique as in our previous work; description of them can be found elsewhere [10, 11]. The output emission was monitored in the time domain using a single-shot streak camera with a temporal resolution of about 1.5 ps. Time-resolved optical spectra were also detected by the streak camera. In this case, the emission from the samples was initially collimated on a diffraction grating having 600 lines/mm. The beam deflected at the third diffraction order was then focused on the input slit of the streak camera. This allows us to observe emission features in both time and frequency domains. The technique is discussed in detail in the book [15].

Because a typical duration of superradiant pulses from semiconductor structures lies in the femtosecond range [13, 14], we also use the autocorrelation technique based on SHG for more precise pulse-width measurements. This method has a femtosecond time resolution and permits measurements of both amplitude and phase relationships of the emission under test with femtosecond accuracy.

Figure 1 shows time-resolved optical spectra of a GaAs/AlGaAs *p–i–n* structure in two typical regimes. Laser emission can be generated from the structure when current pulses with an insufficiently large amplitude are applied on the amplifier sections and zero reverse bias is applied on the absorber section of the structure. A typical time-resolved optical spectrum of lasing is presented in Fig. 1a. Here, a photograph taken from the streak camera screen is shown. The lasing starts from the relaxation oscillations (3–4 pulses in front of the trace), which are quite typical of semiconductor lasers. We note that the trace is quite narrow and its center is located at the same place on the frequency axis almost all the time. This means that the central wavelength of the emission varies in time only slightly. Individual modes of the spectrum are not resolved due to a relatively poor spectral resolution of the diffraction grating. The photon energy of the spectral peak is around 1.424 eV. This value was measured separately using a monochromator.

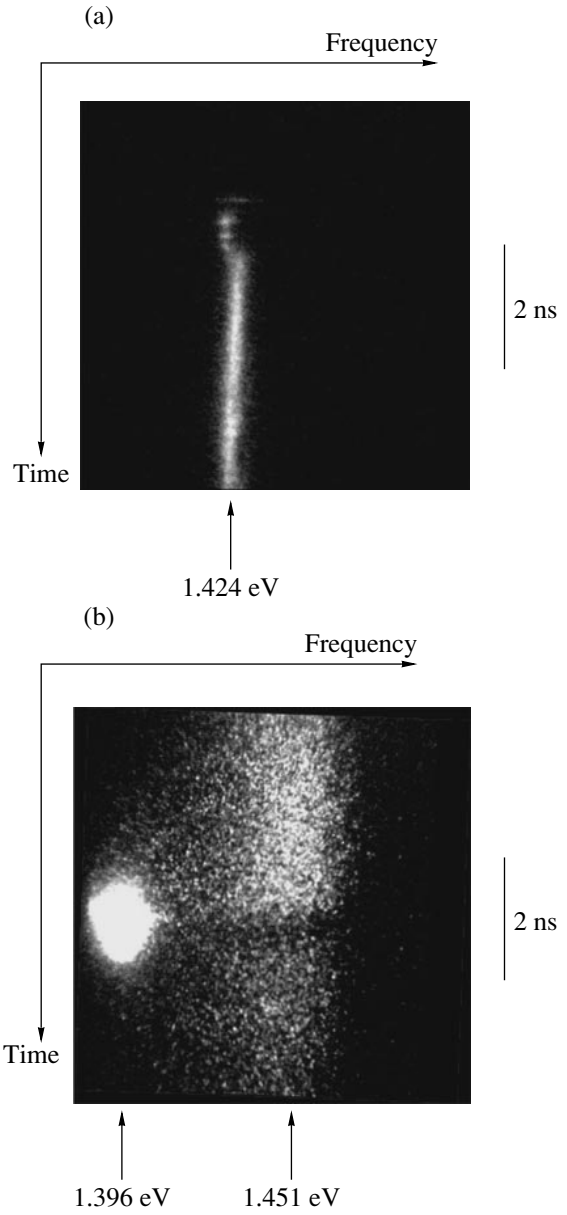


Fig. 1. Time-resolved spectra of (a) lasing and (b) spontaneous emission and superradiant pulse. The width of the spontaneous emission in Fig. 1b is so broad that its high-energy tail is not shown due to the limited input aperture of the streak camera.

A completely different picture is observed when the *e–h* BCS state is formed and femtosecond superradiant pulses are generated as a result of its recombination. Figure 1b corresponds to this regime. The broad vertical stripe represents the ordinary spontaneous emission of nonpaired electrons and holes, whereas the bright spot on the left is the superradiant pulse due to the recombination of the *e–h* BCS state. Because the peak power of the superradiant pulse is typically more than 10^4–10^5 times larger than the spontaneous background, its image on the picture is misshapen due to overexposure. It is clearly seen in Fig. 1b how the femtosecond superradiant pulse develops from the spontaneous emission. As the carrier density increases in time, the spontaneous recombination of electrons and holes occurs at lower and lower frequencies (photon energies) because of shrinkage of the bandgap. When the carrier density is sufficiently high, the de Broglie wavelengths of individual *e–h* pairs start to overlap, the quantum-degeneracy criterion is fulfilled [10], and the *e–h* BCS-like state develops. The phasing of wavefunctions of individual *e–h* pairs occurs via the common electromagnetic emission. The macroscopic polariza-

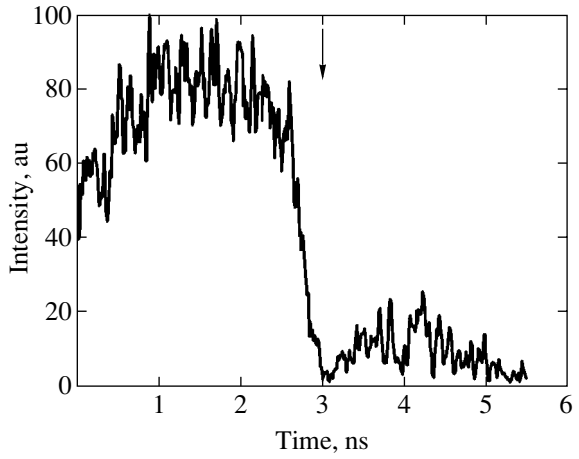


Fig. 2. Intensity of the spontaneous emission against time. The position of the superradiant pulse is shown by the arrow.

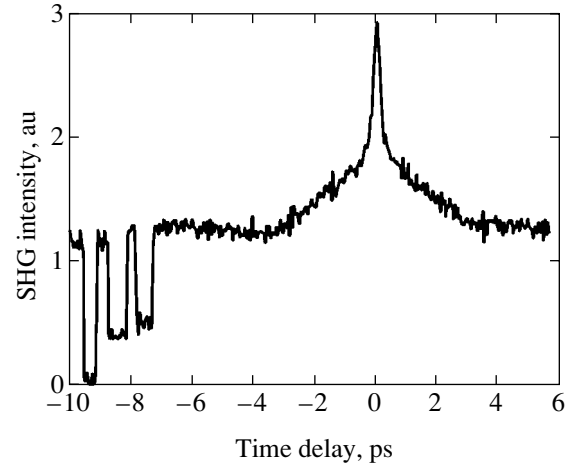


Fig. 3. Intensity autocorrelation of superradiant pulses.

tion builds up at early stages of the leading edge of the superradiant pulse.

Perhaps the most interesting feature of Fig. 1b is the dark stripe across the spontaneous emission at the time when the BCS state recombines. This implies that there are almost no electrons or holes within the bands that can recombine spontaneously. All the electrons and holes are condensed at the very bottoms of the bands and form the BCS state. During its radiative recombination, photons having the minimum possible energy are emitted. This energy is 1.396 eV in our case, while the peak of the spontaneous emission is at 1.451 eV. We recall that the nonrenormalized bandgap in bulk GaAs at room temperature is 1.424 eV.

We calculated the intensity of total spontaneous emission from electrons and holes occupying energy levels inside the bands and plotted it against time, presented in Fig. 2. It is clearly seen how the spontaneous emission intensity almost approaches zero when the *e–h* BCS state is built and recombines. Its temporal position is shown by the arrow.

Such a dynamic behavior is completely different from lasing or spontaneous emission. For example, when lasing starts in a semiconductor laser structure, the spontaneous emission power across the entire frequency range clamps at the threshold level. It is not possible to achieve any dip in the carrier distribution or to quench the spontaneous emission at neighbor frequencies due to ultrafast intraband relaxation processes. On the other hand, it is known [15] that in real semiconductor lasers, *e–h* pairs decohere rapidly with a typical time of 10–100 fs. As a result, we find that for typical operating conditions, the ratio of the number of photons in the sample to the number of *e–h* pairs is $N_{\text{photon}}/N_{e-h} \approx 10^{-5}$–$10^{-4}$ [16], but in the superradiant state, the photon field is linearly coupled to the order parameter and we have $N_{\text{photon}}/N_{e-h} \approx 1$. This implies that when the *e–h* BCS state recombines and photons are emitted from the structure, very few *e–h* pairs should be left in the structure. That is exactly what we experimentally see as the dark horizontal stripe in the middle of Fig. 1b.

Precise measurements of the pulse width of superradiant pulses allow us to estimate the typical lifetime of the *e–h* BCS state. Figure 3 presents an SHG intensity autocorrelation of superradiant pulses. Its full width at half maximum (FWHM) is around 460 fs. This value corresponds to the actual pulse width between 290 and 324 fs depending on the assumed pulse shape (Gaussian, sech, or asymmetric exponential shapes) [15]. It is noteworthy that the achieved pulse width of superradiant pulses is an absolute record among all ultrashort pulses generated by semiconductor lasers, including mode-locked, Q-switched, and gain-switched devices. The SHG trace exhibits a pedestal of 1.5–2.0 ps long, which originates from instabilities of the pulse shape and very large timing jitter. These are likely to be determined by intrinsic quantum-mechanical fluctuations of initial conditions of both the photon field and *e–h* system. Indeed, as noted in [7], the appearance of large noise is strong evidence for the presence of coherence in the *e–h* system. The noise amplitude is known to be inversely proportional to the number of statistically independent entities in an *e–h* system. Thus, large noise observed experimentally implies that only a few entities exist in the macroscopically large regions where the *e–h* BCS state is located.

Unlike the intensity autocorrelations presented in Fig. 3, where all phase information is lost in averaging, fringe-resolved or interferometric autocorrelations can provide some information about phase relationships of the emission under study. This technique enables us to prove experimentally that different spatial regions of the *e–h* BCS state are coherent or have the same or coupled phases. First of all, we point out that all interferometric autocorrelations of ultrashort pulses generated by lasers have a single peak at zero time delay [15]. In the case of mode-locked pulses, there are additional

peaks with fringes separated by the cavity round-trip time.

Figure 4 shows the experimental interferometric autocorrelation of superradiant pulses from a semiconductor structure where two regions with a very high *e–h* density are formed. The separation between the nearest parts of the regions was less than 10 μm. The total length of the structure corresponds to a round-trip time of about 3.1 ps. The coherent beating of the photon field is clearly seen in Fig. 4. The shape of the trace resembles the beating of two oscillators and suggests that the coherency of interaction between the emission and different parts of the BCS state stays on for a few picoseconds while the pulse travels through the medium. Indeed, the very beginning of the recombination pulse induces a small macroscopic polarization while spreading in the medium. This polarization acts as a source to produce an additional photon field that in turn creates more polarization. Dephasing processes prevent establishing the coherency of individual *e–h* pairs and the formation of the macroscopic polarization, but if the number of *e–h* pairs is sufficiently large, the optical gain in the medium can overcome the dephasing [14] and the phase transition of the *e–h* system into a coherent BCS-like state can occur. In this case, the BCS state occupies a sizable portion of the sample and its deexcited parts can then be reexcited by a coherent emission from other regions of the cooperative state. This gives rise to the coherent ringing observed in the output radiation in the form of multiple peaks of the interferometric autocorrelation as presented in Fig. 4.

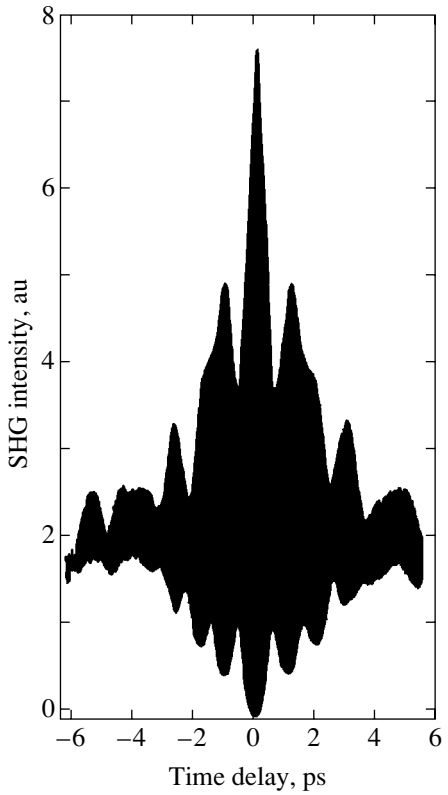


Fig. 4. Interferometric autocorrelation of superradiant pulses illustrating the coherent interaction of the photon field with *e–h* pairs of the BCS state.

We have pointed out earlier [10] that *e–h* interactions within the cooperative state do not affect the coherence of individual electrons and holes, which is quite similar to the Cooper pairs of a superconducting BCS state. This means that relaxation processes must be much slower compared to a system of noncorrelated *e–h* pairs. If we suppose that *e–h* pairs of the BCS state decohere at the same rate as in a normal bulk GaAs (more than 10^{13} s^{-1}), then we would not observe any beating of the photon field on a picosecond scale as shown in Fig. 4. The extremely long phase relaxation time of electrons and holes of the BCS state can be explained as follows.

In contrast to the Bose condensation of excitons, which occurs spontaneously [6–9], the condensation of *e–h* pairs and formation of the BCS state in our case is caused by the resonant electromagnetic emission. The presence of the absorber section in the semiconductor structure results in the absorption of the emission traveling through the structure at all wavelengths except for a narrow region at the very bottom of the bands. In this region of the longest possible wavelengths, we have net gain. The radiative recombination of *e–h* pairs leads to the generation of long-wavelength photons that are amplified and in turn create *e–h* pairs. These bound pairs are bosons and remain coherent with each other and the optical field for some time. Because of very fast intraband relaxation, electrons and holes from upper energy levels in the bands occupy the levels at the bottom that happen to be free almost immediately. At very high carrier concentrations (larger than $(3–4) \times 10^{18}$ cm^{-3}), all energy levels within the band 30–60 meV from the bottom are occupied by electrons. This implies that there is no place within this band for an electron of a bound pair if it becomes free and becomes a fermion again instead of part of a boson. That is why the bound pairs are stable at room temperature. The number of such bound *e–h* pairs increases in time as the optical field travels back and forth between the facets of the crystal. Because the bosons originate from the electrons and holes occupying the lowest possible energy levels, their kinetic energy is very low. This explains the very small value of the Fermi energy of the quasiparticles mentioned above (less than 8 meV). The detailed explanation of this phenomenon is the task of our forthcoming paper.

In conclusion, we present the first direct measurement of the dynamics of a coherent *e–h* BCS state in a semiconductor heterostructure at room temperature. We demonstrate experimentally that almost all electrons and holes are condensed at the very bottoms of the bands when the BCS state is formed. The typical lifetime of the *e–h* BCS state is measured to be about 300 fs. The macroscopic size of the BCS state results in

the coherent interaction of the recombination photon field with *e–h* pairs of the cooperative state on a picosecond time scale, the coherence of individual electrons and holes being unaffected by their collisions.

The authors would like to thank Yu.V. Kopaev for critical comments and fruitful discussions.

REFERENCES

1. *Bose–Einstein Condensation*, Ed. by A. Griffin, D. W. Snoke, and S. Stringari (Cambridge Univ. Press, Cambridge, 1995).
2. L. V. Keldysh and Yu. V. Kopaev, Fiz. Tverd. Tela (Leningrad) **6**, 2791 (1964) [Sov. Phys. Solid State **6**, 2219 (1964)].
3. C. Comte and G. Mahler, Phys. Rev. B **34**, 7164 (1986).
4. T. J. Inagaki, T. Iida, and M. Aihara, Phys. Rev. B **62**, 10852 (2000).
5. P. R. Eastham and P. B. Littlewood, Phys. Rev. B **64**, 235101 (2001).
6. D. Snoke, J. P. Wolfe, and A. Mysyrowicz, Phys. Rev. Lett. **64**, 2543 (1990).
7. L. V. Butov, A. Zrenner, G. Abstreiter, *et al.*, Phys. Rev. Lett. **73**, 304 (1994).
8. L. V. Butov, A. L. Ivanov, A. Imamoglu, *et al.*, Phys. Rev. Lett. **86**, 5608 (2001).
9. A. V. Larionov, V. B. Timofeev, J. Hvam, and K. Soerensen, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 233 (2002) [JETP Lett. **75**, 200 (2002)].
10. P. P. Vasil'ev, H. Kan, H. Ohta, and T. Hiruma, Phys. Rev. B **64**, 195209 (2001).
11. P. P. Vasil'ev, H. Kan, H. Ohta, and T. Hiruma, Quantum Electron. **31**, 870 (2001).
12. P. P. Vasil'ev, Quantum Electron. **24**, 540 (1994).
13. P. P. Vasil'ev, Quantum Electron. **27**, 860 (1997).
14. P. P. Vasil'ev, Quantum Electron. **29**, 842 (1999).
15. P. Vasil'ev, *Ultrafast Diode Lasers: Fundamentals and Applications* (Artech House, Norwood, MA, 1995).
16. P. B. Littlewood and X. Zhu, Phys. Scr. T **68**, 56 (1996).

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 315–320.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 356–361.

SOLIDS
Electronic Properties

Resonance Double Magnetic Bremsstrahlung in a Strong Magnetic Field

P. I. Fomin[a,*] and R. I. Kholodov[b,]**

[a]*Bogolyubov Institute of Theoretical Physics, Ukrainian National Academy of Sciences, Kiev, 03143 Ukraine*
[b]*Institute of Applied Physics, Ukrainian National Academy of Sciences, Sumy, 40030 Ukraine*
**e-mail: pfomin@bitp.kiev.ua*
***e-mail: kholodov@ipfcentr.sumy.ua*

Received August 15, 2002

Abstract—The possibility of resonance double magnetic bremsstrahlung in the approximation of weakly excited electron states in a strong external magnetic field is analyzed. The differential probability of this process in the Breit–Wigner form is obtained. The probability of double magnetic bremsstrahlung (second-order process of perturbation theory) is compared with the probability of magnetic bremsstrahlung (first-order process of perturbation theory). *© 2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

First-order quantum electrodynamic processes in a magnetic field, in particular, magnetic bremsstrahlung (synchrotron emission) have been well studied for quite a long time. Their general relativistic theory has been constructed and investigated, and the general form of expressions describing these processes has been obtained by numerous researchers and included in a number of monographs [1–6]. In feasible experiments, the values of magnetic fields are much less than that of the critical field $H_0 = m^2/e = 4.41 \times 10^{13}$ G. In such a field, charged particles (electrons, positrons) are in highly excited energy states with quasi-continuous energy levels n (n is the number of the Landau level). In this case, the motion of particles is quasi-classical. Note that it is to this approximation that the major part of the available literature is devoted.

No less interesting is the case of electron motion in a strong magnetic field, when an electron finds itself on one of the lowest energy levels [7–9]. It is useful to treat such problems, in particular, in studying the gas of electrons and positrons of the highly magnetized magnetospheres of neutron stars.

In this approximation (weakly excited electron states, strong magnetic field), we treat a second-order process, namely, double magnetic bremsstrahlung.

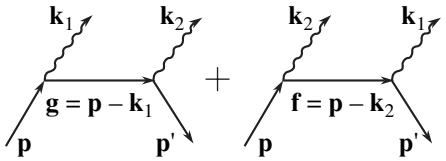
For the first time, double magnetic bremsstrahlung was treated in a quasi-classical (with respect to the electron motion) approximation in terms of solving an auxiliary problem: the emission of a photon by an electron in an external field of Redmond configuration (plane wave + magnetic field along the wave) with subsequent expansion in terms of magnitude of the wave intensity [10, 11]. Sokolov *et al.* [12] studied this process within the second Born approximation in the ultrarelativistic limit. In the studies mentioned above, double synchrotron emission was studied in a region away from the resonance process (away from the poles of the Green function of an intermediate virtual electron).

We will demonstrate in this paper that resonance double magnetic bremsstrahlung is possible when an electron is in a weakly excited energy state and, at the same time, emits photons with an energy equal to the distance between Landau levels.

The best studied of the second-order processes in a strong magnetic field is the process of photon scattering by an electron [13–16]. We will use the results of these studies, because both processes (photon scattering by an electron and emission of two photons by an electron) are described by similar expressions accurate within the replacement of the initial photon by the final one.

2. CONDITIONS FOR INITIATION OF RESONANCES IN THE CASE OF WEAKLY EXCITED ELECTRON STATES IN A STRONG MAGNETIC FIELD

The process of emission of two photons by an electron in a magnetic field is described by the Feynman diagrams given in the figure. The wavy lines indicate photons with four-momenta $k_1 = (\omega_1, \mathbf{k}_1)$ and $k_2 = (\omega_2, \mathbf{k}_2)$, which do not interact with the external field. The external solid lines indicate the exact solutions of the Dirac equation for an electron in a uniform magnetic field with four-momenta $p = (\varepsilon_l, 0, p_y, p_z)$ and $p' = (\varepsilon_{l'}', 0, p_y', p_z')$ (l and l' are the numbers of the Landau levels) [17], and the intermediate solid lines indicate the electron Green function in an external uniform



Feynman diagrams for the process of emission of two photons by an electron in a magnetic field.

magnetic field. Note that these diagrams are similar to the diagrams describing the process of photon scattering by an electron in a magnetic field, with the only difference being that the initial photon in the process of Compton scattering must be replaced by the final one. Therefore, the amplitude of double magnetic bremsstrahlung is obtained from the amplitude of Compton scattering [14, 15] by the following replacement:

$$k \longrightarrow -k_1 = (-\omega_1, -\mathbf{k}_1), \quad k' \longrightarrow k_2 = (\omega_2, \mathbf{k}_2). \quad (1)$$

This amplitude contains three delta functions which correspond to the laws of conservation of energy and projections of momentum onto the direction of the y and z axes,

$$\varepsilon_l = \varepsilon'_{l'} + \omega_1 + \omega_2, \quad p_y = p'_y + k_{1y} + k_{2y}, \quad p_z = p'_z + k_{1z} + k_{2z}, \quad (2)$$

and the poles of the Green functions (intermediate states) of the first and second diagrams are, respectively,

$$g_0^2 - \varepsilon_{gn1}^2 = g_0^2 - (m^2 + 2n_1 hm^2 + g_z^2), \quad (3)$$

$$f_0^2 - \varepsilon_{fn2}^2 = f_0^2 - (m^2 + 2n_2 hm^2 + f_z^2), \quad (4)$$

where n_1 and n_2 are the numbers of the Landau levels of the intermediate states of the first and second diagrams, over which the summation is performed in the amplitude in the general case, and h is the magnetic field in the critical field units m^2/e. For intermediate particles, the four-momenta g and f (except for the x components) are expressed in terms of the momenta of the initial and final particles,

$$g_0 = \varepsilon_l - \omega_1, \quad g_y = p_y - k_{1y}, \quad g_z = p_z - k_{1z}, \quad (5)$$

$$f_0 = \varepsilon_l - \omega_2, \quad f_y = p_y - k_{2y}, \quad f_z = p_z - k_{2z}. \quad (6)$$

The laws of conservation according to Eqs. (2) in view of the laws of dispersion impose the following restriction on the frequency ω_2 (in the frame of reference in which $p_z = 0$):

$$\omega_2 = \frac{\varepsilon_l - \omega_1(1 - vu)}{1 - u^2} \times \left[1 - \sqrt{1 - \frac{\omega_1^2(1 - v^2) - 2\varepsilon_l\omega_1 + 2(l - l')hm^2}{(\varepsilon_l - \omega_1(1 - vu))^2}(1 - u^2)}\right], \quad (7)$$

where v and u are cosines of the angles between the direction along the magnetic field and the directions of photons ω_1 and ω_2,

$$v = \cos\theta_1, \quad u = \cos\theta_2. \quad (8)$$

We will treat the process in a strong external magnetic field, in the case of which individual Landau levels of an electron are experimentally different (ultraquantum approximation), which corresponds to the condition

$$\Delta l = 1, \quad (9)$$

where Δl is the number of levels entering into final states. The distance between the neighboring Landau levels (cyclotron frequency of electron) is on the same order as the photon energy. For example, for photons with an energy on the order of 10^4 eV (X-rays), this requirement is satisfied by magnetic fields $H \sim 10^{12}$ G. On the other hand, the magnetic field is taken to be small compared to the critical field $H_0 = 4 \times 10^{13}$ G,

$$h \equiv eH/m^2 \ll 1. \quad (10)$$

The quantity h is the small parameter of the problem, and the energies ε_l and $\varepsilon_{l'}$ of the initial and final electrons take nonrelativistic values. Under these conditions, expression (7) for ω_2 takes the form

$$\omega_2 = (l - l')hm - \omega_1 - \frac{\omega_1^2}{2m}(v - u)^2 - h\omega_1(l - l')u(v - u) - \frac{h^2 m}{2}(l - l')[l + l' + (l - l')u^2]. \quad (11)$$

The conditions of resonance in the first Feynman diagram imply that the pole according to Eq. (3) is zero, which, after expansion in terms of h, gives the following restriction imposed on the frequency ω_1:

$$\omega_{1r} = (l - n_1)hm\left[1 - hl + \frac{h}{2}(l - n_1)(1 - v^2)\right], \quad l > n_1. \quad (12)$$

One can see in expression (12) that, in order to register the resonance according to the first diagram, the detector registering one of the photons must be tuned to the frequency defined by the integers l, n_1 and the angle of

emission of this same photon. We substitute Eq. (12) into expression (11) for ω_2 to derive

$$\omega_{2r} = (n_1 - l')hm\left\{1 - \frac{h}{2}[n_1 + l' + (n_1 - l')u^2 + 2(l - n_1)vu]\right\}, \quad n_1 > l'. \tag{13}$$

It follows from Eqs. (12) and (13) that the frequencies of emitted photons under resonance conditions are equal, within the first power of the parameter h, to the distance between the Landau levels of the electron and vary only little with the angles of emission of photons, u and v.

We equate expressions (12) and (13) to readily obtain the resonance conditions for a process with emission of two photons of similar frequency,

$$l - n_1 = n_1 - l' = 1, \quad v = u = \pm 1. \tag{14}$$

Conditions (14) imply that the energy level of an intermediate electron is neighboring for the levels of the initial and final electrons, and the photons escape along the direction of the magnetic field, with their frequencies described by the expression

$$\omega_{1,2} = hm - lh^2m. \tag{15}$$

The resonance in the second diagram is realized when expression (4) vanishes, whence follows the restriction on the frequency ω_2,

$$\omega_{2r} = (l - n_2)hm\left[1 - hl + \frac{1}{2}(l - n_2)(1 - u^2)\right], \quad l > n_2. \tag{16}$$

For expressions (16) and (11) to be equivalent up to the second order of smallness with respect to h, the following restriction is imposed on ω_1:

$$\omega_{1r} = (n_2 - l')hm\left\{1 - \frac{h}{2}[n_2 + l' + (n_2 - l')v^2 + 2(l - n_2)vu]\right\}, \quad n_2 > l'. \tag{17}$$

Unlike the previous case (12), the resonance frequency ω_{1r} is defined by the angles of emission v and u of both photons. It is obvious that the resonance frequencies given by Eqs. (16) and (17) are obtained from expressions (12) and (13) by the simple replacement of $(\omega_1, v) \rightleftarrows (\omega_2, u)$.

We equate the frequencies given by Eqs. (16) and (17) to each other to obtain conditions which exactly agree with those given by Eq. (14); in doing so, $n_1 = n_2$. Therefore, in the case of photon emission along a field of the same frequency, the resonance conditions are simultaneously satisfied in both diagrams of the process being treated.

The maximal variation of the frequency ω_1, which is resonant according to the first diagram (12), occurs when the angle θ_1 varies from zero to $\pi/2$ (with v varying from unity to zero),

$$\Delta\omega_{1r} = \omega_{1r}\big|_{v=0} - \omega_{1r}\big|_{v=1} = h^2m(l - n_1)^2/2 \sim h^2m. \tag{18}$$

This quantity is equal to h^2m within a factor of the order of unity for weakly excited electron states. Of the same order of magnitude is the variation of the second photon frequency $\Delta\omega_{2r}$, as well as the variation of these frequencies under resonance conditions for the second diagram.

The Landau level n (n_1 or n_2) of an intermediate electron has nonzero width equal to the double total probability of decay of intermediate state, i.e., the probability of single photon emission [14],

$$\Gamma_n = 2W_n^{\mu_n} = \frac{4}{3}\alpha h^2 m(2n - 1 - \mu_n), \tag{19}$$

where α is the fine structure constant and μ_n is the spin of intermediate state, which has a definite value under resonance conditions (in [14], we used the approximation of the width $\langle\Gamma\rangle = 4(2n - 1)\alpha h^2 m/3$ averaged over the spins of intermediate state). On comparing $\Delta\omega_r$ and Γ_n, we see that, in the approximation according to (9) and (10),

$$\Gamma_n \ll \Delta\omega_r \ll \omega_r, \tag{20}$$

with the ratio $\Gamma_n/\Delta\omega_r$ being independent of the magnitude of the field h and defined only by the numbers of Landau levels of the electron.

We will determine the maximal interval of angles $\Delta\theta_1$ of the first photon emission, which does not take the process beyond the resonance region according to Eq. (12),

$$\left|\omega_{1r}\big|_{v_a} - \omega_{1r}\big|_{v_b}\right| = \Gamma_{n_1}, \tag{21}$$

$$\left|v_a^2 - v_b^2\right| = \sin(\theta_{1a} - \theta_{1b})\sin(\theta_{1a} + \theta_{1b}) = \alpha\frac{8(2n_1 - 1 - \mu_{n_1})}{3(l - n_1)^2}. \tag{22}$$

For definiteness, we will assume that $l = 2$, $n_1 = 1$, and $\mu_{n_1} = -1$ (the process with the lowest energy electron states). We will treat two limiting cases, namely, the region of angles in the neighborhood of zero, $\theta \approx 0$, and in the neighborhood of $\theta \approx \pi/4$. In the first case,

$$\left|v_a^2 - v_b^2\right| \approx \Delta\theta_1^2 \Rightarrow \sqrt{\Delta\theta_1^2} \approx 11^\circ, \tag{23}$$

and in the second case,

$$\left|v_a^2 - v_b^2\right| \approx \Delta\theta_1 \Rightarrow \Delta\theta_1 \approx 2^\circ. \quad (24)$$

The estimates indicate that the resonance conditions of the process are defined by the frequency of one of the photons (the photon emitted by the initial electron) and by the angle of its emission. When u is varied, the maximal variation of the frequency of the second photon emitted by an intermediate electron (for the first diagram according to Eq. (13)) also exceeds the resonance width; however, this does not take the process out of the resonance region.

3. PROBABILITY OF MAGNETIC BREMSSTRAHLUNG UNDER RESONANCE CONDITIONS

As was observed above, the amplitude of the process may be provided by the Compton scattering amplitude [14] where replacement (1) has been made. The Compton scattering cross section is equal to the product of the square of the amplitude by the number of final states divided by the flux j of initial photons and time T,

$$d\tilde{\sigma}_C = \frac{W_C V d^3k_2 S d^2p'}{Tj(2\pi)^5}. \quad (25)$$

The tilde implies that replacement (1) has been made in this expression. In the process of double magnetic bremsstrahlung, in the final state one photon is added, and the probability of such process per unit time is

$$dW_D = \frac{W_D V d^3k_2 S d^2p'}{T(2\pi)^5}\frac{V d^3k_1}{(2\pi)^3}, \quad (26)$$

where W_D is the square of the amplitude of double magnetic bremsstrahlung. Because the squares of amplitudes W_C and W_D are equal to each other, the ratio of the differential probability given by Eq. (26) to the differential cross section given by Eq. (25) has the form

$$\frac{dW_D}{d\tilde{\sigma}_C} = \frac{\omega_1^2 d\omega_1 dv}{4\pi^2}. \quad (27)$$

In the vicinity of resonance, the Compton cross section is defined by a formula of the Breit–Wigner type, in which the partial widths are replaced by the differential (with respect to the photon entrance (departure) angle) probabilities of magnetic bremsstrahlung, dW/dv, dW/du (probabilities of single photon emission) [15]. Replacement (1) in the expression for a cross section in resonance according to the first diagram leads to the replacement of $dW_{n_1, l}$ (electron transition from level n_1 to level l) by dW_{l, n_1} (electron transition from level l to level n_1),

$$\frac{d\tilde{\sigma}_C}{du} = \pi\lambda\!\!\!^{-2}\frac{\dfrac{dW_{l,n_1}}{dv}\dfrac{dW_{n_1,l'}}{du}}{(\omega_1-\omega_{1r})^2 + \Gamma^2/4}, \quad (28)$$

where $\lambda\!\!\!^{-} = 1/\omega_{1r}$, with ω_{1r} preassigned by expression (12). It is obvious that the probability of double magnetic bremsstrahlung in resonance according to the first diagram will have a form similar to that according to Eq. (28) within a factor which may be easily determined using the ratio given by Eq. (27),

$$\frac{dW_{D1}}{d\omega_1 dv du} = \frac{1}{4\pi}\frac{\dfrac{dW_{l,n_1}}{dv}\dfrac{dW_{n_1,l'}}{du}}{(\omega_1-\omega_{1r})^2 + \Gamma^2/4}, \quad l > n_1 > l'. \quad (29)$$

The probability of a process which is resonant in accordance with the second diagram, dW_{D2}, is given by expression (29), in which the following replacement must be made:

$$v \longrightarrow u, \quad u \longrightarrow v, \quad n_1 \longrightarrow n_2. \quad (30)$$

To complete the picture, we will write the explicit form of the differential probabilities of single photon emission by an electron at different values of the spin projection of the latter in the initial and final states (the plus and minus superscripts indicate the spin projection in and against the field direction, respectively) [8, 9],

$$\frac{dW_{ln}^{--}}{dv} = \alpha m A\frac{l}{n}\eta^{l-n-1}h^2(1+v^2), \quad (31)$$

$$\frac{dW_{ln}^{++}}{dv} = \alpha m A\eta^{l-n-1}h^2(1+v^2), \quad (32)$$

$$\frac{dW_{ln}^{+-}}{dv} = \alpha m A\frac{(l-n)^2}{2n}\eta^{l-n-1}h^3(1+v^2), \quad (33)$$

$$\begin{aligned}\frac{dW_{ln}^{-+}}{dv} &= \alpha m A\frac{l(l-n)^2}{8(l-n+1)^2}\eta^{l-n-1}h^5\\ &\times\{1+v^2[1+6(l-n)+4(l-n)^2]\\ &-v^4[2(l-n)+3(l-n)^2]+v^6(l-n)^2\},\end{aligned} \quad (34)$$

where

$$A = \frac{(l-n)(l-1)!}{2(n-1)!(l-n-1)!^2}, \quad \eta = \frac{(l-n)^2h(1-v^2)}{2}. \quad (35)$$

For a process occurring without reorientation of spin, the intermediate electron spin is oriented in the

same manner as are the spins of initial and final electrons. This is associated with the fact that the probabilities $dW_{ln}^{--}/d\nu$ and $dW_{ln}^{++}/d\nu$ contain a smaller power of the small parameter h than $dW_{ln}^{+-}/d\nu$ and, the more so, than $dW_{ln}^{-+}/d\nu$. So, if, in the initial and final states, the spin is against the field, expression (29) has $dW_{ln}^{--}/d\nu$ and $dW_{nl'}^{--}/du$ in the numerator, and this means that the spin of intermediate state is likewise directed against the field.

Most probable are processes with the transition of electrons to the neighboring levels such as $l \longrightarrow n_{1,2} = l-1 \longrightarrow l' = l-2$. The differential probability of double magnetic bremsstrahlung by an electron with the spin directed against the field ($\mu = \mu_{n_1} = \mu' = -1$) at the point of resonance according to the first diagram in view of the expressions for the probability of double magnetic bremsstrahlung (31) and for width (19) has the form

$$\frac{dW_{D1,l,l-2}^{--}}{d\omega_1 d\nu du} = \frac{9l}{2^8\pi(l-1)}(1+\nu^2)(1+u^2), \quad l = 2, 3, \ldots . \tag{36}$$

For a process of emission by electrons with the spins in the field direction ($\mu = \mu_{n_1} = \mu' = +1$), the differential probability in view of Eqs. (32) and (19) is

$$\frac{dW_{D1,l,l-2}^{++}}{d\omega_1 d\nu du} = \frac{9(l-1)}{2^8\pi(l-2)}(1+\nu^2)(1+u^2), \quad l = 3, 4, \ldots . \tag{37}$$

The expression for differential probability at the point of resonance according to the first diagram, dW_{D2}, is derived from expressions (36) and (37) by way of replacement $\nu \rightleftharpoons u$. However, the dependence on ν and u in the expression of probability of the process is the same when the electrons make a transition to the neighboring levels. Therefore, in this case, the probabilities dW_{D1} and dW_{D2} coincide.

We will estimate the total probability of double magnetic bremsstrahlung. The integration of expression (29) with respect to ω_1 is trivial and equivalent to multiplying the differential probability at the resonance point according to Eqs. (36) and (37) by the width $\pi\Gamma/2$. As was observed above, the resonance conditions depend on the angles of photon emission, ν and u, and the variation of these quantities may cause a variation of the resonance frequencies ω_{1r} and ω_{2r} by a value exceeding the resonance width. However, in the integrand with respect to $d\omega_1$, the variation of the angles causes a variation of only the position of the resonance range of integration. Therefore, integration with respect to ν and u may be performed at the resonance point proper. Because the conditions of resonances in the first and second diagrams do not coincide (except for the small range of $\nu \sim u \sim 1$), the total probability is equal to the double probability ΔW_{D1} (Δ indicates that the probability is estimated within the resonance region, because the employed expression (29) is invalid outside of resonance),

$$\Delta W_{Dl,l-2}^{\mu\mu} = \Delta W_{D1l,l-2}^{\mu\mu} + \Delta W_{D2l,l-2}^{\mu\mu} = 2\Delta W_{D1l,l-2}^{\mu\mu} = \alpha h^2 m(2l-1-\mu)/3. \tag{38}$$

In this formula, $\mu = 1$ (-1) corresponds to a process with electrons with the spin in the field direction (against the field). For comparison, we will write the probability of magnetic bremsstrahlung with the transition of an electron from level l to level $l-2$, which is obtained by integrating expressions (31) and (32) with $n = l-2$,

$$W_{l,l-2}^{\mu\mu} = 16\alpha h^3 m(l-1)(l-1-\mu)/5. \tag{39}$$

We see that this expression is less than (38) by an order of h. This means that, in the approximation (9), (10) being treated, a second-order resonance process of perturbation theory exceeds a first-order process with the same electron states.

It must be emphasized that the probability given by Eq. (39) is not total. The most significant contribution to the expression for total probability of magnetic bremsstrahlung is made by the term corresponding to the transition of an electron to the neighboring level $l \longrightarrow l-1$,

$$W^{\mu\mu} = 2\alpha h^2 m(2l-1-\mu)/3, \tag{40}$$

and this probability is twice that of double magnetic bremsstrahlung.

Analysis reveals that, in the ultraquantum case, when one can distinguish between individual Landau levels of an electron, the inclusion of higher-order processes of perturbation theory is important in approximation (9), (10) in calculating the probability of magnetic bremsstrahlung.

ACKNOWLEDGMENTS

We are grateful to S.P. Roshchupkin for valuable discussions and to A.I. Nikishov for his interest in our work.

REFERENCES

1. A. A. Sokolov and I. M. Ternov, *Radiation from Relativistic Electrons* (Nauka, Moscow, 1974; AIP, New York, 1986).
2. I. M. Ternov, V. R. Khalilov, and V. N. Rodionov, *Interaction of Charged Particles with Strong Electromagnetic Field* (Mosk. Gos. Univ., Moscow, 1982).

3. N. P. Klepikov, Zh. Éksp. Teor. Fiz. **26**, 19 (1954).
4. I. M. Ternov, V. G. Bagrov, and R. A. Rzaev, Zh. Éksp. Teor. Fiz. **46**, 374 (1964) [Sov. Phys. JETP **19**, 255 (1964)].
5. V. N. Baĭer, V. M. Katkov, and V. M. Strakhovenko, Zh. Éksp. Teor. Fiz. **67**, 453 (1974) [Sov. Phys. JETP **40**, 225 (1975)].
6. A. I. Nikishov, Tr. Fiz. Inst. im. P. N. Lebedeva Akad. Nauk SSSR **111**, 152 (1979).
7. V. G. Bagrov, D. M. Gitman, V. N. Rodionov, *et al.*, Zh. Éksp. Teor. Fiz. **71**, 433 (1976) [Sov. Phys. JETP **44**, 228 (1976)].
8. I. G. Mitrofanov and A. S. Pozanenko, Zh. Éksp. Teor. Fiz. **93**, 1951 (1987) [Sov. Phys. JETP **66**, 1112 (1987)].
9. R. I. Kholodov and P. V. Baturin, Ukr. Fiz. Zh. **46**, 621 (2001).
10. V. Ch. Zhukovskiĭ and I. Kherman, Yad. Fiz. **14**, 150 (1971) [Sov. J. Nucl. Phys. **14**, 85 (1972)].
11. V. Ch. Zhukovskiĭ and N. S. Nikitina, Zh. Éksp. Teor. Fiz. **64**, 1169 (1973) [Sov. Phys. JETP **37**, 595 (1973)].
12. A. A. Sokolov, A. M. Voloshchenko, V. Ch. Zhukovskiĭ, and Yu. G. Pavlenko, Izv. Vyssh. Uchebn. Zaved., Fiz., No. 9, 46 (1976).
13. R. W. Bussard, S. A. Alexander, and P. Meszaros, Phys. Rev. D **34**, 440 (1986).
14. P. I. Fomin and R. I. Kholodov, Zh. Éksp. Teor. Fiz. **117**, 319 (2000) [JETP **90**, 281 (2000)].
15. P. I. Fomin and R. I. Kholodov, Laser Phys. **10**, 1150 (2000).
16. P. L. Gonthier, A. K. Harding, M. G. Baring, *et al.*, Astrophys. J. **540**, 907 (2000).
17. A. I. Akhiezer and V. B. Berestetskiĭ, *Quantum Electrodynamics*, 4th ed. (Nauka, Moscow, 1981; Wiley, New York, 1965).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 321–330.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 362–372.

SOLIDS
Electronic Properties

Zeeman Effect for Holes in a Ge/Si System with Quantum Dots

A. V. Nenashev, A. V. Dvurechenskiĭ, and A. F. Zinov'eva*
Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva 13, Novosibirsk, 630090 Russia
**e-mail: zinoviev@isp.nsc.ru*
Received August 21, 2002

Abstract—The tight binding approximation is employed to study the Zeeman effect for the hole ground state in a quantum dot. A method is proposed for calculating the *g* factor for localized states in a quantum dot. This method can be used both for hole states and for electron states. Calculations made for a Ge/Si system with quantum dots show that the *g* factor of a hole in the ground state is strongly anisotropic. The dependence of the *g* factor on the size of a germanium island is analyzed and it is shown that anisotropy of the *g* factor increases with the island size. It is shown that the value of the *g* factor is mainly determined by the contribution of the state with the angular momentum component $J_z = \pm 3/2$ along the symmetry axis of the germanium island.

1. INTRODUCTION

The interaction of electronic states having a spin of ±1/2 with an external magnetic field us described by the *g* factor characterizing the spin splitting of a free electron ($g \approx 2$). The interaction with the lattice potential in solids leads to a considerable difference in the *g* factor from that for a free electron. As the system dimensionality decreases from 3D to 2D and lower, size quantization effects lead to new changes in the *g* factor of charge carriers. For example, quantization for electrons in a low-dimensional system leads to a considerable renormalization of the value of the *g* factor [1] and to its strong anisotropy [2]. The *g* factor contains numerical information on the change in the band structure of the semiconductor upon the reduction in its dimensionality. For this reason, a large number of theoretical and experimental studies are devoted to analysis of this parameter. In some publications dealing with electron states, consistent **kp** theories have been developed, which make it possible to calculate the *g* factor in quantum wells and superlattices [3] as well as in quantum dots [4]. For hole states, the Zeeman effect has been studied theoretically and experimentally for structures with quantum wells [5–7].

