Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 1008–1012.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1174–1178.
Original Russian Text Copyright © 2002 by Arkhincheev.

SOLIDS
Electronic Properties

The Effective Conductivity of Two-Phase Two-Dimensional Media as a Function of the Geometric Parameters of the Medium

V. E. Arkhincheev

Buryat Scientific Center, Siberian Division, Russian Academy of Sciences, Ulan-Ude, Republic of Buryatia, 670047 Russian Federation
e-mail: varkhin@bsc.buryatia.ru

Received November 29, 2001

Abstract—The paper deals with the problem of current flow in doubly periodic two-dimensional media whose unit cell is a parallelogram. Local distributions of currents (fields) are found, and the effective conductivity of such media is calculated for equal phase concentrations. The dependence of conductivity on the angle of parallelogram is determined. This dependence is shown to be threshold in media of the metal–insulator type. *© 2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is known that Dykhne [1] established the invariance of two-dimensional equations of direct current with respect to linear transformations of rotation, and, as a consequence, an expression for effective conductivity was derived,

$$\sigma_{\text{eff}} = \sqrt{\sigma_1\sigma_2}, \qquad (1)$$

where σ_i (i = 1, 2) denotes the phase conductivities. This expression is valid for both random and doubly periodic arrangement ("checkerboard") of phases. For the checkerboard case, this result was obtained later using the method of conformal transformations; the local distribution of currents (fields) was also determined [2, 3]. Recently, this result was obtained using a different method [4].

It is the objective of this study to find the distributions of local currents and to calculate the effective conductivity of two-phase doubly periodic media. Treated as the initial model was a periodic structure whose unit cell consists of two contiguous equal parallelograms (Fig. 1). Because of the symmetry of the problem, the field pattern is periodically repeated; therefore, it is sufficient to find the distribution in a unit cell consisting of two parallelograms with different conductivities. We will explain the solution algorithm of the problem. To follow Emets [2, 3], we use the Schwarz–Christoffel integral to construct the mapping of the contiguous parallelograms onto the lower and upper half-planes. Then, the resultant boundary-value problem in the matrix form is transformed to a set of equations for Riemann scalar problems. The solutions to the scalar problems are found by the standard method. We perform inverse transformations to find the local distributions of fields and currents. On performing the necessary averagings, we will finally derive the general expression for the effective conductivity, which depends both on the phase conductivities and on the geometry of the unit cell, i.e., the angle of parallelogram. In the case of square cells, the formula transforms to the known result of Dykhne given by Eq. (1). In the limiting cases of stratified media formed by one-dimensional lines, known expressions for stratified media are obtained. The step dependence of the effective conductivity on the angle of parallelogram for a metal–insulator medium appears to be of the most interest.

The structure of this paper is as follows. In Section 2, a complex representation is introduced for a two-dimensional conductivity problem. In Section 3, a conformal mapping is constructed for contiguous parallelograms of a unit cell onto the lower and upper half-planes. In Section 4, the Riemann boundary-value problem is formulated, and solutions are obtained for scalar problems with preassigned boundary conditions. In Section 5, the general expression for the effective conductivity of the medium is derived, and the limiting cases are analyzed.

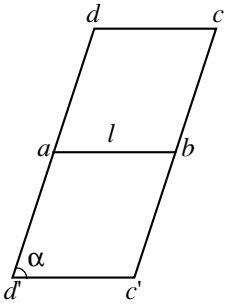

Fig. 1. A two-color doubly periodic structure with a unit cell formed by parallelograms.

2. COMPLEX REPRESENTATION OF TWO-DIMENSIONAL CONDUCTIVITY PROBLEM

We will treat a two-color structure with a unit cell of equilateral contiguous parallelograms with different conductivities. We will mark the physical quantities for the respective phases by the superscripts 1 and 2. The distribution of electric fields (currents) is described by the set of equations

$$\operatorname{div}\mathbf{j} = 0, \quad \operatorname{curl}\mathbf{e} = 0, \quad \mathbf{j} = \hat{\sigma}\mathbf{e} \tag{2}$$

with the boundary conditions of continuity of the normal components of current and tangential components of the electric field

$$j_n^{(1)} = j_n^{(2)}, \quad e_t^{(1)} = e_t^{(2)}. \tag{3}$$

For solving the problem by the methods of the theory of functions of complex variables, we go over to the plane of complex variable $z = x + iy$ and introduce the complex values of current density and electric field strength,

$$j^{(k)}(z) = j_x^{(k)}(x, y) - ij_y^{(k)}(x, y), \tag{4}$$

$$E^{(k)}(z) = E_x^{(k)}(x, y) - iE_y^{(k)}(x, y), \quad k = 1, 2. \tag{5}$$

As is known, this representation is valid for functions satisfying the Cauchy–Riemann conditions. In our case, this is possible, because the conductivity of the medium varies discretely and assumes constant values of σ_i ($i = 1, 2$) in the respective cells. Here, the equations of direct current in the two-dimensional case coincide with the Cauchy–Riemann equations. Indeed, the equation of continuity for current gives one of these conditions,

$$\frac{\partial j_x^{(k)}}{\partial x} = -\frac{\partial j_y^{(k)}}{\partial y}, \tag{6}$$

and the second condition follows from the equation $\operatorname{curl}\mathbf{e} = 0$; for cells with constant value of σ, the latter equation is equivalent to the equation $\operatorname{curl}\mathbf{j} = 0$,

$$\frac{\partial j_x^{(k)}}{\partial y} = \frac{\partial j_y^{(k)}}{\partial x}. \tag{7}$$

Analogously, the equations $\operatorname{curl}\mathbf{e} = 0$ and $\operatorname{div}\mathbf{e} = 0$ (following from the equation $\operatorname{div}\mathbf{j} = 0$) lead to the Cauchy–Riemann conditions for the electric field. Ohm's law in the complex form is written similarly,

$$j(z) = \sigma E(z). \tag{8}$$

3. CONFORMAL TRANSFORMATION ONTO THE ζ PLANE

We will construct a conformal mapping of the internal regions of contiguous parallelograms onto the lower and upper half-planes. In the case of doubly periodic media formed by parallelograms, this is done using the Weierstrass function $w = \Pi(z)$ (see, for example, [5]). In this case, parallelogram *abcd* goes to the lower half-plane, and parallelogram *abc'd'*, to the upper half-plane with a section on the axis $\eta = 0$. The correspondence of the transformation points is shown in Fig. 2. In our case,

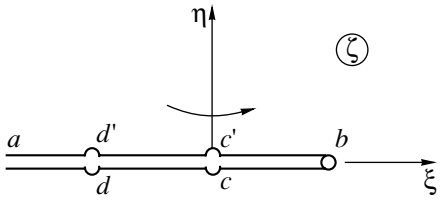

Fig. 2. The correspondence between points under conditions of conformal mapping of parallelograms on a half-plane.

$$b = \Pi(l), \quad c\Pi(l(1 - \cos\alpha) + il\sin\alpha),$$

$$d = \Pi(-l\cos\alpha + il\sin\alpha),$$

$$\omega_1 = l, \ \omega_2 = -i\cos\alpha + il\sin\alpha,$$

where ω_1 and ω_2 are the periods of the doubly periodic Weierstass function.

In the particular case of squares, a lemniscatic case is observed,

$$b = \Pi(l), \quad c = \Pi(l + il), \quad d = \Pi(il).$$

4. SOLUTION OF RIEMANN BOUNDARY-VALUE PROBLEM IN A MATRIX FORM

The boundary conditions in the ζ plane in the notation for complex current,

$$j^{(k)}(\zeta) = j_\xi^{(k)}(\xi, \eta) - ij_{(\eta)}^{(k)}(\xi, \eta),$$

hold and have the following form in different portions of the ξ axis:

$$\begin{aligned} j_1 - \bar{j}_1 &= j_2 - \bar{j}_2, \\ \frac{\sigma_2}{\sigma_1}(j_1 + \bar{j}_1) &= j_2 + \bar{j}_2 \end{aligned} \tag{9}$$

on the segments [*ab*], [*cd*], and [*c'd'*] (here, the bar over the function indicates complex conjugation); and

$$\begin{aligned} &j_1 + \bar{j}_1 - i(j_1 - \bar{j}_1)\cot\alpha \\ &= j_2 + \bar{j}_2 - i(j_2 - \bar{j}_2)\cot\alpha, \\ &\frac{\sigma_2}{\sigma_1}[(j_1 + \bar{j}_1)\cot\alpha - i(j_1 - \bar{j}_1)] \\ &= (j_2 + \bar{j}_2)\cot\alpha - i(j_2 - \bar{j}_2) \end{aligned} \tag{10}$$

on the segments [*bc*], [*ad*], [*bc'*], and [*ad'*].

Therefore, the initial boundary conditions are reduced to a three-body Riemann boundary-value problem relative to the piecewise analytic function $j(\zeta)$. The coefficients of the problem have discontinuities on the ξ axis at points b, c, and d. For solving this problem, we will formulate an equivalent problem for two pairs of functions. As follows from the set of equations defining the boundary conditions, four unknown quantities arise in the problem: j_k and $\bar{j}_k$, where $k = 1, 2$. In accordance with [2], we will express the boundary values of j_2 and $\bar{j}_1$ as functions of j_1 and $\bar{j}_2$. We will introduce the piecewise continuous analytic vector function

$$\Phi(\zeta) = \begin{pmatrix} j_2(\zeta) \\ \bar{j}_1(\zeta) \end{pmatrix}.$$

The introduced function satisfies the symmetry condition

$$\Phi(\bar{\zeta}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\Phi(\zeta).$$

On the ξ axis, it assumes the following values:

$$\Phi^+(\xi) = \begin{pmatrix} j_2(\xi) \\ \bar{j}_1(\xi) \end{pmatrix}$$

and

$$\Phi^-(\xi) = \begin{pmatrix} j_1(\xi) \\ \bar{j}_2(\xi) \end{pmatrix}.$$

As represented, the boundary conditions formulate the Riemann boundary-value problem in the vector-matrix form,

$$\Phi^+(\xi) = G_i\Phi^-(\xi), \tag{11}$$

where the matrices G_i have the following form:

$$G_1 = \frac{1}{\sigma_1 + \sigma_2}\begin{pmatrix} 2\sigma_2 & -(\sigma_1 - \sigma_2) \\ \sigma_1 - \sigma_2 & 2\sigma_1 \end{pmatrix}$$

on the [ab] interval and

$$G_2 = \frac{1}{\sigma_1 + \sigma_2} \times \begin{pmatrix} 2\sigma_2 & -(\sigma_1 - \sigma_2)\dfrac{\cot\alpha + i}{\cot\alpha - i} \\ (\sigma_1 - \sigma_2)\dfrac{\cot\alpha - i}{\cot\alpha + i} & 2\sigma_1 \end{pmatrix}$$

on the [bc] interval.