Let us describe fundamental differences between two-dimensional quantum wells and quantum dots, which must lead to a change in the *g* factor. Wide quantum wells in magnetic fields of energies smaller than the quantization energy (or the energy of band splitting caused by elastic stresses) can be treated in the approximation of a bulk semiconductor to obtain values of *g* factors for hole subbands directly from the exact form of the 8×8 Hamiltonian in the **kp** theory: $g_\parallel = 6k$, $g_\perp = 0$ for a heavy hole and $g_\parallel = 2k$, $g_\perp = 4k$ for a light hole ($g_\parallel$ and $g_\perp$ are the *g* factor components parallel and perpendicular to the principal axis of the structure (*z* axis); *k* and *q* are the Luttinger parameters, the latter parameter being omitted in view of its smallness). In narrower quantum wells, the uncertainty in momentum k_z increases, leading to a modification of the *g* factor for a light hole due to admixture of states of the split-off band and the conduction band (we assume here that *z* is the growth direction of the epitaxial film) [7]. The Lande factor for a heavy hole at the bottom of the band practically does not change, since the heavy hole band does not interact with the nearest bands. In narrow wells, the *g* factors for light and heavy holes change due to the effect of the barriers forming the quantum well [5].

In the case of quantum dots, a considerable renormalization of the *g* factor of hole states must be due to the emergence of a quantizing potential not only to in the growth direction, as in the case of 2D structures, but to equally strong quantization in the lateral direction (in the *xy* plane). This leads to uncertainty in k_x and k_y and, hence, to a strong mixing of light and heavy hole bands with the split-off band [8]. As a rule, this mixing is disregarded in theoretical analysis of the Zeeman effect in 2D systems, since states at the bottom of the band, where $k_x, k_y = 0$, are considered.

In quantum dots created on the basis of stressed heterostructures, the *g* factor may change significantly due to inhomogeneity of strains within quantum dots. If we compare a quantum dot with a quantum well grown in the [100] direction, shear strains ε_{xy}, ε_{xz}, ε_{yz} leading to mixing of light and heavy hole bands are absent in the quantum well [8], while the quantum dot experiences such strains.

Thus, in the case of quantum dots, quantization in all three directions and strain inhomogeneity must lead to a considerable change in the *g* factor of hole states due to energy band mixing.

We propose a method of calculating the *g* factor for hole states in quantum dots using the tight binding approach. This method makes it possible to take into account the specific form of a quantizing potential (described not only by an analytic functions) and calculate the *g* factor for a quantum dot of any shape and of a small size comparable with atomic spacing. The proposed method can be also applied for calculating the *g* factor of electron states in quantum dots.

The paper has the following structure. In Section 2, the method of calculating the *g* factor is described. The *g* factor for hole states in germanium quantum dots in a silicon matrix is calculated in Section 3. Strong anisotropy of the *g* factor of holes is discovered, and the dependence of the *g* factor of a hole on the quantum dot size is established. The probabilities of Zeeman transitions as functions of the magnetic field direction are investigated. Section 4 is devoted to analysis of the results obtained. The main effects determining the magnitude of the *g* factor and its dependence on the island size are revealed using a simplified model of noninteracting bands.

2. COMPUTING METHOD

This method is the evolution of the idea proposed by us earlier in [9], where an atomistic approach was used for calculating the *g* factor of the hole state in a quantum dot. This approach involves the computation of the angular momentum of a hole in an atomic orbital. However, as we pass to the limiting case of a bulk crystal, this approach fails to provide values matching the bulk value of the *g* factor. For this reason, we extend the former approach by taking into account the angular momentum of Bloch functions.

The Zeeman interaction of a particle having a magnetic moment **M** with a magnetic field **H** can be written in the form

$$\hat{H} = -\hat{\mathbf{M}} \cdot \mathbf{H}.$$

Magnetic moment **M** is connected with angular momentum **J** through the relation

$$\mathbf{M} = -g_0 \mu_B \mathbf{J},$$

where μ_B is the Bohr magneton and g_0 is the *g* factor equal to 2 for particles with purely spin electron magnetism and to 1 for those possessing purely orbital electron magnetism.

We introduce the magnetic moment $\mathbf{M}_{QD}$ of a hole (electron) in a quantum dot, which is measured in units of the Bohr magneton:

$$\mathbf{M}_{QD} = -(\mathbf{L} + 2\mathbf{S}),$$

where **L** and **S** are the orbital and spin components of the magnetic moment. We write the Hamiltonian of interaction with the magnetic field in the form

$$\hat{H}_{QD}(\mathbf{H}) = \mu_B \hat{\mathbf{M}}_{QD} \cdot \mathbf{H} = \mu_B(\hat{\mathbf{L}} + 2\hat{\mathbf{S}}) \cdot \mathbf{H}.$$

It follows from symmetry considerations that the ground state in a quantum dot is doubly degenerate and forms a Kramers doublet. The Zeeman interaction energy for states of a Kramers doublet is given by

$$\frac{1}{2}\mu_B \hat{\sigma}_\alpha g_{\alpha\beta} H_\beta,$$

where $\hat{\sigma}_\alpha$ ($\alpha = x, y, z$) are Pauli spin matrices and $g_{\alpha\beta}$ is a tensor which has nine independent components in the general case [10]. In most cases (except in low-symmetry structures), we have $g_{xy} = g_{yx}$, etc., and off-diagonal terms can be eliminated by an appropriate choice of the *x*, *y*, and *z* axes (which are known as principal axes). In these axes, the *g* tensor is characterized by three principal values g_{xx}, g_{yy}, and g_{zz}.

In the first order of perturbation theory, the *g* factor can be determined from the solution of a secular equation, which gives

$$|g| = 2\sqrt{\langle \psi|\mathbf{n} \cdot \hat{\mathbf{M}}_{QD}|\psi\rangle^2 + |\langle \psi|\mathbf{n} \cdot \hat{\mathbf{M}}_{QD}|\psi^*\rangle^2|}, \quad (1)$$

where ψ and ψ^* are the wave functions forming a Kramers doublet for a given level and **n** is a unit vector directed along the magnetic field. Consequently, in order to calculate the *g* factor, we must find the matrix elements of operator $\hat{\mathbf{M}}_{QD}$. To determine the matrix elements, we must find wave eigenfunctions ψ and ψ^* for hole or electron states in a quantum dot. We assume that the magnetic field is quite weak and does not significantly change the waves functions of a hole (electron), which enables us to use wave eigenfunctions of the unperturbed Hamiltonian for calculating the matrix elements.

The wave eigenfunctions ψ and ψ^* for hole states were determined by us in [9], where the energy spectrum of holes in a quantum dot was calculated. For this purpose, we used the tight binding model with basis sp^3 [11]. In this model, each atom is supplied with a set of orbitals *s*, p_x, p_y, and p_z, and the dimensionality of the vector of state of the system is equal to the number of atoms multiplied by the number of orbitals per atom. The interactions between nearest neighbors are taken into account in the two-center approximation [12] as well as the spin–orbit interaction [13]. Deformation effects [14] are taken into account by introducing the dependence of interatomic matrix elements of the Hamiltonian on the orientation of relevant bonds [12] and their length [15]. Vector $|\psi\rangle$ is determined using the free relaxation method [16]. Each vector component ψ_{nN} is the amplitude of probability of finding a particle in the *n*th orbital of the *N*th atom.

Since the vector of the state corresponding to a certain size quantization level is defined as a combination of atomic orbitals, we must determine matrix elements of operator $\hat{\mathbf{M}}_{QD}$ in the representation of atomic orbitals. Let us first define the orbital moment **L**. We can ascribe to each diagonal matrix element $\langle \psi_{nN}|\hat{\mathbf{L}}|\psi_{nN}\rangle$ the physical meaning of the angular momentum of a particle on the corresponding *n*th orbital of the chosen *N*th atom in a quantum dot.

For a hole (electron) with coordinates (*x, y, z)* located with probability $|\psi_{nN}|^2$ in the *n*th orbital around the chosen *N*th atom with coordinates (*X, Y, Z*), we can write the angular momentum operator

$$\hat{L}_\alpha = \frac{1}{\hbar} e_{\alpha\beta\gamma}\hat{p}_\beta\hat{r}_\gamma,$$

where $e_{\alpha\beta\gamma}$ is a unit antisymmetric tensor.

Using the rules for differentiating operators with respect to time [17], we can express the momentum operator $\hat{\mathbf{p}} = m\dot{\hat{\mathbf{r}}}$ in terms of Hamiltonian $\hat{H}_0$ and the coordinate operator $\hat{\mathbf{r}} = (\hat{x}, \hat{y}, \hat{z})$:

$$\hat{\mathbf{p}} = \frac{im}{\hbar}(\hat{H}_0\hat{\mathbf{r}} - \hat{\mathbf{r}}\hat{H}_0),$$

where *m* is the mass of a free electron. Then the angular momentum operator has the form

$$\hat{L}_\alpha = \frac{im}{\hbar^2} e_{\alpha\beta\gamma}\hat{r}_\beta\hat{H}_0\hat{r}_\gamma.$$

The obtained expression cannot be used directly for determining the matrix elements $\langle \psi|\hat{\mathbf{M}}_{QD}|\psi\rangle$ and $\langle \psi|\hat{\mathbf{M}}_{QD}|\psi^*\rangle$, since the wave functions ψ and ψ* have been calculated in the tight binding approximation, while operator $\hat{\mathbf{r}} = (\hat{x}, \hat{y}, \hat{z})$ of the hole (electron) coordinates is meaningless in this approximation. For this reason, we replace it by coordinate operator $\hat{\mathbf{R}} = (\hat{X}, \hat{Y}, \hat{Z})$ of the atom possessing the orbital,

$$\hat{L}_\alpha = \frac{im}{\hbar^2} e_{\alpha\beta\gamma}\hat{R}_\beta\hat{H}_0\hat{R}_\gamma. \quad (2)$$

Carrying out the substitution $\hat{\mathbf{r}} \longrightarrow \hat{\mathbf{R}}$, we lose a fraction of the angular momentum associated with the strongly oscillating Bloch wave function (which can be referred to as the effective spin angular momentum component) and have only a part of the angular momentum associated with a smooth envelope of the wave function of a hole (electron), viz., the orbital component.

If we disregard the interaction between the nearest energy bands, we must simply supplement Eq. (2) with the effective spin component of the angular momentum of a charge carrier in the corresponding energy band (conduction band for electrons and valence band for holes) in order to obtain the total moment $\mathbf{M}_{QD}$. However, the state of a hole (electron) in a quantum dot is formed not only by states from the valence (conduction) band; neighboring bands also make a contribution to the formation of the state. The nearest bands for hole states are the split off (SO) band and the conduction band (CB). For electron states, these are bands of heavy holes (HH) and light holes (LH) as well as the split-off band. The contribution from other bands is negligibly small.

The wave function of a hole (electron) can be presented in the form

$$|\psi\rangle = |A_1(\mathbf{R})CB\rangle + A_2(\mathbf{R})|HH\rangle + A_3(\mathbf{R})|LH\rangle + A_4(\mathbf{R})|SO\rangle,$$

where coefficients A_1, A_2, A_3, A_4 depend on the position of an atom in a quantum dot and reflect the contributions from the corresponding bands to the state of a particle in the quantum dot. Each wave function component possesses its own effective spin and interacts with the magnetic field in accordance with the following expressions.

For a hole in the HH band, effective spin $\mathbf{S}_{HH}$ is often introduced for describing Zeeman sublevels [6]: spin $(S_{HH})_z = -1/2$ is ascribed to one of the sublevels with $J_z = -3/2$, while spin $(S_{HH})_z = 1/2$ is ascribed to the other sublevel with $J_z = 3/2$. In this case, the Zeeman interaction can be written in the form

$$\hat{H}(\mathbf{H}) = \mu_B g_{HH}(\hat{\mathbf{S}}_{HH} \cdot \mathbf{H}), \quad (3)$$

where g_{HH} is the *g* factor for a hole in the HH band. The same can be done for a hole in the LH band: we ascribe spin $(S_{LH})_z = -1/2$ for the sublevel with $J_z = -1/2$ and spin $(S_{LH})_z = 1/2$ for the sublevel with $J_z = 1/2$. Then the Zeeman interaction in the LH band has the form

$$\hat{H}(\mathbf{H}) = \mu_B g_{LH}(\hat{\mathbf{S}}_{LH} \cdot \mathbf{H}), \quad (4)$$

where g_{LH} is the *g* factor for a hole in the LH band.

For a degenerate valence band, the interaction with the magnetic field at point Γ can be described in the form [8]

$$\hat{H}(\mathbf{H}) = 2\mu_B[k(\hat{\mathbf{J}} \cdot \mathbf{H}) + q(\hat{J}_x^3 H_x + \hat{J}_y^3 H_y + \hat{J}_z^3 H_z)],$$

where **J** is the effective angular momentum of a hole ($J = 3/2$). We will use this expression in our subsequent analysis, although expressions (3) and (4) can also be used in principle.

The Hamiltonian of the Zeeman interaction in the split-off band and in the conduction band can also be expressed in terms of the corresponding effective spins $\mathbf{S}_{SO}$ and $\mathbf{S}_{CB}$. For a hole in the split-off band, we have

$$\hat{H}(\mathbf{H}) = \mu_B g_{SO}(\hat{\mathbf{S}}_{SO} \cdot \mathbf{H}), \quad (5)$$

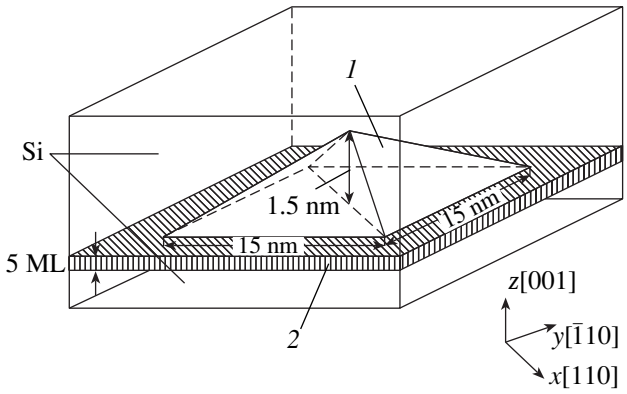


Fig. 1. Schematic diagram of a typical quantum dot (germanium island) in silicon: germanium island (quantum dot) (*1*), germanium film (wetting layer) (*2*), and monolayer (ML).

while for an electron in the conduction band, we have

$$\hat{H}(\mathbf{H}) = \mu_B g_{CB}(\hat{\mathbf{S}}_{CB} \cdot \mathbf{H}), \tag{6}$$

where g_{SO} is the g factor of a free hole in the split-off band, g_{CB} is the g factor in the conduction band, and effective spin operators $\hat{\mathbf{S}}_{SO}$ and $\hat{\mathbf{S}}_{CB}$ are defined in terms of the Pauli spin matrices $\hat{\sigma}_x$, $\hat{\sigma}_y$, and $\hat{\sigma}_z$, acting on the corresponding spin variables, $\hat{S}_\alpha = \hat{\sigma}_\alpha/2$.

The total energy of interaction with the magnetic field taking into account the orbital moment **L** is given by the sum

$$\begin{aligned}\hat{H}(\mathbf{H}) &= 2\mu_B[k(\hat{\mathbf{J}} \cdot \mathbf{H}) \\ &+ q(\hat{J}_x^3 H_x + \hat{J}_y^3 H_y + \hat{J}_z^3 H_z)] \\ &+ \mu_B g_{SO}(\hat{\mathbf{S}}_{SO} \cdot \mathbf{H}) + \mu_B g_{CB}(\hat{\mathbf{S}}_{CB} \cdot \mathbf{H}) + \mu_B(\hat{\mathbf{L}} \cdot \mathbf{H}),\end{aligned} \tag{7}$$

where $\hat{\mathbf{L}}$ is defined by formula (2). It follows hence that

$$\begin{aligned}(\hat{M}_{QD})_\alpha &= -[2k\hat{J}_\alpha + 2q\hat{J}_\alpha^3 \\ &+ g_{SO}(\hat{S}_{SO})_\alpha + g_{CB}(\hat{S}_{CB})_\alpha + \hat{L}_\alpha].\end{aligned} \tag{8}$$

The final formula for calculating the total magnetic moment has the form

$$\begin{aligned}(\hat{M}_{QD})_\alpha &= -\Big[2k\hat{J}_\alpha + 2q\hat{J}_\alpha^3 + g_{SO}(\hat{S}_{SO})_\alpha \\ &+ g_{CB}(\hat{S}_{CB})_\alpha + \frac{im}{\hbar^2} e_{\alpha\beta\gamma}\hat{R}_\beta \hat{H}_0 \hat{R}_\gamma\Big].\end{aligned} \tag{9}$$

Using now this expression, we can determine the matrix elements $\langle \psi|\hat{\mathbf{M}}_{QD}|\psi\rangle$ and $\langle \psi|\hat{\mathbf{M}}_{QD}|\psi^*\rangle$ and calculate the g factor by formula (1).

3. CALCULATION OF THE g FACTOR IN A Ge/Si SYSTEM WITH QUANTUM DOTS

Quantum dots in a Ge/Si system are formed during heteroepitaxy of germanium on a Si(100) substrate under certain conditions of transition from the 2D-layer mechanism of germanium film growth to the 3D growth. The typical size of islands in the familiar experimental studies varied from 10 to 20 nm, their height being 1–2 nm; consequently, the behavior of charge carriers in these islands is determined by quantum size effects [18]. The band offset existing in the Ge/Si heterosystem and deformation effects lead to the formation of a potential well in germanium for holes only. The states in a quantum dot are mainly formed from the states of the valence band, i.e., a superposition of states $|3/2, \pm 3/2\rangle$, $|3/2, \pm 1/2\rangle$, and $|1/2, \pm 1/2\rangle$ (state $|J, J_z\rangle$ is characterized by the angular momentum J and its component J_z along the z axis, viz., growth direction; Fig. 1). It follows from experimental results that a germanium island can be regarded as a square pyramid whose height h is an order of magnitude smaller than the base side l ($h/l \sim 1/10$) [19]. In fact, an island is a quasi-two-dimensional quantum object with a preferred symmetry axis z. The strain distribution in a quantum dot [14] removes the degeneracy existing at point Γ in the valence band. Since the crystal is subjected to uniaxial extension along the z axis within the island, states $|3/2, \pm 3/2\rangle$ of heavy holes are at the bottom of the valence band [20]. Consequently, we can expect that the contribution from heavy holes to the ground state in a quantum dot is predominant. The same conclusion can be drawn taking into account the fact that the effective mass of heavy holes is larger than that of light holes.

Let us consider the case when the magnetic field is parallel to the growth direction (**H** ∥ z). The energy of interaction with the field is determined by the magnetic moment component along the magnetic field, i.e., along z. In order to calculate the g factor, we must know the matrix elements of operators $\hat{J}_z$, $\hat{J}_z^3$, $(\hat{S}_{SO})_z$, $(\hat{S}_{CB})_z$, and $\hat{L}_z$.

Let us first demonstrate that g factor can be estimated only from the wave function expansion in the basis $|J, J_z\rangle$, i.e., in the basis $|3/2, \pm 3/2\rangle$, $|3/2, \pm 1/2\rangle$, and $|1/2, \pm 1/2\rangle$. We will disregard the effect of the conduction band on the hole states in the Ge/Si system with a quantum dot because the contribution from the states of this band to the wave function amounts to only about 0.5%.

The results of expansion of the wave function of the ground state in a quantum dot having a size of $l = 15$ nm and $h = 1.5$ nm are compiled in the table. The component with $J_z = \pm 3/2$ constitutes approximately 84% of the entire wave function. The remaining part corresponds to the component with $J_z = \pm 1/2$. It can be seen from the table that the state with the "up" spin, $|\uparrow\rangle$ (the state with the average angular momentum directed

Results of expansion of wave functions in the basis $|J, J_z\rangle$ for two spin sublevels $|\uparrow\rangle$ and $|\downarrow\rangle$ of the ground state in a germanium island of height $h = 1.5$ nm and base side $l = 15$ nm

$\|J, J_z\rangle$	$\left\|\frac{3}{2}, \frac{3}{2}\right\rangle$	$\left\|\frac{3}{2}, \frac{1}{2}\right\rangle$	$\left\|\frac{3}{2}, -\frac{1}{2}\right\rangle$	$\left\|\frac{3}{2}, -\frac{3}{2}\right\rangle$	$\left\|\frac{1}{2}, \frac{1}{2}\right\rangle$	$\left\|\frac{1}{2}, -\frac{1}{2}\right\rangle$
$\|\uparrow\rangle$	83.67%	2.26%	4.7%	0.08%	1.17%	8.11%
$\|\downarrow\rangle$	0.08%	4.7%	2.26%	83.67%	8.12%	1.17%

along the field) is mainly formed from components with $J_z = 3/2$ and $J_z = -1/2$, while the $|\downarrow\rangle$ state (the state with the average angular momentum opposite to the field) is formed by components with $J_z = -3/2$ and $J_z = 1/2$. The component with $J_z = \pm 1/2$ reflects contributions from states $|3/2, \pm 1/2\rangle$, $|1/2, \pm 1/2\rangle$, these contributions being almost identical and constituting about 8% each. This means that the formation of the ground state in a quantum dot is equally affected by the light hole subband and the split off subband.

If the ground state of a hole in a quantum dot were formed only by states with $J_z = \pm 3/2$, i.e., the $|\uparrow\rangle$ state corresponded to $J_z = 3/2$ and the $|\downarrow\rangle$ state to $J_z = -3/2$, the Zeeman splitting in a magnetic field $\mathbf{H} \parallel z$ would be determined by the expression

$$E(H_z) = 2\mu_B \langle (M_{QD})_z \rangle H_z = 2\mu_B H_z \left(2k \times \frac{3}{2} + 2q \times \frac{27}{8} + \langle L_z \rangle \right), \quad (10)$$

where $\langle (M_{QD})_z \rangle$ and $\langle L_z \rangle$ are the mean values of the z components of the magnetic and orbital moments in the $|\uparrow\rangle$ state.

In obtaining estimates, the term with q can be omitted in view of its smallness ($|q| = 0.06$) [21]. If we take into account the admixture of states with $J_z = \pm 1/2$, expression (10) is transformed to

$$E(H_z) = 2\mu_B H_z \left\{ 2k(a^2 - d^2) \times \frac{3}{2} + [2k(b^2 - c^2) + g_{SO}(e^2 - f^2)] \times \frac{1}{2} + \langle L_z \rangle \right\},$$

where a^2, b^2, c^2, and d^2 are the probabilities of values $J_z = 3/2$, $J_z = 1/2$, $J_z = -1/2$, and $J_z = -3/2$ (for $J = 3/2$) in the $|\uparrow\rangle$ state, respectively, and e^2 and f^2 are the probabilities of the values $J_z = 1/2$ and $J_z = -1/2$ (for $J = 1/2$) in the $|\uparrow\rangle$ state, respectively. For a quantum dot of size $l = 15$ nm and $h = 1.5$ nm, $a^2 \approx 0.84$, $b^2 \approx 0.02$, $c^2 \approx 0.05$, $d^2 \approx 0$, $e^2 \approx 0.01$, and $f^2 \approx 0.08$ (see table).

An estimate obtained disregarding the $\langle L_z \rangle$ term gives the following value for the g factor:

$$g_{zz} \approx 6k \times 0.82 - 2k \times 0.03 + g_{SO} \times 0.07,$$

here, $k = -3.41 \pm 0.03$ [21] and $g_{SO} = -10 \pm 3$ [22], which gives $|g_{zz}| \approx 15.86$.

Calculating the g factor by formula (1) for the same island size taking into account the orbital moment $\langle L_z \rangle$ and using the wave functions determined in the tight binding approximation, we obtain $|g_{zz}| = 15.71$.

A comparison with the bulk value of the longitudinal g factor for a heavy hole, $|g_{HH}| \approx 6k = 20.46$, shows that size quantization reduces the g factor, indicating the suppression of the spin-orbit interaction due to an admixture of the state with a smaller value of J ($J = 1/2$) and a decrease in the effective angular momentum of the particle.

In order to estimate the contribution of the orbital moment, we calculated the g factor using formulas (1) and (8), omitting in Eq. (8) all the terms except $\hat{L}_\alpha$. As a result, we obtained an order-of-magnitude smaller value of g factor: $|g_{zz}| = 0.59$. Thus, the g factor is mainly determined by the effective angular momentum **J** rather than by the orbital moment **L**.

We will give here the principal values of the g factor for the ground state in a quantum dot of size $l = 15$ nm and $h = 1.5$ nm, calculated by formula (1): $|g_{zz}| = 15.71$ (in the growth direction [001]), $|g_{xx}| = 1.14$ (in the [110] direction), and $|g_{yy}| = 1.76$ (in the $[\bar{1}\,10]$ direction).

3.1. Dependence of g Factor on the Size of the Germanium Island

The results obtained clearly demonstrate anisotropy in the values of the g factor: g_{zz} is an order of magnitude larger than the values of g_{xx} and g_{yy}. Anisotropy increases upon an increase in the island base for a constant height (Fig. 2). This tendency can be explained by the fact that the wave function of the ground state is close in composition to the wave function $|3/2, \pm 3/2\rangle$ (to a heavy hole state), in which the transverse components of the g factor are close to zero [6].

We can assume that the wave function of the ground state becomes closer and closer to the wave function of a heavy hole as the size of the island increases, which enhances the anisotropy of the g factor.

Indeed, according to the results of our calculations, the contribution of the state with $J_z = \pm 3/2$ to the wave function of a hole increases with the lateral dimension

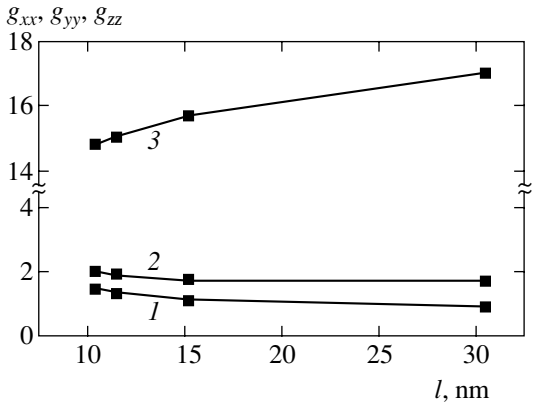


Fig. 2. Dependence of the *g* factor of the ground state of a hole on the lateral size *l* of a germanium island of height h = 1.5 nm: g_{xx} (*1*), g_{yy} (*2*), and g_{zz} (*3*).

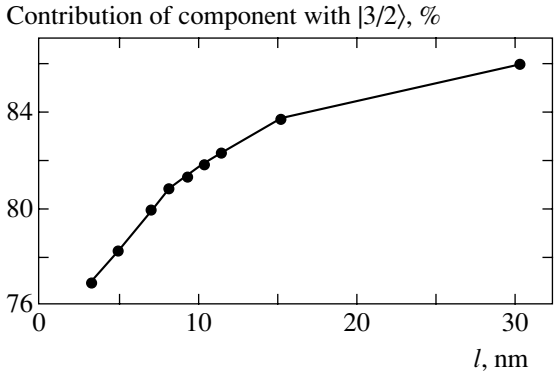


Fig. 3. Contribution of the component with $J_z = \pm 3/2$ to the wave function of the ground state of a hole as a function of the lateral size *l* of a germanium island of height h = 1.5 nm.

of the island, and the wave function tends to "pure" state $|3/2, \pm 3/2\rangle$ (Fig. 3). For example, as the lateral size *l* of the island increases from 15 to 30 nm for a height of h = 1.5 nm, the contribution from the component with $J_z = \pm 3/2$ to the wave function of the ground state increases from 83 to 86%. Anisotropy of the *g* factor increases: the value of $|g_{zz}|$ increases to 17.0, while the transverse components decrease to $|g_{xx}| = 0.91$ and $|g_{yy}| = 1.71$. If we increase the island size to that where the pyramid proportion is preserved (h/l = 1/10), the anisotropy of the *g* factor becomes stronger. For example, for l = 30 nm and h = 3 nm, the principal values of the *g* factor are as follows: $|g_{zz}| = 20.99$, $|g_{xx}| = 0.06$, and $|g_{yy}| = 1.1$. It turns out that the contribution of the component with $J_z = \pm 3/2$ to the wave function of the ground state in this case increases to 90%, leading to such a strong anisotropy.

The obtained dependence of the *g* factor on the island size indicates the correctness of our approach. Indeed, as the lateral size of the island increases, we pass to the limiting case of a pseudomorphic stressed germanium film. Inhomogeneity of strains typical of quantum dots disappears. The uncertainty in k_x and k_y for states at the bottom of the band becomes equal to zero. All these factors suppress the interaction of the HH band with other energy bands; as a result, the *g* factor of the ground state tends to the *g* factor of a heavy hole, which is in accordance with our results.

3.2. Sharpness of Germanium Island Boundaries

The above values of the *g* factor were obtained for an island with sharp boundaries. If we take into account diffuse boundaries of the island, which is always observed in a real Ge island, these values will slightly change.

The diffuse boundaries was taken into account as follows: each atom in the crystal lattice was replaced, with a probability of 2/3, by one of its four nearest neighbors. As a result, we obtained a boundary with a smooth variation in the composition of the substance within three monolayers.

Taking into account the diffuse island boundaries, we found that the transverse components of the *g* factor change significantly (e.g., the value of $|g_{xx}|$ decreases from 1.14 to 0.52 for an island with a lateral size of l = 15 nm, while the value of $|g_{yy}|$ decreases from 1.76 to 0.18. The longitudinal component $|g_{zz}|$ of the *g* factor virtually does not change and amounts to 15.81. Consequently, anisotropy of the *g* factor increases on account of blurring of the heteroboundary. This is probably associated with an effective increase in the island size.

3.3. Probability of Zeeman Transitions

The probability of Zeeman transitions is directly connected with the form of the wave function. For a state with $J_z = \pm 3/2$ in a magnetic field $\mathbf{H} \parallel z$, induced transitions between the Zeeman sublevels with $J_z = 3/2$ and $J_z = -3/2$ are forbidden by the selection rules, since allowed transitions must satisfy the condition $\Delta J_z = \pm 1$. An admixture of a state with $J_z = \pm 1/2$ facilitates transitions between the Zeeman sublevels of the ground state in a germanium island. Consequently, the prohibition imposed on Zeeman transition is released upon an increase in the island size.

For an arbitrary direction **h** of the magnetic field, the energy of interaction with the field is determined by the angular momentum component along **h**. States $|J, J_z\rangle$ are transformed into states $|J, J_h\rangle$ as follows:

$$|J, J_z\rangle \longrightarrow |J, J_h\rangle = \sum_{J_z} R^J_{J_z J_h}(\theta, \varphi)|J, J_z\rangle;$$

here, θ and φ are polar angles of vector **h** in the system of coordinates *x, y, z*, and matrix *R* is connected to the standard matrix of rotations [23] through the variation

$$R^J_{J_z J_h}(\theta, \varphi) = D^J_{J_z J_h}(0, -\theta, -\varphi).$$

In the particular case when $\theta = \pi/2$ and $\varphi = 0$, the magnetic field direction lies in the plane of the island and coincides with the x axis. Let us consider a pure state with $J_z = 3/2$. In the $|J, J_z\rangle$ representation, the wave function of this state can be written in the form

$$|\psi\rangle = a\left|\frac{3}{2}, \frac{3}{2}\right\rangle + b\left|\frac{3}{2}, \frac{1}{2}\right\rangle + c\left|\frac{3}{2}, -\frac{1}{2}\right\rangle + d\left|\frac{3}{2}, -\frac{3}{2}\right\rangle = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

where the squares of the coefficients (a^2, b^2, c^2, and d^2) reflect contributions from the states with corresponding values of J_z ($a^2 + b^2 + c^2 + d^2 = 1$). Under transformation $R^J_{J_z J_h}(\pi/2, 0)$, state $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ is transformed into $\begin{pmatrix} \sqrt{1/8} \\ \sqrt{3/8} \\ \sqrt{3/8} \\ \sqrt{1/8} \end{pmatrix}$. It follows hence that the fraction of each component with $J_h = \pm 1/2$ amounts to 3/8 of the entire wave function. On the whole, they constitute 75%; i.e., the contributions from components with $J_h = \pm 1/2$ for the direction of magnetic field **H** in the basal plane of the area larger as compared to the case when **H** $\parallel z$, and the probability of Zeeman transitions increases. This is also observed in the case when the wave function initially contains an admixture of state with $J_z = \pm 1/2$ as, for example, for the ground state in the quantum dot in question, where it amounts to 16%.

Let us consider some numerical estimates of the probabilities of Zeeman transitions for different directions of the magnetic field.

The probability of an induced transition between Zeeman sublevels is determined by the interaction of the magnetic moment with oscillating magnetic microwave field $H_\perp \cos(2\pi\nu t)$ (field $H_\perp$ is perpendicular to the constant magnetic field) and is proportional to the squared matrix element of the magnetic moment component $\mu_\perp$ of a particle in the direction of this field [24],

$$P_{\uparrow\downarrow} \propto |\langle\downarrow|\hat{\mu}_\perp H_\perp|\uparrow\rangle|^2.$$

If the magnetic field direction is such that **H** $\parallel z$, the magnetic moment component $\mu_\perp$ lies in the island basal plane and is proportional to the principal values of the g tensor, g_{xx} (direction [110]) and g_{yy} (direction $[\bar{1}10]$). In the particular case when the microwave field $H_\perp$ is directed along [110], the transition probability $P_{\uparrow\downarrow}$ is proportional g^2_{xx}.

If the direction of the constant magnetic field is such that **H** $\perp z$, the magnetic moment component lies in the plane perpendicular to the basal plane; in the particular case when the microwave field $H_\perp$ is directed along [100], this component is proportional to the principal value of the g tensor: $\mu_\perp \propto g_{zz}$. In this case, the transition probability is $P_{\uparrow\downarrow} \propto g^2_{zz}$.

It can be seen that, for $g_{zz} = 15.71$, $g_{xx} = 1.14$, and $g_{yy} = 1.76$, the probabilities of induced transitions for two directions of the magnetic field (**H** $\parallel z$ and **H** $\perp z$) differ approximately by two orders of magnitude. If, in addition, we take into account a decrease in the transverse components of the g factor due to diffuse boundaries ($g_{xx} = 0.52$ and $g_{yy} = 0.18$), the difference in the transition probabilities for **H** $\parallel z$ and **H** $\perp z$ will be more than three orders of magnitude.

4. DISCUSSION

The results obtained show that the main factor determining the dependence of the g factor on the size of an island is the change in the contribution to the wave function of a hole from the component with $J_z = \pm 3/2$ upon a change in the island size.

The factors determining the relation between the contributions from the components with $J_z = \pm 3/2$ and $\pm 1/2$ can be grasped from the following simplified model disregarding the interaction between energy bands. Let us consider separately the quantization of the energy spectra of holes with $J_z = \pm 3/2$ and $J_z = \pm 1/2$. In such a model, the deepest energy levels belong to holes with $J_z = \pm 3/2$; in the range of excited states, the levels of holes both with $J_z = \pm 1/2$ and $J_z = \pm 3/2$ are present. In a more realistic model taking into account the interaction between energy bands (e.g., the six- or eight-band **kp** model or the tight binding model), the energy range corresponding to excited states contains some "mixed" states with comparable contributions from both types of holes, while the range corresponding to the deepest layers contains states formed mainly by holes with $J_z = \pm 3/2$. Such a qualitative model is in agreement with the results of our calculations.

Figure 4 shows the contribution from the component with the momentum component $J_z = \pm 3/2$ to the wave functions of states in a quantum dot of size $l = 15$ nm and $h = 1.5$ nm. The results of expansion show that the component with $J_z = \pm 3/2$ constitutes approximately 84% of the wave function of the ground state ($E_0 = 420$ meV). For the first excited state ($E_1 = 377$ meV), the contribution from the component with $J_z = \pm 3/2$ decreases approximately to 79%. As the number of the excited state increases, a tendency towards a decrease in the component with $J_z = \pm 3/2$ is observed. For the

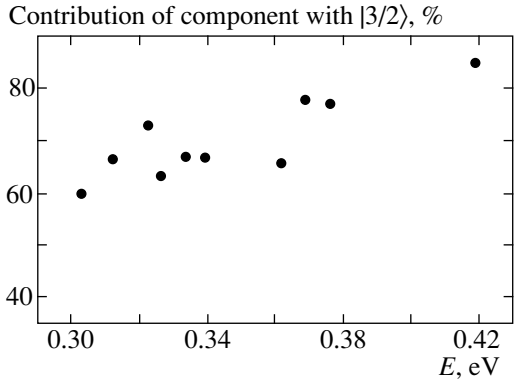


Fig. 4. Contribution of the component with $J_z = \pm 3/2$ to the states of the discrete spectrum of a germanium island. The energy of the state measured from the edge of the valence band of silicon is laid along the abscissa axis. The island size: height $h = 1.5$ nm and base side $l = 15$ nm.

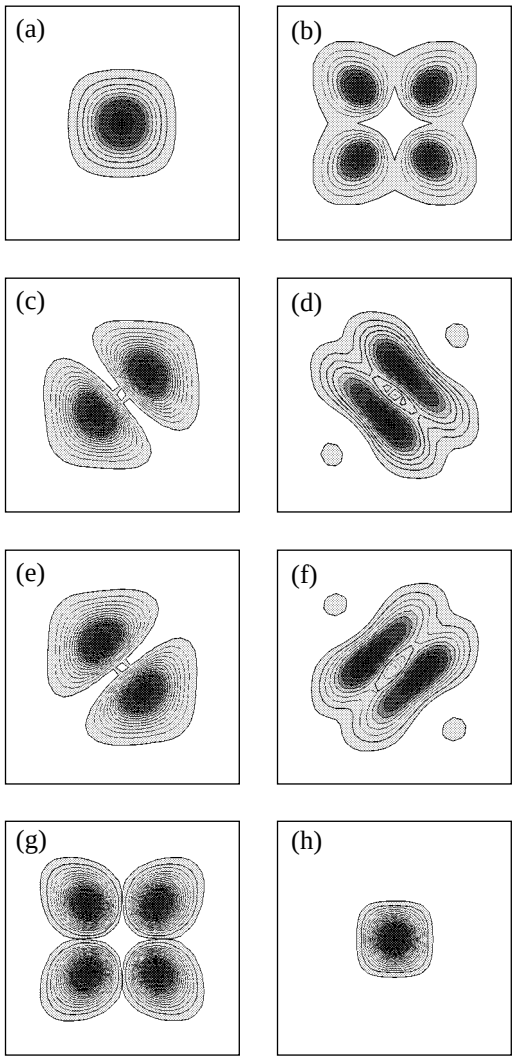


Fig. 5. Wave function density distribution for the component with $J_z = \pm 3/2$ (a, c, e, and g) and $J_z = \pm 1/2$ (b, d, f, and h) (projection on the plane of the pyramid base) in the ground (a, b), first excited (c, d), second excited (e, f), and third excited (g, h) states of a hole in a quantum dot.

ninth excited state ($E_9 = 303$ meV), the contribution from the component with $J_z = \pm 3/2$ amounts approximately to 60%.

The form of the wave function of the ground state is determined by the separation between the ground energy level and the states of holes with $J_z = \pm 1/2$, which lie in the range of excited states in a quantum dot.

The data presented in Fig. 4 show that the dependence of the contribution from the component with $J_z = \pm 3/2$ on the hole energy cannot be described by a smooth function. In order to explain the step form of this dependence, we analyzed the form of wave functions separately for the components with $J_z = \pm 3/2$ and $\pm 1/2$. Figure 5 shows the wave functions for the components with $J_z = \pm 3/2$ and $\pm 1/2$, forming the first four states in the quantum dot. In the ground state, the component with $J_z = \pm 3/2$, which constitutes 84%, is *s*-shaped. However, the component with $J_z = \pm 1/2$, which is admixed on account of interaction between energy bands and constitutes approximately 16%, is *d*-shaped. If we return to the simplified model and consider separately the quantization of the spectra of holes with $J_z = \pm 3/2$ and $\pm 1/2$, each of these two spectra contain *s*-, *p*-, and *d*-shaped states, etc. (Fig. 6). Naturally, the *s*-shaped state will be the lowest state in both spectra, followed by *p*- and *d*-shaped states, etc. The extent of admixture of a state with $J_z = \pm 1/2$ to a state with $J_z = \pm 3/2$ is inversely proportional to the difference in the energies of these states, i.e., to $(E_{\pm 3/2} - E_{\pm 1/2})^{-1}$. It follows from the data presented in Fig. 5 that the *s* state from the spectrum of a hole with $J_z = \pm 3/2$ interacts with the *d* state from the spectrum of a hole with $J_z = \pm 1/2$ and forms the ground energy level. In this case, the contribution from the component with $J_z = \pm 1/2$ is determined by the energy gap $\Delta E_0 = E^s_{\pm 3/2} - E^d_{\pm 1/2}$. The first and second excited states are formed by mixing the *p* state from both spectra. In such cases, the energy gaps $\Delta E_1 = \Delta E_2 = E^p_{\pm 3/2} - E^p_{\pm 1/2}$ coincide; consequently, the contributions from the component with $J_z = \pm 1/2$ are practically identical. The third excited state is formed by mixing the *d* state from the spectrum of a hole with $J_z = \pm 3/2$ and the *s* state from the spectrum of a hole with $J_z = \pm 1/2$. In this case, the contribution from the component with $J_z = \pm 1/2$ is determined by the energy gap $\Delta E_3 = E^d_{\pm 3/2} - E^s_{\pm 1/2}$; i.e., the distance between interacting energy levels is smaller, and the contribution from the component with $J_z = \pm 1/2$ is greater than for deeper levels. Thus, we can establish a relation between the energy gaps in all four cases ($\Delta E_0 > \Delta E_1$, $\Delta E_1 = \Delta E_2$, and $\Delta E_2 > \Delta E_3$) and explain the dependence of the contribution of the component with $J_z = \pm 1/2$ on

the number of the state in a discrete spectrum in a quantum dot.

The interpretation for the next energy levels is complicated on account of the fact that the wave functions of these states have a complex form and cannot be classified as *s*-, *p*-, … shaped states.

Let us now consider the main effects determining the variation of the relation between the contributions of components with $J_z = \pm 1/2$ and $\pm 3/2$ upon the variation of the island size.

If the size of an island increases so that the proportions between dimensions is preserved ($h/l = 1/10$), the distribution and magnitude of strains in the island does not change significantly; consequently, the splitting between the HH and LH bands remains unchanged. The size quantization energy decreases and amounts, for example, to a few millielectronvolts for $l = 100$ nm and $h = 10$ nm. As a result, the ground state in the spectrum of a hole with $J_z = \pm 3/2$ shifts towards the bottom of the potential well. Excited states are less sensitive to a change in the island size since their localization radius is larger, and the wave functions penetrate more strongly under the barrier (into silicon surrounding the germanium island). Consequently, the position of energy levels for holes with $J_z = \pm 1/2$ changes more weakly than for holes with $J_z = \pm 3/2$ upon an increase in the island size. For this reason, the energy gap ΔE_0 between the ground *s* state of a hole with $J_z = \pm 3/2$ and the *d* state of a hole with $J_z = \pm 1/2$ increases. Accordingly, the contribution from the component with $J_z = \pm 1/2$ to the ground state of the hole decreases, and the wave function becomes closer to the state of a hole with $J_z = \pm 3/2$. In this case, the main effect determining the change in the relation between the components with $J_z = \pm 3/2$ and $\pm 1/2$ is a confinement effect (decrease in the size quantization energy).

If only the lateral size of the island increases, the size quantization energy remains practically unchanged since it is mainly determined by the height of the island. However, an increase in the l/h ratio leads to an increase in biaxial strain $\varepsilon_{zz} - (\varepsilon_{xx} + \varepsilon_{yy})/2$ [25], leading to an increase in the splitting between the HH and LH subbands. In this case, the reason for the increasing energy gap ΔE_0 is a strain effect (change in strains in the island).

The experimental value of the *g* factor for a hole in quantum dots is usually a result of indirect measurements. As a rule, the photoluminescence spectrum in a magnetic field is analyzed [26–28], and the *g* factor of a hole is calculated from the experimentally determined *g* factor of an exciton (g_{ex}) and the *g* factor of an electron (g_e) using the relation $g_{ex} = g_{HH} \pm g_e$ (the minus sign is used for bright excitons, and the plus sign, for dark excitons). In order to eliminate an additional systematic error associated with the existence of exchange interaction between the electron and the hole in an exciton, a single hole in a quantum dot must be analyzed instead of a hole bound with an electron in an exciton complex. For this purpose, an experiment involving tunneling of a hole through a quantum dot in a magnetic field, similar to that proposed in [29] for an electron, can be made. In this case, it is important to appropriately choose the magnetic field direction, since it determines the Zeeman splitting and the intensity of the Zeeman transitions. For the $\mathbf{H} \parallel z$ direction, Zeeman transitions are forbidden. For $\mathbf{H} \perp z$, the Zeeman splitting is small. For this reason, it is expedient to carry out experiments in a tilted magnetic field, when the Zeeman splitting is strong enough and the intensity of induced transitions is appreciable.

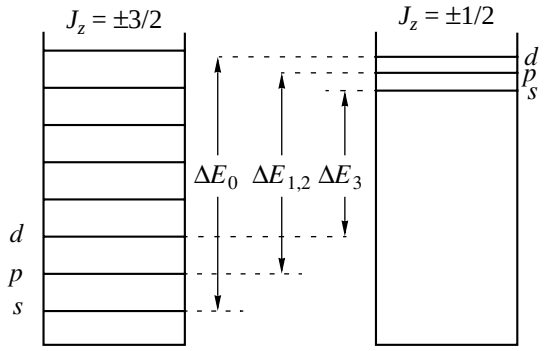


Fig. 6. Schematic diagram of the energy spectra of holes with $J_z = \pm 3/2$ and $\pm 1/2$ in the simplified model of noninteracting energy bands. The spectra are separated in space. Letters *s*, *p*, and *d* mark the energy levels corresponding to *s*-, *p*- and *d*-shaped states.

The method of calculating the *g* factor in quantum dots proposed here will make it possible to compare consistently the theory and available experimental data on *g* factors of hole (electron) states in quantum dots grown in various heterosystems, since its applicability is not limited to quantum dots in the Ge/Si system.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 02-02-16020 and 99-02-39051GFEN), the program Russian Universities (grant no. UR.01.01.019), and INTAS (grant no. 2001-0615).

REFERENCES

1. M. Bayer, V. B. Timofeev, T. Gutbrod, *et al.*, Phys. Rev. B **52**, R11623 (1995).
2. V. K. Kalevich, B. P. Zakharchenya, and O. M. Fedorova, Fiz. Tverd. Tela (St. Petersburg) **37**, 283 (1995) [Phys. Solid State **37**, 154 (1995)].
3. E. L. Ivchenko and A. A. Kiselev, Fiz. Tekh. Poluprovodn. (St. Petersburg) **26**, 1471 (1992) [Sov. Phys. Semicond. **26**, 827 (1992)].
4. A. A. Kiselev and E. L. Ivchenko, Phys. Rev. B **58**, 16353 (1998).

5. A. A. Kiselev and L. V. Moiseev, Fiz. Tverd. Tela (St. Petersburg) **38**, 1574 (1996) [Phys. Solid State **38**, 866 (1996)].
6. X. Marie, T. Amand, P. Jeune, *et al.*, Phys. Rev. B **60**, 5811 (1999).
7. A. A. Kiselev and K. W. Kim, Phys. Rev. B **64**, 125303 (2001).
8. G. L. Bir and G. E. Pikus, *Symmetry and Stain-Induced Effects in Semiconductors* (Nauka, Moscow, 1972; Wiley, New York, 1975), pp. 303, 325, 393.
9. A. V. Dvurechenskii, A. V. Nenashev, and A. I. Yakimov, Nanotechnology **13**, 75 (2002).
10. A. Abragam and B. Bleaney, *Electron Paramagnetic Resonance of Transition Ions* (Clarendon, Oxford, 1970; Mir, Moscow, 1972), Vol. 1.
11. D. J. Chadi and M. L. Cohen, Phys. Status Solidi B **68**, 405 (1975).
12. J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).
13. D. J. Chadi, Phys. Rev. B **16**, 790 (1977).
14. A. V. Nenashev and A. V. Dvurechenskiĭ, Zh. Éksp. Teor. Fiz. **118**, 570 (2000) [JETP **91**, 497 (2000)].
15. J.-M. Jancu, R. Sholz, F. Beltram, and F. Bassani, Phys. Rev. B **57**, 6493 (1998).
16. A. A. Kiselev and U. Rossler, Phys. Rev. B **50**, 14283 (1994).
17. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 4th ed. (Pergamon, New York, 1977; Nauka, Moscow, 1989).
18. A. V. Dvurechenskiĭ, A. I. Yakimov, A. V. Markov, *et al.*, Izv. Ross. Akad. Nauk, Ser. Fiz. **64**, 288 (1998).
19. A. I. Yakimov, A. V. Dvurechenskii, Yu. Yu. Proskuryakov, *et al.*, Appl. Phys. Lett. **75**, 1413 (1999).
20. C. G. Van de Walle, Phys. Rev. B **39**, 1871 (1989).
21. J. C. Hensel and K. Suzuki, Phys. Rev. Lett. **22**, 838 (1969).
22. R. L. Aggarwal, Phys. Rev. B **2**, 446 (1970).
23. V. F. Gantmakher and I. B. Levinson, *Scattering of Carriers in Metals and Semiconductors* (Nauka, Moscow, 1984), p. 37.
24. S. A. Al'tshuler and B. M. Kozyrev, *Electron Paramagnetic Resonance* (Fizmatgiz, Moscow, 1961; Academic, New York, 1964).
25. M. A. Cusack, P. R. Briddon, and M. Jaros, Phys. Rev. B **56**, 4047 (1997).
26. M. Bayer, A. Kuther, F. Forchell, *et al.*, Phys. Rev. Lett. **82**, 1748 (1999).
27. A. Zrenner, M. Markmann, E. Beham, *et al.*, J. Electron. Mater. **28**, 542 (1999).
28. M. Bayer, O. Stern, A. Kuther, and A. Forchel, Phys. Rev. B **61**, 7273 (2000).
29. H.-A. Engel and D. Loss, Phys. Rev. Lett. **86**, 4648 (2001).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 331–334.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 373–377.

SOLIDS
Electronic Properties

The Effect of Colossal Magnetoresistance in $SmCu_3Mn_4O_{12}$ Perovskite-Like Oxide

L. S. Lobanovsky* and I. O. Troyanchuk

Institute of Solid-State and Semiconductor Physics, National Academy of Sciences of Belarus, Minsk, 220072 Belarus

**e-mail: Lobanov@ifttp.bas-net.by*

Received August 27, 2002

Abstract—The magnetoresistance effect in the $SmCu_3Mn_4O_{12}$ compound with a perovskite-like structure is investigated for the first time. It is found that an intragranular magnetoresistance effect shows up in $SmCu_3Mn_4O_{12}$ in the vicinity of the magnetic ordering temperature (T_C = 376 K). In addition to this colossal magnetoresistance, a giant magnetoresistance effect is also observed, which is associated with the scattering of spin-polarized charge carriers by the granule boundaries. The giant magnetoresistance is comparable in magnitude to the effect revealed in substituted lanthanum orthomanganites. However, the coexistence of intragranular and intergranular magnetoresistance in manganese-containing perovskites was previously observed for the concentration of tetravalent manganese ions in the range from 15 to 45%. The concentration of Mn^{4+} ions in $SmCu_3Mn_4O_{12}$ is 75%. © *2003 MAIK "Nauka/Interperiodica".*

The discovery of the effect of colossal magnetoresistance (CMR) in manganites with a perovskite structure stimulated the search for new magnetic materials exhibiting this effect. Such materials are used in devices for information storage and processing and in radioelectronic equipment for conversion of magnetic to electric signals. However, a decrease in resistance by several orders of magnitude occurs in manganites in strong magnetic fields (over 2 T) and is a property of granules [1–3]. In order to attain significant magnetoresistive properties in weak external magnetic fields, numerous researchers investigated the effect of dielectric barrier tunneling of charge carriers [4–6]. The Curie temperature of different compounds did not exceed 150 K. It was recently revealed that A_2FeMoO_6 compounds (A = Ca, Sr, Ba) with a perovskite structure, while being ferrimagnets, demonstrated a high magnetoresistance effect associated with tunneling of charge carriers through the intergranular interlayer [7–9]. Ba_2FeMoO_6 exhibited both intergranular (far below the magnetic ordering temperature) and intragranular (in the vicinity of T_C) magnetoresistive properties. In interpreting the presence of intragranular magnetoresistance in these compounds, Maignan *et al.* [8] proceeded from the binary exchange interactions and high degree of spin polarization of the charge carriers.

The intergranular magnetoresistance was also revealed previously in ferrimagnetic polycrystalline $CaCu_3Mn_4O_{12}$ [10] containing no ions of trivalent manganese. As is known [11, 12], $AC_3B_4O_{12}$ compounds crystallize into a perovskite-like structure with a double unit cell parameter. The unit cell parameter in these compounds is doubled as a result of ordering of A ions and C ions in the 1/3 ratio and because of the inclination of oxygen octahedrons relative to the [100] crystallographic directions. However, no peak of intragranular magnetoresistance was observed in the vicinity of the Curie temperature (T_C = 350 K) of $CaCu_3Mn_4O_{12}$. This behavior is probably associated with the absence of manganese ions of different valences and, as a consequence, with the absence of binary exchange interactions used in describing the effect of colossal magnetoresistance.

It is of interest to study the magnetoresistive properties of compounds from the $ACu_3Mn_4O_{12}$ series, where the A position is taken by an ion of suitable size with a valence of other than two. During the introduction of ions of different valences into the A position, the charge is compensated by way of forming a mixed valence of manganese ions in the B sublattice. With a trivalent A ion, the ratio of manganese ions of different valences in the B position will be Mn^{3+}/Mn^{4+} = 1/3. In connection with this, we measured the magnetoresistive properties of $Sm^{3+}Cu_3Mn_4O_{12}$. In contrast to $CaCu_3Mn_4O_{12}$, the samarium-based compound exhibits a clearly defined magnetoresistance effect in two temperature ranges: a magnetoresistance characteristic of scattering of charge carriers by the boundaries of neighboring granules was observed at temperatures from 80 to 350 K; a peak of magnetoresistance associated with the scattering of charge carriers by magnetic nonuniformities within granules was observed at temperatures close to T_C = 376 K.

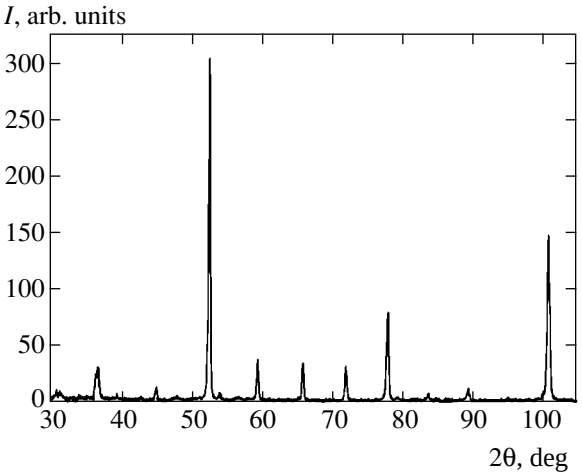


Fig. 1. An X-ray pattern of $SmCu_3Mn_4O_{12}$, taken at room temperature.

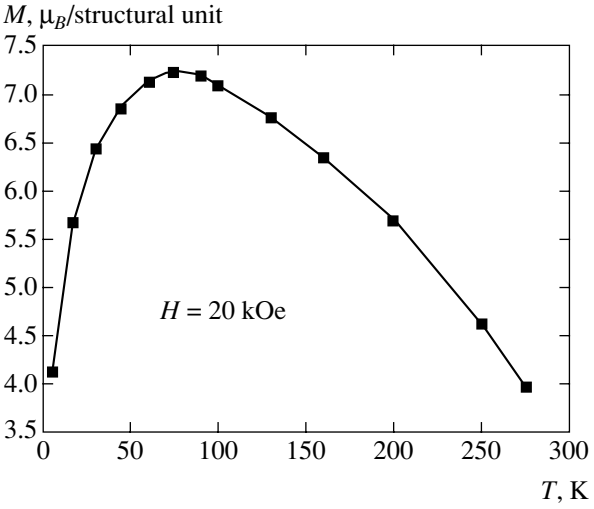


Fig. 2. Temperature dependence of magnetization in $SmCu_3Mn_4O_{12}$, obtained in an external magnetic field.

$ACu_3Mn_4O_{12}$ compounds are characterized by a high temperature of magnetic ordering [12] and are potential candidates for use as magnetoresistive materials operating at room temperature.

A $SmCu_3Mn_4O_{12}$ sample was prepared from oxides of the respective elements by the method of solid-phase reactions with simultaneous exposure to high pressure and temperature. In order to preclude the reduction reaction, the sample in the process of synthesis was insulated from the heater by nickel foil. Synthesis was followed by X-ray diffraction analysis of the prepared compound in K_α radiation of Cr using a DRON-3 diffractometer with a view to determining the degree to which the sample is monophase and to refining the unit cell parameters. The magnetization was measured at different temperatures using a vibrating-sample magnetometer. The data of dynamic magnetic susceptibility were obtained using a mutual induction bridge. The standard four-point method was used to measure the electrical properties and magnetoresistance. The magnetoresistance effect was measured in an external magnetic field of 0.9 T and determined as $MR = \Delta\rho/\rho = 100\%(\rho_{H=0} - \rho_H)/\rho_{H=0}$.

The resultant $SmCu_3Mn_4O_{12}$ compound was characterized by the cubic symmetry of unit cell with an ABO_3 perovskite structure with double parameter a equal to 7.296 Å (Fig. 1). The calculation and refinement of the unit cell parameters were performed in the $Im3$ spatial group (previously identified by Bochu *et al.* [12]) for $CaCu_3Mn_4O_{12}$.

Measurements of the dynamic magnetic susceptibility and magnetization revealed that $SmCu_3Mn_4O_{12}$ exhibited a spontaneous magnetic moment below T_C = 376 K. The temperature dependence of magnetization had a maximum at 75 K. As previously found, $ThCu_3Mn_4O_{12}$ is a ferrimagnet with opposite ordering of magnetic moments of copper and manganese ions [13]. In view of the data of measurements of spontaneous magnetization, one should expect a similar ordering of magnetic moments of ions for $SmCu_3Mn_4O_{12}$ as well. The magnetization peak in the compound investigated by us is apparently associated with the ordering of magnetic moments of samarium ions in the direction opposite to the overall magnetic moment of copper and manganese ions, which leads to a decrease in the magnetization at temperatures below 75 K (Fig. 2). The spontaneous magnetic moment measured at 5 K is 4.15 Bohr magnetons per structural unit. This value corresponds to the assumption of antiparallel ordering of overall magnetic moments of copper and samarium ions in the A and C sublattices and of manganese ions in the B sublattice of an $AC_3B_4O_{12}$ structure.

More striking results, however, were obtained in measuring the electrical properties and magnetoresistance (Fig. 3). The electrical resistance was measured in the temperature range from 77 to 425 K. Two intervals with conductivity of different types may be identified on the curve of temperature dependence of electrical resistance. Below 300 K, the resistance of $SmCu_3Mn_4O_{12}$ increases with temperature, while above that temperature, the compound exhibits a semiconductor dependence of the behavior of electrical resistance on temperature. Moreover, an anomaly is observed on the temperature curve of electrical resistance in the vicinity of T_C.

It is known [14] that substituted lanthanum orthomanganites in the magnetically ordered phase exhibit conductivity of metallic type with the concentration of Mn^{4+} ions ranging from 15 to 45%. Given this content of Mn^{4+} ions, the metal–dielectric transition is observed in the vicinity of the magnetic ordering temperature. Above T_C, these compounds are semiconductors. It was recently demonstrated [15] that, in the case of $La_{0.82}Sr_{0.18}MnO_z$ ($z \approx 3$), the electrical properties strongly depend on the conditions of synthesis. At a moder-

ate temperature of synthesis, the intergranular interlayer of a polycrystalline sample of $La_{0.82}Sr_{0.18}MnO_z$ may have a considerable effect on the electrical conductivity. In this case, the Curie temperature of the intergranular interlayer decreases due to the weakening of exchange interactions on the granule boundary, and, as a result, the temperature of the metal–insulator transition shifts toward lower temperatures. However, the contribution by granules to the electrical properties shows up as a peak on the resistance curve in the vicinity of T_C. The temperature dependence of the electrical resistance of such compounds is a curve with two peaks.

Apparently, the intergranular layer similarly affects the transport properties in polycrystalline $SmCu_3Mn_4O_{12}$. Note, however, that the conductivity of the metallic type below T_C is observed in manganites at a concentration of tetravalent manganese ions ranging from 15 to 45%. As was mentioned above, the concentration of Mn^{4+} ions in $SmCu_3Mn_4O_{12}$ is 75%. In addition, as distinct from lanthanum manganites with a bond angle close to 180°, the Mn–O–Mn angle in $AC_3B_4O_{12}$ structures is about 140°. It is customary to assume that the foregoing facts cause a narrowing of the conduction band in manganites. However, the order of magnitude of the resistance of solid solution of $SmCu_3Mn_4O_{12}$ investigated by us is close to that of $La_{0.82}Sr_{0.18}MnO_z$ ($z \approx 3$). Possibly, this low value of electrical resistance of $SmCu_3Mn_4O_{12}$ is associated with the conditions of synthesis, because samples prepared by the method of solid-phase reactions under high pressure are characterized by lower porosity.

The temperature dependence of magnetoresistance of $SmCu_3Mn_4O_{12}$ is more complex than the previously observed effect in Fe_3O_4 and CrO_2 [16, 17]. At temperatures much lower than the magnetic ordering temperature, the magnetoresistance monotonically decreases as the temperature increases. A similar temperature dependence is predicted by theory [18]. However, in the vicinity of T_C, the magnetoresistance exhibits a peak of up to 2.5% in a field of 0.9 T, which contradicts the theoretically expected behavior. Similar results were obtained for polycrystalline ferrimagnetic compounds of the Ba_2FeMoO_6 series [8].

In order to clarify in more detail the nature of this phenomenon, we measured the field dependences of the magnetoresistance effect at different temperatures. One can see (Fig. 4) that, at low temperatures, the magnetoresistance abruptly increases in weak magnetic fields and is gradually saturated as the external magnetic field increases. A different behavior is exhibited by the high-temperature part of the magnetoresistance effect. The magnetoresistance monotonically increases with the external magnetic field. The observed features of the behavior of resistance in the external magnetic field at different temperatures indicate that these effects are of different nature. One can infer that the magnetoresistance effect consists of two contributions. At low temperatures, the behavior of magnetoresistance reminds one of the effect revealed by Kobayashi *et al.* [7] in Sr_2FeMoO_6 double oxide with a perovskite structure. Kobayashi *et al.* [7] observed that the magnetoresistance of this kind shows up as a result of the effect of the granule boundaries on the electrical properties of compounds under the influence of external magnetic field. The abrupt disappearance of the magnetoresistance effect above T_C is indicative of the effect of the magnetic order on the magnetoresistance. In doing so, a correlation in the behavior of magnetoresistance and magnetization is observed in an external magnetic field both at low temperatures and in the vicinity of the magnetic ordering temperature (Fig. 4, inset). Maignan *et*

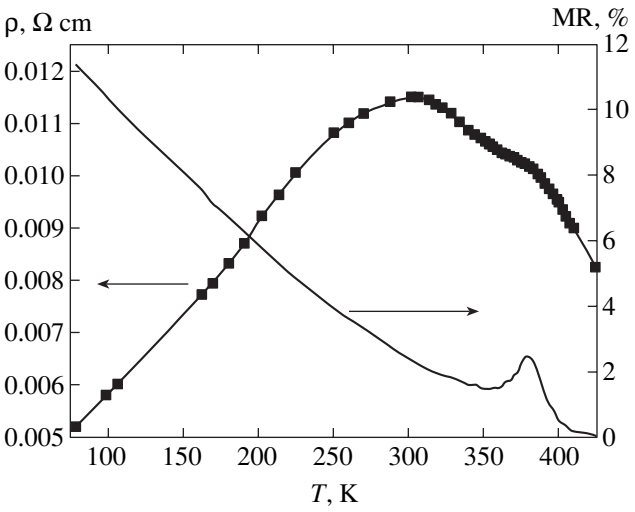


Fig. 3. Temperature dependences of electrical resistance and magnetoresistance effect in $SmCu_3Mn_4O_{12}$.

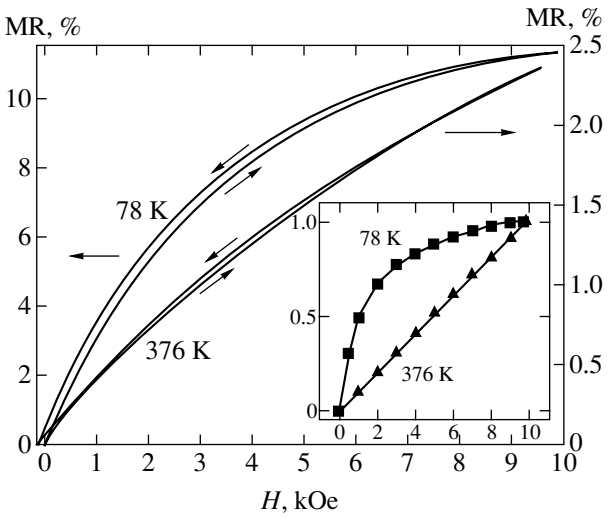


Fig. 4. Field dependence of magnetoresistance in $SmCu_3Mn_4O_{12}$, taken at different temperatures. The inset shows the behavior of reduced magnetization for different temperatures.

al. [8] have demonstrated that the peak of magnetoresistance is a result of the decrease in the scattering of charge carriers within the granules by magnetic fluctuations under the effect of external magnetic field.

Note further that the peak of magnetoresistance in substituted orthomanganites shows up at a concentration of Mn^{4+} of up to 45%. All compounds in the magnetically ordered phase are ferromagnetic in contrast to ferrimagnetic $SmCu_3Mn_4O_{12}$.

So, we have demonstrated that ferrimagnetic polycrystalline $SmCu_3Mn_4O_{12}$ with a perovskite-like structure, prepared by the method of solid-phase reactions under high pressure, exhibits the magnetoresistance effect of two types, namely, (1) of the low-temperature type, which shows up clearly in weak magnetic fields and is a result of the scattering of spin-polarized charge carriers by the boundaries of neighboring granules; and (2) of the high-temperature intergranular type, which shows up in the vicinity of the magnetic ordering temperature (in our case, much higher than room temperature) and is associated with the scattering of charge carriers by magnetic fluctuations. $SmCu_3Mn_4O_{12}$ is a representative of a new family of high-temperature magnetoresistive materials. From both the theoretical and practical standpoints, it is important to perform similar measurements for a number of $ACu_3Mn_4O_{12}$ compounds with different ions in the A position. It is no less important to measure the magnetoresistance effect in the temperature range in which the ordering of magnetic moment of samarium ions is observed for a more detailed understanding of the correlation between the magnetic and electrical properties and magnetoresistance of these compounds.

ACKNOWLEDGMENTS

This study was supported by the Foundation for Basic Research of the Republic of Belarus, grant no. F02M-098.

REFERENCES

1. S. Jin, T. H. Tiefel, M. McCormac, *et al.*, Science **264**, 413 (1994).
2. H. Y. Hwang, S-W. Cheong, P. G. Radaelli, *et al.*, Phys. Rev. Lett. **75**, 914 (1995).
3. I. O. Troyanchuk, L. S. Lobanovsky, D. D. Khalyavin, *et al.*, J. Magn. Magn. Mater. **210**, 63 (2000).
4. M. N. Baibich, J. M. Broto, A. Fert, *et al.*, Phys. Rev. Lett. **61**, 2472 (1988).
5. A. E. Berkowitz, J. R. Mitchell, M. J. Carey, *et al.*, Phys. Rev. Lett. **68**, 3745 (1992).
6. A. Milner, A. Gerber, B. Groisman, *et al.*, Phys. Rev. Lett. **76**, 475 (1996).
7. K.-I. Kobayashi, T. Kimura, H. Sawada, *et al.*, Nature **395**, 677 (1998).
8. A. Maignan, B. Raveau, C. Martin, *et al.*, J. Solid State Chem. **144**, 224 (1999).
9. L. S. Lobanovsky, I. O. Troyanchuk, N. V. Pushkarev, and H. Szymczak, Fiz. Tverd. Tela (St. Petersburg) **43**, 651 (2001) [Phys. Solid State **43**, 677 (2001)].
10. Z. Zeng, M. Greenblatt, M. A. Subramanian, and M. Croft, Phys. Rev. Lett. **82**, 3164 (1999).
11. J. Chenavas, J. C. Joubert, M. Marezo, and B. Bochu, J. Solid State Chem. **14**, 25 (1975).
12. B. Bochu, J. Chenavas, J. C. Joubert, and M. Marezio, J. Solid State Chem. **11**, 88 (1974).
13. B. Bochu, M. N. Deschizeaux, J. C. Joubert, *et al.*, J. Phys. (Paris) **38**, C1-95 (1977).
14. V. M. Loktev and Yu. G. Pogorelov, Fiz. Nizk. Temp. **26**, 231 (2000) [Low Temp. Phys. **26**, 171 (2000)].
15. N. Zhang, S. Zhang, W. P. Ding, *et al.*, Solid State Commun. **107**, 417 (1998).
16. J. M. D. Coey, A. E. Berkowitz, Ll. Balcels, *et al.*, Appl. Phys. Lett. **72**, 734 (1998).
17. J. M. D. Coey, A. E. Berkowitz, Ll. Balcels, *et al.*, Phys. Rev. Lett. **80**, 3815 (1998).
18. H. Y. Hwang, S-W. Cheong, N. P. Ong, and B. Batlogg, Phys. Rev. Lett. **77**, 2041 (1996).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 335–339.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 378–383.

SOLIDS
Electronic Properties

Damping of Coupled Phonon–Plasmon Modes¶

L. A. Falkovsky

Landau Institute of Theoretical Physics, Russian Academy of Sciences, Moscow, 117940 Russia
e-mail: falk@itp.ac.ru

Received October 15, 2002

Abstract—The effect of free carriers on the dispersion and damping of coupled phonon–plasmon modes is considered in the long-wave approximation. The electron and phonon scattering rate, as well as Landau damping, is taken into account. © *2003 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

After the pioneering paper by Migdal [1], the effect of electron–phonon interactions on the phonon dispersion has been continually debated. The problem was how to explain that the many-body approach based on the Fröhlich Hamiltonian gives a strong phonon renormalization. For instance, the sound velocity is renormalized on the order of phonon–electron coupling λ, which is on the order of unity. It follows that the phonon–electron system can be unstable. The phonon softening caused by the electron–phonon interaction has been discussed in many papers. All these results contradict the Born–Oppenheimer conception [2], according to which light electrons must adiabatically follow the slow lattice vibrations. Therefore, the phonon renormalization should involve a small nonadiabatic parameter $\sqrt{m/M}$, where m and M are the electron and ion masses, respectively.

The discrepancy was resolved by Brovman and Kagan [3]. They demonstrated the shortcomings of the Fröhlich model. Employing the adiabatic approximation, they found that there are two terms in second-order perturbation theory which compensate each other, making the result small in terms of the nonadiabatic parameter.

Recently, Alexandrov and Schrieffer [4] again obtained a strong phonon renormalization and predicted extremely large dispersion of optical phonons (on the order of the Fermi velocity) because of coupling with electrons. The large phonon dispersion is typical of the theory [5] using the Frölich Hamiltonian. Reizer [6] pointed out the importance of the screening effects accompanying the longitudinal optical modes, but nonphysical renormalization still remained in his results. For the first time, the screening of the Coulomb field in optical vibrations was treated by Gurevich *et al.* [7].

Only the phonon frequencies were calculated in [4, 6, 7], while no results were available for the attenuation of optical phonons. The collisions of electrons with each other, with defects and phonons were ignored. Besides the electron collision rate γ, the natural phonon width Γ^{nat} was also disregarded. The natural phonon width is caused by the anharmonic processes of phonon decay into two (or more) phonons. Note that the collision processes determine the conductivity and the dielectric permittivity, that is, the electrodynamics of the electron–phonon system. In the optical range, the collision rates $\gamma, \Gamma^{\text{nat}} \sim \sqrt{m/M}\,\omega_O$ are small compared to a typical phonon frequency ω_O and they give the widths of the plasmon and phonon resonances. Experimental studies of this resonances provide information about isotope compositions and quality of semiconductor materials.

Using the Boltzmann equation for electrons and the equation of motion for phonons, we calculated [8] the frequency shift and the width of optical phonons in metals. We take into account all the aforementioned features: the electron–phonon interactions of different types, the Coulomb screening, and the collision rates of electrons and phonons. We find that dispersion and damping of the longitudinal modes can be correctly described if we neglect the direct electron–phonon interaction λ and retain only the screening and the collision rates $\gamma, \Gamma^{\text{nat}}$. It should be emphasized that, in the semiclassical approximation (when the phonon momentum transfer to electrons is small compared to the electron momentum), the method of the Boltzmann equation is completely equivalent to the diagram technique. The corresponding equations can be formulated as equations for the electron and phonon self-energies. In any case, we must properly incorporate adiabatic approximation and screening.

Experimentally, it is convenient to investigate the effect of free carriers on the phonon modes by varying the carrier density, that is, by using doped semiconductors or superconductors (see, e.g., recent work [9], carried out on the HTSC compound $Nd_{1.86}Ce_{0.14}CuO_{4+\delta}$ using IXS). In this paper, the results obtained in [8] are extended to the case of small free-carrier densities,

¶This article was submitted by the author in English.

when the electron plasma frequency ω_{ep} can be on the order of the phonon frequency ω_O and the coupled phonon–plasmon modes can exist. We consider the screening as well as the collision effects γ and Γ^{nat}, focusing on the width of the phonon–plasmon modes. We essentially simplify the problem, ignoring the direct electron–phonon interaction (terms with λ) and assuming that, first, the electron system is degenerate and, second, the momentum transfer k is small compared to the Fermi momentum p_F.

2. ASYMPTOTIC EXPRESSIONS FOR THE DIELECTRIC FUNCTION

Let us find the limiting expressions of the dielectric function

$$\varepsilon(k,\omega) = \varepsilon_\infty\frac{\omega^2-\omega_{LO}^2+i\Gamma^{nat}}{\omega^2-\omega_{TO}^2+i\Gamma^{nat}} - \frac{4\pi e^2\nu_0\langle v_z/\Delta_p(k)\rangle}{k(1-i\langle\gamma/\Delta_p(k)\rangle)}. \quad (1)$$

Here, the ion contribution (first term) is assumed dispersionless, while we are interested in the wave vectors which are small in comparison to the Brillouin zone size. This term has a pole and a zero at the frequencies of the transverse and longitudinal phonons, respectively. The natural phonon width $i\Gamma^{nat}/2$ is added to ω; the high-frequency ion permittivity is denoted by ε_∞.

The second term in Eq. (1) is the electron contribution to permittivity for $k \ll p_F$, written [8] with the help of the Boltzmann equation in the collision rate approximation. In Eq. (1),

$$\Delta_p(k) = \omega - \mathbf{k}\cdot\mathbf{v} + i\gamma$$

and the angular brackets

$$\langle\ldots\rangle = \frac{1}{\nu_0}\int(\ldots)\frac{2dS_F}{v(2\pi)^3}$$

denote averaging over the Fermi surface with the density of states ν_0. For the isotropic case, the density of states on the Fermi surface is $\nu_0 = m^*p_F/\pi^2$ and m^* is the effective electron mass.

Let us rewrite the electron contribution as

$$\varepsilon_e(k,\omega) = \varepsilon_\infty\frac{k_0^2}{k^2}\left[1-\frac{\langle\omega/\Delta_p(k)\rangle}{1-i\langle\gamma/\Delta_p(k)\rangle}\right], \quad (2)$$

where $k_0^2 = 4\pi e^2\nu_0/\varepsilon_\infty$ is the Thomas–Fermi parameter. For the isotropic Fermi surface, we can carry out the integration

$$\left\langle\frac{1}{\Delta_p(k)}\right\rangle = \frac{1}{2kv_F}\ln\frac{\omega+i\gamma+kv_F}{\omega+i\gamma-kv_F}.$$

Separating the imaginary and real parts, we obtain

$$\left\langle\frac{1}{\Delta_p(k)}\right\rangle = \frac{1}{2kv_F}\left[\frac{1}{2}\ln\frac{(\omega+kv_F)^2+\gamma^2}{(\omega-kv_F)^2+\gamma^2} + i\arctan\frac{\omega-kv_F}{\gamma} - i\arctan\frac{\omega+kv_F}{\gamma}\right]. \quad (3)$$

The imaginary part, known as the Landau damping, is pronounced at $kv_F > |\omega + i\gamma|$:

$$\varepsilon_e(k,\omega) = \varepsilon_\infty\left(\frac{k_0}{k}\right)^2\left(1+i\frac{\pi\omega}{2kv_F}\right), \quad kv_F \gg |\omega+i\gamma|. \quad (4)$$

In the range of $\gamma < \omega - kv_F \ll kv_F$, we have

$$\varepsilon_e(k,\omega) = \varepsilon_\infty\frac{k_0^2}{k^2} \times\left\{1-\frac{\omega}{2kv_F}\left[\frac{1}{2}\ln\frac{4k^2v_F^2}{(\omega-kv_F)^2+\gamma^2}-i\frac{\gamma}{\omega-kv_F}\right]\right\}. \quad (5)$$

For a small $kv_F \ll |\omega + i\gamma|$, the expansion in $kv_F/|\omega + i\gamma|$ gives

$$\varepsilon_e(k,\omega) = \varepsilon_\infty\left\{1-\frac{\omega_{pe}^2}{\omega(\omega+i\gamma)}\left[1+\frac{3}{5}\left(\frac{kv_F}{\omega+i\gamma}\right)^2\right]\right\}. \quad (6)$$

Here, the k-independent term represents the Drude conductivity. The electron plasma frequency is given by the integral over the Fermi surface

$$\omega_{pe}^2 = \frac{e^2}{3\pi^2\varepsilon_\infty}\int v\,dS_F.$$

The limiting expressions (4)–(6) are also valid for an arbitrary Fermi surface, but the constant v_F has different values. In Eq. (4), this is an average velocity on the belt $\mathbf{v}\perp\mathbf{k}$, the velocity in the limiting point $\mathbf{v}\parallel\mathbf{k}$ in Eq. (5), and the squared velocity averaged over the whole Fermi surface in Eq. (6). We do not consider the cases when the Fermi surface has flat or cylindrical pieces.

3. FREQUENCY AND DAMPING OF PHONON–PLASMON MODES

The frequencies of the longitudinal phonon–plasmon modes are determined by the equation $\varepsilon(k,\omega) = 0$. In the absence of the electron and phonon collisions ($\gamma = \Gamma^{nat} = 0$), we obtain with the help of Eqs. (1) and (6)

the biquadratic equation. It gives the frequencies of the coupled phonon–plasmon modes at $k = 0$:

$$\omega_{\pm}^2 = \frac{1}{2}(\omega_{pe}^2 + \omega_{\mathrm{LO}}^2) \pm \frac{1}{2}[(\omega_{pe}^2 + \omega_{\mathrm{LO}}^2)^2 - 4\omega_{pe}^2\omega_{\mathrm{TO}}^2]^{1/2}. \quad (7)$$

These frequencies (related to the ω_{TO}) are shown in Fig. 1 (left panel) as functions of the electron density, namely, $\omega_{pe}/\omega_{\mathrm{TO}}$. The upper dashed line begins at ω_{LO} and tends to the electron plasma frequency ω_{pe}. The lower frequency (solid line) starts as $\omega_{pe}\omega_{\mathrm{TO}}/\omega_{\mathrm{LO}}$ and then approaches ω_{TO}. In other words, observing the longitudinal phonon mode in the optical range and adding electrons, we see a transition of the longitudinal phonon frequency from ω_{LO} to ω_{TO}. This is a result of the Coulomb screening.

Since both the collision rates γ and Γ^{nat} in the optical range are small in comparison to ω_O, the damping of the phonon–plasmon modes can be added to their frequencies, $\omega = \omega_{\pm} - i\Gamma_{\pm}/2$. Using Eqs. (1) and (7), we find at $k = 0$

$$\Gamma_+ = [\gamma(\omega_+^2 - \omega_{\mathrm{LO}}^2) + \Gamma^{\mathrm{nat}}(\omega_+^2 - \omega_{pe}^2)]/(\omega_+^2 - \omega_-^2),$$

$$\Gamma_- = [\gamma(\omega_-^2 - \omega_{\mathrm{LO}}^2) + \Gamma^{\mathrm{nat}}(\omega_-^2 - \omega_{pe}^2)]/(\omega_-^2 - \omega_+^2).$$

The behavior of damping as a function of the electron density is shown in Fig. 1 (right panel) for the ratio $\gamma/\Gamma^{\mathrm{nat}} = 3$. We see that the variation of the width corresponds to the type of modes: for a low electron density, the solid line represents mainly the electron plasma mode and the dashed line is associated with the longitudinal phonon. The character of modes reverses at high electron densities.

Now let us consider the dispersion of the coupled phonon–plasmon modes. When the wave vector k increases, the upper mode (see Fig. 2 and Eq. (5)) approaches the asymptote $\omega = kv_F$:

$$\omega_+ = kv_F\left\{1 + 2\exp\left[-2 - \frac{k^2(k^2v_F^2 - \omega_{\mathrm{LO}}^2)}{k_0^2(k^2v_F^2 - \omega_{\mathrm{TO}}^2)}\right]\right\} - i\gamma.$$

The lower mode drops into the domain $kv_F > \omega$ (dashed curve), where the Landau damping (4) appears:

$$\omega_-^2 = \frac{k^2\omega_{\mathrm{LO}}^2 + k_0^2\omega_{\mathrm{TO}}^2}{k^2 + k_0^2} - i\frac{\pi\omega k k_0^2\omega_{pi}^2}{2v_F(k^2 + k_0^2)^2}. \quad (8)$$

Here, the ion plasma frequency is $\omega_{pi}^2 = \omega_{\mathrm{LO}}^2 - \omega_{\mathrm{TO}}^2$. In the imaginary part, we should replace ω with the frequency ω_- defined by the real part. It should be emphasized that a large dispersion arises under the conditions $k_0v_F \gg \omega$ and $k_0 < p_F$. Then, at $k \approx k_0$, the frequency changes from ω_{TO} to ω_{LO}. Again, this is the effect of screening.

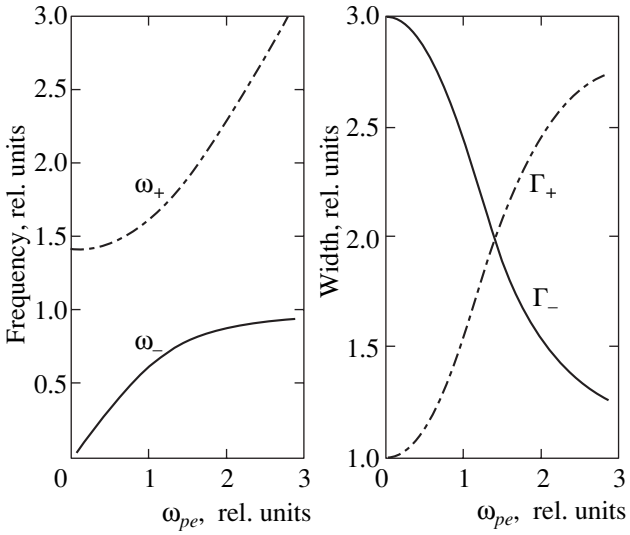


Fig. 1. Frequencies (in units of ω_{TO}, left panel) and widths (in units of Γ^{nat}, right panel) of the phonon–plasmon modes at $k = 0$ as functions of the free carrier density (i.e., of the electron plasma frequency in units of ω_{TO}). We set the ratio of the LO and TO frequencies in the absence of the free carriers $\omega_{\mathrm{LO}}/\omega_{\mathrm{TO}} = \sqrt{2}$, and the ratio of the electron and phonon damping $\gamma/\Gamma^{\mathrm{nat}} = 3$.

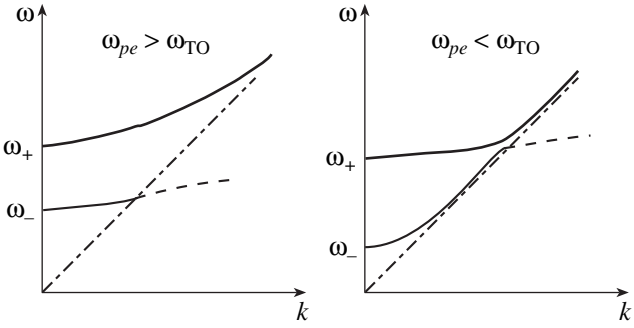


Fig. 2. The dispersion of phonon–plasmon modes for metallic ($\omega_{pe} > \omega_{\mathrm{TO}}$, left panel) and semiconductor ($\omega_{pe} < \omega_{\mathrm{TO}}$, right panel) carrier densities. The straight dashed lines separate the domain $kv_F > \omega$, where the Landau damping exists; the dashed curves represent strongly damped modes.

In the general case, the frequency and the damping of the phonon–plasmon modes can be found by numerically solving the equation $\varepsilon(k, \omega) = 0$. Of particular interest is the function $\mathrm{Im}[-1/\varepsilon(k, \omega)]$, since it gives the intensity of the inelastic Raman or X-ray scattering, where ω and $\mathbf{k}$ have the sense of frequency and momentum transfer, respectively. The plots of the intensity obtained with the help of Eqs. (1)–(3) are shown in Figs. 3–5 for the case of a large electron density ($\omega_{pe} > \omega_{\mathrm{TO}}$). The peak at $\omega/\omega_{\mathrm{TO}} = 2.6$–$2.7$ corresponds to the plasmon excitation; note that $\omega_{pe} = k_0v_F/\sqrt{3}$ for the quadratic electron dispersion.

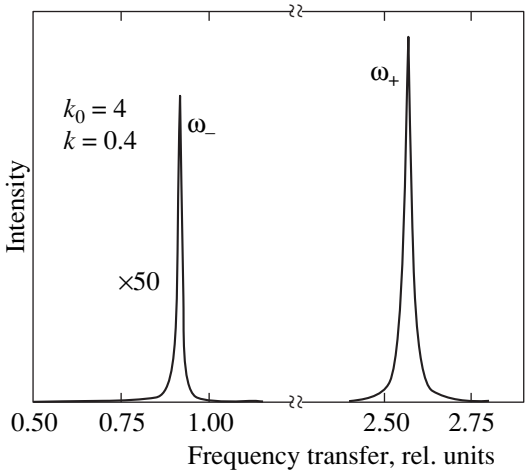


Fig. 3.

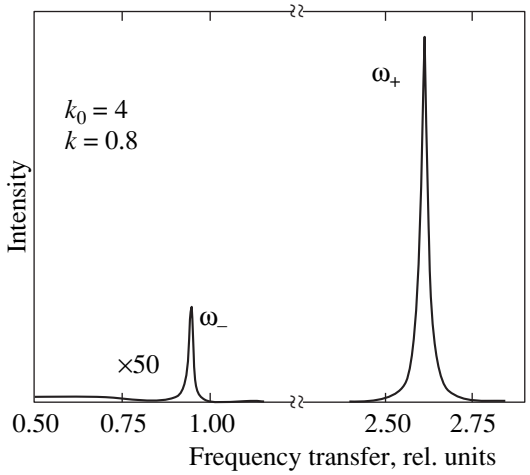


Fig. 4.

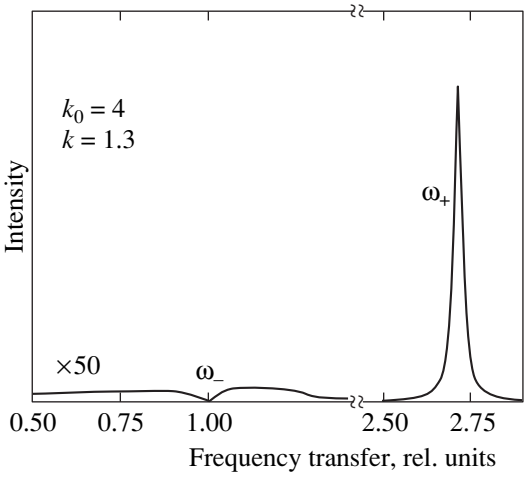


Fig. 5.

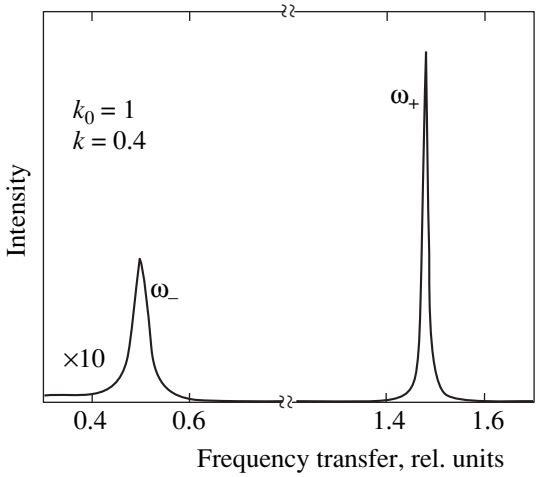


Fig. 6.

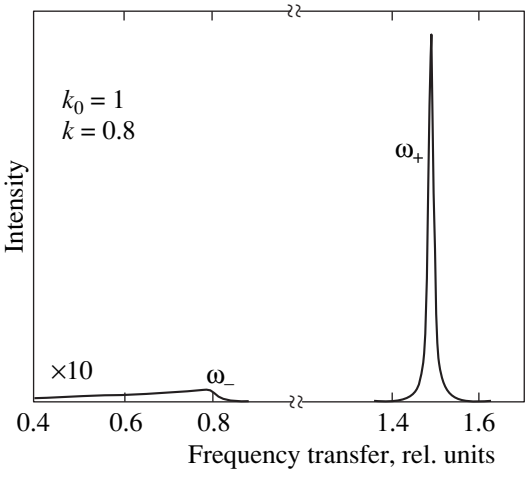


Fig. 7.