The solution of the matrix problem is constructed by reducing it to two independent Riemann scalar problems. For this purpose, it is necessary to simplify the boundary conditions and diagonalize of the matrices of boundary condition G_i. For this, we introduce a new piecewise analytic function

$$\Psi(\zeta) = \begin{pmatrix} \Psi_1(\zeta) \\ \Psi_2(\zeta) \end{pmatrix}.$$

The introduced function is related to the initial vector function $\Phi(\zeta)$ by

$$\Psi(\zeta) = \begin{cases} \Phi(\zeta), & \mathrm{Im}\,\zeta > 0 \\ M\Phi(\zeta), & \mathrm{Im}\,\zeta < 0. \end{cases}$$

Let the matrix M coincide with the matrix G_3 of boundary conditions on the axis (a_3a_1). Then, on this axis, the boundary-value problem take the form

$$\Psi^+(\xi) = \Psi^-(\xi).$$

Consequently, on this interval, the function can be analytically continued from one half-plane to the other. In the remaining part of the real axis, the boundary conditions take the form

$$\Psi^+(\xi) = G_kM^{-1}\Psi^{-1}(\xi). \tag{12}$$

Here, $G = G_1$ on the [a_1a_2] segment and $G = G_2$ on the [a_2a_3] segment. Further simplification consists in the diagonalization of the resultant matrices of boundary conditions (12). For this purpose, we introduce yet another function $F(\zeta)$ by the relation

$$\Psi(\zeta) = SF(\zeta). \tag{13}$$

Then, the boundary-value problem take the form

$$F^+(\xi) = S^{-1}G_kM^{-1}SF^-(\xi). \tag{14}$$

The matrix S is selected such that the relation

$$(\lambda E - G_iM^{-1})S = 0$$

should be valid, where E is a unit matrix and λ denotes the eigenvalues of the characteristic equation

$$\det\left|\lambda E - G_iM^{-1}\right| = 0.$$

After necessary computations, we derive the values of the roots of the equation

$$\lambda_1 = \frac{a + i(\sigma_1 - \sigma_2)}{a - i(\sigma_1 - \sigma_2)}, \quad \lambda_2 = \frac{a - i(\sigma_1 - \sigma_2)}{a + i(\sigma_1 - \sigma_2)}. \tag{15}$$

Here, $a^2 = 4\sigma_1\sigma_2 + (\sigma_1 - \sigma_2)^2\cot^2\alpha$. The sought matrix S is found from the calculations of the eigenvectors and is

$$S = \begin{pmatrix} \dfrac{2\sigma_2[a - i(\sigma_1 - \sigma_2)]}{4\sigma_1\sigma_2 + i[\cot\alpha(\sigma_1 + \sigma_2)^2 - a(\sigma_1 - \sigma_2)]} & \dfrac{4\sigma_1\sigma_2 + i[\cot\alpha(\sigma_1 + \sigma_2)^2 + a(\sigma_1 - \sigma_2)]}{2\sigma_2[a + i(\sigma_1 - \sigma_2)]} \\ i & -i \end{pmatrix}.$$

As a result, we derive the set of independent boundary-value problems

$$F_1^+(\xi) = \lambda_1 F_1^-(\xi), \quad F_2^+(\xi) = \lambda_2 F_2^-(\xi). \tag{16}$$

Next, the solution is constructed in a standard manner using the theory of a Riemann scalar problem [6] and has the form

$$F_1(\zeta) = C_1 X(\zeta), \quad F_2(\zeta) = C_2 X^{-1}(\zeta),$$
$$X(\zeta) = \frac{(\zeta - e_1)^\gamma (\zeta - e_3)^\gamma}{\zeta^\gamma}. \tag{17}$$

Here, $\gamma = (\sigma_2 - \sigma_1)/a$.

One can readily see that, in the particular case of squares, the derived expressions transform to the known formulas derived by Emets [2].

Turning back to the initial vector function and properly determining the phases of the coefficients C_1 and C_2 (see [2]), we derive the following expressions for the distribution of currents:

$$j_1 = \exp\left[\frac{i(\alpha - \gamma)}{2}\right]|C_1|X(z)$$
$$+ \exp\left[-\frac{i(\alpha - \gamma)}{2}\right]|C_2|X^{-1}(z),$$

$$j_2 = A\exp\left[\frac{i(\alpha + \gamma)}{2}\right]|C_1|X(z) \tag{18}$$
$$+ B\exp\left[-\frac{i(\alpha + \gamma)}{2}\right]|C_2|X^{-1}(z).$$

Here,

$$A = \frac{2\sigma_2\sqrt{1 + \cot^2\alpha}}{a - \cot\alpha(\sigma_1 - \sigma_2)}, \quad B = \frac{2\sigma_2\sqrt{1 + \cot^2\alpha}}{a + \cot\alpha(\sigma_1 - \sigma_2)}$$

are constants and moduli $|c_i|$ are defined by specifying the external current.

Thus, we derive the following relations between the average values of currents and fields:

$$\langle j_2\rangle = A\langle j_1\rangle, \quad \langle e_2\rangle = \frac{\sigma_1}{\sigma_2}A\langle e_1\rangle. \tag{19}$$

Here, allowance is made for the fact that, in the case of averaging over the unit cell areas, the average values differ by the phase $\exp(-i\pi\gamma)$.

5. EFFECTIVE CONDUCTIVITY OF TWO-PHASE MEDIA

We use the foregoing results to find the expression for the effective conductivity of the medium being investigated,

$$\sigma_{\text{eff}} = \sigma_1\frac{\sqrt{(\sigma_1 + \sigma_2)^2\cot^2\alpha + 4\sigma_1\sigma_2} - \cot\alpha(\sigma_1 - \sigma_2) + 2\sigma_2\sqrt{1 + \cot^2\alpha}}{\sqrt{(\sigma_1 + \sigma_2)^2\cot^2\alpha + 4\sigma_1\sigma_2} - \cot\alpha(\sigma_1 - \sigma_2) + 2\sigma_1\sqrt{1 + \cot^2\alpha}}. \tag{20}$$

The derived expression is rather cumbersome. Therefore, we will investigate the limiting cases which follow from the derived formula.

(i) Let $\cot\alpha = 0$ ($\alpha = \pi/2$). Then, the expression for conductivity transforms to the known formula of Dykhne,

$$\sigma_{\text{eff}} = \sqrt{\sigma_1\sigma_2}. \tag{21}$$

(ii) In the case of $\cot\alpha = +\infty$ ($\alpha = 0$), we will have the limiting case of stratified medium when the resistance is averaged,

$$\sigma_{\text{eff}} = \frac{2\sigma_1\sigma_2}{\sigma_1 + \sigma_2}. \tag{22}$$

This result may be easily understood because, in the case of the parallelogram angle tending to zero, we have a medium consisting of one-dimensional lines made up of alternating segments of conductivity σ_1 and σ_2.

(iii) With $\cot\alpha = -\infty$ ($\alpha = \pi/2$), we have the other limiting case of a stratified medium, when the electric field is averaged,

$$\sigma_{\text{eff}} = (\sigma_1 + \sigma_2)/2. \tag{23}$$

In this case, the medium is formed by alternation of one-dimensional homogeneous conducting filaments with their conductivities.

The obtained limiting results, namely, the angular dependence of effective conductivity, may further be

interpreted as follows. If the conductivities of the phases differ strongly as regards their properties, the current largely flows in the well-conducting phase, and the flow of current from one phase to another proceeds in the vicinity of the phase contact via a narrow neck. In the case of small angles of the cells of the well-conducting phase, the current density increases and, accordingly, the resistance of the well-conducting regions increases so that the flow via poorly conducting regions becomes more profitable: the area of the necks increases. This corresponds to successive inclusion of resistances (see Eq. (22)).

On the contrary, when the cell angle increases, the region of necks in which the current flows in the poorly conducting phase decreases and, for fairly large angles, the effect of the poorly conducting phase becomes unimportant. In this case, the conductivity is described by expression (23).

(iv) Of the most interest appears to be the step dependence of the effective conductivity on the cell angle in the case when one of the phases is dielectric,

$$\sigma_{\text{eff}} \sim \begin{cases} 0, & \alpha < \pi/2 \\ \sigma_1/2, & \alpha \geq \pi/2. \end{cases}$$

This result follows from the foregoing formulas. With small cell angles and up to the straight angle $\pi/2$, i.e., with "acute" contact of one-color conducting parallelograms, the current density is such that, for flow from one unit cell to another with one and the same conductivity, the other conducting phase must be included. Because the conductivity of this phase is zero (insulator), the effective conductivity of the system is likewise zero. In the case of wide contacts corresponding to the obtuse angle of the parallelograms, the current may flow from one phase to another without involving the other phase. Therefore, the effective conductivity of the medium is other than zero.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 99-02-17355) and by a grant of the sixth Evaluation Competition of Young Scientists, Russian Academy of Sciences.

I am grateful to A.M. Dykhne for the interesting preliminary discussion of the results.

I am also grateful to Abdus Salam ICTP Institute in Trieste, Italy, for their hospitality; it was there that a part of the results described above were obtained.

REFERENCES

1. A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110, 641 (1970) [Sov. Phys. JETP **32**, 63, 348 (1970)].
2. Yu. P. Emets, *Electrical Properties of Composites with Regular Structure* (Naukova Dumka, Kiev, 1986).
3. Yu. P. Emets, Zh. Éksp. Teor. Fiz. **96**, 701 (1989) [Sov. Phys. JETP **69**, 397 (1989)].
4. Yu. N. Ovchinnikov and M. Ya. Ovchinnikova, Zh. Éksp. Teor. Fiz. **118**, 1434 (2000) [JETP **91**, 1242 (2000)].
5. A. M. Markushevich, *Brief Course of the Theory of Analytic Functions* (Nauka, Moscow, 1978).
6. M. A. Lavrent'ev and B. V. Shabat, *Methods of the Theory of Functions of a Complex Variable* (Nauka, Moscow, 1973).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 1013–1025.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1179–1193.
Original Russian Text Copyright © 2002 by Golosova, Bobrovskiĭ, Mitberg, Podlesnyak, Zhdakhin, Mirmelstein.

SOLIDS
Electronic Properties

Superposition Structure of Crystal Field Spectra in High-T_c Superconductors: Overdoped Regime

N. O. Golosova[a], V. I. Bobrovskiĭ[a], É. B. Mitberg[a, b], A. A. Podlesnyak[a, c], I. L. Zhdakhin[a], and A. V. Mirmelstein[a,*]

[a]*Institute of Metal Physics, Ural Division, Russian Academy of Sciences, ul. S. Kovalevskoĭ 18, Yekaterinburg, 620219 Russia*
[b]*Institute of Solid-State Chemistry, Ural Division, Russian Academy of Sciences, ul. Pervomaiskaya 91, Yekaterinburg, 620219 Russia*
[c]*Swiss Federal Technological Institute and Paul Scherrer Institute, CH-5232, Villigen, Switzerland*
**e-mail: neutron@imp.uran.ru*
Received December 13, 2001

Abstract—Experiments on inelastic scattering of neutrons show that the crystal field spectra for high-T_c superconductors $R_{1-y}Ca_yBa_2Cu_3O_{x\approx 7}$ (R = Ho, Er; $0 < y < 0.25$) have two spectral components associated with optimally doped and overdoped clusters, respectively. An increase in the calcium concentration does not affect the local density of charge carriers in clusters, but changes the concentration of clusters themselves and, hence, the spectral weights of the spectral components. In light of such a "two-phase" pattern observed earlier for cuprate-based superconductors with a doping level below optimal, an increase in the charge carrier concentration leads to a smooth transition (crossover) from the underdoped regime to the overdoped one. The obtained results show, however, that these two regions of the phase diagram differ qualitatively in the form of charge distribution in CuO_2 planes responsible for superconductivity. *© 2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It was assumed for a long time that superconductivity in cuprates emerges in a homogeneous electron system through, say, a magnetic mechanism of superconducting pairing. However, numerous experimental results published in recent years indicate the presence of space charge and spin inhomogeneities and lattice effects in cuprates associated with them [1–11]. Taking into account these data, Schneider and Keller [11] proved recently that the anomalous dependence of the isotope effect on the doping level in cuprates can be explained by treating the superconductor–insulator transition as a 2D quantum phase transition in a heterogeneous system, whose critical point depends not only on the charge carrier concentration, but also on the ion mass. This result confirms the opinion according to which charge and spin inhomogeneities emerging for a low doping level as a result of strong electron–electron and electron–lattice interactions determine the properties of cuprates to a considerable extent and are responsible for the emergence of superconductivity in them. In this connection, the experimental study of the spatial distribution of doping-induced charges in CuO_2 planes responsible for superconductivity is undoubtedly of interest.

The crystal field potential at the site of a rare-earth ion located in a cuprate between two adjacent CuO_2 planes is determined by the spatial distribution and magnitude of the electric charge surrounding this rare-earth ion. Consequently, the crystal field spectrum of the rare-earth ion, which is measured with the help of inelastic neutron scattering, is an ideal tool for studying doping processes in copper–oxygen perovskites by charge carriers. The neutron spectroscopy technique makes it possible to directly observe and describe quantitatively the mechanism of charge transfer from CuO chains to superconducting CuO_2 planes during doping and to obtain unique information on cluster formation indicating the emergence of "frustrated phase separation" as well as on the symmetry of the superconducting order parameter [12–16].

The effect of frustrated phase separation reflects the considerable difference between the charge density averaged over the sample volume and the local charge density in the vicinity of a doping site. This difference is manifested in the superposition structure of the crystal field spectra whose different spectral components may be put in correspondence to different local configurations in the environment of rare-earth ions, which coexist in the sample, while the spectral weights of the components are determined by the probability of detecting a given configuration for a given doping level [12, 14]. However, the cluster formation under doping of high-temperature cuprates has been reliably established by the neutron spectroscopy method only for the compound $RBa_2Cu_3O_x$ (R stands for a rare-earth ele-

ment) with R = Er, the doping being carried out by changing the oxygen content ($6 < x < 7$) [12, 14]. It is well known that doping in 123 systems can be carried out not only by varying the oxygen content (off-plane doping), but also by doping the cation sublattice, e.g., upon partial substitution of Ca^{2+} ions for R^{3+} ions (in-plane doping) [17]. The substitution of calcium may transfer the system to the overdoped regime characterized by a decrease in the superconducting transition temperature T_c upon an increase in the electronic hole concentration, thus expanding the range of investigations.