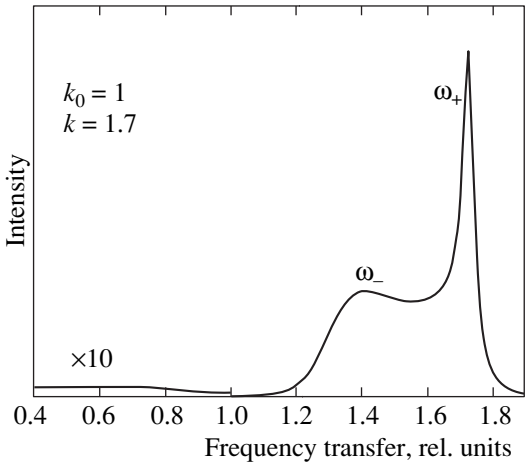


Fig. 8.

Figs. 3–8. The inelastic light scattering intensity $\mathrm{Im}[-1/\varepsilon(k, \omega)]$ as a function of the frequency transfer ω for the momentum transfer k and the Thomas–Fermi parameter k_0 (in the units of ω_{TO}/v_F) indicated in the figures. We set $\omega_{LO}/\omega_{TO} = \sqrt{2}$, $\Gamma^{nat}/\omega_{TO} = 10^{-2}$, $\gamma/\Gamma^{nat} = 3$.

In the domain of $\omega/\omega_{TO} < 1$, there is a phonon peak (see left panel of Fig. 1 and Eq. (7)). Because of the screening, its frequency is smaller than ω_{LO}. For a small k (Fig. 3), the intensity of the phonon peak is 50 times smaller than the plasmon peak intensity. For larger k (see Fig. 4), the broad continuum $\omega < kv_F$ appears on the low-frequency side of the phonon peak. Here, the dielectric constant $\varepsilon(k, \omega)$ has a noticeable imaginary part (4) arising from the electron–hole excitations. The intensity of the phonon peak decreases, and its line shape becomes asymmetric, similarly to the Fano resonance. Finally, the phonon peak broadens, being immersed in the electron continuum (see Fig. 3c and Eq. (8)).

For a relatively small electron density $\omega_{pe} < \omega_{TO}$, the intensities are shown in Figs. 6–8. Now the phonon peak is observed approximately at $\omega/\omega_{TO} = 1.5$. This yields $\omega = \omega_{LO}$, since we set $\omega_{LO}/\omega_{TO} = \sqrt{2}$. There are also the plasmon peak and the electron–hole continuum. In Fig. 6, the plasmon peak is broader than the phonon peak, since we set $\gamma/\Gamma^{nat} = 3$. The plasmon peak becomes still broader in Fig. 7 because of the neighboring electron continuum at $\omega < kv_F$. It disappears completely in Fig. 8.

4. CONCLUSIONS

We have shown that the width of the longitudinal phonon–plasmon modes increases with increasing free carrier density because the mode frequency approaches the region $kv_F > \omega$, where the electron–hole excitations exist. Detailed experimental studies can clarify the comparative role of different mechanisms of the phonon shift and damping in metallic and doped semiconductor materials at various electron densities.

ACKNOWLEDGMENTS

The work was partially supported by the Russian Foundation for Basic Research, project no. 01-02-16211.

REFERENCES

1. A. B. Migdal, Zh. Éksp. Teor. Fiz. **34**, 1438 (1958) [Sov. Phys. JETP **7**, 996 (1958)].
2. M. Born and K. Huang, *Dynamical Theory of Crystal Lattices* (Clarendon Press, Oxford, 1954; Inostrannaya Literatura, Moscow, 1958), Chap. 4.
3. E. G. Brovman and Yu. Kagan, Zh. Éksp. Teor. Fiz. **52**, 557 (1967) [Sov. Phys. JETP **25**, 365 (1967)].
4. A. S. Alexandrov and J. R. Schrieffer, Phys. Rev. B **56**, 13731 (1997).
5. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics*, 2nd ed. (Dobrosvet, Moscow, 1998; Prentice-Hall, Englewood Cliffs, 1963).
6. M. Reizer, Phys. Rev. B **61**, 40 (2000).
7. V. L. Gurevich, A. I. Larkin, and Yu. A. Firsov, Fiz. Tverd. Tela (Leningrad) **4**, 185 (1962) [Sov. Phys. Solid State **4**, 131 (1962)].
8. L. A. Falkovsky, Zh. Éksp. Teor. Fiz. **122**, 411 (2002) [JETP **95**, 354 (2002)].
9. M. d'Adusto, P. K. Mang, P. Giura, *et al.*, E-print archives, cond-mat/0201501.

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 340–354.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 384–398.

MISCELLANEOUS

Quantum Recognition of Eigenvalues, Structure of Devices, and Thermodynamic Properties¶

Yu. I. Ozhigov

Institute of Physics and Technology, Russian Academy of Sciences, Nakhimovsky pr. 34, Moscow, 117218 Russia
e-mail: ozhigov@ftian.oivta.ru

Received April 1, 2002

Abstract—Quantum algorithms that speed up their classical counterparts are proposed for the following problems: recognition of eigenvalues with a fixed precision, recognition of molecular and electronic device structures, and the finding of thermodynamic functions. We mainly consider structures that generate sparse spectra. These algorithms require a time from about the square root to the logarithm of the time of the classical analogues, and they provide exponential savings in memory for the first three problems. For example, the time required for distinguishing two devices with the same given spectrum is about the seventh root of the time of the direct classical method, and about the sixth root for the recognition of an eigenvalue. Microscopic quantum devices can therefore recognize molecular structures and physical properties of environment faster than large classical computers. © *2003 MAIK "Nauka/Interperiodica".*

1. ELECTRONIC DEVICES AND QUANTUM COMPUTATIONS

1.1. Statement of the Problem and Outline of the Paper

The aim of this paper is to build effective quantum algorithms for problems of the following types:

(1) given a quantum gate array generating a unitary operator U and a complex number ω, to determine whether it is an eigenvalue of U with a fixed precision;

(2) to recognize the structure of an unknown electronic or molecular device given only access to its function.

The first problem is an important intermediate step in solving the second.[1] We consider them sequentially.

Recognition of eigenvalues. This problem is closely related to finding the eigenvalue distribution or density of states (DOS), i.e., the energy levels $E_0 < E_1 < \ldots$ and the dimensions of the corresponding subspaces $d_0, d_1, \ldots$. The DOS plays a key role in calculating thermodynamic functions given by

$$F = \sum_j a(j) d_j \exp\left(-\frac{E_j}{k_B T}\right) \quad (1)$$

for some values $a(j)$ such that the summands rapidly converge to zero. For example, this expression gives the partition function Q if all $a(j) = 1$, the average energy if

$$a(j) = E_j/Q$$

and the entropy if

$$a(j) = -\frac{k_B}{Q}\ln\left(\exp\left(-\frac{E_j}{k_B T}\right)/Q\right).$$

Having an efficient method of finding d_j, we would be able to obtain thermodynamic functions and to determine important properties (e.g., heat capacity) of an environment consisting of such molecules. The best known classical method of finding the DOS was proposed by Hams and Raedt in [1]. Their method requires the time on the order of given by the dimension N of the space of states and the memory of the same order (whereas the direct method of calculating eigenvalues requires a time on the order of N^3). The first quantum algorithm for this problem proposed by Abrams and Lloyd in [2] requires the same $O(N)$ time and logarithmic memory. The method proposed in the present work requires a time on the order of that given by the square root of the classical one and memory on the order of $\ln^2 N$.

The idea of our approach is as follows. We use a combination of the Grover search algorithm (GSA) and the Abrams and Lloyd method [2], which reveals eigenvalues and the universal quantum function of application App. The Abrams and Lloyd method of revealing eigenvalues is based on the application of U controlled by ancillary qubit α as

$$U_{\text{cond}}|x, \alpha\rangle \longrightarrow \begin{cases} |Ux, \alpha\rangle & \text{if } \alpha = 1 \\ |x, \alpha\rangle & \text{if } \alpha = 0. \end{cases}$$

¶This article was submitted by the author in English.

[1] A straightforward calculation shows that the simulation of evolution generated by a given Hamiltonian up to a time instant τ with a fixed accuracy requires the number of steps on the order of τ^2 on a quantum computer. This means that all results of this paper can be generalized to arbitrary quantum systems.

We note that it is a direct generalization of Shor's trick, which can be obtained if U is a multiplication by a given integer modulo q [3].

Recognition of device structures. We separate two versions of this general problem: recognition of molecular structures and recognition of electronic circuits.

If we want to determine a molecular structure, it is natural to assume that its functionality is given as the spectrum of its Hamiltonian, e.g., the set of its energy levels. It is therefore required to find a quantum system whose Hamiltonian has a given spectrum.

The problem of recognition of electronic circuits is stated differently. An electronic device is considered as a source of electromagnetic fields that can control some quantum system Q. Let such a field induce evolution of a system with Hamiltonian H in the time frame δt. We then have the correspondence

$$\text{(electronic device)} \longrightarrow \text{(Hamiltonian, } \delta t\text{)}.$$

The evolution of the quantum system Q induced by this Hamiltonian can be represented as a unitary transformation:

$$U = \exp\left(-\frac{i}{h}H\delta t\right).$$

Given a device C and a time instant t, we can then associate with it some unitary transformation U_C. We assume that we have recognized a circuit C if we have found some circuit C_1 such that $U_C = U_{C_1}$ with high accuracy. We write U instead of U_C for the circuit C that we want to recognize. In fact, we solve a more general problem where the tested device C can be used as a black box acting on n qubits as a function U_C such that if x is an input; then $U_C|x\rangle$ is the result of its action on this input. The tested device can contain its own quantum memory and can be entangled with Q in the course of performing the transformation U, but this entanglement must then be eliminated. The existence of such an entanglement implies that this case cannot be described by the Hamiltonian of the system Q. For simplicity, we assume that the unknown circuit is built from elementary functional elements taken from some fixed set. The next natural assumption is that the size of the circuit is limited by some constant c such that the circuit is some unknown combination of c functional elements. We let $\mathscr{C}$ denote all circuits of length c. We can encode such $C \in \mathscr{C}$ by a string $[C]$ of ones and zeroes such that the decoding procedure is easy and we can immediately recreate a circuit given its code. We can therefore look through all circuits by looking through their codes. The same coding can be built for electronic devices.

A straightforward solution of the problems is clear. For the problem of recognition of molecular structures, all we need is to be able to recognize eigenvalues of the transformation generated by a given circuit. Each eigenvalue of a unitary operator has the form $e^{2\pi i\omega}$, where ω is a real number from [0, 1) called the frequency. In what follows, the spectrum is meant to be the set of all frequencies. Let all the frequencies be grouped near points of the form l/M, where M is not very large and $l = 0, 1, \ldots, M-1$. We assume that the acceptable precision of the recognition of frequencies is $1/M$. Having an algorithm for eigenvalue recognition, we can apply it repeatedly in constructing spectra generated by all possible circuits, and thus find the sought for circuit with the given spectrum. If we need to recognize a circuit of an electronic device, we can examine all possible circuits taken in some order. Examination of one circuit means that we run it at all possible inputs one after another and compare the results with the corresponding result of the action of the tested device.

For the problem of the recognition of molecular structures, our method requires a time on the order of the sixth root of the time of the direct classical method, whereas saving memory is exponential. For the problem of the recognition of electronic circuits, our method gives at least square-root time saving in the case where the classical counterparts exist (this is the narrow formulation where the tested device generates a classical mapping). However, the advantage can be greater in the general case. For example, we can distinguish between two devices with the same spectrum in the time about the seventh root of the time of the brute force method.

To recognize devices at the quantum level, we must be able to store and perform operations on codes of different circuits. This possibility is based on the existence of a quantum analog of the universal Klini function. This is a unitary operator App such that for all quantum devices C and all inputs x,

$$\text{App}|x, [C]\rangle = |U_C x, [C]\rangle.$$

We assume that for a wide range of quantum devices C with c particles, C can be encoded as an integer $[C]$ in time $O(c)$ such that the quantum complexity of App is also $O(c)$.

We here consider a particular case of the problem where all eigenvalues of U are known a priori or can be obtained in advance. This restriction is not very constraining. To illustrate the tasks that can be solved by the proposed method, we consider several examples of the problem of recognition of an electronic device whose spectrum is known.

Recognition of quantum algorithms is designed as subroutines. Such an algorithm must restore the input if we apply it twice. Computing a function f, it acts as

$$|x, b\rangle \longrightarrow |x, b + f(x) \mod 2\rangle.$$

All known quantum algorithms can be represented in this form. For such quantum algorithms, the unitary transformation U has only two eigenvalues, 1 and –1. Given a controlling device for such an algorithm (which can also include classical elements and ancillary qubits), we can quickly recognize its construction.

Alternatively, we can quickly find a quantum or classical algorithm for a given task.

We consider the "classical" particular case of the recognition problem where U maps each basic state to a basic state, which means that the matrix of U consists of ones and zeroes and in addition U equals U^{-1}. Here, the evident recognition strategy takes the number of steps on the order of card($\mathscr{E}$). In this case, the problem can be reformulated as finding such t that some given predicate $A(t, s)$ is true for all s. This is the problem of verification of logical formulas. Its quantum solution in a time about the square root of the classical time based plainly on Grover's trick was proposed in [4]. This method is inapplicable in the general case, where U_C is an arbitrary involutive unitary transformation, e.g., such that $U = U^{-1}$. This general case is precisely the subject of this work. Here, we cannot recognize a circuit as easily as in the "classical" case because it is difficult to compare two quantum states $U_C|x\rangle$ and $U|x\rangle$.

The general idea of our approach to the recognition of arbitrary electronic devices is as follows. We include the device C whose structure we want to recognize into the classical controlling part of a quantum computer. We consider the main system of n qubits. The tested device then generates a unitary transformation on this system. We then find the eigenvectors of U using U_{cond} by the above method and compare them with the eigenvectors of circuits from $\mathscr{E}$ choosing a circuit that gives the best approximation. Here, GSA is used at the last step and at the several intermediate steps.

The sparse spectrum assumption. In this paper, we mainly consider circuits generating sparse spectra. This means that the spectra of the operators U_C are designed such that the frequencies are assembled in groups and the minimum distance between frequencies from the different groups is greater than $1/M$ and the maximum distance between frequencies in the same group is less than $1/L$. In the problems of eigenvalue and molecular structure recognition, we require that $L = 16M$, which is not very restricting. In the problem of recognition of electronic devices, we assume that $L \gg M$, which is a stronger limitation. A spectrum is called sparse if $M = \text{const}$ as $N \longrightarrow \infty$. Our algorithms show best performance for sparse spectra.

Spectra that are not sparse are called dense. For dense spectra, our methods give less advantage over the classical algorithms (see Section 3.6). An example of a dense spectrum is given by $\omega_k = k/N$, $k = 0, 1, \ldots, N-1$. Similar problems for dense spectra require additional investigations.

We write $\omega' \approx \omega$ iff ω' and ω belong to the same group. For simplicity, we also assume that for each group of frequencies, there exists a number of the form l/M positioned between some two frequencies of this group, where l is an integer less than M.

1.2. Abstract Model of a Quantum Computer: "Plug and Play" Technique

To build algorithms recognizing circuits, we need an abstract model of a quantum computer. A quantum computer consists of two parts, quantum and classical. The classical part exactly determines what unitary transformation must be performed on the quantum part at each time instant and therefore plays the role of a controller for the quantum part. These unitary transformations are of two types: working transformations, which our computer performs itself, and query transformations, which are induced by a tested device, U or U_{cond}.

We can suppose that the quantum part Q consists of nuclear spins or interacting dipoles (or some other quantum two-level systems) and the classical part is a source of electromagnetic fields determining the evolution of the quantum part. The general form of the state of the quantum part is

$$\chi = \sum_{i=0}^{2^\nu - 1} \lambda_i e_i,$$

where the basic states $e_0, \ldots, e_{2^\nu - 1}$ are simply strings of ones and zeroes of the length $\nu > n$; this length is the size of the quantum part that can contain some auxiliary qubits in addition to the input for U, $N = 2^n$ is the number of all classical input words for U, and

$$\sum_{i=0}^{2^\nu - 1} |\lambda_i|^2 = 1.$$

The classical part determines when the tested device is to be "switched on" (this usually occurs many times) and when the result of the computation is to be observed. Observation of a state χ gives every basic state e_i with the corresponding probability $|\lambda_i|^2$.

The problem of recognition of electronic devices presumes the so-called "plug and play" technique, where the tested device is applied only as a black box. If query transformations are only U, then our model evidently satisfies the requirements of the plug and play technique, where we classically control switching the tested device. An implementation of U_{cond} in the framework of this technique is not so easy because it requires a quantum control on applications of the device.[2]

It is nevertheless possible to implement U_{cond} in the framework of the plug and play technique. This problem requires additional investigations; here, we simply presume that it is possible. This difficulty does not exist in problems of eigenvalue and molecular structure recognition. Here, we can manage without oracles because

[2] This would evidently be possible provided we have access to the internal details of our device and can simultaneously control their work at the quantum level, but this assumption contradicts the plug and play technique.

having an explicit form of a quantum gate array that realizes the universal function of application App, we can control its actions in each element at the quantum level separately and simultaneously, thereby implementing U_{cond}.

Let every basic state be partitioned as

$$e_i = |\text{place for code } [C], R_{\bar{1}}, R_{\bar{2}}, \ldots, R_{\bar{l}}\rangle,$$

where each register $R_{\bar{l}}$ is in turn partitioned into a place for the argument, places for time instants, and places for the corresponding frequencies. A complex index $\bar{i}$ contains one or two integers, and the length of e_i is therefore a polynomial in c and n of at most second degree.

2. OBTAINING NEW ALGORITHMS FROM BASIC QUANTUM TRICKS

2.1. GSA and the Amplitude Amplification

The GSA proposed in [5] is one of two basic quantum tricks. It is used for quickly obtaining a quantum state $\bar{a}$ given the inversion $I_{\bar{a}}$ along this state. The inversion along some state $\bar{a}$ is defined by

$$I_{\bar{a}}|\bar{x}\rangle = \begin{cases} |\bar{x}\rangle & \text{if} \quad x \perp a \\ -|\bar{a}\rangle & \text{if} \quad x = a. \end{cases}$$

We also assume that $I_{\bar{a}}$ acts as the identity if $\bar{a}$ does not exist. A typical situation is where a state is unknown but the inversion along it can be performed easily. For example, let $\bar{a}$ be a solution of the equation $f(x) = 1$ with a simply computable Boolean function f. The inversion $I_{\bar{a}}$ can then be implemented by modulo-2 addition of $f(x)$ to an ancillary qubit initialized by

$$\frac{|0\rangle - |1\rangle}{\sqrt{2}}.$$

This transformation maps the state

$$\left|x, \frac{|0\rangle - |1\rangle}{\sqrt{2}}\right\rangle$$

to the same state with the sign "+" or "–" depending on whether the equality $f(x) = 1$ is satisfied. The transformation is unitary and can easily be performed given a device performing f. All sequential transformations in our formulas are applied from right to left.

The GSA is a sequential application of the transformation $G = I_{\bar{a}} I_{\tilde{0}}$ to a randomly preset state $\tilde{0}$. If we apply this transformation $O(\sqrt{N})$ times, where N is the dimension of the main space, then observation of the quantum part yields $\bar{a}$ with a visible probability, whereas finding $\bar{a}$ without a quantum computer would have required a number of steps on the order of N.

A minor difficulty is that here we do not exactly know the time instant t at which the iterations must be terminated in order to make the probability of error negligible, as required in applying the GSA as a subroutine. The following simple trick helps.

We define the number $B = B(N)$ such that $1/B$ is the average value of $|\langle a|\tilde{0}\rangle|$ for $\tilde{0}$ uniformly distributed on a sphere of radius 1 in the space of inputs. A straightforward calculation shows that $B = O(\sqrt{N})$. Let GenArg_j be operators generating arbitrary vectors $\bar{a}_j$ from the space of inputs belonging to independent uniform distributions, $j \in \{1, 2, \ldots, k\}$, and let GenTimeArg_j be operators generating time instants t_j from independent uniform distributions on integers from the segment $[0, B]$. We arrange k copies of two working registers, for the input and for the storage of a time instant, and apply the corresponding operator

$$(I_{\tilde{0}} I_{\bar{a}})^{t_j} \text{GenArg}_j \text{GenTime}_j$$

to each register. If $\bar{a}$ exists, the probability of obtaining $\bar{a}$ observing any single register is at least one-fourth (see [6]) and the probability of obtaining any other fixed state is negligible because our operators GenArg_j generate independent uniformly distributed samples. If $\bar{a}$ does not exist, which implies that $I_{\bar{a}}$ is the identity, then the probability of obtaining any fixed state is negligible. We let $\bar{a}_j$ denote the contents of the jth register for the argument in the resulting state. We consider the following criterion: if at least one-fifth of $\bar{a}_j$, $j = 1, 2, \ldots, k$, coincide, we decide that $\bar{a}$ is this value, otherwise $\bar{a}$ does not exist. We now calculate the error probability of this criterion. Let K be the number of those j for which $\bar{a}_j = \bar{a}$. By the central limit theorem, the probability that the fraction

$$\frac{(k/4) - K}{\sqrt{(k/4) \times (3/4)}}$$

belongs to the segment $[\alpha_1, \alpha_2]$ converges to

$$\frac{1}{\sqrt{2\pi}} \int_{\alpha_1}^{\alpha_2} e^{-x^2/2} dx.$$

Straightforward calculations then show that the probability that $K \le k/5$ is on the order of

$$\int_{\alpha_1}^{\infty} e^{-x^2/2} dx$$

for α_1 on the order of $\sqrt{k}$. To make the error probability on the order of $1/\sqrt{N}$, it therefore suffices to choose k on the order of $n = \log N$. This method can be used not only for the GSA but also for other algorithms. If the probability of obtaining the correct result for each of the k registers is some positive number p independent of the dimensionality, then to make the error probability $1/N_1$, it suffices to choose k on the order of $\log N_1$. In what follows, we use this simple trick without specially mentioning it and let $\bigotimes_j$ denote simultaneous operations of the same type on all working registers. We assume that all ensembles generated by the different jth copies of operators are taken from independent distributions.

We use the standard norm

$$\|A\| = \sup_{\|\bar{x}\| = 1} \|A\bar{x}\|$$

on operators in a Hilbert space. Given an operator A, we let A_ϵ denote an operator such that $\|A - A_\epsilon\| \le \epsilon$. In what follows, we use the above method requiring copies of registers, thereby raising the accuracy of our operators to the required level. When we must repeat an operator T times, the required accuracy of one application must be $1/T$ and, as shown above, it can be ensured by only linear cost in memory. Instead of A, we therefore always use A_ϵ, where $\epsilon = O(1/T)$, whenever an operator A must be repeated T times; we do not explicitly indicate this in the notation.

2.2. Revealing the Eigenvalues

The second basic quantum trick is used for revealing eigenvalues of a given unitary operator U. We define an operator revealing frequencies in accordance with [2].

Let $M = 2^m$ and $L = 2^p$. We determine frequencies of unitary operators within $1/L$, where L is the number of applications of U required for revealing frequencies with this accuracy, which means that the accuracy $1/M$ is sufficient to distinguish the eigenvalues of U. For the recognition of eigenvalues, we set $p = m + 4$, and therefore, $L = 16M$.

We let $(0.l)_p$ denote the number from [0, 1) of the form l/L. Let the operator U have the eigenvalues $e^{2\pi i\omega_k}$, where the frequencies $\omega_0, \omega_1, \ldots, \omega_{N'-1}$ are different real numbers from [0, 1). Let E_k be the space spanned by all eigenvectors corresponding to ω_k. An arbitrary vector with a length of 1 from E_k is denoted by Φ_k. Every state ξ therefore has the form

$$\xi = \sum_{k=0}^{N'-1} x_k \Phi_k.$$

Let N' be some integer and $\Omega = \{\tilde{\omega}_{k,i}\}$ be some set of integers from $\{0, 1, \ldots, L-1\}$, $0 \le i \le M-1$, $0 \le k \le N'-1$; $\varepsilon, \delta > 0$. We set

$$L_\varepsilon^k(\Omega) = \{i : |(0.\tilde{\omega}_{k,i})_p - \omega_k| \le \varepsilon$$

or $|(0.\tilde{\omega}_{k,i})_p - \omega_k - 1| \le \varepsilon\}$.

Definition 1. *A transformation W of the form*

$$W : |\xi, 0^{m+4}\rangle \longrightarrow \sum_{k=0}^{N'-1}\sum_{i=0}^{L-1} \lambda_{i,k}|\Phi_k, \tilde{\omega}_{k,i}\rangle$$

is called a transformation of type $W_{\delta,\varepsilon}$ if for all k and ξ,

$$\sum_{i \in L_\varepsilon^k(\Omega)} |\lambda_{i,k}|^2 \ge |x_k|^2(1-2\delta).$$

Thus, δ is the error probability of obtaining the correct frequencies ω_k by observation of the second register and ε is the precision of the frequency approximations.

Definition 2. *A unitary operator R is called revealing frequencies of U if R belongs to the type $W_{\frac{1}{K},\frac{K}{L}}$ for any $K \in \{1, 2, \ldots, L\}$.*[3]

The key here is the quantum version of the Fourier transform (QFT), defined by

$$\mathrm{QFT}_L : |s\rangle \longrightarrow \frac{1}{\sqrt{L}}\sum_{l=0}^{L-1} \exp(-2\pi i s l/L)|l\rangle.$$

We also need the following generalization U_{seq} of the operator U_{cond}:

$$U_{\mathrm{seq}}^L|x, a\rangle = |U^a x, a\rangle.$$

This is the result of a sequential applications of U to the main register. To implement this operator by means of U_{cond}, we perform the following cycle. For an integer counter j ranging from 1 to the maximum value $L-1$ of a, we apply U iff $j \le a$. One cycle then consists of U_{cond} with a properly prepared controller and the resulting operator is U_{seq}^L.

We define the operator revealing frequencies by

$$\mathrm{Rev} = \mathrm{QFT}_L U_{\mathrm{seq}}^L \mathrm{QFT}_L,$$

where quantum Fourier transformations are applied to the second register.[4] It was proved in [7] that Rev is a transformation revealing frequencies. We now need more. For a redistribution of amplitudes x_k, we also need the transformation Rest cleaning the second register. The ideal candidate for this role would be Rev^{-1}, but

[3] In what follows, we use this notion only with $K = 16$.

[4] As in [2], the first QFT can be replaced by the Walsh–Hadamard transform because it is equivalent to the QFT on a zero ancilla.

it requires the application of U^{-1}, which is physically unrealizable given only the device fulfilling U, except in evident cases where, e.g., $U = U^{-1}$. We can use this simplest definition of Rest only in the case where we are given a circuit implementing U (e.g., gate array) because U^{-1} is then accessible for us as well as U. However, if C is given only as a black box, the restoring operator must be defined separately.

We find the operator restoring the ancilla in the form

$$\mathrm{Rest} = \mathrm{Rev}D,$$

where D is some operator of turning. Given some integers $\tilde{\omega}_k^L$ of the form q/L, where q is an integer, $\tilde{\omega}_k^L \approx \omega_k$, we can define the operator D of turning by

$$D|\Phi_k, l\rangle = \exp(-2\pi i(L-1)\delta_{k,l})|\Phi_k, l\rangle,$$

where

$$\delta_{k,l} = \tilde{\omega}_k^L - (0.l)_m.$$

It was proved in [7] that

$$\left\|(\mathrm{RestRev}|\chi, \bar{0}\rangle - |\chi, \bar{0}\rangle)\right\| < 7M/L,$$

which implies that the restoring operator thus defined indeed restores zeroes in the second register after the action of Rev if L is sufficiently large. To create these good approximations, we apply a slightly more general construction. We set

$$D = \mathrm{Enh}\,\tilde{D}\,\mathrm{Enh},$$

where the operator Enh calculates an integer function $h(l)$ giving a good approximation $(0.h(l))_p$ of frequencies within $1/L$ given their rough approximation $(0.l)_m$ within $1/M$ and places them into the ancilla; $\tilde{D}$ rotates each eigenvector by an appropriate angle

$$\tilde{D}|\Phi_k\rangle = \exp[-2\pi i(M-1)((0.h(l))_p - (0.l)_p)]|\Phi_k,$$

and the last application of Enh cleans the ancilla. The operator Enh is accessible given good approximations of eigenvalues. Our operator Rest therefore restores zeroes in the ancilla within $1/L$.

We can reach accuracy $1/L$ for all operators of Rest type that are less than $1/t$, where t is the number of all steps in the computation; this accuracy can be guaranteed with $\log L = p$ registers. We emphasize that this difficulty with the eigenvalue precision arises only when U^{-1} is inaccessible, as in the problem of recognition of electronic circuits in Section 3.4, where we must choose $L \gg M$.

The operators Rev and Rest can be built in the form of a quantum gate array using the universal quantum Klini function App, where the code $[C]$ of a circuit generating U is a part of the input. We write the operator U corresponding to these two operators as a superscript.

3. RECOGNITION PROBLEMS

3.1. Obtaining Eigenvectors and Recognition of Eigenvalues

Our assumption about a sparse spectrum is now stated as $L = 16M = \mathrm{const}$. Because Rev reveals frequencies, it belongs to the type $W_{1/16,\,1/M}$. By definition of $W_{\delta,\varepsilon}$, this implies that Rev gives a state

$$\sum_{k=0}^{N'-1}\sum_{i=0}^{M-1} \lambda_{i,k}|\Phi_k, \tilde{\omega}_{k,i}\rangle,$$

where seven-eighths of the probability are concentrated on the pairs i, k such that $(0.\omega_{i,k})_m$ is close to ω_k. This implies that we can obtain eigenvalues with a high probability by observing the second register; the first register then contains the corresponding eigenvector. This procedure for obtaining eigenvectors was proposed in [2, 8]. Its first disadvantage is irreversibility. Observing a state, we lose the complete information about it; we cannot use this state again, which is very important for building nontrivial quantum algorithms. The second disadvantage is that this procedure gives a random eigenvector when it is typically required to obtain the eigenvector corresponding to a given frequency.

We consider a good approximation $\tilde{\omega}^L$ of some frequency ω written as a string of p of its sequential binary digits and let $\mathscr{E}_\omega = \{\Phi_1^\omega, \ldots, \Phi_l^\omega\}$ be a basis of the subspace E_ω of eigenvectors corresponding to all frequencies $\omega' \approx \omega$. We now build the operator State_ω that concentrates the bulk of the amplitude on some superposition of the corresponding eigenvectors

$$\sum_{j=1}^{l} \lambda_j \Phi_j^\omega \in E_\omega.$$

For this, we apply the GSA. Let

$$|\bar{a}\rangle = \sum_{j=1}^{l} \mu_j \Phi_j^\omega + \sum_s \nu_s \Phi_s$$

be some randomly chosen vector from the main space with all eigenvectors in the second sum corresponding to frequencies $\omega' \neq \omega$. Our target state is the vector

$$E_\omega(\bar{a}) = \sum_{j=1}^{l} \lambda_j \Phi_j^\omega,$$

where

$$\lambda_j = \frac{\mu_j}{\sqrt{\sum_{j=1}^{l} |\mu_j|^2}}.$$

The vector is therefore of length 1 and is directed along the projection of $\bar{a}$ to the subspace E_ω.

Let A be some set of vectors. We let I_A denote the operator that changes the sign of all vectors in A and leaves all vectors orthogonal to A unchanged. Our aim is to obtain the operator I_{E_ω} constrained to the two-dimensional subspace $S(\bar{a}, \omega)$ spanned by the vectors $|\bar{a}\rangle$ and $E_\omega(\bar{a})$.

Let Rev_j and Rest_j be jth copies of the respective operators Rev and Rest acting on the corresponding places of the jth register. We let l_j denote the string contained in the place for the frequency of the jth register. Set

$$\tilde{I}_{E_\omega} = \bigotimes_j^v \mathrm{Rest}_j \mathrm{Sign}_\omega \bigotimes_j^v \mathrm{Rev}_j.$$

It follows that Sign_ω changes the sign iff

$$\left|(0.l_j)_p - (0.\tilde{\omega}^L)_p\right| \le 1/L$$

for at least a half of all j.[5] Applying the argument at the end of Section 2.1, we conclude that the actions of I_{E_ω} and $\tilde{I}_{E_\omega}$ restricted to $S(a, \omega)$ differ by less than $1/2^{O(v)}$; this difference can therefore be made very small with only a linear growth of memory. We thus omit the tilde from our notation.

We define

$$\mathrm{St} = \mathrm{GenArg}^{-1}\mathrm{GenTimeArg}^{-1} \circ (I_{\bar{a}} I_{E_\omega})^t \mathrm{GenTimeArg}\,\mathrm{GenArg},$$

where the respective operators GenArg and GenTimeArg generate the pair $\bar{a}$, $[C]$ and the time instant t, with C being a gate array implementing $I_{\bar{a}}$. Here, the actions of $I_{\bar{a}}$ are implemented by the universal function of application App. The result $\xi = \mathrm{St}|\bar{0}\rangle$ of its action on $\bar{0}$ is then close to $E_\omega(\bar{a})$. Indeed,

$$\left|\langle E_\omega(\bar{a})|\xi\rangle\right| = \left|\sin(2t \arcsin\langle\bar{a}|E_\omega(\bar{a})\rangle)\right|$$

(see [6]). The average value of $|\langle\bar{a}|E_\omega(\bar{a})\rangle|$ with the uniformly distributed probability of choosing $\bar{a}$ and t over all space and the time frame $[0, B]$, respectively, is on the order of $1/\sqrt{N}$. Therefore, if t is randomly chosen from the uniform distribution over 1, 2, …, B, then the average value of $|\langle E_\omega(\bar{a})|\xi\rangle|^2$ is not less than one-fourth. Of course, it would be much more convenient to obtain $E_\omega(\bar{a})$ with the error probability converging to zero, which is possible by the method described in Section 2.1. Namely, we arrange h equal registers for the states χ_k, $k = 1, 2, \ldots, h$ in the main space, the corresponding h registers for the frequencies, and associate the variable t_k with each kth register. Let St_k be a pattern of the operator St acting on the kth register. We recall that the operators GenArg_k and $\mathrm{GenTimeArg}_k$ generate independent distributions for different $k = 1, 2, \ldots, h$. We now define

$$\mathrm{State}^\omega = \mathrm{St}_1 \otimes \mathrm{St}_2 \otimes \ldots \otimes \mathrm{St}_h. \quad (2)$$

Applied to zero initial state, this operator gives a state $\chi_1 \otimes \chi_2 \otimes \ldots \otimes \chi_h$, and the average value of $|\langle E_\omega(\chi_k)|\chi_k\rangle|^2$ is close to some number not less than one-fourth with the vanishing probability of error. This also implies that if we then apply the corresponding operators $\mathrm{Rev}_1 \otimes \mathrm{Rev}_2 \otimes \ldots \otimes \mathrm{Rev}_h$ revealing frequencies to this state, then the main part of the amplitude of the resulting state χ is concentrated on the basic states for which at least 5/32 of all registers for the frequencies contain numbers l such that[6]

$$\left|(0.l)_m - (0.\tilde{\omega}^L)_p\right| < 1/L.$$

On the other hand, if ω is not a frequency, the probability of obtaining such a basic state vanishes because the distributions generated by $\mathrm{GenTimeArg}_k$ and GenArg_k are independent for different k.

The time complexity of this algorithm is on the order of $M\sqrt{N}n^2$. The latter factor arises because of copying the registers. We therefore have a solution of the first problem of the recognition of eigenvalues.

3.2. Finding Thermodynamic Functions

Given the structure of the molecule of a gas, we consider the problem of finding its thermodynamic function (1). Because the common term in this sum rapidly converges to zero, it is sufficient to find the first several summands. It is therefore sufficient to find the degree of degeneracy of the subspace corresponding to the frequencies $\omega' \approx \omega$ for any $\omega = l/M$. Let $E_0 < E_1 < \ldots < E_s$ be energy levels of the molecule (the eigenvalues of its Hamiltonian H).

The evolution operator in time frame t is then given by

$$U = \exp\left(-\frac{iH}{h}t\right).$$

[5] We could choose any fixed ρ: $1/8 < \rho < 7/8$ instead of 1/2. Indeed, $\tilde{I}_{E_\omega}$ thus defined would change the sign of all $\bar{a} \in E_\omega$. If $\bar{a} \perp E_\omega$, the probability of obtaining ω in observing the frequency from Rev is less than one-eighth.

[6] We note that in this criterion, 5/32 could be replaced by any ρ such that $0 < \rho < 1/4 \times 7/8 = 7/32$.

Adding the diagonal matrix rI with a constant r to the Hamiltonian does not change the physical picture. Choosing

$$r = -E_s, \quad t = \frac{h}{2\pi E_s},$$

we then obtain a unitary operator U whose frequencies belong to the segment [0, 1). Thus, the problem is reduced to the case considered above.

We assume that M is fixed and we must examine only several frequencies close to 0. We can first recognize all numbers of the form l/M that are frequencies within $1/L$. Let ω be such a number. We now show how to find the degeneracy degree d of the corresponding subspace. This is the dimension of the subspace E_ω spanned by the eigenvectors corresponding to frequencies $\omega' \approx \omega$. Our strategy is as follows. We build the operator I_{E_ω} of reflection along this subspace. Using a counting procedure built in [6], we then evaluate the time required for turning an arbitrary initial vector to this subspace. This time is about $\sqrt{N/d}$, and we thus find d. We fix some $\epsilon > 0$ and show how to obtain the value of d within ϵd.

Let the operators $\mathrm{GenTimeArg}_j^a$ generate time instants t_j from independent uniform distributions on the segment $[0, [a]]$, where a is a nonnegative number. For a from 1 to $\sqrt{N}$, we perform the following three-step loop:

(1) apply the operator

$$\bigotimes_j \left[\bigotimes_k \mathrm{Rev}_{j,k} \right] (I_{\tilde{a}} I_{E_\omega})^{t_j} \mathrm{GenTimeArg}_j^a \mathrm{GenArg}_j;$$

(2) find the fidelity of the result, i.e., the number of all j for which at least $7/8 - \epsilon$ of all k are such that $\omega_{j,k} \approx \omega$; if the fidelity of this step is larger than at the previous step, we proceed with the loop, otherwise we stop;

(3) replace a with $4a/3$.

If we finish the computation at step 2, the current value of a is taken as a rough approximation of d from above. We have $3a/4 \le d \le a$. To find d more exactly, we divide the segment $[3a/4, a]$ into $[1/\epsilon]$ equal parts by points $a_0 < a_1 < \ldots < a_l$ and repeat the above procedure sequentially for all a_i. We thus determine d within $g(\epsilon)d$, where g is a function rapidly converging to zero with ϵ. Thus, our algorithm finds d and thermodynamic functions with an arbitrary relative error in the time $O(\sqrt{N})M$, where the constant depends on the admissible error. A more refined algorithm can be obtained if we apply the method of counting in [9]. In that work, the quantum Fourier transform is used analogously to the Abrams and Lloyd operator Rev only in order to find the time period of the function $G|\xi, t\rangle = |G^t\xi, t\rangle$, which is about $\sqrt{N/d}$. Their method gives an accuracy on the order of $\sqrt{d}$, which implies that the relative error converges to zero as $d \longrightarrow \infty$.

3.3. Recognition of Molecular Structures

We now consider the problem of recognition of molecular structures. Given the spectrum of a molecule, we must recognize its construction. We have no access to the device, but it is sufficient to find an arbitrary device generating this spectrum. To clarify the formulation, we assume the following form of determining the spectrum. Given a set $\bar{w} = \{w_1, \ldots, w_Q\}$ of numbers from [0, 1) of the form $w_i = l_i/M$ with $l_i \in \{0, 1, \ldots, M-1\}$, we let F denote the subspace spanned by vectors of the form $|l_i\rangle$, $i = 1, \ldots, Q$. A spectrum S is determined by this set $\bar{w}$ if

(a) for each $\omega \in S$, there exists its good approximation $w_i \in \bar{w}$, $|w_i - \omega| \le 1/L$, and

(b) each $w_i \in \bar{w}$ is a good approximation of some $\omega \in S$.

We would obtain slightly different formulation of the problem if we wished to find a circuit whose spectrum only contains one given set of frequencies and/or does not contain other sets, or if we permit some more general form of a sparse set for $\bar{w}$ instead of l_i/M. These versions of the problem have similar solutions.

As above, we find the recognizing algorithm in the GSA form

$$(I_{\tilde{0}} I_{\mathrm{cir},\bar{w}})^t, \tag{3}$$

where $\tilde{0}$ is an arbitrarily chosen vector from the space spanned by codes of the circuits, $t = O(\sqrt{T})$, where T is the number of all possible circuits and $I_{\mathrm{cir},\bar{w}}$ is the reflection along all codes $[C]$ such that $\mathrm{Spectr}(U_C)$ is determined by $\bar{w}$. It now suffices to build $I_{\mathrm{cir},\bar{w}}$.

We choose $B_f = O(\sqrt{Q})$ such that a randomly chosen vector $w \in F$ satisfies

$$|\langle w|w_1\rangle| > 1/B_f$$

with probability 0.99. Let $\mathrm{GenFreq}_j$ and $\mathrm{GenTimeFreq}_j$ be the respective operators generating a linear combination of frequencies $\tilde{\omega}_j \in F$ and a time instant $t_{\mathrm{freq},j} \le B_f$; all these objects are taken from the corresponding uniform distributions over all possible values and the code of the gate array generating the inversion along the corresponding state $\tilde{\omega}_j$. These operators generate objects in the corresponding ancillary registers. We let ω_j be the frequency contained in the jth register (initially, $\tilde{\omega}_j$).

We assume that the code of the circuit generating U is fixed and define the operator $I_{\text{cir},\bar{w}}$ by

$$I_{\text{cir},\bar{w}} = \bigotimes_j [\text{GenFreq}_j^{-1}\text{GenTimeFreq}_j^{-1} \circ (I_{\text{BadFreq},\bar{w},j} I_{\tilde{\omega}_j})^{t_{\text{freq},j}}]\text{SignGoodFreq} \bigotimes_j [(I_{\tilde{\omega}_j} I_{\text{BadFreq},\bar{w},j})^{t_{\text{freq},j}}\text{GenFreq}_j\text{GenTimeFreq}_j],$$

where $I_{\text{BadFreq},\bar{w},j}$ inverts the sign of states with "bad frequencies" in the jth register; these are the values of ω_j of the form l/M, $l \in \{0, 1, \ldots, M-1\}$ that either belong to $\bar{w}$ and are not a good approximation of frequencies $\omega \in \text{Spectr}(V)$ or do not belong to $\bar{w}$ but have a close frequency

$$\omega \in \text{Spectr}(V): \quad |\omega_j - \omega| \le \frac{1}{L};$$

on all other frequencies, this operator acts as the identity. Application of the sequence preceding SignGoodFreq concentrates the amplitude on bad frequencies. We note that $I_{\tilde{\omega}_j}$ can be implemented by a given code by means of the quantum Klini operator App. The subsequent application of SignGoodFreq inverts the sign of a state depending on whether bad frequencies are present. Namely, SignGoodFreq changes the sign for codes $[C]$ without bad frequencies and does nothing for codes $[C]$ with bad frequencies. The subsequent operators clean all ancillae. Therefore, $I_{\text{cir},\bar{w}}$ defined this way inverts the sign of exactly those codes C for which $\text{Spectr}(U_C)$ is determined by $\bar{w}$. We need to define two types of operators: SignGoodFreq and $I_{\text{BadFreq},\bar{w},j}$.

With each ω_j contained in the jth register, we associate a family of registers enumerated by two indices j, k and containing the frequencies $\omega_{j,k}$.

Definition 3. *A family of all $\omega_{j,k}$ is called good if the following property is satisfied for at least 1/5 from all j: for at least 1/10 of all k, $\omega_{j,k} \approx \omega_j \in \bar{w}$.*

The registers enumerated by different k for a fixed j are designed for the application of the jth copy of the operator State^{ω} defined in the previous section. Here, it is given by State^{ω_j}. Each k corresponds to the operator St_k in definition (2) such that each $\omega_{j,k}$ is the frequency obtained from the result of the action of St_k.

We first build the operator $I_{\text{BadFreq},\bar{w},j}$. We set

$$I_{\text{BadFreq},\bar{w},j} = \bigotimes_{j,k} [(\text{State}^{\omega_j})^{-1}\text{Rest}_{j,k}] \circ \text{Sign}' \bigotimes_{j,k} [\text{Rev}_{j,k}\text{State}^{\omega_j}],$$

where the operator Sign' changes the sign of only states with bad families of frequencies.

It was shown in the previous section that if a frequency ω_j is bad, we can only have $\omega_{j,k} \approx \omega_j \in \bar{w}$ for the vanishing part of all k, and before Sign', almost the entire probability is concentrated on bad families $\omega_{j,k}$; therefore, $I_{\text{BadFreq},\bar{w},j}$ changes the sign.

If ω_j is good, then it belongs to $\bar{w}$ and has a close $\omega' \in S$. It follows from the previous section that about $7/8 \times 1/4 = 7/32 > 1/5$ of all k satisfy $\omega_{j,k} \approx \omega \in \bar{w}$ and almost the entire probability before Sign' is concentrated on good families; therefore the sign is unchanged.

Hence, $I_{\text{BadFreq},\bar{w},j}$ is defined correctly.