In the present work, neutron spectroscopic study of the effect of in-plane doping on the fine structure of the crystal-field spectra of the overdoped compound $Ho_{1-y}Ca_yBa_2Cu_3O_{x\approx 7}$ is carried out. The holmium system was chosen by us since the crystal field spectrum of Ho^{3+} ions in the orthorhombic phase consists only of singlets, which makes it possible to interpret unambiguously the experimental spectra of inelastic scattering of neutrons. In the analysis of the results, we used the data on neutron inelastic scattering for $Er_{1-y}Ca_yBa_2Cu_3O_{x\approx 7}$ obtained earlier [18, 19]. The crystal field spectra for both compounds ($Ho_{1-y}Ca_yBa_2Cu_3O_x$ and $Er_{1-y}Ca_yBa_2Cu_3O_x$, $6 < x < 7$) were reliably established in the entire oxygen stoichiometry region $6 < x < 7$ [13, 15]. It will be shown below that a combined analysis of neutron spectroscopic data for the holmium and erbium systems leads to the conclusion that, first, the effect of frustrated phase separation is independent of the doping method (in-plane or off-plane doping) and, second, clusters with different charge carrier concentrations are also preserved in the deeply overdoped regime with $T_c = 56$ K, although the form of the charge distribution in CuO_2 planes changes qualitatively in the vicinity of the optimal doping level.

2. SAMPLES AND EXPERIMENTAL TECHNIQUE

Ceramic samples of $Ho_{1-y}Ca_yBa_2Cu_3O_{x\approx 7}$ ($y = 0$, 0.1, 0.25) were prepared by using the standard technique from high-purity oxides Ho_2O_3 and CuO and carbonates $CaCO_3$ and $BaCO_3$. The stoichiometric mixtures of the initial reagents were annealed in a dynamic forevacuum under gradual heating from 820 to 850°C for 20 h until complete decomposition of carbonates. Subsequent synthesis was carried out in air at 900–960°C for 80 h with several intermediate homogenization processes. The samples were then annealed in an oxygen flow at 1000°C for several hours and slowly cooled to 350°C also in oxygen. Such a procedure minimizes the emergence of $BaCuO_2$ and $Ba_4CaCu_3O_8$ impurities, facilitates the most complete substitution of calcium for rare-earth ions at lattice sites, and makes it possible to attain a homogeneous distribution of oxygen over the sample volume with an oxygen concentration close to $x = 7$ [20]. The phase purity of the synthesis products was controlled by using X-ray diffraction.

The superconducting properties of the obtained samples were verified using ac magnetic susceptibility and dc resistivity measurements by the standard four-probe technique, and the crystal structure of the samples was determined using high-resolution neutron powder diffraction. These results were described in detail in [21]. The ac susceptibility measurements revealed that the dependence of T_c on the calcium concentration y ($T_c \approx 92$, 71, and 56 K for $y = 0$, 0.1, and 0.25, respectively) is in accord with the results obtained by other groups of scientists [18, 20]. The absolute values and temperature dependences of resistivity confirm that the samples with substituted calcium are in the overdoped regime [21]. (It should be noted that the static magnetic susceptibility measurements which are often used for verifying the presence of overdoping regime are inapplicable in the present case since the dominating contribution of the rare-earth ion to the susceptibility masks the absence of a pseudogap which is a distinguishing feature of the overdoping regime.) According to the results of neutron diffraction analysis, the substitution of calcium leads to a slight increase in the lattice parameter c and to a decrease in the orthorhombicity $s = 2(b-a)/(a+b)$, where a and b are the lattice parameters, and the chemical composition of the samples (according to the results of profile analysis of diffraction patterns) can be written in the form $HoBa_2Cu_3O_{6.95(1)}$, $Ho_{0.904(6)}Ca_{0.096(6)}Ba_2Cu_3O_{6.99(2)}$, and $Ho_{0.754(8)}Ca_{0.246(8)}Ba_2Cu_3O_{6.86(4)}$ (the error in determining the occupation numbers is given in parentheses; the value of the Ca concentration in the samples can be refined owing to a considerable difference in the neutron scattering amplitudes of Ho and Ca nuclei) [21].

The measurements of the neutron inelastic scattering spectra associated with transitions between the crystal field levels of the lower multiplet $^5I_8$ of Ho^{3+} ions were made on the triple-axis spectrometer Druehal at the neutron source SINQ (Paul Scherrer Institute, Switzerland). The crystal field of the ortho-rhombic symmetry splits the multiplet $^5I_8$ into seventeen singlets, eight of which lie in the low-energy range ($\Delta E <$ 12 meV), and the remaining nine fall in the high-energy region (55 meV $< \Delta E <$ 74 meV) [15]. Since it is known from [15] that lower lying levels of the crystal field are more sensitive to doping, we studied the low-energy part of the spectrum. Measurements were made at temperature $T = 1.5$ K, at which transitions can occur only from the ground state of the multiplet. In the energy transfer region up to $\Delta E = 3.0$ meV, the spectra were recorded for the scattering vector modulus $Q = 0.85$ Å^{-1} with a fixed energy $E_f = 3.5$ meV of scattered neutrons. Neutrons with a higher order energy were suppressed with the help of a beryllium filter. In the energy spectral range up to $\Delta E = 15$ meV, the spectra were measured for $Q = 1.8$ Å^{-1} and $E_f = 7$ meV.

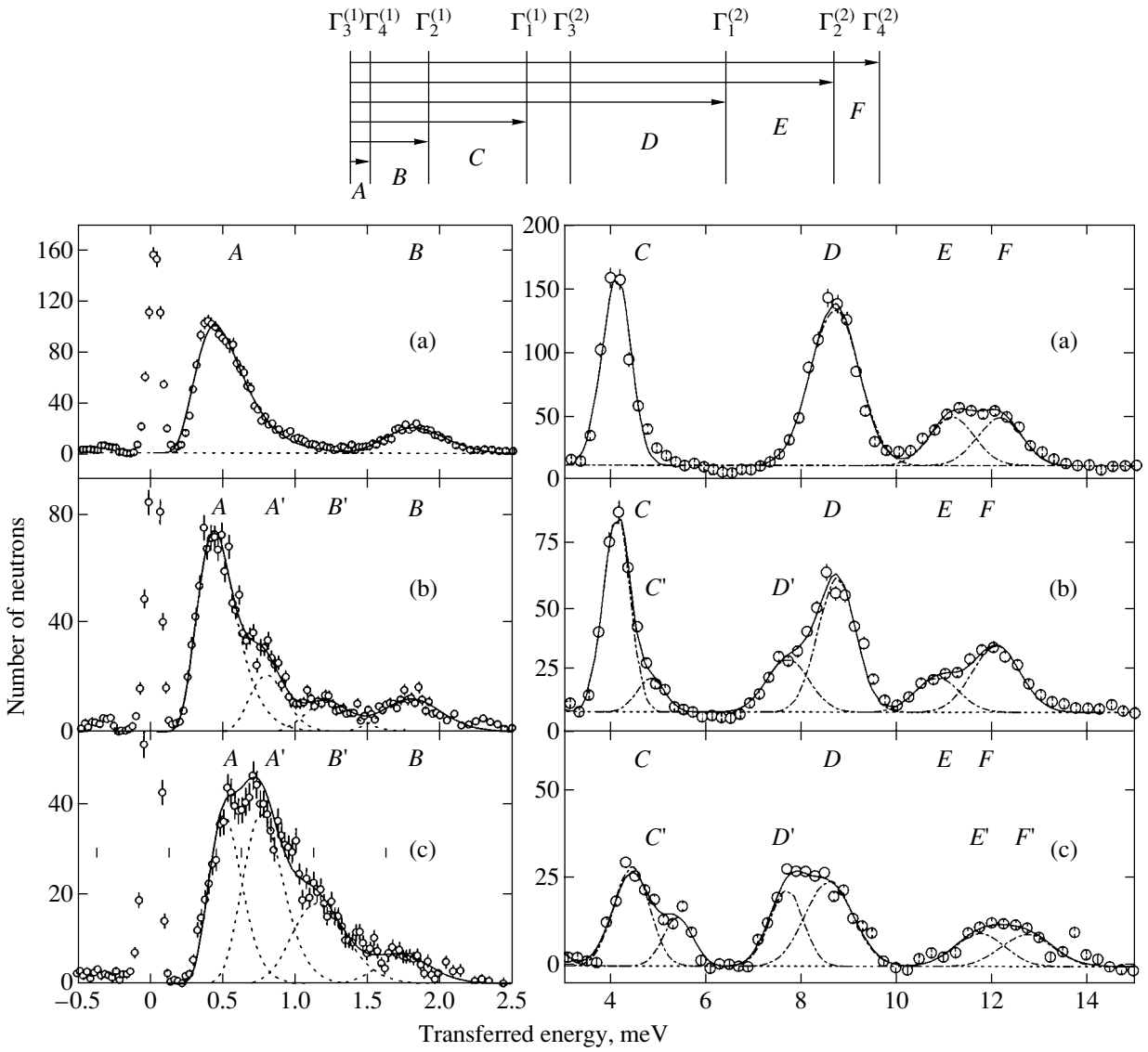

Fig. 1. Energy spectra of neutrons scattered from $Ho_{1-y}Ca_yBa_2Cu_3O_{x\approx 7}$ at $T = 1.5$ K for $y = 0$, 0.1 and 0.25 (a, b, and c, respectively). Left diagrams correspond to recording for $Q = 0.85$ Å^{-1} and $E_f = 3.5$ meV. The right diagrams are recorded for $Q = 1.8$ Å^{-1} and $E_f = 7$ meV. The top diagram shows the splitting of low-lying crystal field levels in $HoBa_2Cu_3O_{6.95}$ [15].

3. EXPERIMENTAL RESULTS

The energy spectra of neutrons scattered from the samples of $HoBa_2Cu_3O_{6.95}$, $Ho_{0.9}Ca_{0.1}Ba_2Cu_3O_{6.99}$, and $Ho_{0.75}Ca_{0.25}Ba_2Cu_3O_{6.86}$ at $T = 1.5$ K are shown in Fig. 1. As expected, the low-energy spectrum of the optimally doped composition $HoBa_2Cu_3O_{6.95}$ consists of six transitions from the ground state $\Gamma_3^{(1)}$ [15]. In Fig. 1, these transitions are denoted by *A*, *B*, *C*, *D*, *E*, and *F*. Partial substitution of calcium for holmium leads to the emergence of four additional spectral components *A*', *B*', *C*', and *D*', whose intensity clearly increases with the calcium concentration. Since all the levels of the crystal field of Ho^{3+} ions are singlets in the orthorhombic symmetry, the observed spectral line splitting indicates the presence of clusters creating electronic inhomogeneity in Ca-substituted samples. It was mentioned above that a similar effect of frustrated phase separation is a distinguishing feature of underdoped cuprates. The results presented in Fig. 1 demonstrate that the local electron inhomogeneity in CuO_2 planes is also preserved in the deeply overdoped regime ($T_c = 56$ K).

Figure 1 shows that the energies and relative intensities of lines *A*–*F* in the spectra of Ca-substituted samples remain virtually unchanged as compared to the optimally doped Ho-123 (although the lines *A*, *C*, *E*, and *F* are slightly shifted upwards). Consequently, these lines can be put in correspondence with optimally doped clusters existing in Ca-substituted samples, while the additional components *A*', *B*', *C*', and *D*' can

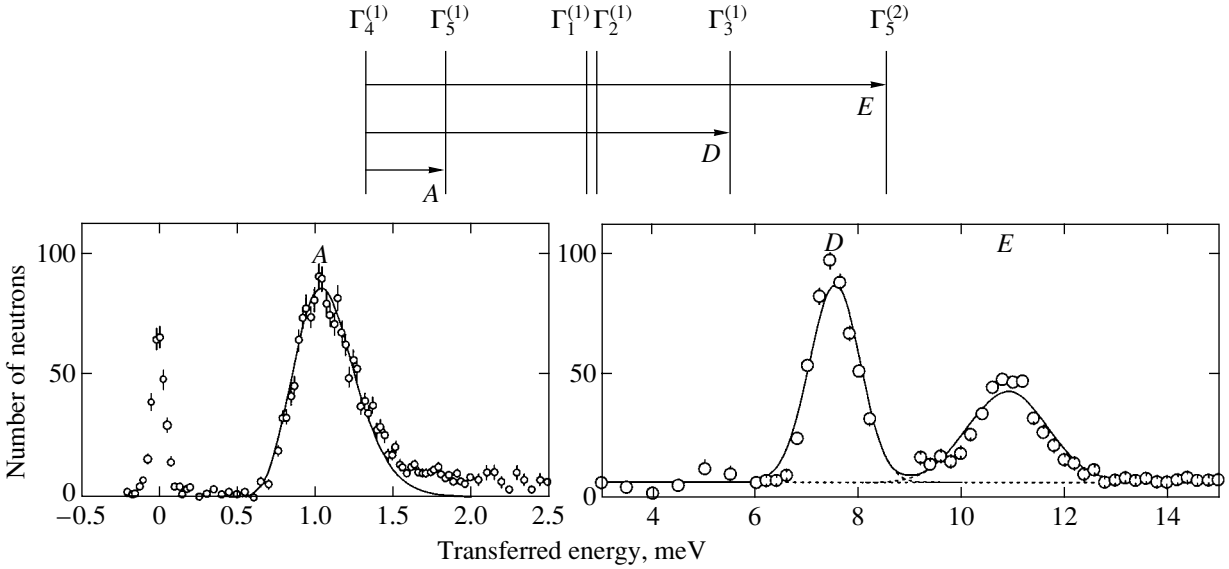

Fig. 2. Energy spectra of neutrons scattered from undoped tetragonal $HoBa_2Cu_3O_{6.3}$ at $T = 1.5$ K. Measurements were made under the same conditions as indicated in Fig. 1. The top diagram shows the splitting of low-lying crystal field levels in a tetragonal field [15].

be juxtaposed to presumably overdoped clusters created by the substitution of calcium in the sublattice of rare-earth ions. Obviously, the ratio $P_{\text{undist}} = [I_A + I_B]/[I_A + I_B + I_{A'} + I_{B'}]$, where I is the integrated intensity of the corresponding line, determines the concentration of optimally doped (i.e., unperturbed by the introduction of calcium) clusters. It should be noted that, in order to describe the strongly asymmetric shape of transitions (see Fig. 1), we used the Naperian logarithmic function. The origin of the asymmetry is discussed in detail in [22].We used the results presented in Fig. 1 to obtain $P_{\text{undist}} = 0.74 \pm 0.03$ and 0.40 ± 0.03 for $y = 0.1$ and 0.25, respectively. Since the crystal field potential at a rare-earth site is a local probe for its surroundings [12], we can assume that only those holmium ions which are the nearest neighbors of an implanted calcium ion are sensitive to the substitution of Ca^{2+} for Ho^{3+}. In the lattice of a 123 compound, the nearest neighborhood of each Ho^{3+} ion contains four sites that can be replaced by Ca^{2+}. Consequently, the statistical probability of discovering k ($0 \le k \le 4$) from the four positions occupied by calcium is equal to [14]

$$P_k^4(y) = \frac{4!}{(4-k)!k!} y^k (1-y)^{4-k}, \qquad (1)$$

where y is the calcium concentration per structural unit. For given values of $y = 0.096 \pm 0.006$ and 0.246 ± 0.008 (see Section 1), relation (1) for $k = 0$ leads to $P_{\text{undist}} = 0.67 \pm 0.03$ and 0.33 ± 0.03, respectively. The expected concentrations of unperturbed clusters calculated in this way are in reasonable agreement with the concentrations obtained as a result of processing of the crystal field spectra. These numerical estimates confirm that the superposition effect discovered in Ca-substituted samples is indeed associated with the substitution of calcium in the sublattice of rare-earth ions.

Figure 2 shows the energy spectrum of neutrons scattered by an undoped sample of $HoBa_2Cu_3O_{6.3}$, which was measured under the same conditions as the spectra presented in Fig. 1. It should be noted that, for a low concentration of oxygen, 123 compounds have a tetragonal structure. Accordingly, the low-energy crystal field spectrum consists of four singlets and two doublets (see Fig. 2) [15]. Note that the transitions $\Gamma_4^{(1)} \longrightarrow \Gamma_1^{(1)}$ and $\Gamma_4^{(1)} \longrightarrow \Gamma_2^{(1)}$ are forbidden by the tetragonal symmetry. This spectrum will be required for an analysis of the obtained results.

4. ANALYSIS AND DISCUSSION OF RESULTS

4.1. Crystal Field Parameters of the Components of the Spectrum of Ho-123 with Calcium Substitutions

Since the total splitting of the lower multiplet $^5I_8$ of Ho^{3+} ions (approximately 74 meV) [15] is quite large as compared to the separation between multiplets (approximately 620 meV for Ho^{3+}), the effects of multiplet mixing and spin–orbit interaction (in the intermediate-bond approximation) must be taken into account for a correct description of the experimentally observed energies and intensities of transitions [13, 15]. Following [15], we took into account in our calculations all multiplets with $J = 8$, 7, and 6, lying at energies 0, 620, and 1067 meV, respectively. In order to find the energy levels and the wave functions of Ho^{3+} ions, we diagonalized simultaneously the Hamiltonians of the crystal field, electrostatic, and spin–orbit interactions. In this

case, the crystal field Hamiltonian can be written in the form of tensor operators [23]:

$$H_{CF} = \sum_{n=1}^{3}\sum_{m=0}^{n} B_{2m}^{2n}(Y_{2m}^{2n} + Y_{-2m}^{2n}), \qquad (2)$$

where Y_{2m}^{2n} are the spherical tensor operators and B_{2m}^{2n} are the crystal field parameters to be determined from the experiment (for the sake of convenience, we will henceforth use the crystal field parameters in the conventional form $B_{2n2m} \equiv B_{2m}^{2n}/D_{2m}^{2n}$, where D_{2m}^{2n} are numerical factors [23]). It should be recalled that, for a crystal field of tetragonal symmetry, the crystal field parameters with an odd m are equal to zero, but for orthorhombic samples, a set of nine parameters has to be determined. The procedure of determining the crystal field parameters involves the application of the least squares technique for fitting the calculated crystal field spectrum to the experimentally observed spectrum of inelastic scattering of neutrons, which is determined by the second derivative of the neutron scattering cross section [24]:

$$\frac{d^3\sigma}{d\Omega d\omega} \sim F^2(Q)\sum_{i,j}\exp\left(-\frac{E_i}{k_BT}\right) \times \left|\langle\Gamma_j|J_p|\Gamma_i\rangle\right|^2\delta(E_j - E_i + \hbar\omega), \qquad (3)$$

where $F(Q)$ is the magnetic form factor, E_i is the energy of the ith level of the crystal field, Γ_i is the corresponding irreducible representation, and J_p is the total angular momentum component perpendicular to the scattering vector Q. Using this procedure, we determined the crystal field parameters (i) for undoped $HoBa_2Cu_3O_{6.3}$ whose spectrum is shown in Fig. 2, (ii) for optimally doped $HoBa_2Cu_3O_{6.95}$ (Fig. 1a), and (iii) for the spectrum component induced by calcium substitutions and presented by the transitions A', B', C', D', E, and F in Fig. 1c. The information on the high-energy part of the spectrum (55 meV < E < 75 meV) was borrowed from [15] since the energy resolution for large energy transfers does not allow us to resolve the superposition structure of the spectra.

The results of fitting are presented in Tables 1 and 2 (it should be noted that, for Ca-free samples $HoBa_2Cu_3O_x$ (x = 6.3 and 6.95), the values of the crystal field parameters obtained by us are very close to the results obtained in [15]). The data presented in Table 2 clearly demonstrate the general tendency in the variation of all main crystal field parameters[1] (n = 1, 2, 3; m = 0, 2) upon doping: these parameters for $HoBa_2Cu_3O_x$ increase as

[1] These parameters are often referred to as "tetragonal" in contrast to the "orthorhombic" parameters corresponding to odd values of m and vanishing in a tetragonal symmetry field.

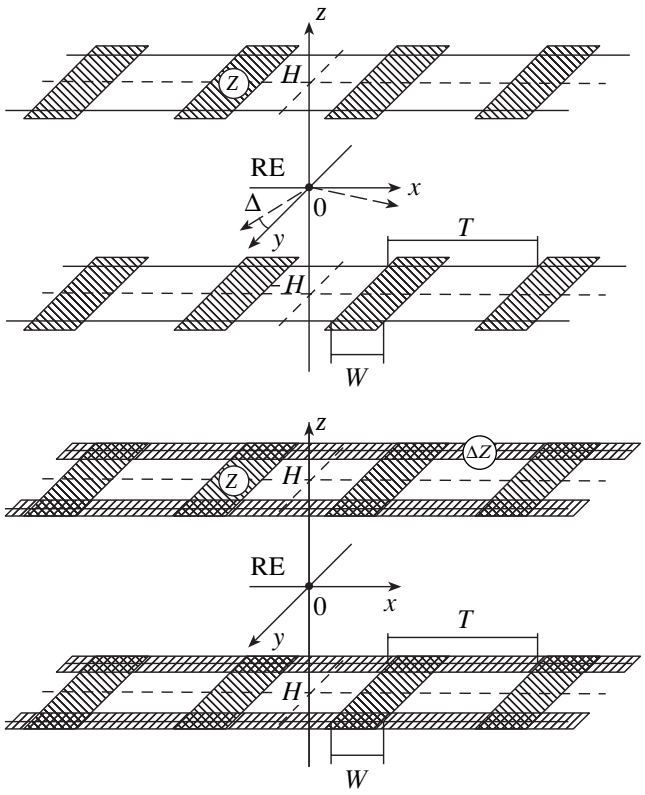

Fig. 3. Schematic diagram of charge distribution in CuO_2 planes of cuprates according to the PACT model below (top) and above (bottom) the optimal doping level.

the value of x increases from 6.3 to 7 and continue to grow with the calcium concentration (except the sixth-order parameters for the "calcium" component of the spectrum; this circumstance will be discussed below). It is important to emphasize that the same tendency associated with an increase in the positive charge in CuO_2 planes upon doping was observed earlier in $ErBa_2Cu_3O_x$ [13, 25]. Thus, we can draw the preliminary conclusion that the calcium component of the spectra is indeed associated with the emergence of overdoped clusters.

For a further analysis of the obtained results, we will use the model of the crystal field generated by a periodic array of charged tapes (PACT) [25, 26] proposed by us for describing the concentration dependence of the crystal field parameters in high-T_c superconductors with the 123 structure.