We set

$$\text{SignGoodFreq} = \bigotimes_{j,k} [(\text{State}^{\omega_j})^{-1}\text{Rest}_{j,k}] \circ \text{Sign} \bigotimes_{j,k} [\text{Rev}_{j,k}\text{State}^{\omega_j}],$$

where the operator Sign changes the sign of only states with good families of frequencies. If a frequency ω_j is not bad, then about $7/8 \times 1/4 = 7/32$ of all k satisfy $\omega_{j,k} \approx \omega_j \in \bar{w}$. If a frequency ω_j is bad, we can only obtain $\omega_{j,k} \approx \omega_j \in \bar{w}$ for the vanishing part of k, as shown in the previous section. Thus, SignGoodFreq acts as required.[7]

We now calculate the complexity of our algorithm for recognizing a molecular circuit. The first factor $\sqrt{T}$ immediately follows from (3). The next factor $\sqrt{Q}$ follows from the definition of $I_{\text{cir},\bar{w}}$. Finally, the definition of $I_{\text{BadFreq},\bar{w}}$ gives the factor $M\sqrt{N}$. The resulting complexity is on the order of $M\sqrt{TNQ}n^2$.

3.4. Distinguishing Eigenvectors of Two Operators with the Same Eigenvalue

We now consider the most difficult of our problems, the problem of recognition of electronic devices. The difficulty is that we need not find a circuit with a given spectrum, but must simulate the action of a given circuit. We recall that we now assume that frequencies can be determined within $1/L$ given their approximation within $1/M$, where $L \gg M$.

As a first step, we consider the following problem: given two operators U and V having the same eigenvalue ω, the difference between the corresponding

[7] Again, we could take arbitrary ρ_1: $0 < \rho_1 < 1$ instead of one-tenth and ρ_2: $0 < \rho_2 < 7/32$ instead of one-fifth in the definition of a good family.

eigenvectors must be found. We let L_ω^U and L_ω^V be the subspaces spanned by the eigenvectors of U and V corresponding to all frequencies $\omega' \approx \omega$. (A particular case is where ω is a frequency of U but not of V. Here, $L_\omega^V = \emptyset$ and our algorithm is applicable in this situation.) We omit the index ω from the notation. For $u \in L^U$, $\|u\| = 1$, we set

$$\mu_u = \min\{\sqrt{1 - |\langle u|v\rangle|^2} \,\big|\, v \in L^V, \|v\| = 1\},$$

which is the sine of the angle between u and the subspace L^V, or the distance between u and this subspace; we define μ_v for $v \in L^V$, $\|v\| = 1$, similarly. We set

$$\mu_U = \max_{u \in U} \mu_u, \quad \mu_V = \max_{v \in V} \mu_v.$$

Then $\mu_U = 0$ implies that $U \subseteq V$. If the dimensions of the spaces L^U and L^V are equal, then $\mu_U = \mu_V$; if they are not equal, e.g., $\dim L^U > \dim L^V$, then $\mu_U = 1$. Let $d = d(N)$ be some function taking values in (0, 1]. We call these subspaces d-distinguishable if one of μ_U, μ_V is not less than d, or if one of the subspaces is empty and the other is nonempty.

We construct a procedure that determines whether these subspaces are the same provided they can be either d-distinguishable or coincident. The smaller the values the function $d(N)$ takes, the more accurate our recognition. Let $L^U \cap L^V = L_0$. Then $L^U = L_0 \oplus L_U'$ and $L^V = L_0 \oplus L_V'$. We note that if $L_U' \neq \emptyset$, then for all vectors from L_U' of length 1, their distances from L^V are exactly μ_U, and the same is true for L^V if L_V' is not empty. Let L' be the linear subspace spanned by vectors from $L_V' \cup L_U'$, and $\mathrm{Proj}_A B$ be the projection of a subspace B to a subspace A. If $\dim L^U > \dim L^V$, we have the decomposition into a sum of orthogonal subspaces,

$$L^U = L_U'' \otimes \mathrm{Proj}_{L^U} L^V,$$

where L_U'' is the subspace in L^U consisting of vectors orthogonal to L^V. Let L_V'' be defined symmetrically. Then either

(1) $L^U = L^V$ or

(2) $\dim L^U = \dim L^V$ and $L' \neq \emptyset$, or

(3) $\dim L^U > \dim L^V$ and $L_U'' \neq \emptyset$, or

(4) $\dim L^U < \dim L^V$ and $L_V'' \neq \emptyset$.

We define the main operator determining the equality of L^U and L^V by

$$\begin{aligned} \mathrm{Difference} &= \mathrm{Differ}^{-1}\mathrm{SignDif}\,\mathrm{Differ}, \\ \mathrm{Differ} &= \mathrm{Dif}_{\text{same dim}}\mathrm{Dif}_{L^U > L^V}\mathrm{Dif}_{L^U < L^V} \\ &\quad \circ \mathrm{Dif}^{\mathrm{ort}}_{L^U > L^V}\mathrm{Dif}^{\mathrm{ort}}_{L^U < L^V}, \end{aligned} \tag{4}$$

where SignDif changes the sign of the main ancilla α_{dif} iff at least one ancilla in the list

$$\bar{\alpha} = \{\alpha_{\text{same dim}}, \alpha_{L^U > L^V}, \alpha_{L^U < L^V}, \alpha^{\mathrm{ort}}_{L^U > L^V}, \alpha^{\mathrm{ort}}_{L^U < L^V}\}$$

contains 1, and each operator of the type Dif changes the corresponding ancilla from $\bar{\alpha}$ in the following cases:

(1) $\dim L^U = \dim L^V$ and $L^U \neq L^V$,

(2) $\dim L^U > \dim L^V$ and $\mu_V < \sqrt{2/3}$,

(3) $\dim L^U < \dim L^V$ and $\mu_U < \sqrt{2/3}$,

(4) $\dim L^U > \dim L^V$ and $\mu_V > \sqrt{1/3}$, or $L^V = \emptyset$, $\dim L^U < \dim L^V$ and $\mu_U > \sqrt{1/3}$, or $L^U = \emptyset$;

these operators do nothing if $L^U = L^V$. In view of symmetry, it is sufficient to define the Dif operators in the first, second, and fourth cases. We note that the first case, $\dim L^U = \dim L^V$, is the only nondegenerate case and the corresponding definition of Dif is more difficult.

Definition of $\mathrm{Dif}_{\text{same dim}}$. We suppose that $\dim L^U = \dim L^V$. Our first aim is to build an operator Inv that acts as the identity if L^U and L^V are coincident and that acts as $I_{L'}$ if they are d-distinguishable. We arrange the first two ancillary qubits α_U and α_V that signal whether a given state has the projection to L^U or correspondingly, to L^V of a length of at least one-third. We consider the operator

$$\begin{aligned} \mathrm{Check} &= \bigotimes_s \mathrm{Rest}_s^V \mathrm{Anc}_V \bigotimes_s \mathrm{Rev}_s^V \\ &\quad \bigotimes_s \mathrm{Rest}_s^U \mathrm{Anc}_U \bigotimes_s \mathrm{Rev}_s^U, \end{aligned}$$

where Anc inverts the corresponding ancilla iff at least nine-tenths of the copies for the respective frequencies are equal to ω within $1/M$. It coincides with the inverse operator Check^{-1}.

Let t be some random integer from the segment $[0, [2/d]]$. We define the operator

$$\mathrm{Turn}_t = (I_{L^U} I_{L^V})^t \tag{5}$$

of Grover's type. Two subspaces L^U and L^V are said to be almost orthogonal iff $\sqrt{1 - \mu^2} \leq 1/30$ for some $\mu \in$

$\{\mu_U, \mu_V\}$. If L^U and L^V are not almost orthogonal, then given some $a \in L'_U$ ($a \in L'_V$), the average distance between $\mathrm{Turn}_t|a\rangle$ and L_U (L_V) is at least one-half if L^U and L^V are d-distinguishable and zero if these subspaces are coincident. To distinguish the close location and near-orthogonality cases, we build two operators, $\mathrm{Dist}_{\mathrm{ort}}$ and $\mathrm{Dist}_{\mathrm{closed}}$.

We first suppose that L^U and L^V are almost orthogonal. Then $\alpha_U = 1$ implies that $\alpha_V = 0$. We introduce the notation

$$L(\alpha_U, \alpha_V) = \begin{cases} L^V & \text{if} \quad \alpha_U = 1 \\ L^U & \text{if} \quad \alpha_V = 1. \end{cases}$$

Let $\bar{a}$ be a vector from the space of inputs. We note that $L^U \neq L^V$ implies $\alpha_U = \alpha_V$ for each $\bar{a} \perp L'$ because $\bar{a}$ then belongs to the subspace spanned by L_0 and the orthogonal subspace to $L^U \cup L^V$. The first operator $\mathrm{Dist}_{\mathrm{ort}}$ does nothing if $\alpha_U = \alpha_V$ and changes the sign and special ancilla α_{ort} if the projection of $\bar{a}$ to $L(\alpha_U, \alpha_V)$ is less than 1/30.

The second operator $\mathrm{Dist}_{\mathrm{closed}}$ acts as the identity if $\alpha_U = \alpha_V$ and changes the sign if the following conditions are satisfied simultaneously: $\bar{a} \in L'$, L^U and L^V are distinguishable, and $\alpha_{\mathrm{ort}} = 0$.

We set

$$\mathrm{Dist}_{\mathrm{ort}} = \bigotimes_j \mathrm{Res}_j \mathrm{Si}_{\neq\omega} \bigotimes_j \mathrm{Re}_j,$$

where Re (Res) denotes

$$\mathrm{Rev}^V(\mathrm{Rest}^V) \quad \text{if} \quad \alpha_U = 1, \quad \alpha_V = 0,$$

$$\mathrm{Rev}^U(\mathrm{Rest}^U) \quad \text{if} \quad \alpha_V = 1, \quad \alpha_U = 0,$$

and the identity if $\alpha_U = \alpha_V$; $\mathrm{Si}_{\neq\omega}$ changes the sign and simultaneously inverts α_{ort} iff at least half the frequencies ω_j are such that $|\omega_j - \omega| > 1/M$ and $\alpha_U \neq \alpha_V$. If we want to clean the second ancilla after the action of $\mathrm{Dist}_{\mathrm{ort}}$ and keep the sign change, we can use the operator

$$\mathrm{Dist}^-_{\mathrm{ort}} = \bigotimes_j \mathrm{Res}_j S_{\neq\omega} \bigotimes_j \mathrm{Re}_j,$$

where S acts as Si but without changing the sign.

The second operator is defined by

$$\mathrm{Dist}_{\mathrm{closed}} = D_1^{-1} \ldots D_n^{-1} S' D_n D_{n-1} \ldots D_1,$$

$$D_j = (\mathrm{GenTimeArg}_j)^{-1} (\mathrm{Turn}^j_{t_j})^{-1} \circ [\bigotimes_k \mathrm{Rest}^U_{j,k}] \mathrm{Sig}^j_{\neq\omega} [\bigotimes_k \mathrm{Rev}^U_{j,k}] \mathrm{Turn}^j_{t_j} \mathrm{GenTimeArg}_j,$$

$$j = 1, 2, \ldots, n,$$

where the operator $\mathrm{Sig}^j_{\neq\omega}$ changes the corresponding ancilla β_j only in one of two cases:

(1) $\alpha_U = 1$ and at least a half of $\omega_{j,k}$ are such that $|\omega_{j,k} - \omega| \geq 1/M$, or

(2) $\alpha_U = 0$, $\alpha_V = 1$ and at least a half of $\omega_{j,k}$ are such that $|\omega_{j,k} - \omega| < 1/M$.

The operator S' changes the sign iff one of α_U, α_V is nonzero and at least 1/20 of all β_j contain 1.

We consider the action of $\mathrm{Dist}_{\mathrm{closed}}$ following Check on an input vector $\bar{a}$. We first consider the case where $L^U \neq L^V$, which implies that these subspaces are distinguishable.

If $\bar{a} \perp L^U, L^V$, then $\alpha_U = \alpha_V = 0$ and $\mathrm{Dist}_{\mathrm{closed}}$ does nothing.

If $\bar{a} \in L_0$, then $\alpha_U = \alpha_V = 1$ and all $\mathrm{Sig}^j_{\neq\omega}$ do nothing because for almost all j, about three-fourths of $\omega_{j,k}$ are close to ω, $|\omega_{j,k} - \omega| \leq 1/M$, and hence, S' and $\mathrm{Dist}_{\mathrm{closed}}$ does nothing.

Let $\bar{a} \in L'$. We prove that $\mathrm{Dist}_{\mathrm{closed}}$ changes the sign. We decompose L' into the sum of orthogonal subspaces, $L' = L'_U \oplus L'^{\mathrm{ort}}_U$, and let $\bar{a}_j$ denote the result of the action of $\mathrm{Turn}^j_{t_j}$ on $\bar{a}$.

If $a \in L'_U$, then $\alpha_U = 1$ and for more than one-tenth of all $\bar{a}_j$, the revealed frequencies are not close to ω with a probability of about $3/4 \times 9/10$, and the sign is therefore changed in accordance with case 1.

If $\bar{a} \in L'^{\mathrm{ort}}_U$, then by the same reason we obtain the change of sign in accordance with case 2. Hence, $\mathrm{Dist}_{\mathrm{closed}}$ changes the sign for all $\bar{a} \in L'$.

We can now define Inv as

$$\mathrm{Inv} = \mathrm{Check}\,\mathrm{Dist}^-_{\mathrm{ort}}\,\mathrm{Dist}_{\mathrm{closed}}\,\mathrm{Dist}_{\mathrm{ort}}\,\mathrm{Check}.$$

For $a \perp L^U, L^V$, we have $\mathrm{Inv}|a\rangle = |a\rangle$ because Check gives zero in the ancilla α_U, α_V, thereby depriving the subsequent operators of the ability to change the state

vector. If $a \in L_0$, then $\mathrm{Inv}|a\rangle = |a\rangle$ because $\mathrm{Dist}_{\mathrm{ort}}$ does nothing and $\mathrm{Dist}_{\mathrm{closed}}$ does nothing as well. Thus,

$$\mathrm{Inv}|a\rangle = |a\rangle \quad \text{for} \quad \bar{a} \perp L',$$

and

$$\mathrm{Inv}|a\rangle = -|a\rangle \quad \text{for} \quad a \in L'.$$

We are now ready to construct the operator $\mathrm{Dif}_{\text{same dim}}$ inverting the ancilla $\alpha_{\text{same dim}}$ iff L^U and L^V are distinguishable. Let Gen generate the list y, $[I_y]$, $[C_Z]$, where $[C_Z]$ is the code of a circuit generating some unitary operator $Z = Z^{-1}$ whose only eigenvalues are 1 and –1 (that is, its frequencies are 0 and 1/2), the space corresponding to frequency 0 is one-dimensional, and y is a basic vector of this space. As usual, the index j means that the corresponding vectors y_j are taken from the uniform distribution on all possible vectors. We assume that operators of the form Gen^{-1} are also accessible, and we set

$$\begin{aligned}&\mathrm{Dif}_{\text{same dim}}\\ &= \bigotimes_j [\mathrm{GenTimeArg}_j^{-1}\mathrm{Gen}_j^{-1}(\mathrm{Inv}_j I_{y_j})^{t_j}\mathrm{Rest}_j^{Z_j}] \qquad (6)\\ &\circ \mathrm{Change}\bigotimes_j [\mathrm{Rev}_j^{Z_j}(I_{y_j}\mathrm{Inv}_j)^{t_j}\mathrm{Gen}_j\mathrm{GenTimeArg}_j],\end{aligned}$$

where each copy of Inv acts on the register where y_j is placed initially and Change makes the desired change in the resulting qubit $\alpha_{\text{same dim}}$ if at least 5/32 of all frequencies differ from 0 by more than $1/M$.

The group $(I_{y_j}\mathrm{Inv}_j)^{t_j}$ of the GSA type turns the vector y_j generated by Gen_j substantially iff L^U and L^V are d-distinguishable.

If $L^U = L^V$, then y_j remains unchanged and at least seven-eighths of all frequencies are close to 0.

If $L^U \neq L^V$, then at least $7/8 \times 1/4 = 7/32$ of frequencies for the result of the turn of y_j are far from 0 because they must be close to 1/2.[8]

Definition of $\mathrm{Dif}_{L^U > L^V}$. We suppose that $\dim L^U > \dim L^V$ and $\mu_V < \sqrt{2/3}$, and recall the decomposition

$$L^U = L_U'' \oplus \mathrm{Proj}_{L^U} L^V$$

into the sum of orthogonal subspaces with $L_U'' \neq \emptyset$. We define the operator Dif very similarly to the previous case,

$$\begin{aligned}&\mathrm{Dif}_{L^U > L^V}\\ &= \bigotimes_j [\mathrm{GenTimeArg}_j^{-1}\mathrm{Gen}_j^{-1}(\mathrm{Inv}''_{j,U} I_{y_j})^{t_j}\mathrm{Rest}_j^{Z_j}]\\ &\circ \mathrm{Change}\bigotimes_j [\mathrm{Rev}_j^{Z_j}(I_{y_j}\mathrm{Inv}''_{j,U})^{t_j}\mathrm{Gen}_j\mathrm{GenTimeArg}_j],\end{aligned}$$

where the definition of Inv_U'' (which inverts L_U'') is similar to the definition of $\mathrm{Dist}_{\mathrm{ort}}$ with L_U'' playing the role of L',

$$\mathrm{Inv}_U'' = \mathrm{Check}[\bigotimes_k \widetilde{\mathrm{Res}}_k^V]\tilde{\mathrm{Si}}_{\neq\omega}[\bigotimes_k \widetilde{\mathrm{Re}}_k^V]\mathrm{Check}.$$

Here, $\widetilde{\mathrm{Re}}^V$ and $\widetilde{\mathrm{Res}}^V$ act as Rev^V and Rest^V only if $\alpha_U = 1$; if $\alpha_U = 0$, they do nothing, and the operator $\tilde{\mathrm{Si}}_{\neq\omega}$ changes the sign in only one case, if $\alpha_U = 1$ and at least three-fourths of all frequencies ω_k are far from ω: $|\omega_k - \omega| \geq 1/M$. In the operator Dif, we therefore use a set of ancillary registers enumerated by the pairs of indices j, k.

For $\bar{a}_j \in \mathrm{Proj}_{L^U} L^V$, in view of the inequality $\mu_V < \sqrt{2/3}$, the operator $\tilde{\mathrm{Si}}_{\neq\omega}$ does not change the sign because the fraction of all frequencies close to ω is then $7/8 \times 1/3 = 7/24 > 1/4$.

For $\bar{a}_j \perp \mathrm{Proj}_{L^U} L^V$, the operator Inv_U'' does nothing.

Definition of $\mathrm{Dif}^{\mathrm{ort}}_{L^U > L^V}$. We suppose that $\dim L^U > \dim L^V$ and $\mu_V > \sqrt{1/3}$. The definition of Dif is similar to the previous case but with the entire subspace L_U playing the role of L',

$$\begin{aligned}&\mathrm{Dif}^{\mathrm{ort}}_{L^U > L^V}\\ &= \bigotimes_j [\mathrm{GenTimeArg}_j^{-1}\mathrm{Gen}_j^{-1}(\mathrm{Inv}_{j,U} I_{y_j})^{t_j}\mathrm{Rest}_j^{Z_j}]\\ &\circ \mathrm{Change}\bigotimes_j [\mathrm{Rev}_j^{Z_j}(I_{y_j}\mathrm{Inv}_{j,U})^{t_j}\mathrm{Gen}_j\mathrm{GenTimeArg}_j],\end{aligned}$$

where

$$\mathrm{Inv}_U = \mathrm{Check}[\bigotimes_k \widetilde{\mathrm{Res}}_k^V]\tilde{\mathrm{Si}}^{\mathrm{ort}}_{\neq\omega}[\bigotimes_k \widetilde{\mathrm{Re}}_k^V]\mathrm{Check}.$$

Here, $\tilde{\mathrm{Si}}^{\mathrm{ort}}_{\neq\omega}$ changes the sign if more than half the frequencies are far from ω: $|\omega_j - \omega| > 1/M$. The conditions

[8] Thus, we could take any number ρ: $1/8 < \rho < 7/32$ instead of 5/32 in the definition of Change.

required for the operator Dif are satisfied because $7/8 \times 2/3 = 7/12 > 1/2$ and can be checked straightforwardly.

We finally estimate the complexity of the procedure constructed. The operator Turn in (5) requires the number of elementary steps on the order of

$$\mathrm{Turn}_{\mathrm{complexity}} = M\sqrt{1/d}.$$

The operator Difference in (4) then requires the number of elementary steps on the order of $\mathrm{Turn}_{\mathrm{complexity}}\sqrt{N}$, that is, $O(M\sqrt{N/d})$. We note that there exists a similar form of the operator Difference that does not act on the resulting qubit α_{dif} but changes the sign instead; such an operator can be constructed similarly. We let it be denoted by $\mathrm{Difference}_{\mathrm{sign}}$ and assume that its input contains the frequency ω.

3.5. Recognition of Electronic Device Circuits

We are now ready to consider the recognition of circuits. We assume that for every pair of circuits with the transformations U_1 and U_2, the subspaces spanned by the corresponding eigenvalues are either coincident or d-distinguishable. We also assume that our coding procedure gives a one-to-one correspondence between circuits and the T basic states $e_0, e_1, \ldots, e_{T-1}1$ in the space H_{cir}. The recognition procedure is denoted by Rec and has the GSA form,

$$\mathrm{Rec} = (I_{\tilde{0}}I_U)^t, \quad t = O(\sqrt{T}). \tag{7}$$

This operator acts on states of the form $|\chi\rangle$, where the basic states for χ are codes of circuits. Here, $\tilde{0} \in H_{\mathrm{cir}}$ is chosen arbitrarily and I_U inverts the sign of every code whose circuit induces a given operator U. The implementation of $I_{\tilde{0}}$ is straightforward, and all we need to do is to construct I_U.

We define I_U as

$$I_U = \bigotimes_j [\mathrm{Conc}^{-1}_{\mathrm{freq},\,j}\mathrm{Difference}_j]\mathrm{Sign}\bigotimes_j [\mathrm{Difference}_j\mathrm{Conc}_{\mathrm{freq},\,j}],$$

where for every basic state C of the argument, $\mathrm{Conc}_{\mathrm{freq}}$ generates some arbitrary distribution of the amplitude on ancillary registers with Q basic states and then concentrates a substantial part of the amplitude on a frequency ω for which L^U and L^V are distinguishable (if such a frequency exists). The operator $\mathrm{Difference}_j$ then changes the resulting qubit for the jth copy iff these subspaces are distinguishable on this frequency. The next operator Sign changes the sign iff at least one fifth of the resulting qubits α_{dif} contain 1, e.g., iff U and U_C are the same operator. The subsequent applications of $\mathrm{Difference}_j$ to each copy of the register then clean the corresponding resulting qubits and the inverse operators to Conc_j restore the initial state of the ancillary register. The difference was constructed in the previous section, and it only remains to construct $\mathrm{Conc}_{\mathrm{freq},\,j}$. This transformation can be defined as

$$\mathrm{Conc}_{\mathrm{freq},\,j} = \mathrm{GenTimeFreq}_j^{-1}\mathrm{GenFreq}_j^{-1} \circ (\mathrm{Difference}_{\mathrm{sign}}I_{\omega_j})^{t_j}\mathrm{GenFreq}_j\mathrm{GenTimeFreq}_j. \tag{8}$$

If U and U_C are different, then their subspaces L^U and L^V are d-distinguishable for some ω by our assumption and Conc_j concentrates a substantially large part of the amplitude over all j on some combination of such values ω. Thus, we have constructed the required procedure Rec that gives the target code with a substantial probability as the result of an observation of the register for the code C. After the observation, we can verify the fitness of the code C, which is found by a straightforward procedure. This procedure is similar to I_U with a single change: Sign is to be replaced by a change in a special ancilla that can be observed after the procedure; we thus determine whether the code C fits.

To find the complexity of our procedure Rec, we note that the complexity $Mn^2\sqrt{N/d}$ of Difference must be multiplied by $\sqrt{Q}$ following from (8) and by $\sqrt{T}$ following from definition (7). The resulting complexity is $Mn^2\sqrt{TQN/d}$.

3.6. Advantages of the Recognition Algorithms

Advantages of the proposed algorithms are their high speed and small memory. In particular, the algorithm for molecular structure recognition allows recognizing molecular circuits using microscopic memory, whereas classically this task requires exponentially large memory. We now compare the proposed algorithms with their classical counterparts; we omit logarithmic multipliers.

1. Recognition of eigenvalues and finding thermodynamic functions. We fix some value of M that determines the precision of the eigenvalue approximation. We first consider the case where the number of ancillary qubits in a quantum gate array is small. By the direct classical method, we must then construct the matrix of the unitary transform induced by the gate array. This requires an order of N^3 steps and at least an order of N^2 bits. The known quantum algorithm given by Travaglione and Milburn in [8], based on the Abrams and Lloyd operator Rev, contains repeated measurements of frequencies and therefore requires time on the order of NM; for sparse spectra, it is of the same order as for the Hams–Raedt algorithm and its only advantage over the latter is exponential memory saving.

Our algorithm recognizes an eigenvalue in $\sqrt{N}Mn$ steps. This time for the sparse area of the spectrum is about the square root of the time of the best known algorithms. Here, the memory is on the order of g^2 qubits (g is the size of the gate array), that is, about the squared memory used in [2], but still exponentially smaller than in classical methods. The proposed algorithm therefore gives an essential speedup over the known methods in the case where the number of ancillary qubits in a given gate array is small (as in the case of a molecular structure simulated by the gate array) and the area of the spectrum is sparse. The same advantage is possessed by the proposed method of finding thermodynamic functions.

If the spectra are dense, we assume that $M = N$, which means that eigenvalues differ by at least $1/N$. The time of our algorithm is then $O(N)$.

We next consider the case where the number a of ancillary qubits involved in the gate array simultaneously is greater than the length n of the input. The direct classical method then requires more than 2^{2a} steps and at least 2^m bits, whereas our algorithm requires only about $g2^n$ steps and gn^2 memory and the quantum speed-up can be more than the square root.

2. Recognition of molecular structures. We first assume that the spectra are sparse. To be able to compare our method with the evident classical algorithm, we assume that the code of a molecular circuit of length n is a string of ones and zeroes of this length. Therefore, $M = N$. The next natural assumption that can also be presumed for electronic circuits is that the sampling of the code of a circuit from the uniform distribution induces a sampling of all possible spectra from the uniform distribution. Then the number of all possible choices of spectrum approximations (or parts of the spectrum subject to the statement of the recognition problem) within $1/L$ consisting of frequencies of the form l/M is about $2^M = N$. This implies that M and Q must be logarithmic in N in our assumption. Our method therefore has the time complexity $O(N)$. With these assumptions, the time complexity of the classical direct algorithm that examines all codes and calculates the corresponding spectra is about $N^3 \cdot N = N^4$, whereas our algorithm requires a time of about N and logarithmic memory. The quantum time for this problem is therefore about the fourth root of the time of the classical direct method, and the quantum space is logarithmic.

If the spectra are dense, then Q and M are on the order of N and our method requires a time of $O(N^{2.5})$, to be compared with $O(N^4)$ of the direct classical method.

3. Recognition of electronic devices. There are no classical analogs of this problem in the general case. We compare the two algorithms constructed above with their classical and known quantum counterparts. We first consider a single quantum recognition algorithm that can easily be deduced from the previously known technique. This is the algorithm of recognizing a circuit that realizes a classical involutive function of the form f:

$$Q \longrightarrow Q, \quad f = f^{-1}.$$

This task can be reduced to the search of y such that the following logic formula is true: $\forall x\ A(x, y)$, where $A(x, y)$ is some predicate. Indeed, if we take $Y(x) = U(x)$ instead of $A(x, y)$, where Y is a function whose code is y, we obtain the problem of recognition of the circuit generating U. The algorithm for such formulas, given in [4], has time complexity on the order of $\sqrt{TN}$. This task is a particular case of our algorithm for involutive devices and has the same complexity. In this particular case, quantum time is on the order given by the square root of the classical time. However, if we consider a slightly more general, but still restricted, problem of the recognition of involutive devices producing linear combinations of basic states (like quantum subroutines), the advantage over the classical method of recognition increases. For example, we consider the restricted problem where we must choose between two alternative constructions of a tested device, inducing a nonclassical unitary transformation. The naive method of observing the results of the action of the tested device on the different inputs requires steps on the order of $(1/\epsilon)N^3$ to restore the matrix of the operator U_C within ϵ. This ϵ must then be less than $1/\sqrt{N}$ to give a vanishing difference between operators in the Hilbert space. Therefore, the time complexity of the naive method of recognition is roughly $N^{7/2}$. On the other hand, the method proposed in Section 3.4 requires choosing d that only converges to zero as N tends to infinity. The time required by our method is therefore slightly more than $\sqrt{N}$. We thus have almost the seventh degree speed-up for the problem of distinguishing electronic circuits generating transformations with nonclassical matrices.

4. CONCLUSIONS

The main conclusion is that the molecular structure and physical properties of environment can be quickly recognized on the microscopic level, whereas the classical methods require much time and especially memory. The new algorithms of recognizing eigenvalues with a fixed precision, recognizing the molecular structure, and finding thermodynamic functions give a quadratic speed-up and an exponential memory saving compared with the best classical algorithms. The new method based on quantum computing was proposed for fast recognition of electronic devices. By this method, two devices with the same given spectrum can be distinguished in a time of about the seventh root of the time of direct measurements. All these algorithms show

significant potential advantages of microscopic-size quantum devices compared to their classical counterparts with much larger memory. The advantages pertain to intellectual tasks like recognition of the structure of other devices and important properties of environment. The proposed algorithms are constructed from the standard known subroutines; they have a simple structure and are entirely within the framework of the conventional quantum computing paradigm.

ACKNOWLEDGMENTS

I am sincerely grateful to K. Valiev for creating the conditions for investigations in quantum computing at the Institute of Physics and Technology and for his attention and valuable advice concerning my work.

REFERENCES

1. A. Hams and H. de Raedt, E-print archives, physics/0004016.
2. D. S. Abrams and S. Lloyd, E-print archives, quant-ph/9807070.
3. P. W. Shor, Siam J. Sci. Stat. Comp. **26**, 1484 (1997).
4. H. Buhrman, R. Cleve, and A. Wigderson, in *Proceedings of STOC 98* (Los Angeles, CA, USA, 1998), p 63.
5. L. K. Grover, in *Proceedings of STOC 96* (Santa Fe, NM, USA, 1996), p. 212.
6. M. Boyer, G. Brassard, P. Hoyer, and A. Tapp, Fortschr. Phys. **46**, 493 (1998).
7. Y. Ozhigov, E-print archives, quant-ph/0004021.
8. B. C. Travaglione and G. J. Milburn, E-print archives, quant-ph/0008053.
9. G. Brassard, P. Hoyer, and A. Tapp, E-print archives, quant-ph/9805082.

Journal of Experimental and Theoretical Physics, Vol. 96, No. 2, 2003, pp. 355–382.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 123, No. 2, 2003, pp. 399–428.

REVIEWS

Density of Prelocalized States in Mesoscopic NS Systems

P. M. Ostrovsky*, M. A. Skvortsov, and M. V. Feigel'man

Landau Institute for Theoretical Physics, Russian Academy of Sciences,
ul. Kosygina 2, Moscow, 117940 Russia
**e-mail: ostrov@itp.ac.ru*

Received September 27, 2002

Abstract—The semiclassical theory of the proximity effect predicts the formation of a gap $E_g \sim \hbar D/L^2$ in the excitation spectrum of a diffusive contact between a normal metal and a superconductor (NS). Mesoscopic fluctuations lead to the emergence of states localized anomalously in the normal metal and weakly linked with the superconducting bank, creating a nonzero density of states for energies lower than E_g. In this review, the behavior of the density of quasiparticle states below a quasi-classical gap is considered for various geometries of the NS system (special attention is paid to SNS junctions) and for the problem of a superconductor with a low concentration of magnetic impurities, in which a similar effect is observed. Analysis is mainly carried out on the basis of a fully microscopic method of the supermatrix σ model; in this method, a nonzero density of states emerges due to instanton configurations with broken supersymmetry. In addition, the results of an alternative approach proceeding from the idea of universality of the spectra of random Hamiltonians with the given symmetry are reviewed. In situations studied using both methods, the results are identical. They include the exact expression for the mean density of states of an NS system in the vicinity of E_g. In the framework of 1D and 2D σ models, the subgap density of states is determined with an exponential accuracy. The contacts with a poor transparency of the NS interface are also considered. It is shown that the number of subgap states in the case of low transparency is much greater than unity. © *2003 MAIK "Nauka/Interperiodica".*

CONTENTS

1. INTRODUCTION

Mesoscopic properties of metals are manifested when the coherence length of conduction electrons is equal to the characteristic size of the sample [1]. These properties are observed most clearly in small samples and are hence accompanied by strong mesoscopic fluctuations.

Superconductivity is another coherent (but not size) effect. It is due to Cooper attraction between electrons. Such an attraction leads to a rearrangement of the ground state of the electron system and radically changes low-energy properties of a metal. In modern experiments, both these conditions can be satisfied simultaneously, which leads to mesoscopic superconductivity. The most interesting and diverse effects are observed for hybrid structures formed by superconducting and normal parts. In the mesoscopic limit, such structures exhibit global coherence leading to phenomena known as the "proximity effect." These phenomena are reduced qualitatively to superconductivity suppression in the superconducting parts and to the emergence of certain superconducting properties in normal regions.

1063-7761/03/9602-0355$24.00

A typical example of such phenomena is the Josephson effect. Cooper pairs tunneling through an insulating layer partly preserve their coherence and, in this way, can carry supercurrent through such a layer. If two superconductors are linked via a normal metal region, an effect of this type has a more complex microscopic structure. In this case, Andreev reflection is a fundamental effect [2].

When an electron is incident on a superconductor–normal metal interface from the side of the normal metal, it cannot penetrate the superconductor since the given energy corresponds to a gap in the spectrum of the superconductor. However, Andreev reflection is possible in this case: the electron is reflected from the normal metal and becomes a hole, while a Cooper pair starts moving in the superconductor. Alternately, this process can be treated as tunneling of a Cooper pair from the superconductor to the normal metal. Although the attraction between electrons vanishes in this case, their combined state is partly coherent. If the normal layer is thin, such a pair may get in the second superconductor, carrying supercurrent in this way. The situation corresponds to a certain electron trajectory connecting two superconducting banks: an electron moving along this trajectory is transformed into a hole upon Andreev reflection, the hole repeating the electron path in the opposite direction, and the trajectory becomes closed after the second Andreev reflection. Such trajectories are allowed when they accommodate an integral number of wavelengths, giving rise to Andreev states [3]. These states form a discrete spectrum and are arranged symmetrically relative to the Fermi level (in the absence of current).

It can be seen from this example that the proximity of the superconductor changes the low-energy spectrum of the normal metal. Similar phenomena, which are also associated with Andreev reflection, may also occur in a simpler case of a single contact between a superconductor and a normal metal. In this case, Andreev states changing the spectrum of the normal metal also appear in the normal region. These changes are determined to a considerable extent by the classical dynamics of electrons in the normal part of the contact. If, for example, the normal region is rectangular in shape and contains no impurities, electron trajectories between two Andreev reflections existing in this region can be infinitely long. This leads to the formation of levels with an arbitrarily low energy and, hence to the absence of a gap in the spectrum. However, the density of states still linearly tends to zero as the Fermi energy is approached [4, 5]. In the general case, this type of spectrum appears when the classical dynamics of electrons in the normal region is integrable.

The opposite limit of chaotic dynamics is realized, for example, in the case of a high density of potential (nonmagnetic) scattering centers (impurities). Under such conditions, the motion of electrons is of the diffusive type. However, a naive attempt to determine the form of the spectrum by analyzing the probability of trajectories of various lengths leads to an erroneous result [6]. Indeed, in the case of random motion, we can always find infinitely long trajectories, but the spectrum will have a gap. The reason for this error lies in the disregard of quantum interference. As a matter of fact, semiclassical diffusive trajectories are broken lines: an electron is scattered successively by a large number of impurities. For two sufficiently long trajectories, we can always find an impurity in common. This means that, in addition to the two corresponding Andreev states, there exist at least two more states: an electron moving along the first trajectory and experiencing scattering by a common impurity passes to the second trajectory, while a hole after Andreev reflection returns to the first trajectory at the same impurity, and vice versa. In view of quantum interference between the above processes, low-lying Andreev levels cannot be described in the naive language of simple trajectories. An appropriate semiclassical technique is well known [7, 8] and is based (in the diffusion case) on the Usadel equation [9]. The result is reduced qualitatively to the emergence of a gap in the density of states on the order of $\hbar/\tau_c$, where τ_c is the characteristic time of diffusion between two Andreev reflections [4, 10–12]. It is determined by the strength and concentration of impurities, the size of the normal region, and the transparency of the interface with the superconductor.

However, the semiclassical theory disregards mesoscopic fluctuations. We can assume, on a qualitative level, that the diffusion coefficient fluctuates, leading to a deviation in the gap width in each specific sample from its mean value. As a result of averaging over possible configurations of impurities, the density of states decreases sharply for a certain energy instead of vanishing, its value being exponentially small for lower energies.

We can formulate the following general statement. If the position of the spectrum edge is determined by a fluctuating physical quantity, averaging over such fluctuations leads to the emergence of a "tail" in the density of states in the forbidden gap. This type of a tail was considered for the first time by Lifshits for an ordinary semiconductor [13]. At present, a large number of fluctuation effects have been observed in various systems (see, for example, [14–16]).

Disordered systems can be treated using a purely phenomenological approach known as the random matrix theory [17–19]. In the framework of this theory, a Hamiltonian is a random matrix and different matrix elements are regarded as uncorrelated (except the relation associated with additional symmetries of the Hamiltonian). In the main order in the large dimension of the matrix, the average density of states of a random Hamiltonian is a "Wigner semicircle,"

$$\langle \rho(E) \rangle = \delta^{-1} \sqrt{1 - E^2/E_0^2},$$

where E_0 is the bandwidth and δ is the mean distance between energy levels at the center of the band.

Owing to its universal nature, the random matrix theory has found wide applications [20] for describing spectral properties of mesoscopic systems. This is manifested in the fact that, in spite of their difference at the level of a microscopic Hamiltonian, the spectra of mesoscopic systems with chaotic dynamics and the spectra of random matrices with identical values of δ are statistically identical. This was demonstrated for the first time by Efetov [21] for a pair correlator of energy levels of a diffusive metallic grain. In this case, both systems are considered at a large distance from the band edge, when the mean density of states can be regarded as independent of energy.

The mean density of states in the Wigner–Dyson ensemble near the band edge vanishes in the semiclassical approximation in proportion to the square root. When corrections are taken into account, an exponentially decreasing tail appears in the semiclassical approximation for energies $|E| > E_0$. In the diffusive NS system, the edge of the spectrum near the gap is also of the root type. If we assume that the shape of the tail is completely determined by the semiclassical behavior of the density of states near the band edge (universality hypothesis), the result of the random matrix theory can be extended to the case of a diffusive NS system. This was done in [22].

Another case when an exponentially small tail appears in the density of states is a superconductor with magnetic impurities. The presence of magnetic impurities suppresses superconductivity. If their concentration is not very high, the gap in the spectrum becomes smaller than in a superconductor without impurities, but does not vanish. However, the impurity concentration may fluctuate in space; consequently the probability of finding an energy level below the mean value of the gap differs from zero. Such a tail in the density of states was calculated in [23, 24] using the method of the nonlinear supermatrix σ model.

Several methods have been developed for calculating electronic properties of systems with disorder. In traditional statistical physics, the properties of a system are mainly determined by the generating functional $Z[J] = \int e^{-S[\Phi, J]} D\Phi$. Various correlation functions, including the density of states, can be expressed in terms of logarithmic derivatives of this functional with respect to sources J. If the system is disordered, all correlation functions should be averaged over the disorder; i.e., the mean value of $\ln Z$ is required. However, the logarithm is a nonlinear function and its averaging is complicated in the general case. One of the methods for overcoming this difficulty, viz., replica trick, was proposed in [25]. In this method, n copies (replicas) are considered instead of a system. The generating functional in this case is Z^n. If the functional can be averaged over disorder for an arbitrary value of n and an analytic continuation in n to point $n = 0$ can be carried out, the formula $\ln Z = \lim_{n \to 0} (Z^n - 1)/n$ can be used.

However, the calculation of the generating functional in the general form for an arbitrary number of the replicas is often a complicated problem. This difficulty is removed in the method of a supermatrix σ model [26]. The essence of this method lies in the addition to physical fields of the same number of Grassmann (anticommuting) fields. For an arbitrary action of the system, the generating functional is equal to unity, and correlation functions are defined by conventional variational derivatives of this functional instead of logarithmic derivatives.

The semiclassical approximation (Usadel equation) corresponds to the evaluation of the generating functional by the steepest descent method. The corresponding saddle point of the action of the nonlinear σ model is supersymmetric; i.e., it has the same form in commuting and Grassmann variables. An exponentially small contribution from low-frequency mesoscopic fluctuations corresponds to other (nonsupersymmetric) saddle points, viz., instantons. Such a calculation was made for the first time in [16] for the density of states at a high Landau level in a 2D system in a magnetic field. The applicability of the steepest descent method in the vicinity of instantons was ensured by the large number of the Landau level. In the case of diffusive NS systems, the corresponding large parameter is the number of conducting channels at the interface between the normal metal and the superconductor.

A generalization of the σ model for diffusive NS systems was proposed in [27] and will be described briefly below. In the same publication, it was pointed out that the subgap density of states corresponds to instantons in this model. The instanton configuration responsible for the emergence of a tail in the density of states in a homogeneous superconductor with magnetic impurities [28] was determined in [23, 24]. Analysis of instantons and calculation of the density of prelocalized states in hybrid systems was carried out in [29]. It was found, among other things, that the contribution to the subgap density of states comes from two instantons. For energies not very close to the threshold, one of these instantons plays the major role, which makes it possible to determine the density of states with an exponential accuracy. In the case when the σ model becomes effectively zero-dimensional (for not very low energies), the preexponential factor was also calculated; the energy dependence of the density of states obtained in this case was found to be the same as in the random matrix theory. The applicability of these results is ensured by a large conductance of the normal region.

We developed this method further in [30]. In particular, we managed to obtain an exact expression for the density of states describing the entire transition region above as well as below the critical energy without using

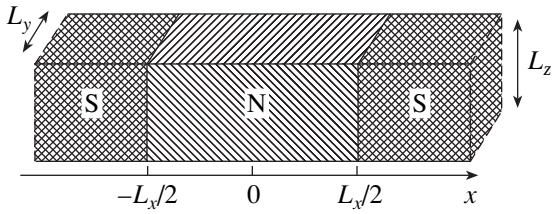


Fig. 1. SNS junction of length L_x. Transverse dimensions are L_y and L_z.

the instanton approximation. In this case also, the form of energy dependence of the density of states is the same as predicted phenomenologically by the random matrix theory. We also considered a long one-dimensional SINIS contact with nonideal interfaces between the normal and superconducting parts. Poor transparency of the interfaces suppresses the proximity effect and reduces the semiclassical gap. When the conductance of an interface becomes smaller than a certain critical value, a transition to another class of universality occurs. The density of states increases as the inverse root function as the energy approaches the threshold from above. In the fluctuation region near the threshold, this divergence is smoothed and transforms into an exponentially decreasing tail described by an expression given in [30]. In contrast to all the cases mentioned above, when the number of states in the region of the fluctuation tail is on average of the order of unity, the number of subgap states in the latter case is parametrically large, and the tail is referred to as strong. In [30], we wrote only the final result for an strong tail due to lack of space. This gap will be made up for in this paper.

This paper is devoted to a review of the results obtained in the framework of the nonlinear supermatrix σ model for mesoscopic superconducting systems. The structure of the review is as follows.

In Section 2, we first consider a semiclassical approach to the evaluation of the density of states. The properties of the semiclassical solution will be required on a later stage for describing possible instantons of the σ model. Then a brief derivation of the σ model for a superconductor is considered. One of the main results of this section is parametrization of the saddle manifold and classification of possible instantons.

The problem of an NS system with absolutely transparent boundaries is solved in Section 3. An exact expression for the density of states near the spectrum edge is derived. The application of the random matrix theory to the problem of an NS contact is considered briefly at the end of the section.

Section 4 generalizes the results to the case of a boundary with an arbitrary transparency. A classification of possible tails of the density of states is constructed depending on the transparency. The case of an strong tail is analyzed separately.

Section 5 is devoted to a system that cannot be described in the framework of the zero-dimensional σ model. These are SNS junctions with a large transverse size, for which the instanton radius is much smaller than the size of the system. A superconductor with magnetic impurities is also considered. At the end of the section, the density of prelocalized states in an SNS junction deep in the gap is calculated with a logarithmic accuracy.

2. GENERAL THEORY

2.1. Semiclassical Approach

Let us first consider a semiclassical method for calculating the density of states in an NS system with transparent boundaries. The Green function of the superconductor provides information on fast oscillations of electrons forming Cooper pairs as well as on a relatively slow motion of a pair as a whole. Averaging the retarded Green function over fast modes in the diffusive case leads to the Usadel equation[1] [9]:

$$\begin{gathered} D\nabla(\hat{g}_R(\mathbf{r})\nabla\hat{g}_R(\mathbf{r})) + i[\tau_z E + i\tau_x\Delta, \hat{g}_R(\mathbf{r})] = 0, \\ \hat{g}_R^2(\mathbf{r}) = 1. \end{gathered} \tag{2.1}$$

Here, $\hat{g}_R(\mathbf{r})$ is a 2×2 matrix in the Nambu space, τ_i are the Pauli matrices acting in the Nambu space, D is the diffusion coefficient, and energy E is measured from the Fermi level. In the angular parametrization $\hat{g}(\mathbf{r}) = \tau_z\cos\theta + \tau_x\sin\theta$, the Usadel equation has the form

$$D\nabla^2\theta + 2iE\sin\theta + 2\Delta\cos\theta = 0. \tag{2.2}$$

We disregard the proximity effect in the superconductor and fix Δ = const in it, while in the region of the normal metal, we assume that Δ is equal to zero. If the size of the normal region exceeds the superconducting coherence length ξ, the gap in the spectrum is on the order of the Thouless energy, which is assumed to be much smaller than Δ; consequently, we can set $\theta = \pi/2$ in the superconductor.

In the normal region of the NS system, the Usadel equation has the form

$$D\nabla^2\theta + 2iE\sin\theta = 0. \tag{2.3}$$

The boundary conditions require that $\theta = \pi/2$ at the interface with the superconductor (ideal interface) and $\nabla_{\mathbf{n}}\theta = 0$ at the free boundary of the normal metal.

The density of states averaged over disorder can be expressed as (ν is the density of states per spin component)

$$\begin{aligned} \langle\rho(E,\mathbf{r})\rangle &= \nu\mathrm{Re}\,\mathrm{tr}(\tau_z\hat{g}(\mathbf{r})) \\ &= 2\nu\mathrm{Re}\cos\theta = 2\nu\mathrm{Im}\sinh\psi, \end{aligned} \tag{2.4}$$

[1] We assume that the phase of the order parameter is equal to zero.

where the substitution $\theta = \pi/2 + i\psi$ has been made, which transforms Eq. (2.3) into the following equation with real coefficients:

$$D\nabla^2\psi + 2E\cosh\psi = 0. \tag{2.5}$$

By way of an example, we consider a one-dimensional SNS junction of length L_x, which is depicted in Fig. 1. In this case, Eq. (2.5) can be integrated easily, which gives the expression for energy in terms of the value of ψ at the middle of the junction:

$$\sqrt{\frac{E}{E_{Th}}} = \int_0^{\psi(0)} \frac{d\psi}{\sqrt{\sinh\psi(0) - \sinh\psi}}, \qquad E_{Th} = \frac{D}{L_x^2}. \tag{2.6}$$

This function is plotted in Fig. 2. It can be seen that the quantity ψ is real (i.e., the density of states is zero) only for energy values smaller than a certain threshold value $E_g = 3.12E_{Th}$. This is exactly the Thouless gap [10, 11].

The example of a 1D junction can be used to establish the following general properties of solutions to the Usadel equation. Equation (2.5) has two real solutions for $E < E_g$. We denote the smaller of these solutions by $\psi_1(\mathbf{r})$ and the larger solution by $\psi_2(\mathbf{r})$. For $E = E_g$, these solutions coincide: $\psi_{1,2}(\mathbf{r}) = \psi_0(\mathbf{r})$. For $E > E_g$, the Usadel equation has two complex solutions, from which we choose the one leading to a positive density of states. From physical considerations, we choose $\psi_1(\mathbf{r})$ under the gap since $\psi_2(\mathbf{r})$ increases indefinitely as the energy tends to zero. It will be shown below, however, that the solution with $\psi_2(\mathbf{r})$ is possible as a fluctuation, which is responsible for a nonzero density of states for energies $E < E_g$.

In the subsequent analysis, we will need a normalized difference of the solutions to the Usadel equation for energy tending to the threshold value (V is the volume of the normal region):

$$f_0(\mathbf{r}) = \lim_{E\to E_g} \frac{\psi_2(\mathbf{r}) - \psi_1(\mathbf{r})}{\sqrt{\int (\psi_2(\mathbf{r}) - \psi_1(\mathbf{r}))^2 \frac{d\mathbf{r}}{V}}}. \tag{2.7}$$

Function $f_0(\mathbf{r})$ satisfies a linear equation which can be derived taking the limit in energy of the difference between the Usadel equations for ψ_1 and ψ_2:

$$D\nabla^2 f_0 + 2E_g f_0 \sinh\psi_0 = 0. \tag{2.8}$$

Expressions for density of states depend on the system geometry only via the following two numerical parameters:

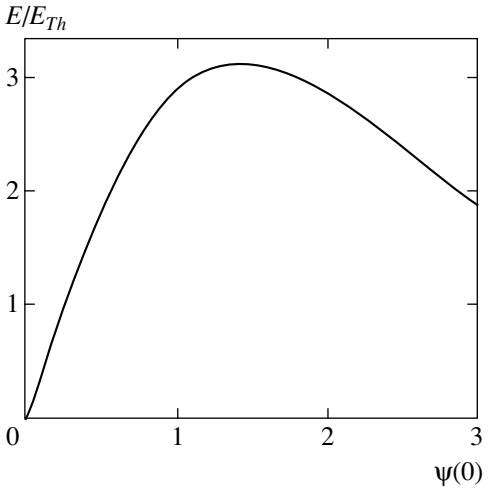


Fig. 2. Dependence of E/E_{Th} on the value of ψ at the middle of a one-dimensional junction (formula (2.6)). The maximum of the function corresponds to the threshold energy value $E_g \approx 3.12E_{Th}$. Below the threshold, the Usadel equation has two solutions, one of which diverges as $E \longrightarrow 0$.

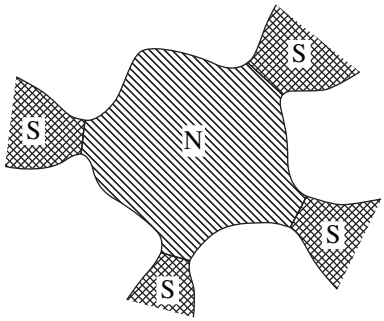


Fig. 3. Contact between a normal grain and superconductors. The main results of Sections 2 and 3 were obtained in the framework of the zero-dimensional σ model for a contact of an arbitrary shape.

$$c_1 = \int \frac{d\mathbf{r}}{V} f_0(\mathbf{r})\cosh\psi_0(\mathbf{r}), \qquad c_2 = \int \frac{d\mathbf{r}}{V} f_0^3(\mathbf{r})\cosh\psi_0(\mathbf{r}). \tag{2.9}$$

In particular, $c_1 \approx 1.15$ and $c_2 \approx 0.88$ for a 1D SNS junction (Fig. 1).

We will consider below a contact of an arbitrary geometry (Fig. 3), assuming that it has only one characteristic scale of length. In this case, the density of states near the threshold can be determined in the approximation of the zero-dimensional σ model (see Subsection 3.2). In the case of a contact whose size along the boundary with superconductors is much larger than their separation, solutions in the form of an

instanton along the larger dimension also exist in addition to the two solutions to the Usadel equation described above. This case requires separate analysis, which will be carried out in Subsection 5.1.

2.2. Derivation of the σ Model

In order to find instanton corrections to the semiclassical density of states, we will briefly describe the derivation of the effective supersymmetric field theory (Efetov σ model [26]) for superconducting systems [27]. The states of an electron in a superconductor are described by the Bogoliubov–De Gennes Hamiltonian, which is a 2×2 matrix in the Nambu space:

$$\mathcal{H} = \tau_z\left(\frac{\mathbf{p}^2}{2m} - \mu + U(\mathbf{r})\right) + \tau_x\Delta(\mathbf{r}). \tag{2.10}$$

We express the density of states in terms of the Green function:

$$\rho(E, \mathbf{r}) = -\frac{1}{\pi}\mathrm{Im}\,\mathrm{tr}\int d\mathbf{r}\,G^R(\mathbf{r}, \mathbf{r}; E). \tag{2.11}$$

The Green function will be calculated using the functional integral[2]

$$G^R(\mathbf{r}, \mathbf{r}'; E) = -i\frac{\int u(\mathbf{r})u^+(\mathbf{r}')e^{-\mathcal{S}[u]}\mathcal{D}u^*\mathcal{D}u}{\int e^{-\mathcal{S}[u]}\mathcal{D}u^*\mathcal{D}u}, \tag{2.12}$$

where the action is defined as

$$\mathcal{S}[u] = -i\int d\mathbf{r}\,u^+(\mathbf{r})(E + i0 - \mathcal{H})u(\mathbf{r}). \tag{2.13}$$

In order to carry out subsequent averaging over disorder, we must get rid of the normalization integral in the denominator of Eq. (2.12). For this purpose, we introduce, in addition to field u, the Grassmann (anticommuting) field χ:

$$\Phi = \begin{pmatrix} \chi \\ u \end{pmatrix}, \tag{2.14}$$

$$G^R(\mathbf{r}, \mathbf{r}'; E) = -i\int u(\mathbf{r})u^+(\mathbf{r}')e^{-\mathcal{S}[\Phi]}\mathcal{D}\Phi^*\mathcal{D}\Phi, \tag{2.15}$$

$$\mathcal{S}[u] = -i\int d\mathbf{r}\Phi^+(\mathbf{r})(E + i0 - \mathcal{H})\Phi(\mathbf{r}). \tag{2.16}$$

Field Φ is formed by four components and belongs to the product of a Nambu space and a supersymmetric Fermi–Bose (FB) space. After averaging over disorder, we must write the effective action for slow modes in $\langle\Phi\Phi^+\rangle$. In addition, we must take into account slow modes in the Cooper channel $\langle\Phi\Phi^T\rangle$ and $\langle\Phi^*\Phi^+\rangle$. For this purpose, we carry out an additional field doubling [26] just now (following notation adopted in [31]):

$$\psi = \frac{1}{\sqrt{2}}\begin{pmatrix} \Phi \\ i\tau_y\Phi^* \end{pmatrix}. \tag{2.17}$$

We will refer to the obtained space as a particle–hole (PH) space. Vector ψ consists of a particle (upper) block and a hole block. We denote by σ_i the Pauli matrix in the PH space and introduce the operation of charge conjugation of supervectors and supermatrices:

$$\bar{\psi} = (C\psi)^T, \quad \bar{A} = CA^TC^T,$$
$$C = -\tau_x\begin{pmatrix} i\sigma_y & 0 \\ 0 & \sigma_x \end{pmatrix}_{FB}. \tag{2.18}$$

Averaging over impurities in Eq. (2.15) leads to an action of the form

$$\mathcal{S}[\psi] = -i\int d\mathbf{r}$$
$$\times\left[\bar{\psi}\left(\Lambda(E + i0) - \frac{\mathbf{p}^2}{2m} + \mu - i\tau_y\Delta(\mathbf{r})\right)\psi + \frac{(\bar{\psi}\psi)^2}{4\pi\nu\tau}\right], \tag{2.19}$$

where we have introduced notation $\Lambda = \sigma_z\tau_z$. The fourth-order term must be decoupled with the help of the Hubbard–Stratonovich transformation. For this purpose, we introduce an 8×8 matrix superfield Q. The transformation leads to the following action:

$$\mathcal{S}[\psi, Q] = -i\int d\mathbf{r}\left[\bar{\psi}\left(\Lambda(E + i0) - \frac{\mathbf{p}^2}{2m} + \mu - i\tau_y\Delta(\mathbf{r}) + \frac{iQ}{2\tau}\right)\psi + \frac{\pi\nu}{8\tau}\mathrm{str}\,Q^2\right]. \tag{2.20}$$

The measure of functional integration over the new field Q is determined from the supersymmetry condition, $\int\exp(-\mathrm{str}\,Q^2)\mathcal{D}Q = 1$.

As a result, the integral over fields ψ becomes a Gaussian integral and can be evaluated. However, there is another difficulty lying in the fact that not all components of ψ and $\bar{\psi}$ are independent. The introduction of the PH space (formula (2.17)) has not resulted in the addition of new variables; old variables have just been regrouped. The action averaged over ψ has the form

$$\mathcal{S}[Q] = \int d\mathbf{r}\,\mathrm{str}\left\{\frac{\pi\nu}{8\tau}Q^2 - \frac{1}{2}\ln\left[\sigma_z(E + i0) - \tau_z\left(\frac{\mathbf{p}^2}{2m} - \mu\right) - \tau_x\Delta + \frac{i\tau_z}{2\tau}Q\right]\right\}. \tag{2.21}$$

This circumstance resulted in the emergence of coefficient 1/2 in front of the logarithm and necessitated the

[2] Symbol $\mathcal{D}u^*\mathcal{D}u$ must be interpreted as follows: $\mathcal{D}u^*\mathcal{D}u = \prod_n\prod_{k=1,2}\pi^{-1}d\mathrm{Re}\,u_n^{(k)}d\mathrm{Im}\,u_n^{(k)}$. Here, $u_n^{(k)}$ are the expansion coefficients of the kth component of vector $u(\mathbf{r})$ in the orthonormal basis of functions.

imposition of the self-conjugate condition on Q:

$$Q = \bar{Q}. \tag{2.22}$$

The obvious saddle point of action (2.21) is $Q = \Lambda$. The approximation of the σ model involves determining a saddle solution which is a slow function of coordinates. To this end, we assume that $Q = e^{-iU/2}\Lambda e^{iU/2}$ and retain in the action the principal terms in gradients Q, in energy E, and in the value of the order parameter Δ. The result of such an expansion is the action of the σ model [27, 29]

$$\mathcal{S}[Q] = \frac{\pi\nu}{8}\int d\mathbf{r} \times \mathrm{str}[D(\nabla Q)^2 + 4iQ(\Lambda(E + i0) + i\tau_x\Delta)], \tag{2.23}$$
$$Q^2 = 1,$$

and the expression for density of states assumes the form

$$\langle \rho(E, \mathbf{r})\rangle = \frac{\nu}{4}\mathrm{Re}\int \mathrm{str}(k\Lambda Q(\mathbf{r}))e^{-\mathcal{S}[Q]}\mathcal{D}Q. \tag{2.24}$$

Here, we have introduced the following notation for the matrix violating supersymmetry [26]:

$$k = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_{FB}. \tag{2.25}$$

2.3. Parametrization of the Q Matrix Manifold

Supermatrix Q has a size of 8×8 and contains 32 complex commuting parameters and the same number of anticommuting (Grassmann) parameters in the general case. Condition $Q = \bar{Q}$ reduces the number of these parameters by half. In accordance with the construction of the σ model, matrix Q has the same structure of eigenvalues as Λ,

$$Q = e^{-iU/2}\Lambda e^{iU/2},$$

which reduces the number of independent parameters to eight complex and eight Grassmann parameters. In this subsection, we construct parametrization of the commuting part of the Q matrix.

The self-conjugate condition for Q leads to anti-self-conjugate for U: $U + \bar{U} = 0$. In addition, we impose the natural condition $\{\Lambda, U\} = 0$, since only such generators can "rotate" matrix Λ. In the absence of Grassmann variables, matrices Q and U split into two independent sectors (FF and BB). Charge conjugation operates in different ways in these two sectors, leading to their different topologies. Matrix U contains the following generators:

$$U^{FF}: \ \sigma_x \quad \sigma_y \quad \sigma_z\tau_x \quad \sigma_z\tau_y$$
$$U^{BB}: \ \sigma_x\tau_z \quad \sigma_y\tau_z \quad \sigma_z\tau_x \quad \sigma_z\tau_y.$$

The FF sector is generated by four pairwise anticommuting generators. Consequently, it forms topologically a 4-dimensional complex sphere S_4. Generators of the BB sector can be divided into two pairs in which they anticommute, while commutativity holds between pairs; i.e., the BB sector is the product of two two-dimensional complex spheres. We choose spherical angles in the form

$$Q^{FF} = \tau_z\cos\theta_F[\sigma_z\cos k_F + \sin k_F(\sigma_x\cos\chi_F + \sigma_y\sin\chi_F)] + \sin\theta_F(\tau_x\cos\varphi_F + \tau_y\sin\varphi_F), \tag{2.26}$$

$$Q^{BB} = [\sigma_z\cos k_B + \tau_z\sin k_B(\sigma_x\cos\chi_B + \sigma_y\sin\chi_B)] \times [\tau_z\cos\theta_B + \sigma_z\sin\theta_B(\tau_x\cos\varphi_B + \tau_y\sin\varphi_B)]. \tag{2.27}$$

An additional symmetry of manifold Q^{BB} is worth noting: the matrix does not change under the simultaneous inversion of both spheres: $(\theta_B, \varphi_B, k_B, \chi_B) \longrightarrow (\pi - \theta_B, \varphi_B + \pi, \pi - k_B, \chi_B + \pi)$. As a result, the BB sector is topologically equivalent to the factorized product $S_2 \times S_2/Z_2$.

The general requirement of the σ model convergence imposes the condition of compactness on the FF sector and noncompactness on the BB sector, which reduces the number of independent variables to four real variables in the FF and BB sectors.

Substituting matrix Q into Eq. (2.23), we obtain the explicit representation of action in terms of the angles introduced above:

$$\mathcal{S} = \frac{\pi\nu}{2}\int d\mathbf{r}(\mathcal{L}^{FF} - \mathcal{L}^{BB}),$$

$$\mathcal{L}^{FF} = D[(\nabla\theta_F)^2 + \sin^2\theta_F(\nabla\varphi_F)^2 + \cos^2\theta_F(\nabla k_F)^2 + \cos^2\theta_F\sin^2 k_F(\nabla\chi_F)^2] + 4iE\cos\theta_F\cos k_F - 4\Delta\sin\theta_F\cos\varphi_F, \tag{2.28}$$

$$\mathcal{L}^{BB} = D[(\nabla\theta_B)^2 + \sin^2\theta_B(\nabla\varphi_B)^2 + (\nabla k_B)^2 + \sin^2 k_B(\nabla\chi_B)^2] + 4iE\cos\theta_B\cos k_B - 4\Delta\sin\theta_B\cos k_B\cos\varphi_B.$$

Angles θ and φ in both sectors have the meaning of the Usadel angle and the phase of the order parameter. In order to determine the saddle configurations of this action for zero phase difference at the contact, we can immediately set $\varphi_F = \varphi_B = 0$. On the saddle solution, angles $\chi_{F,B}$ are independent of coordinates and are completely cyclic: action is independent of $\chi_{F,B}$.

Angle k_F is also equal to zero on a saddle solution. In the BB sector, it is convenient to carry out the substitution of variables:

$$\theta_B = \frac{\alpha+\beta}{2}, \quad k_B = \frac{\alpha-\beta}{2}. \tag{2.29}$$

In terms of angles θ_F, α, and β, action has the simple form

$$\mathcal{S}[\theta_F, \alpha, \beta] = 2S_0[\theta_F] - S_0[\alpha] - S_0[\beta], \tag{2.30}$$

$$S_0[\theta] = \frac{\pi\nu}{4}\int d\mathbf{r} \times [D(\nabla\theta)^2 + 4iE\cos\theta - 4\Delta\sin\theta]. \tag{2.31}$$

2.4. Saddle Points

Variation of action (2.31) leads to Eq. (2.2). Thus, a saddle point can be described by the Usadel equation in each of the three variables (θ_F, α, and β). In accordance with Subsection 2.1, the Usadel equation has two solutions: $\theta_1(\mathbf{r})$ and $\theta_2(\mathbf{r})$. (The only exception is the situation when the contact size along the interface with superconductors is much larger than their separation; this case will be treated separately in Subsection 5.1.) Since solutions $\theta_1(\mathbf{r})$ and $\theta_2(\mathbf{r})$ coincide for $E = E_g$, the mode transforming $\theta_1(\mathbf{r})$ into $\theta_2(\mathbf{r})$ becomes softer near the threshold. As a result, the functional integral over $Q(\mathbf{r})$ becomes an ordinary integral over supermatrix Q; i.e., a transition to the zero-dimensional σ model occurs.

Thus, in the zero-dimensional case, there exist $2^3 = 8$ saddle solutions in all. If we choose solution θ_1 in all three variables θ_F, α, and β, we automatically obtain $\theta_B = \theta_1$, $\theta_F = \theta_1$, and $k_B = 0$; i.e., the FF and BB sectors are identical. Expanding action (2.28) up to the second order in fluctuations in the vicinity of such a saddle point, we find that the superdeterminant of this quadratic form is equal to unity, which is a direct consequence of the FF–BB symmetry of the solution. The functional integral (2.24) in this case is reduced to (2.4). Thus, the semiclassical approximation can be obtained from the σ model in the steepest descent approximation in the vicinity of a supersymmetric saddle point.

Higher orders of the expansion of action near a supersymmetric saddle point correspond to perturbation corrections to a semiclassical result. These corrections were analyzed in [27], where it was shown that their inclusion leads to renormalization of E_g. In this case, the density of states below the renormalized value of the gap vanishes as before. The average density of states is found to be finite in the entire energy range only if we take into account other saddle points (instantons). In other words, we must use the second solution of the Usadel equation in one or several variables θ_F, α, and β.

The density of states vanishes when $\mathrm{Re}\,\theta_{1,2} = \pi/2$. This equality is valid for both Usadel solutions for $E < E_g$. How is it that an instanton may contribute to the subgap density of states? As a matter of fact, a quadratic action in the vicinity of an instanton saddle point contains a negative eigenvalue; consequently, an imaginary unity appears in the integral over fluctuations in the vicinity of the instanton and, as a result, a nonzero density of states appears below the gap.

Let us consider instantons, which can appear in the zero-dimensional case. First, we note the obvious inequality

$$S_0[\theta_1] > S_0[\theta_2]. \tag{2.32}$$

In order to obtain a positive action (2.30) at an instanton, we must fix $\theta_F = \theta_1$. It will be shown below (see Subsection 3.2) that the saddle point $\theta_F = \theta_2$ cannot be attained by deforming the integration contour in the variables of the FF sector under the condition of convergence of the σ model. Thus, we are left with three nonstandard saddle solutions in the BB sector: $(\alpha, \beta) = (\theta_2, \theta_1)$, $(\alpha, \beta) = (\theta_1, \theta_2)$, and $(\alpha, \beta) = (\theta_2, \theta_2)$. At first glance, the two first solutions break the symmetry in angle χ_B. In actual practice, this symmetry is restored due to the fact that there exists a full saddle ring containing both these points. Different points of this ring differ in angle χ_B. In particular, points $(\alpha, \beta) = (\theta_2, \theta_1)$ and $(\alpha, \beta) = (\theta_1, \theta_2)$ can be obtained from each other by changing angle χ_B by π. This ring will be referred to as the first instanton. The third solution, $(\alpha, \beta) = (\theta_2, \theta_2)$, is an isolated saddle point that will be referred to as the second instanton.

Both instanton solutions can be presented in the form

$$Q_0 = e^{-iU_0/2}\Lambda e^{iU_0/2} = \Lambda e^{iU_0}, \tag{2.33}$$

where

$$U_0^{FF} = \sigma_z\tau_y\theta_F,$$
$$U_0^{BB} = \sigma_z\tau_y\theta_B + \tau_z k_B(\sigma_y\cos\chi_B - \sigma_x\sin\chi_B). \tag{2.34}$$

On the ring of the first instanton, we have

$$\theta_F = \frac{\pi}{2} + i\psi_1(\mathbf{r}), \quad \theta_B = \frac{\pi}{2} + i\frac{\psi_1(\mathbf{r}) + \psi_2(\mathbf{r})}{2},$$
$$k_B = i\frac{\psi_2(\mathbf{r}) - \psi_1(\mathbf{r})}{2}, \quad \chi_B \in [0, 2\pi), \tag{2.35}$$

while on the second instanton we have

$$\theta_F = \frac{\pi}{2} + i\psi_1(\mathbf{r}), \quad \theta_B = \frac{\pi}{2} + i\psi_2(\mathbf{r}),$$
$$k_B = 0. \tag{2.36}$$

The instanton action can be easily determined in the vicinity of the semiclassical edge E_g of the spectrum. We make use of the fact that the two solutions of the

Usadel equation coincide at the threshold: $\theta_{1,2}(\mathbf{r}) = \theta_0(\mathbf{r}) = \pi/2 + i\psi_0(\mathbf{r})$. For $E \longrightarrow E_g$, their difference is proportional to function $f_0(\mathbf{r})$ defined by formula (2.7).

In order to evaluate the action at an instanton, we substitute $\theta(\mathbf{r}) = \pi/2 + i\psi_0(\mathbf{r}) + igf_0(\mathbf{r})$ into expression (2.31) and expand it in g and dimensionless energy ε measured from the threshold,

$$\varepsilon = \frac{E_g - E}{E_g}. \tag{2.37}$$

This gives

$$\begin{aligned} S_0[\theta] &= S_0[\theta_0] \\ &+ \frac{\pi\nu}{4}\int d\mathbf{r}\Big[2gf_0(D\nabla^2\psi_0 + 2E_g\cosh\psi_0) \\ &+ g^2 f_0(D\nabla^2 f_0 + 2E_g f_0\sinh\psi_0) \\ &- 4E_g\varepsilon g f_0\cosh\psi_0 + \frac{2}{3}E_g g^3 f_0^3\cosh\psi_0\Big]. \end{aligned} \tag{2.38}$$

The first term in the integrand vanishes in accordance with the Usadel equation, and the second vanishes in accordance with Eq. (2.8). Thus, we can represent S_0 in the form of a cubic polynomial in g:

$$S_0[\theta] = \text{const} + \frac{\pi E_g}{2\delta}\Big[-2c_1\varepsilon g + \frac{c_2}{3}g^3\Big]. \tag{2.39}$$

Here, we have introduced the mean level spacing $\delta = (\nu V)^{-1}$. The expression obtained has two extrema,

$$g_\pm = \pm\sqrt{\tilde{\varepsilon}}, \quad \tilde{\varepsilon} = \frac{2c_1}{c_2}\varepsilon, \tag{2.40}$$

corresponding to two solutions to the Usadel equation (g_+ corresponds to solution $\theta_2(\mathbf{r})$ and g_- to solution $\theta_1(\mathbf{r})$). Substituting these solutions into Eq. (2.30), we obtain the following expression for the action of the first instanton:

$$\mathcal{S}_1 = S_0[\theta_1] - S_0[\theta_2] = \frac{4}{3}\tilde{G}\tilde{\varepsilon}^{3/2}; \tag{2.41}$$

here, we have introduced the notation

$$\tilde{G} = \frac{\pi c_2 E_g}{2\delta}. \tag{2.42}$$

For a planar junction (see Fig. 1), this quantity is of the order of the dimensionless conductance of the normal region: $\tilde{G} \approx 0.34G_N$, where $G_N = 4\pi\nu DL_yL_z/L_x$.

In accordance with Eq. (2.30), action $\mathcal{S}_2$ of the second instanton is twice as large as action $\mathcal{S}_1$ of the first instanton. Consequently, their relative contribution to the density of states is determined by the value of $\mathcal{S}_1$. For $\mathcal{S}_1 \gg 1$, the contribution of the second instanton is exponentially suppressed as compared to the contribution of the first instanton, which is also exponentially small. This regime, which corresponds to energies differing considerably from the threshold value, will be considered in Subsection 3.1. In the case of exact equality $E = E_g$, action $\mathcal{S}_1$ vanishes. For this reason, there exists a fluctuating region in the vicinity of the threshold, which is determined by inequality $|\varepsilon| \lesssim \tilde{G}^{-2/3}$, where $\mathcal{S}_1 \lesssim 1$, so that the contributions from both instantons are of the same order of magnitude and, hence, cannot be separated.

The exact solution taking into account both instantons in the entire energy range will be given in Subsection 3.2.

2.5. Parametrization of Fluctuations

In order to apply the steepest descent method with the saddle point determined by us, we must evaluate the integral over all possible fluctuations of the Q matrix in the vicinity of the instanton. For this purpose, we expand the action in the matrix form up to the second order in the vicinity of the saddle and then propose a parametrization diagonalizing the quadratic form of the action.

We introduce matrix W describing fluctuations:

$$\begin{aligned} &Q = e^{-iU_0/2}e^{-iW/2}\Lambda e^{iW/2}e^{iU_0/2}, \\ &\{\Lambda, W\} = 0, \quad W + \overline{W} = 0. \end{aligned} \tag{2.43}$$

We must now substitute matrix Q expressed in terms of W into action (2.23) and expand it up to the second order in W. We assume that $\Delta = 0$ since we are going to operate with this action only in the normal region. The quadratic form of the action can be written as

$$\begin{aligned} \mathcal{S}^{(2)}[W] &= \frac{\pi\nu}{8}\int d\mathbf{r} \\ &\times \text{str}\Big[D(\nabla W)^2 + \frac{D}{4}[\nabla U_0, W]^2 - 2iE\Lambda Q_0 W^2\Big]. \end{aligned} \tag{2.44}$$

Matrix W, as well as Q, contains eight commuting parameters and the same number of Grassmann parameters. The complete parametrization of this matrix, in which action (2.44) is diagonal, is given in Appendix A. Quadruples a, b, c, d and m, n, p, q of real variables parametrize the FF and BB sectors of matrix W, respectively, while eight Grassmann variables (λ, μ, ζ, κ, η, γ, ξ, ω) parametrize the anticommuting component of

matrix *W*. The quadratic part of the action in terms of new variables assumes the form

$$\begin{aligned}\mathcal{S}^{(2)} &= (a\hat{\mathbb{O}}^{+}_{\theta_F\theta_F}a) + (b\hat{\mathbb{O}}^{-}_{\theta_F\theta_F}b) + (c\hat{\mathbb{O}}^{+}_{\theta_F\theta_F}c)\\ &+ (d\hat{\mathbb{O}}^{+}_{\theta_F\theta_F}d) + (m\hat{\mathbb{O}}^{+}_{\alpha\beta}m) + (n\hat{\mathbb{O}}^{-}_{\alpha\beta}n)\\ &+ (p\hat{\mathbb{O}}^{-}_{\beta\beta}p) + (q\hat{\mathbb{O}}^{-}_{\alpha\alpha}q) + (\lambda\hat{\mathbb{O}}^{+}_{\alpha\theta_F}\eta)\\ &+ (\mu\hat{\mathbb{O}}^{+}_{\beta\theta_F}\gamma) + (\kappa\hat{\mathbb{O}}^{-}_{\beta\theta_F}\omega) + (\zeta\hat{\mathbb{O}}^{-}_{\alpha\theta_F}\xi).\end{aligned} \tag{2.45}$$

Here, we have introduced operators $\hat{\mathbb{O}}^{\pm}_{\alpha\beta}$ acting in accordance with the rule

$$\begin{aligned}(a\hat{\mathbb{O}}^{\pm}_{\alpha\beta}b) &= \frac{\pi\nu}{8}\int d\mathbf{r}\,a(\mathbf{r})\\ &\times\left[-D\nabla^2 - \frac{D}{4}(\nabla\alpha \pm \nabla\beta)^2 - iE(\cos\alpha + \cos\beta)\right]b(\mathbf{r}).\end{aligned} \tag{2.46}$$

Operators $\hat{\mathbb{O}}$ possess a discrete spectrum since fluctuations occur in a bounded space of the normal region. We denote the eigenvalues of operator $\hat{\mathbb{O}}^{\pm}_{\alpha\beta}$ by $(\mathscr{E}^{\pm}_{\alpha\beta})_n/\delta$, where n runs through values from 0 to ∞. It follows from Eq. (2.46) that the difference between the first excited state and the ground state of any operator $\hat{\mathbb{O}}$ is on the order of $(\mathscr{E}^{\pm}_{\alpha\beta})_1 - (\mathscr{E}^{\pm}_{\alpha\beta})_0 \approx E_g$. The ground state energy of operators $\hat{\mathbb{O}}^{+}$ has the scale of E_g. The ground state of operator $\hat{\mathbb{O}}^{-}_{\theta_1\theta_2}$ has zero energy, and its eigenfunction is equal to $\sin((\theta_1 - \theta_2)/2)$ to within normalization. Exactly at the threshold, for $E = E_g$, the ground state energies of operators $\hat{\mathbb{O}}^{-}_{\theta_1\theta_1}$ and $\hat{\mathbb{O}}^{-}_{\theta_2\theta_2}$ are also equal to zero. As we move away from the threshold towards lower energies, $(\mathscr{E}^{-}_{\theta_1\theta_1})_0$ becomes positive and $(\mathscr{E}^{-}_{\theta_2\theta_2})_0$ becomes negative. In the limit $E_g - E \ll E_g$, the inequality $\left|(\mathscr{E}^{-}_{\alpha\beta})_0\right| \ll E_g$ holds.

The spectrum of operators $\hat{\mathbb{O}}$ determines the masses $\mathscr{E}_n$ of various fluctuations in the vicinity of instanton saddle solutions. Depending on the value of these masses, we can single out the following three types of fluctuations.

(i) Zero modes. Strictly zero modes include Grassmann Goldstone modes restoring the supersymmetry broken by a saddle solution (mode $\zeta\xi$ for the first instanton and modes $\zeta\xi$ and $\kappa\omega$ for the second instanton) as well as the Goldstone mode n restoring the symmetry of the first instanton in angle χ_B. Zero modes correspond to the ground state of operator $\hat{\mathbb{O}}^{-}_{\theta_1\theta_2}$.

(ii) Soft modes. These include fluctuations of variables b, p, and q as well as mode $\kappa\omega$ (in the case of the first instanton), which correspond to the ground states of operators $\hat{\mathbb{O}}^{-}_{\theta_1\theta_1}$ and $\hat{\mathbb{O}}^{-}_{\theta_2\theta_2}$. If $E \longrightarrow E_g$, the mass of soft modes tends to zero.

(iii) Hard modes. These modes have a mass on the order of the Thouless energy and above, so that their fluctuations are small in parameter $E_g/\delta \sim G_N \gg 1$. This inequality ensures the validity of the steepest descent method. Hard modes include all eigenstates of operators $\hat{\mathbb{O}}^{+}$ and excited states of operators $\hat{\mathbb{O}}^{-}$.

3. CONTACT WITH IDEAL INTERFACES

We will consider a contact with ideal interfaces between a superconductor and a normal metal. Let us first consider a contact for energies close to the threshold energy, but still differing from it, $\tilde{G}^{-2/3} \ll (E_g - E)/E_g \ll 1$, for which the approximation of the zero-dimensional σ model is applicable. In this energy range, the main contribution to the density of states comes from the first instanton. The main exponent in the expression for density of states is defined by formula (2.41). We will now calculate the preexponential factor and then construct a more complete theory, taking into account the contributions from the second instanton. As a result, we will obtain an exact (naturally, in the steepest descent approximation in hard modes, which is controlled by parameter $\tilde{G} \gg 1$) expression describing the average density of states both below and above the threshold energy, including the entire fluctuation region. The form of the obtained expression coincides with the predictions of the random matrix theory for the edge of the spectrum, which may serve as a microscopic substantiation of the hypothesis put forth in [22].

3.1. Single-Instanton Solution

In this subsection, we determine the contribution of the first instanton to the average density of states. We consider a region in the vicinity of the semiclassical edge E_g of the spectrum, but outside the fluctuation region: $\tilde{G}^{-2/3} \ll \varepsilon \ll 1$, in which we can disregard the contribution from the second instanton. According to the classification given in Subsection 2.4, action (2.45) at the first instanton $(\theta_F, \alpha, \beta) = (\theta_1, \theta_2, \theta_1)$ has a zero mode in variable n, a Grassmann zero mode $\zeta\xi$, and soft modes in variables b, p, q, and $\kappa\omega$. All the remaining modes are hard.

Inequality $\tilde{G} \gg 1$ guarantees that, while integrating over hard fluctuations, we can disregard the preexponential factor in formula (2.24). In this case, integration

becomes trivial and gives the superdeterminant of the quadratic form (2.45):

$$\sqrt{\frac{\det\hat{\mathbb{O}}^{+}_{\theta_1\theta_2}\det'\hat{\mathbb{O}}^{-}_{\theta_1\theta_2}}{\det\hat{\mathbb{O}}^{+}_{\theta_1\theta_1}\det'\hat{\mathbb{O}}^{-}_{\theta_1\theta_1}}} = 1 + O(\varepsilon). \tag{3.1}$$

Here, the prime denotes elimination of the lowest eigenvalue, and the equality follows from the estimate $(\mathscr{E}^{\pm}_{\theta_1\theta_2})_n - (\mathscr{E}^{\pm}_{\theta_1\theta_1})_n = O(\varepsilon)$ (inequality $n > 0$ holds for eigenvalues $\mathscr{E}^{-}$) and from the asymptotic form $(\mathscr{E}^{\pm}_{\alpha\beta})_n \propto n^2E_g$, which is valid for $n \gg 1$. Thus, hard fluctuations make zero contribution to the density of states in the limit in question.

It remains for us to consider soft and zero modes. It should be noted that zero modes in variables n and $\zeta\xi$ behave differently upon a deviation from the instanton solution. The zero mode in variable n corresponds to rotation through angle χ_B and remains massless for any (not necessarily saddle) noncoinciding values of α and β. On the other hand, the Grassmann zero mode $\zeta\xi$ acquires a mass upon a deviation from the instanton solution. Such a behavior is determined by the necessity of satisfying condition $\int e^{-S[Q]}DQ = 1$, which would be impossible if the Grassmann mode remained a strictly Goldstone mode upon deviation from the saddle solution.

Thus, in variables ζ and ξ, it is insufficient to confine analysis to the quadratic action (2.45); it is necessary to continue the expansion to the next order in fluctuations. We can prove[3] that, in the third order, ζ and ξ are entangled only with variable q. With respect to the remaining soft modes b, p, and $\kappa\omega$, we can take a Gauss integral, making use of the fact that our analysis is carried out outside the fluctuation region. The contribution from commuting variables in the emerging superdeterminant exactly cancels the contribution from Grassmann variables.

As a result, we are left with the integral with respect to variables n, $\zeta\xi$, and q. The eigenfunction of the corresponding operators $\hat{\mathbb{O}}^{-}$ near the threshold is $f_0(\mathbf{r})$; consequently, we can single out the coordinate dependence: $n = \tilde{n}f_0(\mathbf{r})$, etc. We must retain in the action the term $\tilde{q}^2(\mathscr{E}^{-}_{\theta_2\theta_2})_0/\delta$ originating from Eq. (2.45) as well as the term proportional to $\tilde{\zeta}\tilde{\xi}\tilde{q}$ and removing degeneracy of the Grassmann zero mode.

In order to calculate the minimal eigenvalue of operator $\hat{\mathbb{O}}^{-}_{\theta_2\theta_2}$ in the main order in deviation from the threshold, we can use function f_0, since it is an eigenfunction of operators $\hat{\mathbb{O}}^{-}$ for $E = E_g$, which corresponds to a zero eigenvalue. Substituting $\theta_2 = \pi/2 + i\psi_0 + i\sqrt{\tilde{\varepsilon}}f_0$ into formula (2.46) and expanding in $\sqrt{\tilde{\varepsilon}}$, we obtain

$$\frac{(\mathscr{E}^{-}_{\theta_2\theta_2})_0}{\delta} = \frac{\pi\nu}{8}\int d\mathbf{r} \times f_0[-D\nabla^2 - 2E_g\sinh(\psi_0 + \sqrt{\tilde{\varepsilon}}f_0)]f_0 = -\frac{\tilde{G}}{2}\sqrt{\tilde{\varepsilon}}. \tag{3.2}$$

In order to calculate the term proportional to $\tilde{\zeta}\tilde{\xi}\tilde{q}$ in the action, it is sufficient to use expression (2.45), which gives $\tilde{\zeta}\tilde{\xi}(\mathscr{E}^{-}_{\theta_1\alpha})_0/\delta$; while calculating this term, we must take into account the difference between angle α and the instanton solution θ_2 associated with fluctuation of q. Using formula (A.4), we assume that $\alpha = \theta_2 - i\tilde{q}/\sqrt{2}f_0$. Expanding in $\tilde{q}$ analogously to relation (3.2), we obtain

$$\frac{(\mathscr{E}^{-}_{\theta_1\alpha})_0}{\delta} = \frac{\tilde{G}\tilde{q}}{4\sqrt{2}}. \tag{3.3}$$

Let us now calculate the preexponential factor in Eq. (2.24). For the density of states averaged over the volume, we must evaluate the integral

$$\begin{aligned}\frac{\nu}{4}\int d\mathbf{r}\,\mathrm{str}(k\Lambda Q) &= \frac{\nu}{2}\int d\mathbf{r}[2\cos\theta_F + \cos\alpha + \cos\beta] \\ &= -\frac{i\nu}{2}\int d\mathbf{r}[3\sinh(\psi_0 - \sqrt{\tilde{\varepsilon}}f_0) \\ &+ \sinh(\psi_0 + (\sqrt{\tilde{\varepsilon}} - \tilde{q}/\sqrt{2})f_0)] = \mathrm{const} + \frac{ic_1\tilde{q}}{2\sqrt{2}\delta}.\end{aligned} \tag{3.4}$$

Here, we have singled out the imaginary constant that makes zero contribution to the density of states.[4] While evaluating the preexponential factor (3.4), we have omitted the Grassmann variables ζ and ξ, which are thus retained only in the action. It can be proved that the contribution from the omitted terms is small in parameter $\sqrt{\tilde{\varepsilon}}$.

[3] In fact, eigenvalue $(\mathscr{E}^{-}_{\alpha\theta_F})_0$ is a function of both q and b (see formula (A.4)). If, however, we take into account the term on the order of $b\zeta\xi$ in the action, this will lead to the emergence of term $\tilde{b}\tilde{q}$ in the preexponential factor of formula (3.5), which is not an imaginary quantity as required and makes zero contribution to the density of states.

[4] The coefficient of $\tilde{q}$ is also imaginary; however, the subsequent integration with respect to $\tilde{q}$ (see relation (3.5)) will be carried out along the imaginary axis; consequently, the density of states will be due to the retained term.

Taking into account relations (3.2)–(3.4), we obtain the following expression for the density of states:

$$\langle\rho\rangle = e^{-\mathcal{S}_0}\mathrm{Re}\int\frac{d\tilde{n}d\tilde{q}}{\pi}d\tilde{\zeta}d\tilde{\xi}\frac{ic_1\tilde{q}}{2^{3/2}\delta} \times\exp\left[-\frac{\tilde{G}}{2}\left(-\sqrt{\tilde{\varepsilon}}\tilde{q}^2+\frac{\tilde{q}\tilde{\zeta}\tilde{\xi}}{2\sqrt{2}}\right)\right] = -\frac{c_1\tilde{G}}{16\pi\delta}e^{-\mathcal{S}_0}\mathrm{Im}\int d\tilde{n}d\tilde{q}\tilde{q}^2\exp\left(\frac{\tilde{G}}{2}\sqrt{\tilde{\varepsilon}}\tilde{q}^2\right). \quad (3.5)$$

In order to integrate with respect to the zero mode $\tilde{n}$, we use relation (A.4):

$$n = 2i\sin k_B\chi_B = (\psi_1-\psi_2)\chi_B = -2\sqrt{\tilde{\varepsilon}}f_0\chi_B,\quad \tilde{n} = -2\sqrt{\tilde{\varepsilon}}\chi_B,\quad \int d\tilde{n} = 4\pi\sqrt{\tilde{\varepsilon}}. \quad (3.6)$$

For the convergence of the integral with respect to $\tilde{q}$, we must choose the integration contour along the imaginary axis, which gives us the required imaginary unity. Substituting the instanton action (2.41), we finally obtain the density of states in the form

$$\langle\rho\rangle = \frac{c_1}{\delta}\sqrt{\frac{\pi}{8\tilde{G}\sqrt{\tilde{\varepsilon}}}}\exp\left(-\frac{4}{3}\tilde{G}\tilde{\varepsilon}^{3/2}\right), \quad (3.7)$$

where quantities $\tilde{\varepsilon}$ and $\tilde{G}$ are defined by formulas (2.37), (2.40), and (2.42). Formula (3.5) describes the behavior of the average density of states outside the fluctuation regions for $\tilde{G}^{-2/3} \ll \varepsilon \ll 1$.