4.2. Model of Periodic Array of Charged Tapes (PACT)

According to the PACT model, extra charges emerging in CuO_2 planes as a result of doping form a periodic array of homogeneously charged tapes extended along one of the main crystallographic directions in the lat-

Table 1. Energies E_i and relative intensities I_i/I_A of spectral lines of the crystal field in Ho-123 at different doping levels obtained in the present work

1. Undoped $HoBa_2Cu_3O_{6.3}$						
	A	*B*	*C*	*D*	*E*	*F*
E^{exp}, meV	1.1(1)		–	7.50(5)	10.9(1)	
E^{calc}, meV	1.10		4.89	7.47	10.85	
$(I_i/I_A)^{exp}$	1		–	0.43(4)	0.30(4)	
$(I_i/I_A)^{calc}$	1		0	0.62	0.36	
2. Optimally doped $HoBa_2Cu_3O_{6.95}$						
E^{exp}, meV	0.42(2)	1.8(1)	4.04(2)	8.61(2)	11.1(3)	12.1(3)
E^{calc}, meV	0.45	1.72	4.10	8.33	11.02	12.11
$(I_i/I_A)^{exp}$	1	0.23(3)	0.25(5)	0.35(5)	0.11(3)	0.11(3)
$(I_i/I_A)^{calc}$	1	0.22	0.30	0.47	0.15	0.19
3. Overdoped component $Ho_{0.75}Ca_{0.25}Ba_2Cu_3O_{6.86}$						
	A'	*B*'	*C*'	*D*'	*E*	*F*
E^{exp}, meV	0.76(3)	1.15(5)	5.4(2)	7.7(2)	11.9(3)	12.8(3)
E^{calc}, meV	0.76	1.12	5.67	7.56	11.9	12.7
$(I_i/I_{A'})^{exp}$	1	0.64(6)	–	–		
$(I_i/I_{A'})^{calc}$	1	0.59	–	–		
4. $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{6.93}$						
	A	*B*	*C*	*A*'	*B*'	*C*'
E^{exp}, meV	9.5(3)	10.1(2)	11.2(4)	9.5(3)	12.4(4)	11.2(4)
E^{calc}, meV	9.5	10.3	11.2	9.5	12.4	11.2
$(I_i/I_A)^{exp}$, $(I_{i'}/I_{A'})^{calc}$	1	0.25	0.69	1	0.41	0.95

Note: Rows 1–3 contain the results obtained by processing the inelastic neutron scattering spectra using the fitting procedure described in the text. The notation of the transitions *A*–*F* correspond to splitting diagrams shown in Figs. 1 and 2. Row 4 contains parameters of the optimally doped and overdoped components of the crystal field spectrum in $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{6.93}$ [19] (Fig. 5).

tice, either along direction *a* or along *b* (upper part of Fig. 3). In the transverse direction, the array of tapes is characterized by the width *W* and period *T* ($W < T$) equal to the corresponding crystal lattice parameter so that the axes of the tapes coincide with lines of chemical bonds Cu–O–Cu. Each tape carries a charge *Z* which is constant for a given doping level, positive for holes and negative for electrons (in units of $|e|/Å^2$). The crystal field parameters $B_{2n2m}(x)$ for a sample with the oxygen concentration *x* can be presented in the form [25]

$$B_{2n2m}(x) = B_{2n2m}(x = 6) + B_{2n2m}(Z), \qquad (4)$$

where parameters $B_{2n2m}(x = 6)$ for an undoped sample are assumed to be known from experiments, *Z* is the surface charge density on a tape, and the additional contribution $B_{2n2m}(Z)$ associated with doping is calculated with the help of the following formulas [25, 26]:

$$B_{20} = \frac{1}{2}b_2, \quad B_{60} = \frac{1}{16}b_6,$$

$$B_{22} = \frac{1}{2}\cos(2\Delta)b_2, \quad B_{62} = \frac{15}{32}\cos(2\Delta)b_6,$$

$$B_{40} = \frac{1}{8}b_4, \quad B_{64} = \frac{3}{16}\cos(4\Delta)b_6, \qquad (5)$$

$$B_{42} = \frac{1}{2}\cos(2\Delta)b_4, \quad B_{66} = \frac{1}{32}\cos(6\Delta)b_6,$$

Table 2. Crystal field parameters B_{2n2m} for compounds Ho-123 and Er-123, obtained in the present work

$(2n, 2m)$	(2,0)	(2,2)	(4,0)	(4,2)	(4,4)	(6,0)	(6,2)	(6,4)	(6,6)
1. Undoped $HoBa_2Cu_3O_{6.3}$									
B_{2n2m}^{calc}, meV	15.3(8)	0	–34.0(2)	0	158.3(1.5)	3.74(5)	0	121.6(3)	0
2. Optimally doped $HoBa_2Cu_3O_{6.95}$									
B_{2n2m}^{calc}, meV	25.6(8)	14.3(1.6)	–33.98(2)	3.6(1.5)	160.3(1.8)	3.97(4)	–2.5(1.0)	123.5(4)	–0.4(3)
B_{2n2m}^{PACT}, meV	25.6	10.3	–32.2	7.2	160.1	3.8	0.2	121.7	0.01
n_{opt}	0.19(1)								
3. Overdoped component $Ho_{0.75}Ca_{0.25}Ba_2Cu_3O_{6.86}$									
B_{2n2m}^{calc}, meV	30.9(9)	8.4(1.2)	–33.2(3)	2.6(1.2)	164.3(1.9)	3.02(4)	1.6(1.0)	123.3(4)	0.5(3)
B_{2n2m}^{PACT}, meV	30.9	8.9	–33.1	0	161.2	3.98	–2.5	123.5	–0.4
Δn	0.096(5)								
4. Optimally doped component $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{6.93}$									
B_{2n2m}^{calc}, meV	17.9(8)	10.6(1.5)	–33.1(3)	10.3(3)	158.5(1.0)	3.67(4)	–0.6(2)	103.5(4)	0.6(2)
Overdoped component $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{6.93}$									
B_{2n2m}^{calc}, meV	23.6(7)	5.2(1.7)	–31.6(4)	5.5(2.3)	157.8(1.2)	3.75(4)	–0.1(2)	103.6(4)	0.2(2)
B_{2n2m}^{PACT}, meV	23.6	4.9	–32.1	6.5	159.5	3.68	–0.7	103.5	0.6
Δn	0.10(1)								

Note: While calculating the parameters according to the PACT model for overdoped spectral components, we took for the initial values in relation (4) the crystal field parameters for optimally doped components of the corresponding spectra; n_{opt} and Δn are the hole concentrations per copper ion, calculated by using the model.

$$B_{44} = \frac{1}{8}\cos(6\Delta)b_4.$$

Here, Δ is the azimuth angle between the directions of the tapes and a chosen crystallographic direction. For the geometry of a rare-earth position depicted in the upper part of Fig. 3, coefficients b_n are defined by the following relations [25, 26]:

$$b_2 = -1.44\times 4Z\frac{\langle r^2\rangle}{H}\sum_{i=1}^{N}\{\nu_2(\beta_i)-\nu_2(\alpha_i)\},$$

$$b_4 = -1.44\times 4Z\frac{\langle r^4\rangle}{2H^3}\sum_{i=1}^{N}\{\nu_4(\beta_i)-\nu_4(\alpha_i)\}, \quad (6)$$

$$b_6 = -1.44\times 4Z\frac{\langle r^6\rangle}{3H^5}\sum_{i=1}^{N}\{\nu_6(\beta_i)-\nu_6(\alpha_i)\},$$

where H is the distance between a rare-earth site and the plane of tapes and $\langle r^n\rangle$ is the corresponding moment of the radial distribution of 4*f* electrons [27]. Functions $\nu_n(\lambda)$ are defined as [25, 26]

$$\nu_2(\lambda) = \frac{\lambda}{1+\lambda^2}, \quad \nu_4(\lambda) = \frac{\lambda(3-\lambda^2)}{3(1+\lambda^2)^3},$$

$$\nu_6(\lambda) = \frac{\lambda(5-10\lambda^2+\lambda^4)}{5(1+\lambda^2)^5},$$

$$\lambda = \{\alpha,\beta\}, \quad \alpha_i = \frac{iT-(T+W)/2}{H}, \quad (7)$$

$$\beta_i = \frac{iT-(T-W)/2}{H},$$

where α and β are the values of variable λ, corresponding to the boundaries of the tapes. The number of charge carriers (holes in our case) n per in-plane copper ion Cu(2) is defined by the relation

$$n = ZaW, \quad (8)$$

where a is the lattice parameter along the direction of a tape. All lengths in relations (6)–(8) are in angstroms, and parameters B_{2n2m} are in electronvolts. In all calculations carried out by us in this work, the period T is assumed to be constant and equal to $a = b = 3.85$ Å (which, according to the results obtained in [21], corre-

sponds to the mean value of the in-plane crystal lattice parameter for all the samples under investigation, and a high degree of accuracy for this parameter is not required in our calculations). The distance H between a rare-earth site and the plane of the tapes was calculated for each composition on the basis of structural data [21] by using the formula $H = (0.5 - z)c$, where c is the crystal lattice parameter and z is the coordinate of the Cu(2) site. We assumed that the linear size of clusters is approximately equal to 20 Å [28], so that the summation in relation (6) was carried out to $N = 3$ (it was proved earlier [25] that the final result of model calculations of hole concentration n is weakly sensitive to the number of terms in the sum). Following Morrison [29], while calculating the crystal field parameters on the basis of relation (6), we replaced the moments $\langle r^n \rangle$ of the radial distribution of 4*f* electrons by $\langle r^n \rangle (1 - \sigma_n)/\tau^n$ (where σ_n and τ^n are parameters depending only on the number of 4*f* electrons of the rare-earth ion). This is done to take into account the screening of the 4*f* shell by external electrons and the variation in the radial distribution of the ion implanted in the solid as compared to a free ion. Thus, the model calculations can be reduced to fitting the crystal field parameters B_{2n2m}^{PACT} calculated on the basis of relations (5)–(7) to the values of parameters B_{2n2m}^{calc} of Hamiltonian (2), determined from experiments on inelastic scattering of neutrons. There are only two fitting parameters. These are the tape width W and the surface charge density Z in a tape (the hole concentration n is determined automatically). All remaining features of model calculations are given in [25]. It is also shown in [25] that, in the geometry of the doping-induced charge presented in Fig. 3 and for a positive charge of the tapes, Eqs. (5)–(7) lead to a positive correction $B_{2n2m}(Z)$ in Eq. (4), i.e., to an increase in the crystal field parameters relative to the initial values of $B_{2n2m}(6.3)$, which is actually observed in the experiments.

4.3. Analysis of Crystal Field Spectra in Ho-123 in the Framework of the PACT Model

The best fitting of the parameters B_{2n2m}^{PACT} to the experimental values of B_{2n2m}^{calc} for the spectrum of optimally doped Ho-123 is attained for the tape width $W = 1.4$ Å and the hole concentration per copper ion $n \approx 0.19 \pm 0.1$ (see Table 2), which is in good agreement with the generally accepted value of the hole concentration near the optimal doping level, i.e., for $x = 6.95 \pm 0.01$ [20]. In our opinion, this circumstance is a strong argument supporting the realistic nature of the PACT model. Another qualitatively important argument in favor of the model is the correlated variation of second-order parameters predicted by it. Since $B_{22} = 0$ for a tetragonal sample, relations (4) and (5) lead to

$$\Delta B_{20} \equiv [B_{20}(x) - B_{20}(6)] \approx B_{22}(x). \tag{9}$$

The second-order parameters B_{2n2m}^{calc} (see Table 2) undoubtedly satisfy this relation within the experimental error. Relation (9), which is valid for the compound $ErBa_2Cu_3O_x$ ($6 < x < 7$) [25], is a direct consequence of the charge geometry under investigation. It can be seen from relation (5) that, as the charge density Z increases, all crystal field parameters vary in accordance with the same law, which is on the whole in accord with the experimental data forming precisely the basis of the PACT model [25, 26]. However, it is more difficult to establish the regularities predicted by relation (5) for the variations of the fourth- and sixth-order parameters (which are important for a correct description of the spectra) in view of their small magnitude as well as the error in their experimental determination than the regularities for the second-order parameters whose variation upon doping is relatively large. This remark primarily concerns the orthorhombic parameters of the fourth and sixth orders, whose experimental indeterminacy is quite significant since their variation weakly affects the position of the crystal field level as well as the intensities of transitions between them. In addition, the fourth-order and especially sixth-order parameters are very sensitive to fine details in the charge distribution in a unit cell (e.g., to quite probable periodic modulation of the surface charge density along the tape), which could be taken into account in our model, but this would lead to unjustified complications of calculations and would in fact be reduced to the emergence of additional fitting parameters. In all probability, a certain discrepancy in the behavior of the model and experimental sixth-order parameters for the "calcium" component of the spectra of Ca-substituted samples is associated with just this circumstance.

In order to describe the crystal field parameters of the spectrum component induced by the substitution of calcium in the framework of the PACT model, we assume that the tape width is constant, $W = 1.4$ Å, and take the values of B_{2n2m}^{calc} for the optimally doped component for the initial values in relation (4) (to avoid the accumulation of computational errors). The best convergence is attained for the extra hole concentration $\Delta n = 0.1 \pm 0.01$ per copper atom; i.e., the total hole concentration in clusters formed as a result of calcium substitution is $n = 0.29 \pm 0.01$ holes per copper atom. In other words, it follows from our analysis that these clusters are indeed overdoped since the hole concentration in them exceeds the optimal doping level.

Table 2 shows, however, that the orthorhombic crystal field parameters for the overdoped spectral component are smaller than the corresponding values for the optimally doped component, which can be seen espe-

cially clearly for B_{22}. Moreover, it can be observed that the empirical correlation

$$[B_{20,\,\mathrm{over}} - B_{20,\,\mathrm{opt}}] \approx -[B_{22,\,\mathrm{over}} - B_{22,\,\mathrm{opt}}], \qquad (10)$$

which is similar to relation (9), but has the opposite sign, holds in this case. How can this be explained if the charge density increases in the tapes? This apparent contradiction is easily resolved using relation (5) if we assume that the extra holes injected in CuO_2 planes by the substitution of calcium form a new array of tapes oriented at right angles to the existing array, which continues to carry the optimally doped charged $n \approx 0.19$ per copper atom (Fig. 3). In accordance with relation (5), these two mutually perpendicular arrays are characterized by the azimuth angles $\Delta = 0$ and $\pi/2$. In this case, the tetragonal crystal field parameters are the sums of the contributions from two arrays, while the orthorhombic parameters are the differences of these contributions.