We will also determine the density of states above the threshold. In this region, we have $\sqrt{\tilde{\varepsilon}} = i\sqrt{|\tilde{\varepsilon}|}$. We do not need the instanton solution since the main contribution to integral (2.24) comes from the neighborhood of the supersymmetric saddle point $\theta_F = \alpha = \beta = \theta_1 = \pi/2 + i\psi_0 - \sqrt{|\tilde{\varepsilon}|}f_0$. The integral over fluctuations becomes equal to unity, and the preexponential factor has the form

$$\langle\rho\rangle = \frac{\nu}{4}\mathrm{Re}\int d\mathbf{r}\,\mathrm{str}(k\Lambda Q) = 2\nu\mathrm{Im}\int d\mathbf{r}\sinh(\psi_0+i\sqrt{|\tilde{\varepsilon}|}f_0) = \frac{2c_1}{\delta}\sqrt{|\tilde{\varepsilon}|}. \quad (3.8)$$

Two results for the density of states above and below the threshold energy are "sewn together": the values on the left and right of the fluctuation region turn out to be of the same order of magnitude:

$$\langle\rho\rangle|_{\varepsilon\sim\pm\tilde{G}^{-2/3}} \sim \frac{1}{\tilde{G}^{1/3}\delta}. \quad (3.9)$$

Another characteristic property of the subgap density of states (3.7) is that the total number of energy levels under the gap is independent of $\tilde{G}$ and is of the order of unity:

$$\mathcal{N} \sim \int_0^{E_g}\langle\rho(E)\rangle dE \sim \frac{E_g}{\tilde{G}\delta} \sim 1. \quad (3.10)$$

It will be shown in Section 4 that this property may be violated for a contact with tunnel boundaries in the limit of a so-called strong tail.

3.2. Exact Solution near the Threshold

Let us now calculate the density of states taking into account the second instanton. The result obtained will be exact in the entire fluctuation region in the vicinity of the critical energy.

It was proved above that hard modes make zero contribution to the density of states. For this reason, in the general case, we are left with eight quasi-zero modes corresponding to the minimal eigenvalues of operators $\hat{\mathbb{O}}^-$: $b, n, p, q, \zeta\xi$, and $\kappa\omega$ (as we seek an exact solution, the division into zero and soft modes, which was made for saddle solutions, becomes meaningless). We take into account the coordinate dependence using the representation $n = \tilde{n}f_0(\mathbf{r})$, etc., and introduce, instead of variables $\tilde{b}$, $\tilde{p}$, and $\tilde{q}$, symmetric variables interpolating between the two solutions to the Usadel equation:

$$\alpha = \frac{\pi}{2}+i\psi_0+iuf_0,\quad \beta = \frac{\pi}{2}+i\psi_0+iVf_0,\quad \theta_F = \frac{\pi}{2}+i\psi_0+iwf_0. \quad (3.11)$$

We must now expand the action up to cubic terms in quasi-zero variables. Calculations similar to those used while deriving Eqs. (2.39), (3.2), and (3.3) lead to the following expression for action:

$$\mathcal{S} = \tilde{G}\left[\tilde{\varepsilon}(u+V-2w)-\frac{u^3+V^3-2w^3}{3} -\tilde{\zeta}\tilde{\xi}\frac{u+w}{4}-\tilde{\kappa}\tilde{\omega}\frac{V+w}{4}\right]. \quad (3.12)$$

Eight extrema of this action (two for each of variables u, V, and w) correspond to a supersymmetric saddle and to different instantons. Analogously to relation (3.4), we obtain the following expression for the preexponential factor:

$$\frac{\nu}{4}\int d\mathbf{r}\,\mathrm{str}(k\Lambda Q) = -\frac{ic_1}{2\delta}(u+V+2w). \quad (3.13)$$

We must now establish a correct measure of integration. The relation between variables u, V, w and $\tilde{p}$, $\tilde{q}$, $\tilde{b}$ follows from relations (A.4). Since the measure of

integration with respect to $\tilde{p}$, $\tilde{q}$, $\tilde{b}$ is trivial (see relation (A.5)), we conclude that

$$\mathcal{D}Q = \frac{1}{\pi^2} d\tilde{b} d\tilde{p} d\tilde{q} d\tilde{n} d\tilde{\zeta} d\tilde{\xi} d\tilde{\kappa} d\tilde{\omega} = \frac{4i}{\pi^2} du dv dw d\tilde{n} d\tilde{\zeta} d\tilde{\xi} d\tilde{\kappa} d\tilde{\omega}. \tag{3.14}$$

The integral with respect to the zero mode $\tilde{n}$ is taken analogously to the corresponding expression from (3.6): $\int d\tilde{n} = 2\pi|u - v|$. After integration with respect to $\tilde{n}$ and Grassmann variables, we obtain the following expression for the density of states:

$$\langle \rho \rangle = \frac{c_1 \tilde{G}^2}{4\pi\delta} \times \mathrm{Re} \int du dv dw |u - v|(u + w)(v + w)(u + v + 2w) \times \exp\left[-\tilde{G}\left(\tilde{\varepsilon}(u + v - 2w) - \frac{u^3 + v^3 - 2w^3}{3}\right)\right]. \tag{3.15}$$

Let us carry out the scaling transformation of the integration variables with a view to eliminate $\tilde{G}$: $(u, v, w) \longrightarrow (2\tilde{G})^{-1/3}(u, v, w)$, and pass to new variables $l = (u + v)/2$ and $m = (u - v)^2/2$. This leads to

$$\langle \rho \rangle = \frac{1}{8\pi^2 \Delta_g} \mathrm{Re} \int_0^\infty dm \times \int dl dw (w + l)(w^2 + 2lw + l^2 - m) \times \exp\left[-\frac{w^3}{3} + \epsilon w + \frac{l^3}{3} + ml - \epsilon l\right], \tag{3.16}$$

where

$$\epsilon = (2\tilde{G})^{2/3}\tilde{\varepsilon} = \frac{E_g - E}{\Delta_g}, \quad \Delta_g = \frac{c_2^{1/3}(\delta^2 E_g)^{1/3}}{2\pi^{2/3} c_1}. \tag{3.17}$$

At this stage, we must choose the contours of integration with respect to w and l. Integral (3.16) converges if the contours of integration with respect to l and w tend to infinity in hatched and light regions, respectively (see Fig. 4). This, however, is valid only when we can confine our analysis to the third-order terms in w and l in the expansion of action. In a region far from zero, convergence is determined by the properties of expression (2.31): the contour for w must tend to infinity along the imaginary axis (compactness of the FF sector), while the contour for l, along the real axis (noncompactness of the BB sector). Nevertheless, the main contribution to the integral comes from a neighborhood of zero; consequently, we will take this integral with a cubic action within infinite limits by displacing the contour for w to the left (C_1 in Fig. 4). For l, we have two alternatives: C_2 and C_3. A correct choice is determined by the condition of positiveness of the density of states, which corresponds to contour C_3 for l.

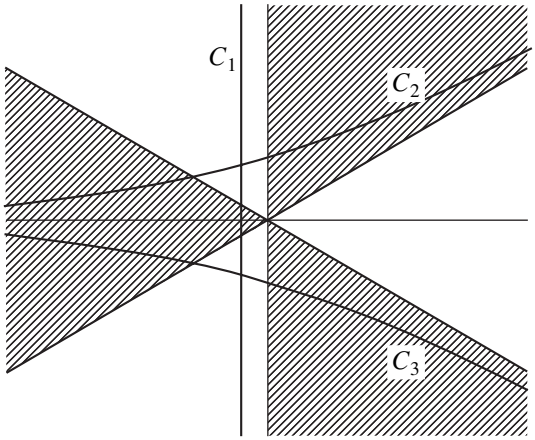


Fig. 4. Possible contours of integration with respect to w and l. Correct choice corresponds to C_1 for w and C_3 for l.

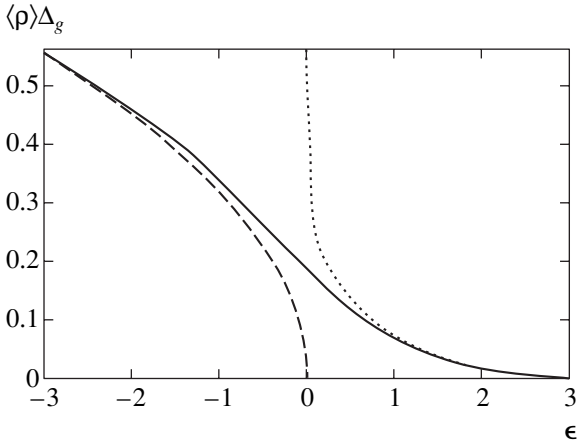


Fig. 5. Exact dependence (3.18) of the density of states $\langle\rho\rangle$ (in units of $1/\Delta_g$) on reduced energy ϵ (3.17) (solid curve). The size of the fluctuation region is on the order of unity. The asymptotic form of the density of states above the gap is given by expression (3.8) (dashed curve) and below the gap, by expression (3.7) (dotted curve).

Integral (3.16) is evaluated in Appendix B. The result of calculation is the following expression for the density of states,

$$\langle \rho \rangle = \frac{1}{\Delta_g}\left[-\epsilon \mathrm{Ai}^2(\epsilon) + [\mathrm{Ai}'(\epsilon)]^2 + \frac{\mathrm{Ai}(\epsilon)}{2}\left(1 - \int_\epsilon^\infty dy \mathrm{Ai}(y)\right)\right], \tag{3.18}$$

where ϵ is defined by the expression from (3.17). This dependence is depicted in Fig. 5 together with the asymptotic forms of relations (3.8) and (3.7).

The single-instanton approximation is valid in the limit $\epsilon \gg 1$. If we substitute into formula (3.18) the asymptotic expression for the Airy function,[5] the first two terms in the brackets cancel out and the integral is found to be much smaller than unity. As a result, Eq. (3.18) exactly coincides with expression (3.7).

Functional dependence (3.18) coincides with the predictions (formula (3.21) below) of the random matrix theory for the spectral edge in the case of an orthogonal ensemble [32]. Such a coincidence is not surprising in view of the equivalence of the random matrix theory and the zero-dimensional σ model [26]. In the case under investigation, the σ model becomes zero-dimensional in the vicinity of the threshold at the stage of introducing quasi-zero-dimensional variables (3.11) with a fixed coordinate dependence specified by function f_0. The application of the random matrix theory for NS systems will be discussed in greater detail in Subsection 3.3.

The results of this and preceding subsections are valid in the case when the normal region has an arbitrary shape and is linked via ideal contacts to an arbitrary number of superconductors with identical phases (see Fig. 3). It is only necessary that the approximation of the zero-dimensional σ model be satisfied. The criterion determining the effective dimensionality of the problem in an important special case of a planar SNS junction will be formulated in Subsection 5.1.

3.3. Method of Random Matrices

The random matrix theory was proposed by Wigner [17] in 1951 for describing properties of nuclear spectra. The idea of this theory is that the statistical properties of the spectrum of a complex and unknown nuclear Hamiltonian do not change upon replacement of all matrix elements by random numbers. Dyson [18] and Mehta [19] subsequently developed the random matrix theory.

In the simplest case, the random matrix theory is formulated for so-called Gaussian ensembles, for which the Hermitian random $N \times N$ matrix H acquired the statistical weight,

$$\mathcal{P}(H) \propto \exp\left(-\frac{\pi^2}{2N\delta^2}\mathrm{tr}H^2\right). \tag{3.19}$$

The average density of states for distribution (3.19) in the limit $N \longrightarrow \infty$ has the form of a Wigner semicircle,

$$\langle \rho(E)\rangle_{RMT} = \frac{1}{\delta}\sqrt{1-\frac{\pi^2E^2}{4N^2\delta^2}}, \tag{3.20}$$

vanishing for $|E| > E_0 \equiv (2/\pi)N\delta$. In accordance with the random matrix theory, the statistical properties of the spectrum at the center of the band for $|E| \ll E_0$ are determined only by the mean level spacing δ and by the symmetry of Hamiltonian H (orthogonal, unitary, or symplectic [19]).

In 1965, Gor'kov and Éliashberg [33] suggested that the random matrix theory may be applied for describing the statistical properties of energy levels in disordered metallic grains. However, the theory remained phenomenological up to 1982, when Efetov [21] used the supermatrix σ model for computing a pair correlator $R_2(\omega) = \delta^2\langle\rho(E)\rho(E+\omega)\rangle$ of energy levels for a metallic grain and proved that it coincides with the predictions of the random matrix theory in the zero-dimensional limit $\omega \ll E_{Th}$. Subsequently, the universality hypothesis, according to which local spectral properties of random systems near the center of a band are determined only by the symmetry of the Hamiltonian and not by its microscopic properties, was proved for a wide class of ensembles of random matrices [34].

In 1997, Altland and Zirnbauer [35] proposed a generalization of three standard Wigner–Dyson ensembles to the case of mixed NS systems by introducing four more symmetry classes, which take into account the mirror symmetry of the Bogoliubov–De Gennes equations. However, they confined analysis only to systems in which the average phase factor $\langle e^{i\phi}\rangle = 0$, where ϕ is the electron phase gained due to Andreev reflection processes. Such a situation is realized, for example, in the core of a superconducting vortex or in an SNS junction with a phase difference of π. Under such a condition, the proximity effect does not lead to the formation of a gap in the spectrum of the normal region, but specific superconducting correlations may appear for energies very close to the Fermi energy (at "distances" on the order of δ).

If, however, the superconducting order parameter averaged over the system does not vanish, the density of states is suppressed to zero on the Fermi surface, and a gap is formed for $E = E_g$ (in the semiclassical approximation). The states emerging near the gap cannot be described with the help of the random matrix theory at the center of the Wigner semicircle (3.20) even if we use new symmetry classes [35]. In 2001, Vavilov *et al.* [22] paid attention to the fact that the semiclassical density of states in NS systems near the threshold, as well as the density of states for a Wigner semicircle, exhibits identical (root) dependences on the distance to the threshold. They were the first to propose, in this connection, using the random matrix theory at the spectral

[5] The asymptotic form of the Airy function for $x \longrightarrow \infty$ has the form $\mathrm{Ai}(x) \sim \frac{1}{2\sqrt{\pi}x^{1/4}}e^{-2x^{3/2}/3}$.

edge for describing close-to-threshold states in NS systems.

The Wigner semicircle (3.20) describes the average density of states far away from thresholds $\pm E_0$. Corrections to it become significant at a distance of a single energy level from the threshold, i.e., for $|E| - E_0 \sim \Delta_0$, where $\Delta_0 = (\delta^2 E_0/2\pi^2)^{1/3}$. In the forbidden gap $|E| > E_0$, a nonzero average density of states appears; in this case,

$$\int_{E_0}^{\infty} \langle \rho(E) \rangle dE \sim 1,$$

which means that almost the entire density of states outside the Wigner semicircle is ensured by fluctuations of the position of the lowest energy level. Universal properties of the spectrum of random matrices near its edge were investigated in [32]. In particular, the exact profile of the average density of states near the threshold $E = E_0$ assumed the universal form in terms of the dimensionless quantity $x = (E - E_0)/\Delta_0$:

$$\langle \rho(E) \rangle = \frac{1}{\Delta_0}\left[-x\mathrm{Ai}^2(x) + [\mathrm{Ai}'(x)]^2 + \delta_{\beta,1}\frac{\mathrm{Ai}(x)}{2}\left(1 - \int_x^{\infty} dy\,\mathrm{Ai}(y)\right)\right], \quad (3.21)$$

where $\beta = 1$ or 2 for the orthogonal (unitary) class of symmetry.

The applicability of the random matrix theory for describing the edge of the spectrum in an NS system was proved for the first time in [30] with the help of the method of the supermatrix σ model. In the limit of the zero-dimensional σ model, which is valid in a close neighborhood of the threshold, we derived expression (3.18), exactly coinciding with result (3.21) obtained for the orthogonal symmetry.

4. CONTACT WITH TUNNEL INTERFACES

In this section, we consider an NS contact (Fig. 3) with nonideal boundaries. As the transparency of the interface decreases, quasiparticles stay for a longer time in the normal region between two Andreev reflections; accordingly, the gap in the density of states becomes smaller. In the tunnel limit, when the transparency of each channel between superconductors and the normal part of the contact is low, a gap is formed at energy $E_g = G_T\delta/8\pi$. In the vicinity of E_g, the semiclassical density of states vanishes in accordance with the law $\sqrt{|\varepsilon|}$; as the distance from the threshold increases, it attains its maximum value and then decreases in accordance with the law $1/\sqrt{|\varepsilon|}$ (see Eq. (4.30) below), resembling more and more a BCS-type singularity.

The shape of the tail in this case also changes, but the total number of quasi-localized states remains on the order of unity as before. The pattern changes qualitatively for $G_T \ll G^{1/4}$ (where G is the characteristic conductance of the normal part of the contact, which will be defined rigorously below). In this limit, the average number of energy levels in the region of the root increase in the density of states above energy E_g is on the order of unity, so that the entire region falls to the domain of strong fluctuations. The number of subgap states starts increasing simultaneously. This regime, which will be referred to as an strong tail, will be considered in Subsection 4.4.

4.1. Action for the Boundary

If the interface between a superconductor and a normal metal is ideal, matrix Q is continuous upon a transition from one region to another. If, however, the interface is not ideal, the boundary conditions become more complicated [36, 37]. These conditions can be taken into account automatically if we supplement the action of the σ model with an additional boundary term. It has the form [26, 38]

$$\mathcal{S}_{\text{boundary}} = -\frac{1}{4}\sum_i N_i \mathrm{str}\ln\left[1 - \frac{\Gamma_i}{2} + \frac{\Gamma_i}{4}\{Q, Q^{(i)}\}\right]. \quad (4.1)$$

Index i labels superconductors in contact with the normal region, and N_i is the total number of conducting channels in the ith contact per spin component. For a boundary of area S, the number of channels is $N = \pi\nu v_0 S$, where v_0 is the Fermi velocity; Γ_i is the transparency of a channel, which is assumed for simplicity to be identical for all the channels at the given interface; $Q^{(i)}$ is the value of matrix Q in the normal region in the vicinity of the ith contact; and Q_S is the value of matrix Q in the superconductor. Expression (4.1) can be used only if matrix $Q^{(i)}$ is constant along each interface. This condition is satisfied automatically for a planar SNS junction (see Fig. 1); in the general case (see Fig. 3), we must require that the size of the contacts is small as compared to the characteristic size of the normal region.

In our parametrization, the separation of variables (2.30) is preserved as before and the expression for S_0 acquires an additional term,

$$S_0[\theta] = \frac{\pi\nu}{4}\int d\mathbf{r}[D(\nabla\theta)^2 + 4iE\cos\theta] - \frac{1}{2}\sum_i N_i \ln(1 + \gamma_i \sin\theta^{(i)}), \quad (4.2)$$

where $\gamma_i = \Gamma_i/(2 - \Gamma_i)$. Here, $\theta^{(i)}$ is the value of angle θ in the normal region in the vicinity of the ith contact. It

was noted above that this value is independent of the coordinates along the boundary. This allows us to use the coordinate representation for the first term in the action and the channel representation for the second term. Henceforth, we will use the superscript "*i*" for values of various fields in the vicinity of the *i*th contact.

The entire classification of instantons made in Subsection 2.4 is preserved. In the expansion in fluctuations, we must supplement action (2.44) with the boundary term:

$$\mathcal{S}^{(2)}[W] = \frac{\pi\nu}{8} \times \int d\mathbf{r}\,\mathrm{str}\Big[D(\nabla W)^2 + \frac{D}{4}[\nabla U_0, W]^2 - 2iE\Lambda Q_0 W^2\Big] + \frac{1}{4}\sum_i N_i T_i \mathrm{str}\Big(\frac{\tau_x Q_0}{1 + T_i\tau_x Q_0}\tilde{W}^{(i)}\frac{1}{1 + T_i\tau_x Q_0}\tilde{W}^{(i)}\Big). \tag{4.3}$$

Here, we have introduced the notation $\tilde{W} = e^{-iU_0/2}We^{iU_0/2}$ and $T_i = (2 - \Gamma_i - 2\sqrt{1-\Gamma_i})/\Gamma_i = \gamma_i - 2\sqrt{1-\gamma_i^2}/\gamma_i$. Parametrization of W diagonalizing this action remains unchanged (Appendix A), but operator $\hat{\mathbb{O}}^{\pm}_{\alpha\beta}$ also acquires a boundary term:

$$(a\hat{\mathbb{O}}^{\pm}_{\alpha\beta}b) = \frac{\pi\nu}{8}\int d\mathbf{r}\Big[D(\nabla a(\mathbf{r}))(\nabla b(\mathbf{r})) - a(\mathbf{r})\Big(\frac{D}{4}(\nabla\alpha \pm \nabla\beta)^2 + iE(\cos\alpha + \cos\beta)\Big)b(\mathbf{r})\Big] + \sum_i \frac{N_i\gamma_i}{16} \times \Bigg(a^{(i)}\frac{(\sin\alpha^{(i)} + \sin\beta^{(i)}) + \gamma_i(1 \mp \cos(\alpha^{(i)} \pm \beta^{(i)}))}{(1 + \gamma_i\sin\alpha^{(i)})(1 + \gamma_i\sin\beta^{(i)})}b^{(i)}\Bigg). \tag{4.4}$$

Thus, the inclusion of nonideality of the boundary does not basically change the strategy of calculating the density of states near the threshold. We can repeat all calculations made for an ideal boundary in Section 3, taking into account the boundary term in the action. The only difference will be redefinition of functions $\psi_{1,2}$ and f_0 as well as constants $c_{1,2}$. The new definitions have the form

$$D\nabla^2\psi + 2E\cosh\psi = 0,\qquad D\nabla_n\psi^{(i)} + \gamma_i v_0\frac{\sinh\psi^{(i)}}{1 + \gamma_i\cosh\psi^{(i)}} = 0; \tag{4.5}$$

$$D\nabla^2 f_0 + 2E_g f_0\sinh\psi_0 = 0,\qquad D\nabla_n f_0^{(i)} + \gamma_i v_0 f_0^{(i)}\frac{\gamma_i\cosh\psi_0^{(i)}}{(1 + \gamma_i\cosh\psi_0^{(i)})^2} = 0; \tag{4.6}$$

$$c_1 = \frac{1}{V}\int d\mathbf{r}\, f_0\cosh\psi_0; \tag{4.7}$$

$$c_2 = \frac{1}{V}\int d\mathbf{r}\, f_0^3\cosh\psi_0 - \frac{\delta}{2\pi E_g} \times \sum_i N_i\gamma_i\sinh\psi_0^{(i)}\frac{1 - \gamma_i^2 - \cosh\psi_0^{(i)}}{(1 + \gamma_i\cosh\psi_0^{(i)})^3}[f_0^{(i)}]^3. \tag{4.8}$$

With these definitions, the results obtained in Subsections 3.1 and 3.2 are preserved.

It should be recalled once again that all formulas in this subsection are valid only when functions $\psi_{1,2}$ and f_0 are constant at the boundaries with a superconductor.

4.2. Zero-Dimensional Action

Here, we consider the case when the gradient terms in action (4.2) have the meaning of a small correction. Such a situation takes place either in the limit of tunnel contacts $\Gamma \approx 2\gamma \ll 1$, or for contacts whose size is smaller (Fig. 3) than the mean free path: $S_i \ll l^2$. In the latter case, diffusion in the normal part of the contact is not necessarily required; the mean free path can be on the order of the size of the system. Taking into account the gradient terms as a small correction, we can determine the optimal coordinate dependence $\psi(\mathbf{r})$ and obtain a zero-dimensional action.

Thus, if $\psi(\mathbf{r})$ changes in space insignificantly, we can single out a large constant,

$$\psi = A + \phi(\mathbf{r}), \tag{4.9}$$

and expand action (4.2) into a series up to the first order in ϕ and to the second order in $\nabla\phi$; in addition, we expand the logarithm in the action up to the second order in $\gamma\cosh A$ (omitting an insignificant constant). The validity of the inequality $\gamma\cosh A \ll 1$ will be proved below:

$$S_0[\psi] = \frac{\pi E}{\delta}\sinh A - \frac{G_T}{8}\cosh A + \frac{\pi\nu}{4}\int d\mathbf{r}(-D(\nabla\phi)^2 + 4E\phi\cosh A) + \frac{1}{4}\Big(\sum_i N_i\gamma_i^2\Big)\cosh^2 A - \frac{1}{2}\sum_i N_i\gamma_i\phi^{(i)}\sinh A, \tag{4.10}$$

where we have introduced the total tunnel conductance

of the contacts,

$$G_T = 2\sum_i N_i\Gamma_i \approx 4\sum_i N_i\gamma_i. \tag{4.11}$$

Disregarding the terms in the second and third lines in Eq. (4.10) and varying the action in A, we obtain $8\pi E/(G_T\delta) = \tanh A$. The effect of the omitted terms becomes significant for large values of A; in this case, corrections to $\tanh A$ are negative. This ensures the emergence of a peak in the dependence of energy on A, which determines the gap width. In the first approximation, we can assume that

$$E_g = \frac{G_T\delta}{8\pi}. \tag{4.12}$$

In the framework of the approximation used by us here, we can replace energy E in the fourth term in Eq. (4.10) by E_g. Using also the fact that A is large, we can replace the hyperbolic functions of A in the second and third lines of Eq. (4.10) by $P/2$, where $P = e^A$. As a result, we obtain action in the form

$$S_0[P,\phi] = -\frac{G_T}{16}\left[\varepsilon P + \frac{2}{P}\right] + \frac{P^2}{16}\sum_i N_i\gamma_i^2 + \frac{\pi\nu}{4}\int d\mathbf{r}\left(-D(\nabla\phi)^2 + \frac{G_T\delta}{4\pi}P\phi\right) - \frac{P}{4}\sum_i N_i\gamma_i\phi^{(i)}. \tag{4.13}$$

We have also introduced here the dimensionless energy measured from the edge of the gap:

$$\varepsilon = \frac{E_g - E}{E_g} = 1 - \frac{8\pi E}{G_T\delta}. \tag{4.14}$$

Varying the obtained action with respect to ϕ, we obtain the following equations:

$$D\nabla^2\phi + \frac{\delta G_T P}{8\pi} = 0, \qquad 2D\nabla_n\phi^{(i)} + P\nu_0\gamma_i = 0. \tag{4.15}$$

These equations define function ϕ to within a constant term since we have not fixed the constant in Eq. (4.9). For the sake of convenience, we introduce the function $\Phi(\mathbf{r}) = 2D\phi(\mathbf{r})/P\nu_0$, which has the dimension of length and satisfies the equations

$$V\nabla^2\Phi + \sum_i S_i\gamma_i = 0, \quad \nabla_n\Phi^{(i)} + \gamma_i = 0. \tag{4.16}$$

Here, S_i is the area of the ith contact. In such a form, the compatibility of the equation and the boundary conditions becomes obvious. Indeed, having integrating the first equation over the volume and applying the Gauss theorem, we obtain $\sum_i S_i(\nabla_n\Phi^{(i)} + \gamma_i) = 0$, which matches the boundary conditions on Eqs. (4.16).

Using Eqs. (4.16), we can simplify action (4.13), which becomes now a function of P only:

$$S_0(P) = -\frac{G_T}{16}\left[\varepsilon P + \frac{2}{P}\right] + \frac{P^2}{16}\sum_i N_i\gamma_i^2 + \frac{3\pi\nu\nu_0 P^2}{16l}\int d\mathbf{r}(\nabla\Phi)^2. \tag{4.17}$$

This expression indicates once again that the arbitrary constant in Φ is deprived of physical meaning. For subsequent calculations, we require only one quantity characterizing the geometry of the system and the properties of the junctions:

$$\tilde{\gamma} = \left(\sum_i S_i\gamma_i\right)^{-1}\left[\sum_i S_i\gamma_i^2 + \frac{3}{l}\int d\mathbf{r}(\nabla\Phi)^2\right]. \tag{4.18}$$

Let us consider some particular cases.

(i) **One-dimensional junction.** Let the normal region be a rectangle of length L (see Fig. 1) and be connected with two superconductors by identical junctions with a transparency of $2\gamma \ll 1$ (for each junction). In this case, the solution to Eq. (4.16) has the form $\Phi(x) = -\gamma x^2/L$, while parameter (4.18) becomes $\tilde{\gamma} = L\gamma/2l + \gamma \approx L\gamma/2l$. Thus, the first term in the square brackets in Eq. (4.18) can be neglected in comparison with the second term. If a superconductor is fixed only at one side, the corresponding parameter $\tilde{\gamma}$ becomes twice as large: $\tilde{\gamma} = L\gamma/l$.

(ii) **Two-dimensional junction.** If the normal region has the shape of a circle of radius R, bordering a superconductor, then $\Phi(r) = -\gamma r^2/2R$, and we ultimately have $\tilde{\gamma} = 3R\gamma/4l$.

(iii) **Three-dimensional junction.** Similarly, if the normal region has the shape of a sphere of radius R, function $\Phi(r)$ is the same as in the previous case and the result is $\tilde{\gamma} = 3R\gamma/5l$.

(iv) **Zero-dimensional junction.** By a zero-dimensional junction, we assume a normal metal of an arbitrary shape, connected to superconductors by narrow junctions (see Fig. 3). In this case, we can disregard the second term in the square brackets of Eq. (4.18) as compared to the first term; if all the channels have the same transparency Γ, we have $\tilde{\gamma} = \Gamma/2$. Indeed, if a junction has a shape similar to that depicted in Fig. 3, the main contribution to integral $\int d\mathbf{r}(\nabla\Phi)^2$ comes from regions in the vicinity of the superconductors. If the distance r from the ith junction with the superconductor is much larger than the contact size ($r \gg \sqrt{S_i}$), but much smaller than the size of the normal metal, we

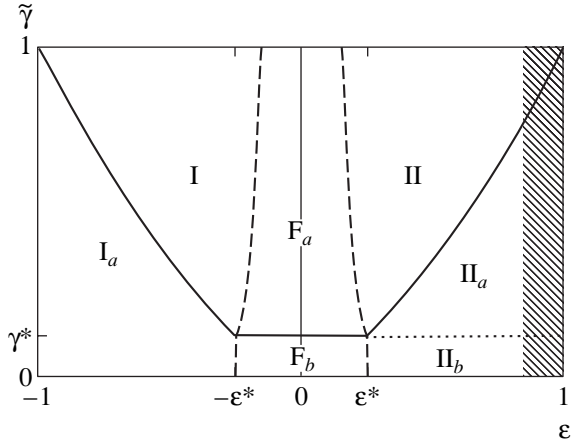


Fig. 6. Approximate "phase diagram" for possible energy dependences of the density of states. Formula (4.25) corresponds to regions I, (4.30) to I_a, (4.26) to II, and (4.44) to II_a and II_b. The density of states in regions I, F_a, and II is described by the universal formula (3.18) taking into account relations (4.24). The total number of prelocalized states in region II_b is large (see relation (4.46)). Parameters $\varepsilon^* = G^{-1/2}$ and $\gamma^* = G^{-3/4}$.

can approximately assume that $|\nabla\Phi| \approx S_i\gamma_i/2\pi r^2$. The integral of the square of this quantity is proportional to $S_i^{3/2}\gamma_i^2$ and can be disregarded in expression (4.18) provided that $S_i \ll l^2$.

The value of $\tilde{\gamma}$ depends on the geometry of the system and, in addition, is proportional to the "average" transparency of a channel in a junction. The tunnel conductance G_T is also proportional to the transparency of a channel. In order to analyze the behavior of various parameters of the system upon a change in the resistance of the junctions, we introduce a quantity describing the shape of the system only,

$$G = G_T/\tilde{\gamma}. \tag{4.19}$$

For a planar junction (see Fig. 1), it is proportional to the conductance of the normal part: $G = 12G_N$.

As a result, we obtain the zero-dimensional action

$$S_0(P) = \frac{G\tilde{\gamma}}{16}\left[-\varepsilon P - \frac{2}{P} + \frac{\tilde{\gamma}P^2}{4}\right]. \tag{4.20}$$

4.3. Classification of Tails

Here, we consider various limiting cases of variation of parameters $\tilde{\gamma}$ and ε, calculate the actions of instantons, and construct a rough pattern of possible behavior of the subgap density of states in various limiting cases.

By varying $S_0(P)$, we obtain the following equation[6] connecting P and ε:

$$\varepsilon = \frac{2}{P^2} + \frac{\tilde{\gamma}P}{2}. \tag{4.21}$$

The minimal value of ε is attained for $P = P_0 = 2\tilde{\gamma}^{-1/3}$ and is equal to $3\tilde{\gamma}^{2/3}/2 \ll 1$. This quantity has a nonzero value since the threshold energy E_g is in fact smaller than $G\tilde{\gamma}\delta/8\pi$. However, this correction is small and should be taken into account only while deriving the next formula.

For $E < E_g$, Eq. (4.21) has two real positive roots $P_{1,2}$ corresponding to two solutions $\theta_{1,2}$ to the Usadel equation. These roots can be found by expanding Eq. (4.20) in small parameters ε and $\delta P = P - P_0$:

$$S_0(P) = S_0(P_0) + \frac{G\tilde{\gamma}}{16}\left[-\varepsilon\delta P + \frac{2(\delta P)^3}{P_0^4}\right]. \tag{4.22}$$

Here, we have taken into account in ε the correction which was omitted in Eq. (4.14): $\varepsilon = 1 - 3\tilde{\gamma}^{2/3}/2 - 4\pi E/G\tilde{\gamma}\delta$. For our subsequent analysis, the accuracy of expression (4.14) is sufficient.

Expansion (4.22) leads to the conclusion that two solutions to the Usadel equation correspond to $P_{1,2} = P_0(1 \mp P_0\sqrt{\varepsilon/6})$, while the exponent for the density of states has the form

$$\mathcal{S} = S_0(P_1) - S_0(P_2) = \frac{G\tilde{\gamma}^{1/3}}{3\sqrt{6}}\varepsilon^{3/2}. \tag{4.23}$$

The same result can be obtained by comparing expansions (4.22) and (2.39). To this end, we must first establish the relation between g and δP. In the regime $\tilde{\gamma} \ll 1$, function $\psi_0(\mathbf{r})$ differs from a constant insignificantly; for this reason, in the main order, we can assume that $f_0(\mathbf{r}) \approx 1$. Thus, $\delta\psi = g$, and on the other hand, $\delta\psi = \delta\ln P = \delta P/P_0$. Equating the right-hand sides of Eqs. (4.22) and (2.39) termwise, we ultimately obtain

$$c_1 = \frac{P_0}{2} = \tilde{\gamma}^{-1/3}, \quad c_2 = \frac{6}{P_0} = 3\tilde{\gamma}^{1/3}, \quad \tilde{G} = \frac{3}{16}G\tilde{\gamma}^{4/3}. \tag{4.24}$$

If we substitute these constants into Eq. (2.41), we again arrive at result (4.23). Using expressions (4.24), we can generalize all results obtained in Subsections 3.1 and 3.2. In particular, for the density of states above the gap, we obtain from Eq. (3.8)

$$\langle\rho\rangle = \frac{1}{\tilde{\gamma}^{2/3}\delta}\sqrt{\frac{8|\varepsilon|}{3}}. \tag{4.25}$$

[6] For an ideal boundary, an analog of this equation was Eq. (2.6).

Similarly, we can determine the preexponential factor for the density of states under the gap from Eq. (3.7):

$$\langle\rho\rangle = \frac{1}{\delta}\sqrt{\frac{\pi}{G\tilde{\gamma}^{5/3}}}\sqrt{\frac{2}{3\varepsilon}}\exp\left(-\frac{G\tilde{\gamma}^{1/3}}{3\sqrt{6}}\varepsilon^{3/2}\right). \quad (4.26)$$

The last two formulas are valid outside the fluctuation region provided that $|\varepsilon| \gg G^{-2/3}\tilde{\gamma}^{-2/9}$, when we can disregard the contribution of the second instanton as compared to the first one. Figure 6 shows a rough "phase diagram" of possible behaviors of density of states for various values of $\tilde{\gamma}$ and ε. Formulas (4.25) and (4.26) correspond to regions I and II, respectively. Fluctuation region F_a lies between these regions. A universal expression for the density of states in the three regions (I, F_a, and II) can be obtained from Eq. (3.18) using Eq. (4.24) also.

Expansion (4.22) is valid if the energy differs from the threshold energy insignificantly. The validity of this expansion requires the fulfillment of inequality $|\delta P| \ll P_0$ or, which is the same, $|\varepsilon| \ll \tilde{\gamma}^{2/3}$.

In the opposite limit $\tilde{\gamma}^{2/3} \ll \varepsilon \ll 1$, the cubic parabola approximation (4.22) is inapplicable, and we must solve Eq. (4.21) to find $P_{1,2}$. In order to find the smaller root (P_1), we can disregard the last term, while the second root (P_2) can be found by disregarding the second term:

$$P_1 = \sqrt{\frac{2}{\varepsilon}}, \quad P_2 = \frac{2\varepsilon}{\tilde{\gamma}}. \quad (4.27)$$

Substituting these values into the expression for action (first part of relation (4.22)), we obtain

$$S_0(P_1) = -\frac{G\tilde{\gamma}}{4\sqrt{2}}\sqrt{\varepsilon}, \quad S_0(P_2) = -\frac{G\varepsilon^2}{16}. \quad (4.28)$$

Since $S_0(P_1) \ll S_0(P_2)$, the exponent for density of states is determined by $S_0(P_2)$ only:

$$\mathcal{S} = -S_0(P_2) = \frac{G\varepsilon^2}{16}. \quad (4.29)$$

The density of states with such an exponent corresponds to regions II_a and II_b. The preexponential factor for this regime will be obtained in the next subsection.

Formula (4.25) is also inapplicable in region I_a above the gap. In order to determine the density of states in this regime, we must solve Eq. (4.21) for a large negative ε. In order to find pairs of complex conjugate roots, we can disregard the last term in Eq. (4.21): $P_{1,2} = \pm i\sqrt{2/|\varepsilon|}$. In this case, density of states has the form

$$\langle\rho\rangle = \frac{2}{\delta}\mathrm{Re}\cos\theta = \frac{1}{\delta}\mathrm{Im}(P - P^{-1}) \approx \frac{1}{\delta}\sqrt{\frac{2}{|\varepsilon|}}. \quad (4.30)$$

Here, we have used root P_1 since the density of states must be positive.

As transparency $\tilde{\gamma}$ decreases, regions I and II become narrower and vanish when $\tilde{\gamma} \sim G^{-3/4}$. The single-instanton contribution to the density of states in this case is described only by an exponential function with exponent (4.29). It will be shown in Subsection 4.4 that the total number of subgap states increases as the value of $\tilde{\gamma}$ decreases below $G^{-3/4}$ (region II_b). Such a behavior will be referred to as an strong tail.

In the vicinity of the Fermi energy ($\varepsilon = 1$), a crossover to a non-zero-dimensional σ model takes place. The corresponding density of states will be considered in Subsection 5.3. This region is shaded in the figure.

It should be emphasized once again that a crossover from one dependence to another takes place near the boundaries of the regions, and the above formulas are valid only in the bulk of the corresponding region.

4.4. Strong Tail

Let us now consider the case when $\tilde{\gamma} \ll G^{-3/4}$. It corresponds to region II_b in Fig. 6. The density of state was determined with an exponential accuracy in the preceding subsection. We will determine now the preexponential factor.

The fluctuation region for the case under study is so large that the difference between functions ψ_1 and ψ_2 outside this region cannot be regarded as small, but their dependence on coordinates is still weak. While evaluating the Gaussian integral over hard modes, we obtain, as in Subsection 3.1, the factor

$$\sqrt{\frac{\det\hat{\mathbb{O}}^{+}_{\theta_1\theta_2}\det'\hat{\mathbb{O}}^{-}_{\theta_1\theta_2}}{\det\hat{\mathbb{O}}^{+}_{\theta_1\theta_1}\det'\hat{\mathbb{O}}^{-}_{\theta_1\theta_1}}} = \sqrt{\frac{(\mathscr{E}^{+}_{\theta_1\theta_2})_0}{(\mathscr{E}^{+}_{\theta_1\theta_1})_0}}, \quad (4.31)$$

where we have confined our analysis only to minimal eigenvalues of operators $\hat{\mathbb{O}}^{+}_{\theta_1\theta_2}$ and $\hat{\mathbb{O}}^{+}_{\theta_1\theta_1}$, since higher eigenvalues are mainly determined by the gradient term and exhibit a weak dependence on energy as before.

However, we will start with determining the quantities $(\mathscr{E}^{-}_{\theta_1\theta_1})_0$ and $(\mathscr{E}^{-}_{\theta_2\theta_2})_0$, which will be required anyway in subsequent analysis. For this purpose, we use relations (A.4) and formulas (2.30) and (4.20):

$$\frac{(\mathscr{E}^{-}_{\theta_1\theta_1})_0}{\delta} = \frac{1}{2}\frac{\partial^2\mathcal{S}}{\partial p^2} = -\frac{1}{4}\frac{\partial^2\mathcal{S}}{\partial\beta^2} \quad (4.32)$$

$$= \frac{1}{4}S_0''(P)\left(\frac{\partial P}{\partial\theta}\right)^2\Bigg|_{P=P_1} = \frac{G\tilde{\gamma}}{16}\left[\frac{1}{P_1} - \frac{\tilde{\gamma}P_1^2}{8}\right] = \frac{G\tilde{\gamma}}{16}\sqrt{\frac{\varepsilon}{2}}.$$

Analogous differentiation with respect to variable q gives

$$\frac{(\mathscr{E}^-_{\theta_2\theta_2})_0}{\delta} = -\frac{G\varepsilon^2}{32}. \tag{4.33}$$

The eigenvalue $(\mathscr{E}^-_{\theta_1\theta_2})_0$ corresponds to rotation of angle χ_B and is equal to zero. We will calculate $(\mathscr{E}^+_{\theta_1\theta_2})_0$ using perturbation theory. The corresponding eigenfunction differs insignificantly from a constant and can be set equal to unity:

$$\begin{aligned}(\mathscr{E}^+_{\theta_1\theta_2})_0 &= (1[\hat{\mathbb{O}}^+_{\theta_1\theta_2} - \hat{\mathbb{O}}^-_{\theta_1\theta_2}]1)\\ &= \frac{\pi\nu D}{8}\int d\mathbf{r}(\nabla\psi_1)(\nabla\psi_2)\\ &+ \sum_i \frac{N_i\gamma_i^2}{8}\cosh\psi_1^{(i)}\cosh\psi_2^{(i)} = \frac{\pi\nu\nu_0 P_1P_2}{32}\\ &\times\left[\frac{3}{l}\int d\mathbf{r}(\nabla\Phi)^2 + \sum_i S_i\gamma_i^2\right] = \frac{G\tilde{\gamma}}{32}\sqrt{\frac{\varepsilon}{2}}.\end{aligned} \tag{4.34}$$

Here, we have used expression (4.18) for $\tilde{\gamma}$.

In the same way, we can also calculate $(\mathscr{E}^+_{\theta_1\theta_1})_0$, but the perturbation theory correction to formula (4.32) obtained in this case will be small, so we can assume that

$$(\mathscr{E}^+_{\theta_1\theta_1})_0 = (\mathscr{E}^-_{\theta_1\theta_1})_0 = \frac{G\tilde{\gamma}}{16}\sqrt{\frac{\varepsilon}{2}}. \tag{4.35}$$

Thus, the contribution to the preexponential factor from hard fluctuations can be reduced to

$$\sqrt{\frac{(\mathscr{E}^+_{\theta_1\theta_2})_0}{(\mathscr{E}^+_{\theta_1\theta_1})_0}} = \frac{1}{\sqrt{2}}. \tag{4.36}$$

Subsequent calculations are similar to those from Subsection 3.1. Fluctuations of soft modes of variables b, p, and $\kappa\omega$ mutually cancel out, and we are left with an integral with respect to variables n, $\zeta\xi$, and q. The pair of Grassmann variables[7] ζ, ξ appears in the action in the form $\zeta\xi(\mathscr{E}^-_{\theta_1\alpha})_0/\delta$. In order to evaluate this eigenvalue (which differs from zero when $\alpha \neq \theta_2$), we will use the eigenfunction $\sinh((\psi_2-\psi_1)/2)$, which assumes, after normalization, the form

$$a(\mathbf{r}) = 1 + \frac{3(P_2-P_1)}{4l}\Phi(\mathbf{r}). \tag{4.37}$$

Substituting this expression into formula (4.4), using Eqs. (4.16) and (4.18), and carrying out cumbersome, but elementary transformations, we obtain

$$\begin{aligned}\frac{(\mathscr{E}^-_{\theta_1\alpha})_0}{\delta} &= \frac{G\tilde{\gamma}}{128}\frac{P_2+P_1}{P_2-P_1}\\ &\times\left[P_1\left(-\varepsilon + \frac{2}{P_1^2} + \frac{\tilde{\gamma}P_1}{2}\right) - P_2\left(-\varepsilon + \frac{2}{P_2^2} + \frac{\tilde{\gamma}P_2}{2}\right)\right].\end{aligned} \tag{4.38}$$

Here, P_2 stands not for the root of Eq. (4.21), but a variable parametrizing angle α: $\alpha \approx \pi/2 + i\ln P_2$. The quantity P_2 satisfies Eq. (4.21) only for $\alpha = \theta_2$; in this case, expression (4.38) obviously vanishes. The same was observed for ideal boundaries. At a saddle point, mode $\zeta\xi$ is zero, but its mass increases linearly upon a deviation in variable q. In our case, this corresponds to a deviation in P_2:

$$\delta P_2 = -iP_2\delta\alpha = -\frac{P_2}{\sqrt{2}}q. \tag{4.39}$$

Differentiating expression (4.38) with respect to P_2 (we must differentiate only the second term in the square brackets), we obtain

$$\frac{(\mathscr{E}^-_{\theta_1\alpha})_0}{\delta} = -\frac{G\tilde{\gamma}\varepsilon}{128}\delta P_2 = \frac{G\varepsilon^2 q}{64\sqrt{2}}. \tag{4.40}$$

Let us now calculate the preexponential factor in Eq. (2.24) in our approximation. Analogously to Eq. (3.4), we have

$$\begin{aligned}\frac{\nu}{4}\int d\mathbf{r}\,\mathrm{str}(k\Lambda Q) &= \mathrm{const} - \frac{\nu}{2}\delta\alpha\int d\mathbf{r}\cosh\psi_2\\ &= \mathrm{const} + \frac{i}{2\sqrt{2}\delta}\frac{\varepsilon q}{\tilde{\gamma}}.\end{aligned} \tag{4.41}$$

It remains for us only to determine the measure of integration with respect to variable n corresponding to the zero mode of rotation in angle χ_B. Using relations (A.4), we obtain, analogously to (3.6),

$$\begin{aligned}n &= 2i\sin k_B\chi_B = -2\sinh\frac{\psi_2-\psi_1}{2}\chi_B\\ &= -2\sqrt{\frac{P_2}{P_1}}\chi_B = -\frac{(2\varepsilon)^{3/4}}{\sqrt{2\tilde{\gamma}}}\chi_B,\\ \int dn &= \pi\sqrt{\frac{2}{\tilde{\gamma}}}(2\varepsilon)^{3/4}.\end{aligned} \tag{4.42}$$

[7] In Subsection 3.1, we used the notation $\zeta(\mathbf{r}) = \tilde{\zeta}f_0(\mathbf{r})$, etc., where the amplitude of fluctuations and their spatial profile are separated explicitly. In the limit we are dealing with now, the normalized eigenfunctions of operators $\hat{\mathbb{O}}$, corresponding to the lowest eigenvalues, differ from unity only slightly; consequently, we may not distinguish between $\tilde{\zeta}$ and ζ within the required accuracy.