Since the positions of the spectral lines of the crystal field, which correspond to optimally doped and overdoped clusters, remain virtually unchanged for the three compositions investigated by us (see Fig. 1), we conclude that an increase in the Ca concentration does not affect the local density of charge carriers in clusters, but changes only the concentrations of clusters themselves and, hence, the spectral weights of the components. However, the average concentration of holes in the samples naturally changes and can be easily calculated for each composition using the following simple "two-phase" relation:

$$\langle n \rangle = P_{\mathrm{undist}} n_{\mathrm{opt}} + (1 - P_{\mathrm{undist}}) n_{\mathrm{over}}, \qquad (11)$$

where $P_{\mathrm{undist}} = 0.74 \pm 0.03$ and 0.40 ± 0.03 for $y = 0.1$ and 0.25, respectively (see above). Relation (11) gives $\langle n \rangle = 0.19 \pm 0.01$ ($y = 0$), 0.22 ± 0.02 ($y = 0.1$), and 0.25 ± 0.01 ($y = 0.25$) holes per copper atom, which is in good agreement with the hole concentration V_- [30] calculated in [21] from the structural data obtained for the same samples using the valence bond method [31] ($V_- = -0.024(3)$ for $HoBa_2Cu_3O_{6.3}$, while, for Ca-substituted samples, $V_- = 0.188(4)$, 0.234(5), and 0.258(6) for $y = 0$, 0.1, and 0.25, respectively). It should be noted that the average concentrations $\langle n \rangle$ for an optimally doped sample and for a sample with $y = 0.1$ are in good agreement with the in-plane hole concentrations obtained from the data on position-sensitive X-ray absorption spectroscopy for the Y-123 compound with a partial substitution of Ca for Y [32].

Thus, the PACT model makes it possible to determine the charge carrier concentration in CuO_2 planes directly from the neutron crystal field spectra. It should be emphasized that it is not the total concentration of charge carriers introduced by doping in the sample, but only its part localized in the planes. At the same time, it should be borne in mind that the resultant hole concentration in the planes determined from the results of neutron spectroscopy of the crystal field, is a function of both the calcium concentration and the oxygen content, the latter being slightly different for different samples. This can explain, in particular, insignificant mutual energy shifts of identical spectral lines in different samples (see Fig. 1) and also the virtual nonlinearity in the values of the above-mentioned hole concentration averaged over the samples. For example, an increase in $\langle n \rangle$ as a result of an increase in the calcium concentration from $y = 0.1$ to 0.25 is partly compensated by a decrease in the oxygen content in a sample with a higher calcium concentration. Thus, in our opinion, the PACT model establishes a reliable scale of concentrations, which will be used in our subsequent analysis.

Figure 4 shows the concentration dependences of the tetragonal crystal field parameters for Er- and Ho-based 123 cuprates. It can be seen from the figure that, for $n < n_{\mathrm{opt}}$ (the change in n in this case is attained by varying oxygen nonstiochiometry), the crystal field parameters in both systems behave almost identically and in accord with the results of model calculations on the whole (the reasons for the discrepancy were discussed above). It should be emphasized once again that, in this concentration range, the relation between the change in the crystal field parameters and the increase in the positive charge in CuO_2 planes can be assumed to be established reliably [12, 13, 15, 25]. Since the crystal field parameters in the overdoped region continue to follow the same tendency (and, hence, could be obtained by simple extrapolation from the region lying below the optimal doping level if the concentration scale is known), we conclude once again that the additional component of the crystal field spectra in Ca-substituted Ho-123 samples is indeed associated with local regions in which the hole concentration exceeds the optimal doping level. In addition, we may conclude that the changes in the crystal field spectra reflect the increase in the positive charge in CuO_2 planes irrespective of the method of doping.

4.4. PACT Model and the Crystal Field Spectrum in $Er_{1-y}Ca_yBa_2Cu_3O_x$

The above algorithm of the analysis of the crystal field parameters for the Ho-123 system can be applied to the crystal field spectra of Er^{3+} ions (the lower J multiplet $^4I_{15/2}$) in the analogous system $Er_{1-y}Ca_yBa_2Cu_3O_x$ ($x \approx 7$, $0 < y < 0.2$) [18, 19], which have not been interpreted adequately so far. It is well known that the low-energy component of the spectrum of optimally doped $ErBa_2Cu_3O_{x\approx 7}$ consists of Kramers doublets A, B, and C with energies $E_A \approx 9.2$ meV, $E_B \approx 10.1$ meV, and $E_C \approx 10.9$ meV [13]. Since only one additional transition with an energy approximately equal to 12.4 meV appears in Ca-doped samples of this compound (Fig. 5), whose intensity definitely increases with the calcium concentration, while the energy and the shape of the spectrum at lower energies do not display any noticeable changes, Böttger *et al.* [18, 19] rightfully inter-

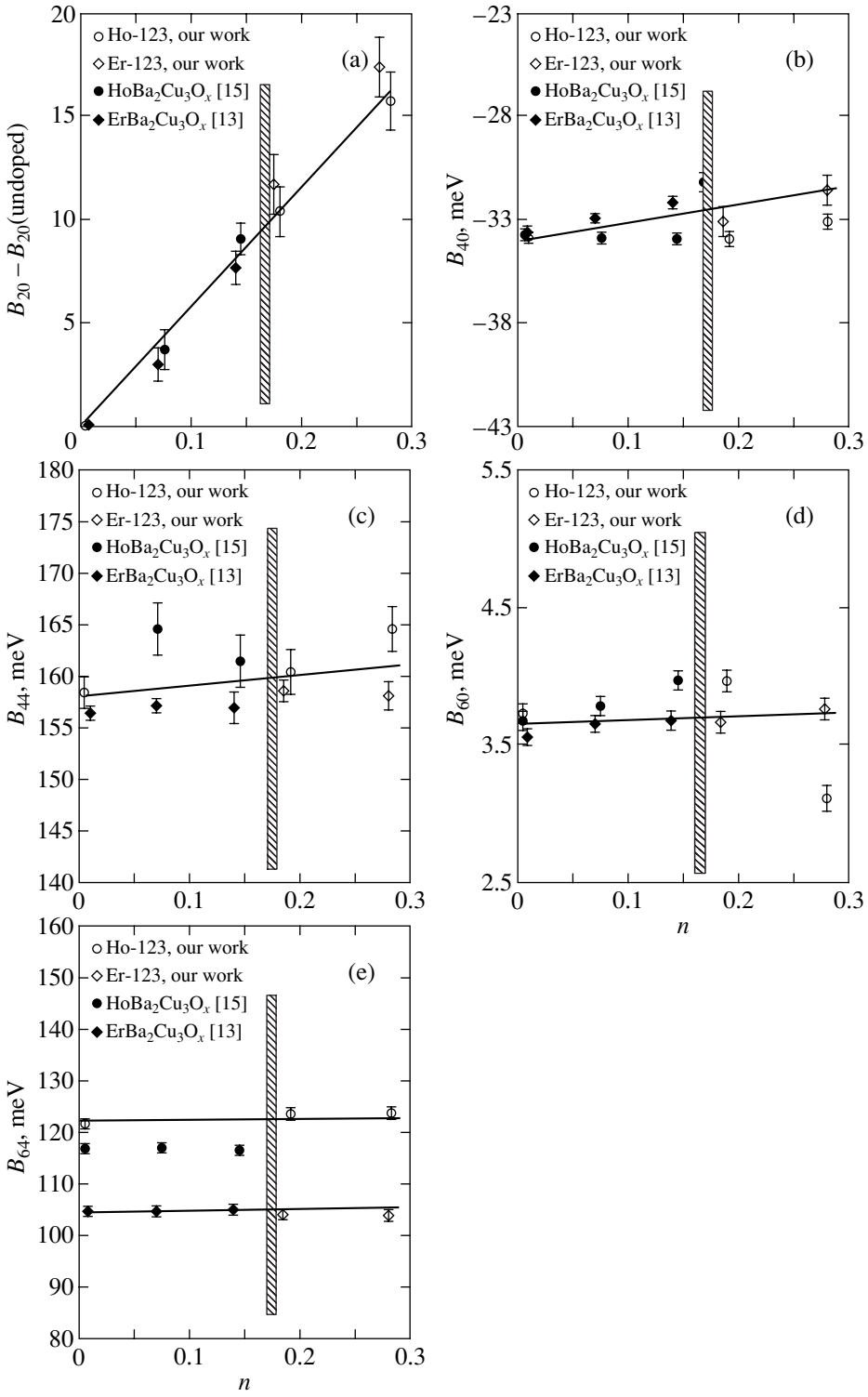

Fig. 4. Basic (tetragonal) crystal field parameters in Er- and Ho-123 systems as functions of the hole concentration in CuO_2 planes. The results obtained in [13, 15], where the hole concentration was varied by changing the oxygen stoichiometry, are given for comparison. Solid lines describe the behavior according to the PACT model. Vertical hatched regions denote the optimal doping level.

preted the transition at 12.4 meV as the only response of the crystal field spectrum to calcium substitutions, which is associated with the formation of overdoped clusters in the samples. However, they could not prove this assumption, the more so that the concentration dependence of the intensity of the additional transition with an energy approximately equal to 12.4 meV, which was interpreted by them as the *C* transition in over-

doped clusters, considerably differs from the dependence predicted by relation (1). It will be proved below that the crystal field spectrum depicted in Fig. 5 can be processed exactly in the same way as the spectrum of $Ho_{0.75}Ca_{0.25}Ba_2Cu_3O_{6.86}$ (see Fig. 1c).

For this purpose, it is sufficient to assume that the spectrum of $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{x\approx 7}$ also consists of two components associated with optimally doped and overdoped clusters. In this case, the lines *A*, *B*, and *C* in Fig. 5 must correspond to optimally doped regions since exactly these lines are present in the spectrum of pure $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{x\approx 7}$ (although the energies corresponding to these lines in the latter spectrum are lower approximately by 0.3 meV) [13]. This set of lines corresponds to a certain set of crystal field parameters for Er^{3+} ions, which can be obtained by diagonalizing Hamiltonian (2). The overdoped component of the spectrum of $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{x\approx 7}$ must correspond to its own set of crystal field parameters; it follows from the PACT model and Fig. 4 that the tetragonal parameters $B_{nm,\,\mathrm{over}}$ must be larger than $B_{nm,\,\mathrm{opt}}$, while the corresponding orthorhombic parameters $B_{nm,\,\mathrm{over}}$ must be smaller than $B_{nm,\,\mathrm{opt}}$. The results of direct calculations show that, for any significant increase in parameter B_{20} relative to $B_{20\,\mathrm{opt}}$ and for a relatively weak variation of the remaining tetragonal parameters, the structure of the crystal field spectrum for Er^{3+} ions changes considerably (see, for example, Fig. 2 in [25]): the position of the *B* line changes the most strongly, its energy becoming higher than the energy of the virtually fixed *C* line, while the intensities of the *A* and *C* lines are comparable. It may turn out in this case that the energies corresponding to the *A* lines for both components are almost identical. Thus, we can logically assume that the spectrum in Fig. 5 consists of two components with energies $E(A)_{\mathrm{over}} \approx E(A)_{\mathrm{opt}} \approx 9.5$ meV, $E(B)_{\mathrm{opt}} \approx 10.2$ meV, $E(B)_{\mathrm{over}} \approx 12.4$ meV, and $E(C)_{\mathrm{over}} \approx E(C)_{\mathrm{opt}} \approx 11.2$ meV. In order to determine the relative intensities of transitions, we must diagonalize Hamiltonian (2), fitting only the energies of the levels (in our calculations, we took into account all possible states with $J = 15/2$ ($E = 0$), $J = 13/2$ ($E = 800$ meV), and $J = 11/2$ ($E = 1240$ meV) [13]). The energies of the levels obtained as a result of fitting and the relative intensities of transitions for optimally doped and overdoped components are given in Table 1, while the corresponding parameters B^{calc}_{2n2m} are presented in Table 2 and in Fig. 4. In order to reconstruct the experimentally observed spectrum of $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{x\approx 7}$, we must fit the background, the line widths including the energy resolution of the spectrometer and the proper line widths (the latter are not necessarily identical for all lines), and the concentration of clusters, i.e., P_{undist}. The best convergence of such a numerical reconstruction to the experimental crystal field spectrum of $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{x\approx 7}$ (see Fig. 5) was obtained for the concentration $P_{\mathrm{undist}} = 0.43 \pm 0.02$,

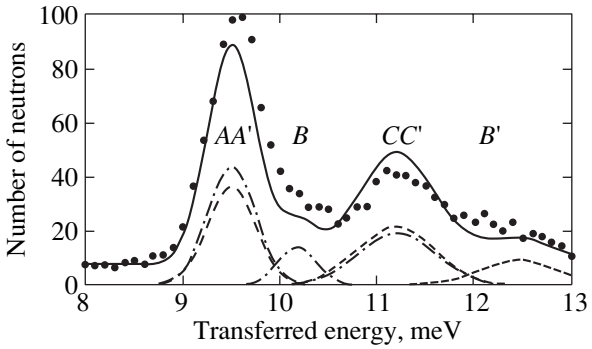

Fig. 5. Spectrum of inelastic neutron scattering for $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{6.93}$ (circles) [15]. The solid curve shows the numerical reconstruction of the experimental spectrum in the form of a superposition of optimally doped (transitions A, B, C; dot-and-dash curve) and overdoped (transitions A', B', C'; dashed curve) components (see Table 2).

which is in excellent agreement with the value of $P^4_0 = 0.4$ expected for the calcium concentration $y = 0.2$.