We have now everything required for calculating the density of states. Using relations (4.33), (4.36), and (4.40)–(4.42), we obtain

$$\langle\rho\rangle = \frac{e^{-\mathcal{S}_0}}{\sqrt{2}}\mathrm{Re}\int\frac{dndq}{n}d\zeta d\xi\frac{i}{2\sqrt{2}\delta}\frac{\varepsilon q}{\tilde{\gamma}}$$
$$\times\exp\left[\frac{G\varepsilon^2}{32}\left(q^2-\frac{q\zeta\xi}{2\sqrt{2}}\right)\right] = -\frac{2^{3/4}}{256\delta}G\frac{\varepsilon^{15/4}}{\tilde{\gamma}^{3/2}} \qquad (4.43)$$
$$\times e^{-\mathcal{S}_0}\mathrm{Im}\int dqq^2\exp\left(\frac{G\varepsilon^2q^2}{32}\right).$$

Integrating along the imaginary axis, we finally obtain

$$\langle\rho\rangle = \frac{1}{\delta}\sqrt{\frac{\pi}{G\tilde{\gamma}^3}}\sqrt{\frac{\varepsilon^3}{8}}\exp\left(-\frac{G\varepsilon^2}{16}\right). \qquad (4.44)$$

This result describes the behavior of density of states in regions II_a and II_b in Fig. 6. In region II_a, the obtained dependence is matched with Eq. (4.26) for $\varepsilon\sim\tilde{\gamma}^{2/3}$:

$$\langle\rho\rangle|_{\varepsilon\sim\tilde{\gamma}^{2/3}}\sim\frac{1}{\delta\sqrt{G\tilde{\gamma}}}\exp(-G\tilde{\gamma}^{4/3}). \qquad (4.45)$$

Matching in region II_b is more complicated. The semiclassical density of states vanishes according to root law (4.25) at the boundary of the gap, attains its peak value for $|\varepsilon|\sim\tilde{\gamma}^{2/3}$, and then decreases in accordance with inverse root law (4.30). As the transparency of the boundary decreases, the size of the fluctuation region increases, and this region covers the peak of the density of states for $\tilde{\gamma}\sim G^{-3/4}$.

The density of states increases as we approach fluctuation region F_b on both sides; consequently, there must be no formal matching of these two dependences: the peak of the density of states lies in the fluctuation region.

Finally, let us estimate the total number of quasi-localized states:

$$\mathcal{N}\sim\int_0^1\frac{E_gd\varepsilon}{\delta\sqrt{G\tilde{\gamma}^3}}\varepsilon^{3/4}e^{-G\varepsilon^2}\sim\tilde{\gamma}^{-1/2}G^{-3/8}\gg 1. \qquad (4.46)$$

Thus, for $\tilde{\gamma}\ll G^{-3/4}$, the number of states under the gap becomes large, which explains the term "strong tail" used in this case. These are apparently most suitable conditions for experimental verification of the above theory.

5. NONUNIVERSAL DENSITY OF STATES

In this section, we consider various cases in which density of states cannot be described by the zero-dimensional σ model. Subsection 5.1 deals with the situation when the instanton describing subgap states has a finite spatial size smaller than the size of the normal part of the contact. Fluctuations near such an instanton solution have a quasi-continuous spectrum, which considerably complicates the calculation of the preexponential factor in the density of states. For this reason, all results of this subsection (and the entire Section 5) will be obtained with an exponential accuracy.

In Subsection 5.2, the problem of subgap states in a superconductor with magnetic impurities is considered. In contrast to the rest of the material, it deals with a spatially homogeneous sample and not with a hybrid system. Quasi-localized states appear owing to spatial fluctuations of magnetic impurities. The problem of subgap states in such a system was solved in [23, 24]. The results are derived, and their relation with quasi-localized states in hybrid systems is demonstrated.

Subsection 5.3 is devoted to the density of states deep in the gap in the vicinity of the Fermi energy. This density of states cannot be described by the zero-dimensional σ model either and will be determined with an exponential accuracy.

5.1. Broad SNS Junction

This and the next subsections are devoted to analysis of a planar SNS junction (see Fig. 1). If the junction is long (the exact criterion will be formulated later), the results obtained in Subsections 3.1 and 3.2 are valid. It should be recalled that the corresponding numerical parameters have the following values in the given case: $c_1\approx 1.15$, $c_2\approx 0.88$, $E_g = 3.12E_{Th}$, and $\tilde{G} = \pi c_2E_g/2\delta = 0.34G_N$. Here, $E_{Th} = D/L_x^2$ and $G_N = 4\pi\nu DL_yL_z/L_x = 4\pi E_{Th}/\delta$. In the case when the resistance of the contacts is finite, we can use values (4.24) (naturally, if $\varepsilon\ll\tilde{\gamma}^{2/3}$).

The action of an instanton defined by formula (2.41) is proportional to the conductance of the normal region and increases with its sizes in the y and z direction parallel to the NS interfaces. Consequently, for large values of L_y and L_z, it is more expedient to form an instanton with finite sizes along the y and z axes. In this case, a loss appears due to gradient terms, but it is compensated by the independence of action of the size of the normal region for large values of L_y and L_z. It will be shown below that the characteristic size of such a field configuration depends on the closeness to the threshold and is on the order of $L_\perp(E)\sim L_x\varepsilon^{-1/4}$.

We will calculate here the action of an instanton for a quasi-two-dimensional and three-dimensional contacts (one-dimensional and two-dimensional σ models, respectively), when one or both transverse sizes exceed $L_\perp(E)$, and obtain a result for the subgap density of states with an exponential accuracy. The power of energy in the exponent will differ from the case of a long one-dimensional junction.

In order to calculate the action for an instanton, we repeat the arguments leading to formula (2.39), but

assume that g is a function of transverse coordinates: $\theta = \pi/2 + i\psi_0(x) + ig(\mathbf{r}_\perp)f_0(x)$:

$$S_0[\theta] = \text{const} + \frac{\pi}{2\delta}\int\frac{dydz}{L_yL_z} \times\left[-\frac{D}{2}(\nabla_\perp g)^2 + E_g\left(-2c_1\varepsilon g + \frac{c_2}{3}g^3\right)\right]. \quad (5.1)$$

This action has two obvious saddle points $g = \pm\sqrt{\tilde{\varepsilon}}$ independent of the transverse coordinates, as in the case of the zero-dimensional problem. We retain for function θ_1 the solution $\theta_1 = \pi/2 + i\psi_0 - i\sqrt{\tilde{\varepsilon}}f_0$ and seek a solution for θ_2 as a function of transverse coordinates. For the sake of convenience, we introduce a function $\tilde{g}$ which describes the dependence of $\theta_2 - \theta_1$ on transverse coordinates and, in addition, transform the variables to the dimensionless form:[8]

$$g(\mathbf{r}_\perp) = \sqrt{\tilde{\varepsilon}}[2\tilde{g}(\tilde{\mathbf{r}}_\perp) - 1], \quad \tilde{\mathbf{r}}_\perp = \frac{\mathbf{r}_\perp}{L_\perp}, \quad L_\perp = \sqrt{\frac{D}{2c_2E_g}}\tilde{\varepsilon}^{-1/4}. \quad (5.2)$$

In this case, action assumes the form

$$\mathcal{S} = S_0[\theta_1] - S_0[\theta_2] = \frac{8L_\perp^2}{L_yL_z}\tilde{G}\tilde{\varepsilon}^{3/2} \times\int_0^{L_\perp/L_y}d\tilde{y}\int_0^{L_\perp/L_z}d\tilde{z}\left[\frac{(\tilde{\nabla}_\perp\tilde{g})^2}{2} + \frac{\tilde{g}^2}{2} - \frac{\tilde{g}^3}{3}\right]. \quad (5.3)$$

By varying this action, we obtain an equation for $\tilde{g}$, which describes the shape of an instanton in transverse directions:

$$\tilde{\nabla}_\perp^2\tilde{g} = \tilde{g} - \tilde{g}^2. \quad (5.4)$$

Substituting the expression for the Laplacian into the action again and integrating by parts, we finally obtain

$$\mathcal{S} = \frac{4}{3}\frac{L_\perp^2}{L_yL_z}\tilde{G}\tilde{\varepsilon}^{3/2}\int d\tilde{y}d\tilde{z}\tilde{g}^3(\tilde{y},\tilde{z}). \quad (5.5)$$

(i) **One-dimensional junction** ($L_{y,z} \ll L_\perp$). When the transverse sizes of the normal part of the junction are small, the results obtained in Section 3 are valid. In this case, $\tilde{g} \equiv 1$, the integral in Eq. (5.5) is equal to $L_yL_z/L_\perp^2$, and we return to expression (2.41) or (4.23) (in the case of tunnel contact). In this limit, the approximation of the zero-dimensional σ model is applicable and, hence, our previous results (3.7) and (4.26) for tunnel contacts are valid.

[8] The value of $L_\perp$ determined in this way is half the value used in [29].

(ii) **Two-dimensional junction** ($L_z \ll L_\perp \ll L_y$). We must solve Eq. (5.4) under the condition that $\tilde{g}$ is a function of y only. This equation describes a soliton of the Korteweg–de Vries type; its solution is well known:

$$\tilde{g}(\tilde{y}) = \frac{3}{2\cosh^2(\tilde{y}/2)}. \quad (5.6)$$

Substituting $\tilde{g}$ into Eq. (5.5), we obtain the following expression for density of states:

$$\langle\rho\rangle \sim \frac{1}{\delta}\exp\left(-\frac{48}{5}\frac{L_\perp}{L_y}\tilde{G}\tilde{\varepsilon}^{3/2}\right) = \frac{1}{\delta}\exp(-4.67G_\square\varepsilon^{5/4}). \quad (5.7)$$

Here, the large dimensionless parameter in the exponent, which determines the applicability of the steepest descent method, is the dimensionless conductance per unit area of the film of the normal metal connecting the superconducting banks: $G_\square = 4\pi\nu DL_z$.

In the case of tunnel contacts, using Eq. (4.24), we obtain

$$\langle\rho\rangle \sim \frac{1}{\delta}\exp(-2.17G_\square\tilde{\gamma}^{-1/6}\varepsilon^{5/4}). \quad (5.8)$$

Here, $\tilde{\gamma} = L_x\Gamma/4l \ll 1$.

(iii) **Three-dimensional junction** ($L_\perp \ll L_{y,z}$). The two-dimensional equation (5.4) cannot be solved analytically. We can only give the numerical result for integral (5.5):

$$\int d\tilde{y}d\tilde{z}\tilde{g}^3(\tilde{y},\tilde{z}) \approx 46.5. \quad (5.9)$$

The density of states can be written in the form

$$\langle\rho\rangle \sim \frac{1}{\delta}\exp(-10.1G_\square\varepsilon). \quad (5.10)$$

In this formula, $G_\square$ denotes the dimensionless conductance per unit area of the film, oriented parallel to the superconducting banks: $G_\square = 4\pi\nu DL_x$.

For tunnel contacts, we obtain

$$\langle\rho\rangle \sim \frac{1}{\delta}\exp(-2.58G_\square\tilde{\gamma}^{-2/3}\varepsilon), \quad (5.11)$$

where $\tilde{\gamma} = L_x\Gamma/4l \ll 1$ as before.

The criterion determining whether or not the dependence of an instanton solution on a transverse coordinate should be taken into account is based on the comparison of $L_\perp$ and $L_{y,z}$. Since the quantity $L_\perp$ is itself a function of energy, dimensions are "frozen out" as we approach the threshold: a crossover occurs from the two-dimensional σ model to the one-dimensional and then to zero-dimensional model.

5.2. Superconductor with Magnetic Impurities

In this section, we consider quasi-localized states in a superconductor with magnetic impurities. In the given case, we are speaking of a macroscopically homogeneous system in contrast to all the cases considered above (and below). The method of nonlinear σ model for determining anomalously localized states in such a system was developed in [23, 24].

A superconductor with magnetic impurities was studied in detail in the mean field approximation in the famous work by Abrikosov and Gor'kov [28]. Among other things, they proved that superconductivity is suppressed by magnetic impurities, the order parameter Δ being larger than the gap E_g in the excitation spectrum. Thus, the gapless superconductivity regime is possible, when $E_g = 0$, while $\Delta > 0$. Following [23, 24], we will consider, however, the case of a finite gap and will determine the instanton correction to the density of states under the gap with an exponential accuracy.

The introduction of magnetic impurities requires a refining of the σ model. The Bogoliubov–De Gennes Hamiltonian now contains an additional term describing scattering at magnetic impurities:[9]

$$\mathcal{H} = \tau_z\left(\frac{\mathbf{p}^2}{2m} - \mu + U(\mathbf{r})\right) + \tau_x\Delta(\mathbf{r}) + J\mathbf{S}(\mathbf{r})\cdot\hat{\boldsymbol{\sigma}}. \quad (5.12)$$

Since magnetic impurities suppress superconductivity, the order parameter Δ appearing in Hamiltonian (5.12) should be determined self-consistently. We assume that the magnetic impurity concentration is quite low and does not destroy superconductivity completely. In all the formulas of this subsection, we assume that the effect of impurities has already taken into account in Δ.

We assume that the random field $\mathbf{S}(\mathbf{r})$ in Eq. (5.12) is delta-correlated and introduce the time τ_s of scattering by magnetic impurities in the standard manner:

$$\langle JS_\alpha(\mathbf{r})JS_\beta(\mathbf{r}')\rangle = \frac{\delta(\mathbf{r}-\mathbf{r}')\delta_{\alpha\beta}}{6\pi\nu\tau_s}. \quad (5.13)$$

Thus, our system has acquired another dimensionless parameter,

$$\zeta = \frac{1}{\tau_s\Delta}. \quad (5.14)$$

The derivation of the σ model is generalized in a trivial manner to the case of Hamiltonian (5.12). In this case, action assumes the form

$$\mathcal{S}[Q] = \frac{\pi\nu}{8}\int d\mathbf{r} \times \mathrm{str}\left[D(\nabla Q)^2 + 4iQ(\Lambda E + i\tau_x\Delta) - \frac{(Q\tau_z\boldsymbol{\sigma})^2}{3\tau_s}\right]. \quad (5.15)$$

Here, the diffusion coefficient is determined by scattering by nonmagnetic impurities only. The Usadel equation can be derived without introducing complete parametrization of matrix Q, but using a simple expression in terms of angle θ alone: $Q = \sigma_z\tau_z\cos\theta + \tau_x\sin\theta$. Separating, as usual, the imaginary part $\theta = \pi/2 + i\psi$, we obtain

$$D\nabla^2\psi + 2E\cosh\psi - 2\Delta\sinh\psi + \frac{1}{\tau_s}\sinh 2\psi = 0. \quad (5.16)$$

Let us first consider homogeneous solutions to this equation. As before, we have two such solutions for energies below a certain threshold E_g. Omitting the gradient term, we can write the equation in a form resembling Eq. (2.6):

$$\frac{E}{\Delta} = \tanh\psi - \zeta\sinh\psi. \quad (5.17)$$

This equation has two solutions. The energy for which these solutions coincide determines the gap width. The maximum of the function on the right-hand side is attained for $\cosh\psi = \zeta^{-1/3}$, which corresponds to

$$E_g = \Delta(1-\zeta^{2/3})^{3/2}. \quad (5.18)$$

This remarkable result was obtained for the first time in [28].

In the vicinity of the threshold, we can expand the action in the dimensionless energy $\varepsilon = (E_g - E)/\Delta$ and in the deviation of the angle $\psi = \psi_0 + g(\mathbf{r})$. As a result, we obtain the following action for $g(\mathbf{r})$:

$$S_0[\theta] = \frac{\pi\nu\Delta}{4}\int d\mathbf{r} \times [\xi^2(\nabla g)^2 + 4\varepsilon\zeta^{-1/3}g - 2\zeta^{1/3}\sqrt{1-\zeta^{2/3}}g^3]. \quad (5.19)$$

Here, we have introduced the coherence length $\xi = \sqrt{D/\Delta}$. The obtained action is completely analogous to expression (5.1). Thus, subsequent calculations just repeat the preceding subsection and lead to the following result [23, 24]:

$$\langle\rho\rangle \sim \delta^{-1}\exp\left[-\frac{16}{3}a_d\pi\nu^{(d)}\Delta\xi^d\zeta^{-2/3} \times (24\sqrt{1-\zeta^{2/3}})^{-(2+d)/4}\varepsilon^{(6-d)/4}\right]. \quad (5.20)$$

Here, d is the effective spatial dimension of the superconductor, $\nu^{(d)}$ is the d-dimensional density of states,

[9] The symbol $\hat{\boldsymbol{\sigma}}$ for the electron spin operator is supplied with a hat to distinguish it from the Pauli spin matrices σ_i acting in the PH space. The spin operator can be expressed in terms of Pauli matrices as $\hat{\boldsymbol{\sigma}} = \tau_z\boldsymbol{\sigma}$.

and the quantity $a_d = \int d^d\mathbf{r}\tilde{g}^3$ was evaluated in Subsection 5.1 (function $\tilde{g}$ satisfies the equation $\nabla^2\tilde{g} = \tilde{g} - \tilde{g}^2$). This quantity assumes the following values:

$$a_0 = 1, \quad a_1 = 36/5, \quad a_2 \approx 46.5, \quad a_3 \approx 262. \quad (5.21)$$

5.3. Low-Energy Limit

In this subsection, we consider the range of energies close to the Fermi energy. For such energies, the density of states is suppressed the most strongly. As before, we will seek a solution with an exponential accuracy; for this purpose, we must solve the Usadel equation (2.2) and calculate action (2.30) based on the solution.

Our calculations almost repeat the calculations made by Muzykantskii and Khmelnitskii [39] (see also [40, 41]), who applied the instanton method in the diffusive σ model for determining the long-time asymptotic form of conductance $G(t)$ of a mesoscopic sample. The saddle equation appearing in this case and describing the instanton solution has the form $D\nabla^2\theta + i\omega\sinh\theta = 0$ with the boundary condition $\theta = 0$, where the quantity ω defined by the self-consistency equation [39] is inversely proportional to t. In the limit $\omega \ll E_{Th}$, the nontrivial solution of this equation almost coincides with solution ψ_2 to Eq. (2.5) for an SNS junction for $E \ll E_{Th}$. As a result, the actions of instantons in our problem and in the problem solved in [39] turn out to be identical.

We begin with a long planar SNS junction (see Fig. 1). The difference $\psi_2 - \psi_1$ is not small for energies $E \longrightarrow 0$; consequently, we must solve Eq. (4.5) directly. It should be noted that, as the energy decreases, $\psi_1(0) \longrightarrow 0$, while $\psi_2(0) \longrightarrow \infty$ (see Fig. 2). This enables us to set $\psi_1 = 0$ and to use the "triangular" approximation for ψ_2 [39],

$$\psi_2(x) = A + B\left(1 - \frac{2|x|}{L_x}\right), \quad (5.22)$$

which is based on the assumption that, for low energies, we can disregard the second term in Eq. (4.5) almost in the entire range of x except in the vicinity of zero. Substituting expression (5.22) into action (4.2), integrating, and expanding the boundary logarithm, we obtain

$$S_0 = \frac{\pi\nu D L_y L_z}{L_x} \times\left[-B^2 + \frac{E}{BE_{Th}}(\cosh(A+B) - \cosh A) - 6\tilde{\gamma}\cosh A\right]. \quad (5.23)$$

Let us first consider ideal boundaries, for which we can set $A = 0$. Carrying out the variation in B (considering that $B \gg 1$), we obtain

$$2B = \frac{Ee^B}{2BE_{Th}}. \quad (5.24)$$

To a logarithmic accuracy, we obtain the following equation for B:

$$B = \ln\frac{E_{Th}}{E} = \ln\frac{E_g}{E}. \quad (5.25)$$

Substituting this result back into Eq. (5.23), we find the density of states,

$$\langle\rho\rangle \sim \frac{1}{\delta}\exp\left(-\frac{G_N}{4}\ln^2\frac{E_g}{E}\right), \quad (5.26)$$

where $G_N = 4\pi\nu D L_y L_z/L_x$.

In the case of tunnel boundaries between a normal metal and superconductors, we must vary Eq. (5.23) in both variables in the limit $A, B \gg 1$. This gives

$$e^B = 6\tilde{\gamma}\frac{E_{Th}}{E}B = \frac{E_g}{E}B, \quad e^{A+B} = \frac{4E_{Th}}{E}B^2. \quad (5.27)$$

Here, we have used the equality $E_g = 6\tilde{\gamma}E_{Th}$. To a logarithmic accuracy, we obtain the following solution:

$$B = \ln\frac{E_g}{E}, \quad e^A = \frac{2}{3\tilde{\gamma}}\ln\frac{E_g}{E}. \quad (5.28)$$

In view of the smallness of E/E_g and $\tilde{\gamma}$, the first term in action (5.23) predominates. In this case, the density of states has form (5.26), as for a contact with ideal boundaries. This is not surprising, since strongly localized states, which are linked to the superconducting banks very weakly, lie deep in the gap. The density of such states exhibits a weak dependence on the properties of the boundary.

For low energy values, the result obtained becomes inapplicable. As a matter of fact, the gradient of ψ_2 tends to infinity, and the diffusion equation forming the basis of the σ model disregards nonlocal (ballistic) effects which become important in this case. The applicability of the local theory requires the fulfillment of the condition $|\nabla\psi_2| \sim B/L_x \ll 1/l$, setting a limit on the energy:

$$E \gg E_g\exp\left(-\frac{L_x}{l}\right). \quad (5.29)$$

Let us now consider the two-dimensional case. For the sake of simplicity, we assume that the normal region has the shape of a disc of radius R and thickness L_z, connected to a superconductor along the entire boundary. As in the previous example, $\psi_1 = 0$. We write Eq. (4.5) for ψ_2 in polar coordinates, assuming that the

radius is equal to unity and replacing the hyperbolic cosine by the exponential:[10]

$$\psi_2'' + \frac{\psi_2'}{r} + \frac{E}{E_{Th}} e^{\psi_2} = 0, \quad \psi_2'(0) = 0. \tag{5.30}$$

This equation has the following solution:

$$\psi_2(r) = 2\ln\frac{B\sqrt{8E_{Th}/E}}{B^2 + r^2}, \tag{5.31}$$

where B must be chosen from the boundary condition in (4.5).

If the boundary between the superconductor and the normal metal is transparent, we must require that $\psi_2(1) = 0$. For $E \ll E_{Th}$, we obtain $B = \sqrt{E/8E_{Th}}$. As before, disregarding the term with energy in Eq. (2.31), we obtain the following expression for action:

$$\mathcal{S} = -S_0[\theta_2] = \frac{\pi\nu D}{4} R^2 L_z \times \int\limits_{\sqrt{E/E_g}}^{1} (\nabla\psi_2)^2 2\pi r dr = 4\pi^2 \nu D L_z \ln\frac{E_g}{E}. \tag{5.32}$$

The density of states is a power function of energy [39]:

$$\langle\rho\rangle \sim \left(\frac{E}{E_g}\right)^{4\pi^2\nu D L_z} = \left(\frac{E}{E_g}\right)^{\pi G_\square}. \tag{5.33}$$

In the case of a tunnel boundary, replacing the hyperbolic sine in the boundary condition (4.5) by an exponential, we obtain

$$B = \frac{1}{2}\sqrt{\frac{E}{\tilde{\gamma}E_{Th}}} = \sqrt{\frac{E}{E_g}}. \tag{5.34}$$

In this equality, we have used the expression $E_g = 4\tilde{\gamma}E_{Th}$, which is valid in the two-dimensional case. Thus, result (5.33) remains valid to a logarithmic accuracy in the case of the tunnel boundary also; i.e., the resistance of the contacts does not affect the density of states lying deep in the gap as in the one-dimensional case.

For a two-dimensional SNS junction of a different shape, the results will have the same order of magnitude, since only E_g depends on the longitudinal size, and this will only change the general numerical factor in the expression for the density of states.

Let us determine the region of applicability of the result. The maximum gradient of ψ_2 is attained in the vicinity of the center, $|\nabla\psi_2| \propto \sqrt{E_g/E}/R$, and must be smaller than $1/l$. Thus, the power dependence of density of states is valid for

$$E \gg E_g\left(\frac{l}{R}\right)^2. \tag{5.35}$$

This condition is much more stringent than condition (5.29). In the three-dimensional case, the line of reasoning similar to that used in this subsection is completely impossible since the corresponding condition turns out to be too stringent. In this case, as well as in the two-dimensional case, the subgap density of states for energies lower than (5.35) can be found with the help of the ballistic σ model [42] (see also [43]).

It should be noted that, after the substitution $E \longrightarrow t_\phi^{-1}$, expressions (5.26) and (5.33) coincide (to within a factor) with the distribution function $\mathcal{P}(t_\phi)$ of the relaxation times for a corresponding mesoscopic system [44, 45]. The long-time asymptotic form of conductance can be expressed in terms of this distribution function as $G(t) \sim \int dt_\phi e^{-t/t_\phi}\mathcal{P}(t_\phi)$. Evaluating the integral by the steepest descent method, we obtain

$$G(t) \sim \exp\left[-\frac{G_N}{4}\ln^2(\delta t)\right],$$

for a one-dimensional system, which corresponds to the substitution $i\omega \longrightarrow G_N/t$ in the Muzykantskii–Khmelnitskii formalism. This explains the difference in the energy scales in our problem and in the problem on the asymptotic form of conductance. Results (5.26) and (5.33) are transformed to the asymptotic expressions [39] for $G(t)$ via the substitution $E_{Th}/E \longrightarrow \delta t$.

6. CONCLUSIONS

We have considered the density of quasiparticle states in diffusive NS systems for energies close to the semiclassical gap $E_g \ll \Delta$, appearing due to mesoscopic fluctuations.

For a system with ideal contacts, exact expression (3.18) for the averaged density of states is obtained in the framework of the zero-dimensional σ model with a large parameter $\tilde{G} \propto E_g/\delta \gg 1$ for energies E in the vicinity of $E_g \sim E_{Th} \sim D/L^2$. For $\varepsilon = (E_g - E)/E_g \geq \tilde{G}^{-2/3}$, the density of states decreases exponentially (see Eq. (3.7)), which resembles to a certain extent the behavior of density of states in semiconductors with impurities (the so-called Lifshits tail [13]). The total number of states in the tail region is found to be on the order of unity for a one-dimensional system; i.e., the equation obtained for density of states has the meaning of the probability of an anomalous location of the lower energy level below E_g. The results are in perfect agreement with the phenomenological random matrix theory [22] and make it possible to determine

[10] Here, $E_{Th} = D/R^2$.

the range of its applicability as well as the values of microscopic parameters used in it.

For planar SNS junctions with a large transverse size $L_{y,z} \geq L_x\varepsilon^{-1/4}$ (see Fig. 1), analogous results were obtained with an exponential accuracy (see relations (5.7) and (5.10)). In view of the large area of the contacts, the averaged (integrated) density of states has direct physical meaning in this case. The problem of the fluctuation tail of the density of states in a superconductor with a low concentration of magnetic impurities, which was considered for the first time in [23, 24], is also found to be formally similar.

Physical factors responsible for the emergence of low-lying electron states in our problem are mesoscopic fluctuations in the normal region of the contact, which lead to the diffusion electron–hole (Andreev) trajectories with a period larger than the characteristic diffusion time L^2/D. A similar effect is observed during analysis of asymptotic forms of the distribution functions for density of states, conductance, and relaxation times for normal systems [44]. Due to mesoscopic fluctuations of impurities, almost localized states can also be found with an exponentially low probability in the energy range corresponding to well delocalized states. Such states ensure an anomalously slow relaxation of current to its equilibrium value over very long times (larger than the inverse mean spacing νV); i.e., such states play the role of "electron traps." Naturally, such traps must appear in the problem of an NS system also, since it is these traps that lead to the emergence of trajectories diffusing between Andreev reflection for an anomalously long time. A direct relation between "anomalously localized states" and low-lying states in an SNS junction follows from our results for the density of states in the low-energy range $E \ll E_g$ (see relations (5.26) and (5.33) for one- and two-dimensional case, respectively). Upon the substitution $E \longrightarrow t_\phi^{-1}$, these formulas coincide (to within a normalization) with the distribution function for relaxation times t_ϕ of a mesoscopic system [39, 44, 45].

We have also considered hybrid NIS structures with a poor transparency of NS interfaces. Such systems can be described by two characteristic parameters: the total tunnel conductance G_T between the normal and superconducting parts and the effective average transparency $\tilde{\gamma}$ of a channel, which depends on the properties of the contacts and on the system geometry (see Eq. (4.18)). In the limit of a low effective transparency of a channel, $\tilde{\gamma} \ll 1$, the semiclassical edge of the spectrum is displaced towards low energies, $E_g = G_T\delta/8\pi$. The behavior of the density of states above the threshold simultaneously changes qualitatively: as the energy decreases, the density of states first increases as in the BCS theory, $\langle\rho(E)\rangle \propto |\varepsilon|^{-1/2}$, and starts decreasing in accordance with the law $\langle\rho(E)\rangle \propto \sqrt{|\varepsilon|}$ only for $|\varepsilon| < \tilde{\gamma}^{2/3}$.

In the limit $\tilde{\gamma} \ll G_T^{-3}$, the root region $\langle\rho(E)\rangle \propto \sqrt{|\varepsilon|}$ near the threshold disappears altogether: the semiclassical number of states becomes smaller than unity in this region. As a result, the subgap density of states, which is determined by the zero-dimensional σ model as before, cannot be described by formula (3.18) any longer, but is given by expression (4.44). This is in agreement with the relation between the root dependence ($\langle\rho(E)\rangle \propto \sqrt{|\varepsilon|}$) of the density of states above the threshold and the applicability of the random matrix theory in zero-dimensional systems, which was noted in [24]. In the limit $\tilde{\gamma} \ll G_T^{-3}$, the tail becomes strong: it contains many energy levels instead of one, as in the case of transparent boundaries. In the intermediate case, $G_T^{-3} \ll \tilde{\gamma} \ll 1$, the asymptotic form (3.18) is realized in the immediate vicinity of the threshold, while the asymptotic form (4.44) is observed for $\tilde{\gamma}^{2/3} \ll \varepsilon \ll 1$; the number of subgap states in this case is on the order of unity.

ACKNOWLEDGMENTS

The authors thank C. Beenakker, D. Ivanov, V. Lebedev, A. Mirlin, Yu. Nazarov, B. Simons, and A. Shitov for numerous fruitful discussions.

This study was supported financially by the Swiss and Dutch science foundations, Russian Foundation for Basic Research (project no. 01-02-17759), the program "Quantum Macrophysics" of the Russian Academy of Sciences, and the Foundation for Assistance to Russian Science (M. A. S.).

APPENDIX A

Parametrization of W Matrix

Matrix W contains eight commuting and eight Grassmann parameters and satisfies Eqs. (2.43). We denote real variables by a, b, c, d, m, n, p, and q. The first four variables parametrize the FF sector, and the last four, the BB sector of W:

$$W^{FF} = \frac{1}{2}(-a\sigma_z\tau_x + b\sigma_z\tau_y - c\sigma_x + d\sigma_y), \qquad \text{(A.1)}$$

$$W^{BB} = \frac{i}{2}\left(m\sigma_z\tau_x + n\sigma_x\tau_z + p\frac{\sigma_y\tau_z - \sigma_z\tau_y}{\sqrt{2}} - q\frac{\sigma_y\tau_z + \sigma_z\tau_y}{\sqrt{2}}\right). \qquad \text{(A.2)}$$

Grassmann variables (λ, μ, ζ, κ, η, γ, ξ, and ω) parametrize the FB sector:

$$W^{FB} = \frac{1}{4}\begin{pmatrix} 0 & \lambda-\mu+\zeta-\kappa & \lambda+\mu+\zeta+\kappa & 0 \\ \lambda-\mu-\zeta+\kappa & 0 & 0 & \lambda+\mu-\zeta-\kappa \\ -\gamma-\eta+\xi+\omega & 0 & 0 & \gamma-\eta+\xi-\omega \\ 0 & -\gamma-\eta-\xi-\omega & \gamma-\eta-\xi+\omega & 0 \end{pmatrix}. \quad \text{(A.3)}$$

The anti-self-conjugate condition $W + \overline{W} = 0$ makes it possible to express the BF sector of W in terms of the FB sector: $W^{BF} = i\tau_x\sigma_x(W^{FB})^T\sigma_y\tau_x$.

We can establish a relation between commuting parameters of W and fluctuations of angles in matrix Q. This relation has the form

$$\begin{aligned} a &= 2\sin\theta_F\delta\varphi_F, \quad b = 2\delta\theta_F, \\ c &= 2\cos\theta_F\delta k_F|_{\chi_F=\pi/2}, \\ d &= 2\cos\theta_F\delta k_F|_{\chi_F=0}, \\ m &= 2i\sin\theta_B\delta\varphi_B, \quad n = 2i\sin k_B\chi_B, \\ p &= i\sqrt{2}\delta\beta, \quad q = i\sqrt{2}\delta\alpha. \end{aligned} \quad \text{(A.4)}$$

It should be noted that variable n at the first instanton corresponds to the zero mode (rotation of angle χ_B).

In order to pass from integration over Q to integration over W, we must calculate the Jacobian of the corresponding transformation. However, this Jacobian is equal to unity, which can easily be established from the following equality:

$$\begin{aligned} \mathrm{str}(dQ)^2 = \mathrm{str}(dW)^2 = da^2 + db^2 + dc^2 + dd^2 \\ + dm^2 + dn^2 + dp^2 + dq^2 + d\lambda d\eta + d\mu d\gamma \\ + d\zeta d\xi + d\kappa d\omega. \end{aligned} \quad \text{(A.5)}$$

APPENDIX B

Airy Type Integrals

In this section, we evaluate integral (3.16). Integration with respect to w can be carried out immediately (contour C_1 in Fig. 3). As a result, the density of states will be reduced to the sum of two contributions:

$$\langle\rho\rangle = -\frac{1}{4\pi\Delta_g}\mathrm{Im}[\mathrm{Ai}(\epsilon)\rho_1(\epsilon) + \mathrm{Ai}'(\epsilon)\rho_2(\epsilon)], \quad \text{(B.1)}$$

$$\rho_1(\epsilon) = \int_0^\infty dm \int_{C_3} dl(l^3 + 3\epsilon l - lm + 1) \times \exp\left(\frac{l^3}{3} + ml - \epsilon l\right), \quad \text{(B.2)}$$

$$\rho_2(\epsilon) = \int_0^\infty dm \int_{C_3} dl(3l^2 - m + \epsilon) \times \exp\left(\frac{l^3}{3} + ml - \epsilon l\right). \quad \text{(B.3)}$$

In order to calculate $\rho_{1,2}(\epsilon)$, we use the following approach. We first eliminate m from the preexponential factor, expressing it in terms of $\partial/\partial l$, then get rid of derivatives by integrating by parts, express l in terms of $\partial/\partial m$, and, finally, integrate with respect to m wherever possible. As a result, the integral with respect to l becomes trivial. Thus, we begin with $\rho_1(\epsilon)$:

$$\begin{aligned} \rho_1(\epsilon) &= \int_0^\infty dm \int_{C_3} dl\left(-l\frac{\partial}{\partial l} + 2l^3 + 2\epsilon l + 1\right) \times \exp\left(\frac{l^3}{3} + ml - \epsilon l\right) \\ &= \int_0^\infty dm \int_{C_3} dl(2l^3 + 2\epsilon l + 2)\exp\left(\frac{l^3}{3} + ml - \epsilon l\right) \\ &= \int_0^\infty dm \int_{C_3} dl\left(2\frac{\partial^3}{\partial m^3} + 2\epsilon\frac{\partial}{\partial m} + 2\right)\exp\left(\frac{l^3}{3} + ml - \epsilon l\right) \\ &= \left[\left(-2\epsilon - 2\frac{\partial^2}{\partial m^2}\right)\Big|_{m=0} + 2\int_0^\infty dm\right] \times \int_{C_3} dl\exp\left(\frac{l^3}{3} + ml - \epsilon l\right). \end{aligned} \quad \text{(B.4)}$$

Similarly, for $\rho_2(\epsilon)$, we have

$$\begin{aligned} \rho_2(\epsilon) &= \int_0^\infty dm \int_{C_3} dl\left(4l^2 - \frac{\partial}{\partial l}\right)\exp\left(\frac{l^3}{3} + ml - \epsilon l\right) \\ &= \int_0^\infty dm \int_{C_3} dl\left(-4\frac{\partial^2}{\partial m^2}\right)\exp\left(\frac{l^3}{3} + ml - \epsilon l\right) \\ &= -4\frac{\partial^2}{\partial m^2}\Big|_{m=0}\int_{C_3} dl\exp\left(\frac{l^3}{3} + ml - \epsilon l\right). \end{aligned} \quad \text{(B.5)}$$

The integral with respect to l appearing in the expressions for $\rho_{1,2}$ is given by

$$\int_{C_3} dl \exp\left(\frac{l^3}{3} + ml - \epsilon l\right) = \pi[\mathrm{Bi}(\epsilon - m) - i\mathrm{Ai}(\epsilon - m)]. \quad \text{(B.6)}$$

Substituting $\rho_{1,2}$ into expression (B.1) and using the functional relations for the Airy functions, we arrive at expression (3.18).

REFERENCES

1. *Mesoscopic Phenomena in Solids*, Ed. by B. L. Altshuler, P. A. Lee, and R. A. Webb (Elsevier, Amsterdam, 1991).
2. A. F. Andreev, Zh. Éksp. Teor. Fiz. **49**, 65 (1965) [Sov. Phys. JETP **22**, 47 (1966)].
3. I. O. Kulik and A. N. Omel'yanchuk, Fiz. Nizk. Temp. **3**, 945 (1977) [Sov. J. Low Temp. Phys. **3**, 459 (1977)].
4. J. A. Melsen *et al.*, Europhys. Lett. **35**, 7 (1996); Phys. Scr. **69**, 223 (1997).
5. A. Lodder and Yu. V. Nazarov, Phys. Rev. B **58**, 5783 (1998).
6. D. Taras-Semchuk and A. Altland, Phys. Rev. B **64**, 014512 (2001).
7. G. Eilenberger, Z. Phys. **214**, 195 (1968).
8. A. I. Larkin and Yu. N. Ovchinnikov, Zh. Éksp. Teor. Fiz. **55**, 2262 (1968) [Sov. Phys. JETP **26**, 1200 (1968)].
9. K. Usadel, Phys. Rev. Lett. **25**, 507 (1970).
10. A. A. Golubov and M. Yu. Kupriyanov, Zh. Éksp. Teor. Fiz. **96**, 1420 (1989) [Sov. Phys. JETP **69**, 805 (1989)].
11. F. Zhou *et al.*, J. Low Temp. Phys. **110**, 841 (1998).
12. S. Pilgram, W. Belzig, and C. Bruder, Phys. Rev. B **62**, 12462 (2000).
13. I. M. Lifshits, Usp. Fiz. Nauk **83**, 617 (1964) [Sov. Phys. Usp. **7**, 571 (1965)].
14. L. B. Ioffe and M. V. Feĭgel'man, Zh. Éksp. Teor. Fiz. **89**, 654 (1985) [Sov. Phys. JETP **62**, 376 (1985)].
15. E. Brezin, D. J. Gross, and C. Itzykson, Nucl. Phys. B **235**, 24 (1984).
16. K. B. Efetov and V. G. Marikhin, Phys. Rev. B **40**, 12126 (1989).
17. E. P. Wigner, Ann. Math. **53**, 36 (1951).
18. F. J. Dyson, J. Math. Phys. **3**, 140 (1962); **3**, 157 (1962); **3**, 166 (1962).
19. M. L. Mehta, *Random Matrices* (Academic, New York, 1991).
20. C. W. J. Beenakker, Rev. Mod. Phys. **69**, 731 (1997).
21. K. B. Efetov, Zh. Éksp. Teor. Fiz. **83**, 883 (1982) [Sov. Phys. JETP **56**, 467 (1982)].
22. M. G. Vavilov, P. W. Brower, V. Ambegaokar, and C. W. J. Beenakker, Phys. Rev. Lett. **86**, 874 (2001).
23. A. Lamacraft and B. D. Simons, Phys. Rev. Lett. **85**, 4783 (2000).
24. A. Lamacraft and B. D. Simons, Phys. Rev. B **64**, 014514 (2001).
25. S. F. Edwards and P. W. Anderson, J. Phys. F **5**, 965 (1975); A. Nitzan, K. H. Freed, and M. H. Cohen, Phys. Rev. B **15**, 4476 (1977).
26. K. B. Efetov, *Supersymmetry in Disorder and Chaos* (Cambridge Univ. Press, New York, 1997).
27. A. Altland, B. D. Simons, and D. Taras-Semchuk, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 21 (1997) [JETP Lett. **67**, 22 (1997)]; Adv. Phys. **49**, 321 (2000).
28. A. A. Abrikosov and L. P. Gor'kov, Zh. Éksp. Teor. Fiz. **39**, 1781 (1960) [Sov. Phys. JETP **12**, 1243 (1961)].
29. P. M. Ostrovsky, M. A. Skvortsov, and M. V. Feigel'man, Phys. Rev. Lett. **87**, 027002 (2001).
30. P. M. Ostrovsky, M. A. Skvortsov, and M. V. Feigel'man, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 407 (2002) [JETP Lett. **75**, 336 (2002)].
31. M. A. Skvortsov, V. E. Kravtsov, and M. V. Feĭgel'man, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 84 (1998) [JETP Lett. **68**, 84 (1998)].
32. C. A. Tracy and H. Widom, Commun. Math. Phys. **159**, 151 (1994); **177**, 727 (1996).
33. L. P. Gor'kov and G. M. Éliashberg, Zh. Éksp. Teor. Fiz. **48**, 1407 (1965) [Sov. Phys. JETP **21**, 940 (1965)].
34. T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller, Phys. Rep. **299**, 189 (1998).
35. A. Altland and M. R. Zirnbauer, Phys. Rev. B **55**, 1142 (1997).
36. A. V. Zaĭtsev, Zh. Éksp. Teor. Fiz. **86**, 1742 (1984) [Sov. Phys. JETP **59,** 1015 (1984)].
37. M. Yu. Kupriyanov and V. F. Lukichev, Fiz. Nizk. Temp. **8**, 1045 (1982) [Sov. J. Low Temp. Phys. **8**, 526 (1982)].
38. W. Belzig and Yu. V. Nazarov, Phys. Rev. Lett. **87**, 197006 (2001).
39. B. A. Muzykantskii and D. E. Khmelnitskii, Phys. Rev. B **51**, 5480 (1995).
40. V. I. Fal'ko and K. B. Efetov, Europhys. Lett. **32**, 627 (1995).
41. A. D. Mirlin, Phys. Rev. B **53**, 1186 (1996).
42. B. A. Muzykantskii and D. E. Khmelnitskii, E-print archives, cond-mat/9601045.
43. A. D. Mirlin, Pis'ma Zh. Éksp. Teor. Fiz. **62**, 583 (1995) [JETP Lett. **62**, 603 (1995)].
44. B. L. Altshuler, V. E. Kravtsov, and I. V. Lerner, in *Mesoscopic Phenomena in Solids*, Ed. by B. L. Altshuler, P. A. Lee, and R. A. Webb (Elsevier, Amsterdam, 1991), p. 449.
45. A. D. Mirlin, Phys. Rep. **326**, 259 (2000).

Translated by N. Wadhwa