Parameters B^{calc}_{2n2m} for the overdoped spectral component for $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{x\approx 7}$ can be calculated from relations (4)–(7) of the PACT model if we take for the initial values of the crystal field parameters in relation (4) the values for the optimally doped component from Table 2 and assume the formation of an array of charged tapes (see lower part of Fig. 3) in the overdoped regime. It can be seen from Table 2 that the extra hole concentration $\Delta n = 0.1 \pm 0.01$ per copper atom required for the reproduction of parameter $B^{\mathrm{calc}}_{2n2m,\,\mathrm{over}}$ and associated with the substitution of calcium is the same as for Ca-substituted samples of Ho-123, and the crystal field parameters themselves behave identically in both systems in the entire range of *n* values (see Fig. 4). It should be emphasized that the transformation of the crystal field spectrum as a result of substitution of calcium in the Er-123 system is manifested so clearly that it cannot be interpreted as a purely structural effect. The estimation of the shift in the crystal field levels as a result of a change in structural parameters associated with the substitution of calcium, which was carried out by us in the point charge approximation [12, 13, 15] on the basis of structural data on $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{x\approx 7}$ [18, 19], gives only a more or less homogeneous upward shift for the lines *A*, *B*, and *C* by a value smaller than 0.2 meV, which is comparable with the shift of these lines in the spectrum of $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{x\approx 7}$ relative to that for pure $ErBa_2Cu_3O_{x\approx 7}$, but does not match the "jump" in energy corresponding to line *B* from 10.2 to 12.4 meV. This transformation can be interpreted only as an effect associated with a change in the charge surroundings of a rare-earth ion.

Consequently, the superposition structures of the spectra shown in Figs. 1 and 2, on the one hand, and in

Fig. 5, on the other hand, reflects the same physical effect associated with the coexistence of optimally doped and overdoped clusters in the samples.

5. CONCLUSIONS

The splitting of the lower J multiplet $^5I_8$ of Ho^{3+} ions by the crystal field in overdoped high-temperature superconductors $Ho_{1-y}Ca_yBa_2Cu_3O_{x\approx 7}$ is investigated with the help of the method of inelastic scattering of neutrons. It is established reliably for the first time that, in this concentration range, the crystal field spectra of superconducting cuprates consist of two components whose spectral weights are determined by the calcium concentration y, while the energy and relative intensities of transitions in both components are virtually independent of y. An analysis of inelastic neutron scattering spectra in the framework of the PACT model made it possible to determine the hole concentrations in clusters. The hole concentration in clusters whose crystal field spectra are similar to the spectra of the samples of $HoBa_2Cu_3O_{6.95}$ (T_c = 92 K) free of calcium substitutions lies near the optimal doping level $n_{opt} = 0.19 \pm 0.01$ per copper atom. In clusters formed as a result of calcium substitutions in the rare-earth sublattices, the hole concentration $n_{over} = 0.29 \pm 0.01$ per copper atom. These values of concentrations are in good agreement with the results obtained from independent experiments (see Subsection 4.3). We proved that the crystal field spectrum of $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{x\approx 7}$ measured earlier also consists of two components, and the features of this spectrum are such that they can be due only to charge effects in the vicinity of doping sites. By applying the PACT model, we proved that the extra hole concentration $\Delta n \sim 0.1$ per copper atom in overdoped local regions of $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{x\approx 7}$ is the same as in analogous clusters of the Ho-123 system. The corresponding numerical results and the established systematic and identical nature of variation of the main crystal field parameters as functions of the hole concentration for the Ho-123 and Er-123 systems lead to the conclusion that the superposition structure of the crystal field spectra, which is obvious for $Ho_{1-y}Ca_yBa_2Cu_3O_{x\approx 7}$, and the emergence of an additional transition with an energy about 12.4 meV in the spectrum of $Er_{0.8}Ca_{0.2}Ba_2Cu_3O_{x\approx 7}$ are the same physical effect. The essence of this effect is that extra holes injected in CuO_2 planes as a result of calcium substitutions are not distributed uniformly over the sample volume, but are localized directly near doping centers. An increase in the calcium concentration does not affect the local density of charge carriers in clusters, but changes the concentration of clusters themselves and, hence, the spectral weights of spectral components.

It was mentioned in the Introduction that, according to earlier observations, a similar effect of spatial segregation of charge carriers is a characteristic feature of cuprates for $n < n_{opt}$, but this was observed experimentally only for off-plane doping (change in the oxygen stoichiometry). Our experiments on inelastic neutron scattering show that the local charge inhomogeneity in CuO_2 planes of cuprate superconductors is a characteristic feature of the doping process, which depends neither on the method nor on the level of doping. Consequently, from the view point of spatial charge inhomogeneity, a smooth transition (crossover) takes place between underdoped and overdoped regions of the phase diagram for cuprates. This conclusion is in qualitative agreement with the results of recent investigations of high-energy longitudinal phonons in $YBa_2Cu_3O_x$ by the method of inelastic neutron scattering, indicating a considerable charge inhomogeneity even at the optimal doping level $x = 6.93$ [10]. However, according to our results, the charge distribution in CuO_2 planes changes for a doping level higher than optimal (this change is shown schematically in Fig. 3). In the framework of the PACT model, this effect is interpreted as a gradual suppression of the charge ordering existing in CuO_2 planes for $n < n_{opt}$.

ACKNOWLEDGMENTS

The authors express their gratitude to A. Furrer for fruitful discussions and to B. Goshchitskiĭ for his support of this research.

This work was supported financially by the Russian Foundation for Basic Research (project no. 00-02-17370); state contract nos. 107-19(00)-P-D01 and 107-1(00)-P-D01; the Russian State Program Supporting the Leading Scientific Schools (project no. 00-15-96581); INTAS (grant no. 99-00256); and the 6th Competition–Expertise of Young Scientists, Russian Academy of Sciences (1999) (grant no. 67).

REFERENCES

1. J. M. Tranquada, B. J. Sternlieb, J. D. Axe, *et al.*, Nature **375**, 561 (1995).
2. J. Tranquada, Physica B (Amsterdam) **241–243**, 745 (1997).
3. G. Aeppli, T. E. Mason, S. M. Hayden, *et al.*, Science **278**, 1432 (1997).
4. S.-W. Cheong, H. Y. Hwang, C. H. Chen, *et al.*, Phys. Rev. B **49**, 7088 (1994).
5. K. Yamada, C. H. Lee, Y. Endoch, *et al.*, Physica C (Amsterdam) **282–287**, 85 (1997).
6. V. Kataev, B. Rameer, A. Vavilov, *et al.*, Phys. Rev. B **58**, R11876 (1998).
7. A. Bianconi, A. Valetta, A. Perali, and N. L. Saini, Solid State Commun. **102**, 369 (1997).
8. N. L. Saini, A. Lanzara, A. Bianconi, and H. Oyanagi, Phys. Rev. B **58**, 11768 (1998).
9. L. Pintschovius and M. Braden, Phys. Rev. B **60**, R15039 (1999).
10. Y. Petrov, T. Egami, R. J. McQueeney, *et al.*, cond-mat/0003414.

11. T. Schneider and H. Keller, Phys. Rev. Lett. **86**, 4899 (2001).
12. J. Mesot and A. Furrer, J. Supercond. **10**, 623 (1997).
13. J. Mesot, P. Allenspach, U. Staub, *et al.*, Phys. Rev. B **47**, 6027 (1993).
14. J. Mesot, P. Allenspach, U. Staub, *et al.*, Phys. Rev. Lett. **70**, 865 (1993).
15. U. Staub, J. Mesot, M. Guillaume, *et al.*, Phys. Rev. B **50**, 4068 (1994).
16. D. Rubio Temprano, J. Mesot, S. Janssen, *et al.*, Phys. Rev. Lett. **84**, 1990 (2000).
17. Y. Tokura, J. B. Torrance, T. C. Huang, and A. I. Nazzal, Phys. Rev. B **38**, 7156 (1988).
18. G. Böttger, *Investigation of High-Temperature Superconductors of the R-Ba-Cu-O Family: Substitution Effects and Magnetic Properties*, Ph. D. Thesis (ETH, Zürich, 1996).
19. G. Böttger, J. Mesot, P. Fischer, and A. Furrer, Physica B (Amsterdam) **234–236**, 843 (1997).
20. J. L. Tallon, C. Bernhard, H. Shaked, *et al.*, Phys. Rev. B **51**, 12911 (1995).
21. N. Golosova, A. Podlesnyak, E. Mitberg, *et al.*, J. Phys.: Condens. Matter **14**, 1923 (2002).
22. W. Henggeler, M. Guillaume, P. Allenspach, *et al.*, J. Phys.: Condens. Matter **11**, 2921 (1999).
23. B. G. Wybourne, *Spectroscopic Properties of Rare Earths* (Interscience, New York, 1965), p. 352.
24. G. T. Trammell, Phys. Rev. B **92**, 1387 (1953).
25. A. Mirmelstein, A. Podlesnyak, V. Bobrovskii, and I. Zhdakhin, J. Phys.: Condens. Matter **11**, 7155 (1999).
26. V. Bobrovskii, A. Mirmelstein, A. Podlesnyak, and I. Zhdakhin, cond-mat/0012400.
27. K. N. R. Taylor and M. I. Darby, *Physics of Rare-Earth Solids* (Chapman and Hall, London, 1972; Mir, Moscow, 1974), p. 45.
28. J. Mesot, *Etude par diffusion neutronique de l'action du champ cristallin sur l'ion Er^{3+} dans les systèmes supraconducteurs à haute température critique*, Ph. D. Thesis (ETH, Zürich, 1992).
29. C. A. Morrison, in *Angular Momentum Theory Applied to Interactions in Solids*, Ed. by G. Berthier (Springer-Verlag, Berlin, 1988), Lecture Notes in Chemistry, Vol. 47, p. 122.
30. J. Tallon, Physica C (Amsterdam) **168**, 85 (1990).
31. D. Altermatt and I. D. Brown, Acta Crystallogr. B **41**, 240 (1985).
32. M. Merz, N. Nücker, P. Schweiss, *et al.*, Phys. Rev. Lett. **80**, 5192 (1998).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 1026–1034.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1194–1203.
Original English Text Copyright © 2002 by Gurevich, Muradov.

SOLIDS
Electronic Properties

The Theory of Shot Noise in the Space-Charge-Limited Diffusive Conduction Regime¶

V. L. Gurevich* and M. I. Muradov**

Solid State Physics Department, Ioffe Physicotechnical Institute, St. Petersburg, 194021 Russia

**e-mail: vadim.gurevich@pop.ioffe.rssi.ru*

***e-mail: mag.muradov@pop.ioffe.rssi.ru*

Received December 18, 2001

Abstract—As is well known, fluctuations from a stable stationary nonequilibrium state are described by the linearized inhomogeneous Boltzmann–Langevin equation. The stationary state itself can be described by the nonlinear Boltzmann equation. The ways of its linearization sometimes seem to be not unique. We argue that there is actually a unique way to obtain a linear equation for the fluctuations. As an example, we consider an analytical theory of nonequilibrium shot noise in a diffusive conductor under the space-charge-limited regime. Our approach is compared to that in [11]. We find some difference between the present theory and the approach in [11] and discuss a possible origin of the difference. We believe that it is related to the fundamentals of the theory of fluctuation phenomena in a nonequilibrium electron gas. © *2002 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

The present paper is devoted to the theory of shot noise in the space-charge-limited diffusive conduction regime. The motivation can be formulated as follows. It is well known that fluctuations from a stable stationary nonequilibrium state are described by the linearized inhomogeneous Boltzmann–Langevin equation (see, e.g., [1–7]). At the same time, the stationary state itself is described by the nonlinear Boltzmann equation. There are instances where the ways of linearization of the nonlinear Boltzmann equation seem to be not unique. We believe, however, that, in each such case, there is a unique way to obtain the linearized Boltzmann equation for the fluctuations, and we give general considerations to find this linearization and indicate it for the particular case treated in the present paper.

We develop a theory of nonequilibrium shot noise in a nondegenerate diffusive conductor under the space-charge-limited regime. This regime is extensively discussed in the literature (see, e.g., [8, 9]). The current noise under such a regime was recently studied by Monte Carlo simulation by González *et al*. [10]. Quite recently, the noise was analytically studied under the same conditions by Schomerus, Mishchenko, and Beenakker [11]. Their general finding was that, because of the Coulomb correlation between electrons, the shot noise is reduced below the classical Poisson value. The authors of both [10] and [11] came to the conclusion that under certain conditions, the suppression factor in the nondegenerate 3D case can be close to 1/3.

Later on, Nagaev [12] showed in a special example that, unlike the 1/3 noise reduction in degenerate systems, the noise suppression by the Coulomb interaction is nonuniversal in nondegenerate systems. The noise suppression in such systems may depend on the details of the electron scattering.

We agree with the conclusion in [10, 11] that the reduction of the shot noise power in nondegenerate diffusive conductors can sometimes be close to the value of 1/3 theoretically predicted for a three-dimensional degenerate electron gas. As mentioned above, we also arrive at some conclusions that may prove important for the general theory of fluctuations in nonequilibrium systems. As is well known, the fluctuation phenomena in nonequilibrium stable systems are described by a linearized Boltzmann equation. We use the example analyzed in detail in the present paper to show that the linearization must be performed with care. In particular, there is a difference between the analytical procedures used in [11] and in the present paper for the calculation of the shot noise power. We discuss the origin of this difference and its implications. Because the point leading to the discrepancy is very subtle, it demands a rather detailed analysis, which we perform in the present paper partly repeating the calculations in [11] with some modifications. Our starting point is the Boltzmann equation formulated for the description of the stationary state; it is then applied to the analysis of fluctuations.

2. BOLTZMANN EQUATIONS

We consider the simplest model, used in [11], for the diffusion-controlled and space-charge-limited trans-

¶This article was submitted by the authors in English.

port. As the starting point, we use the Boltzmann equation in the presence of an electric field,

$$\left(\frac{\partial}{\partial t} + \mathcal{J}_{\mathbf{p}}\right) f_{\mathbf{p}} = 0, \tag{2.1}$$

$$\mathcal{J}_{\mathbf{p}} f_{\mathbf{p}} \equiv \left(\mathbf{v}\frac{\partial}{\partial \mathbf{r}} + e\mathbf{E}\frac{\partial}{\partial \mathbf{p}} + I_{\mathbf{p}}\right) f_{\mathbf{p}}, \tag{2.2}$$

where we have introduced the collision integral $I_{\mathbf{p}}$ describing the electron scattering,

$$I_{\mathbf{p}} f_{\mathbf{p}} = \sum_{\mathbf{p}'} (W_{\mathbf{p}'\mathbf{p}} f_{\mathbf{p}} - W_{\mathbf{p}\mathbf{p}'} f_{\mathbf{p}'}) \tag{2.3}$$

(we deal with the nondegenerate statistics, and therefore $f_{\mathbf{p}} \ll 1$).

Splitting the distribution function into even and odd parts with respect to $\mathbf{p}$, we obtain

$$f^{\pm}_{\mathbf{p}} = \frac{1}{2}(f_{\mathbf{p}} \pm f_{-\mathbf{p}}).$$

We assume that the collision operator acting on the even (odd) part of the distribution function gives an even (odd) function. This can be the case either because of the central symmetry of the crystal itself and the scatterers or because of the possibility of using the Born approximation in calculating the scattering probability. The first split equation is

$$\frac{\partial f^{-}_{\mathbf{p}}}{\partial t} + \mathbf{v}\frac{\partial f^{+}_{\mathbf{p}}}{\partial \mathbf{r}} + e\mathbf{E}\frac{\partial f^{+}_{\mathbf{p}}}{\partial \mathbf{p}} = -I_{\mathbf{p}} f^{-}_{\mathbf{p}}. \tag{2.4}$$

Being interested in relatively small frequencies of fluctuations $\omega\tau_{\mathbf{p}} \ll 1$, where $\tau_{\mathbf{p}}$ is the characteristic value of $I^{-1}_{\mathbf{p}}$, we can neglect the time derivative and express $f^{-}_{\mathbf{p}}$ as

$$f^{-}_{\mathbf{p}} = -I^{-1}_{\mathbf{p}}\left(\mathbf{v}\frac{\partial f^{+}_{\mathbf{p}}}{\partial \mathbf{r}} + e\mathbf{E}\cdot\mathbf{v}\frac{\partial f^{+}_{\mathbf{p}}}{\partial \varepsilon_{\mathbf{p}}}\right). \tag{2.5}$$

Inserting this expression into the second split equation for $f^{+}_{\mathbf{p}} \approx f(\varepsilon, \mathbf{r}, t)$ and averaging over the constant-energy surface in the quasimomentum space, we arrive at

$$\nu(\varepsilon)\frac{\partial f}{\partial t} - \left(\frac{\partial}{\partial x_\alpha} + eE_\alpha\frac{\partial}{\partial \varepsilon}\right)\nu(\varepsilon)D_{\alpha\beta}(\varepsilon)\left(\frac{\partial}{\partial x_\beta} + eE_\beta\frac{\partial}{\partial \varepsilon}\right) f = -\sum_{\mathbf{p}} \delta(\varepsilon - \varepsilon_{\mathbf{p}}) I^{(\mathrm{inel})}_{\mathbf{p}} f, \tag{2.6}$$

where the term on the right-hand side describes the inelastic collisions, while the density of states $\nu(\varepsilon)$ and the diffusion tensor $D_{\alpha\beta}(\varepsilon)$ are defined as

$$\nu(\varepsilon)D_{\alpha\beta}(\varepsilon) = \sum_{\mathbf{p}} \delta(\varepsilon - \varepsilon_{\mathbf{p}}) v_\alpha I^{-1}_{\mathbf{p}} v_\beta, \qquad \nu(\varepsilon) = \sum_{\mathbf{p}} \delta(\varepsilon - \varepsilon_{\mathbf{p}}). \tag{2.7}$$

The electric field obeys the Poisson equation

$$\kappa\nabla\mathbf{E} = 4\pi e[n(\mathbf{r}, t) - n^{\mathrm{eq}}], \qquad n(\mathbf{r}, t) = \int_0^\infty d\varepsilon \nu(\varepsilon) f(\varepsilon, \mathbf{r}, t), \tag{2.8}$$

where κ is the dielectric susceptibility and n^{eq} is the equilibrium concentration (equal to the concentration of donors). In what follows, we neglect n^{eq} compared to the nonequilibrium concentration n.

The part of the distribution function contributing to the current consists of two terms that are proportional to the spatial and energy derivatives of $f(\varepsilon, \mathbf{r}, t)$, respectively,

$$j_\alpha = e\sum_{\mathbf{p}} \mathbf{v} f^{-}_{\mathbf{p}} = -e\nu(\varepsilon)D_{\alpha\beta}(\varepsilon)\left(\frac{\partial}{\partial x_\beta} + eE_\beta\frac{\partial}{\partial \varepsilon}\right) f. \tag{2.9}$$

We consider the case where $D\tau_\varepsilon \gg L^2$, where L is the sample length and τ_ε is the energy relaxation time (of the order $[I^{(\mathrm{inel})}_{\mathbf{p}}]^{-1}$). On the right-hand side of Eq. (2.6), we can then omit the term that describes the energy relaxation. Under the same conditions, we obtain the Boltzmann equation for the fluctuations of the distribution function (we remind the reader that here we consider low-frequency fluctuations with

$$\omega \ll I_{\mathbf{p}} \approx 1/\tau_{\mathbf{p}},$$

where $\tau_{\mathbf{p}}$ is the characteristic time of elastic collisions),

$$\left(\frac{\partial}{\partial x_\alpha} + eE_\alpha\frac{\partial}{\partial \varepsilon}\right)\delta j^\alpha_\omega + e\delta E^\alpha_\omega\frac{\partial}{\partial \varepsilon} j_\alpha = e y_\omega(\varepsilon, x), \tag{2.10}$$

$$\delta j^\alpha_\omega = e\sum_{\mathbf{p}} v_\alpha \delta f^{-}_{\mathbf{p}} = g^\alpha_\omega - e\nu(\varepsilon)D_{\alpha\beta}(\varepsilon) \times\left(\left[\frac{\partial}{\partial x_\beta} + eE_\beta\frac{\partial}{\partial \varepsilon}\right]\delta f_\omega + e\delta E^\beta_\omega\frac{\partial}{\partial \varepsilon} f\right), \tag{2.11}$$

and the source of the current fluctuations g_ω^α is related to the Langevin forces $y_\mathbf{p}^\omega$ as

$$g_\omega^\alpha = e\sum_\mathbf{p}\delta(\varepsilon-\varepsilon_\mathbf{p})v_\alpha I_\mathbf{p}^{-1}y_\mathbf{p}^\omega, \tag{2.12}$$

$$y_\omega(\varepsilon, x) = \sum_\mathbf{p}\delta(\varepsilon-\varepsilon_\mathbf{p})y_\mathbf{p}^\omega = 0. \tag{2.13}$$

The last equality is a consequence of the elasticity of scattering, which leads to the particle conservation within the constant-energy surface in the quasimomentum space.

The correlation function of the Langevin forces is well known [7],

$$\langle y_\mathbf{p}(\mathbf{r})y_{\mathbf{p}'}(\mathbf{r}')\rangle_\omega = (\mathcal{I}_\mathbf{p}+\mathcal{I}_{\mathbf{p}'})\delta_{\mathbf{r}\mathbf{r}'}\delta_{\mathbf{p}\mathbf{p}'}f_\mathbf{p}. \tag{2.14}$$

Integrating Eq. (2.10) over ε, we obtain the continuity equation

$$A\frac{d}{dx}\int_0^\infty d\varepsilon\delta j_\omega(\varepsilon, x) = \frac{d}{dx}\delta J_\omega(x) = 0 \tag{2.15}$$

which implies that the low-frequency current fluctuations are spatially homogeneous.

3. THE DISTRIBUTION FUNCTION

We consider a semiconductor with a uniform cross section A connecting two identical metallic electrodes. The length L of the sample is assumed to be much larger than the elastic scattering length l and much smaller than the inelastic one. We use the 1D versions of the Boltzmann equations describing the distribution function evolution along the dc current direction.

To obtain the stationary solution of Eq. (2.6) in the accepted approximation, we rewrite it as

$$\left(\frac{\partial}{\partial x}+eE\frac{\partial}{\partial\varepsilon}\right)j(\varepsilon, x) = \delta(x)j(\varepsilon). \tag{3.1}$$

We assume here that the current density at $x = 0$, $j(\varepsilon)$, is nonvanishing only for $\varepsilon > 0$. In the absence of tunneling, $j(\varepsilon)$ must have the property that

$$j(\varepsilon) \longrightarrow 0 \quad \text{as} \quad T \longrightarrow 0 \tag{3.2}$$

at the contact $x = 0$, with T being the temperature. This condition must be valid, irrespective of whether a Schottky barrier or an ohmic contact occurs. Evidently, the total current J given by Eq. (3.3) below must have the same property.

The solution of Eq. (3.1) is a function of the total energy $\mathscr{E}$,

$$\mathscr{E} = \varepsilon + U(x),$$

where

$$U(x) = e\varphi(x) - e\varphi(0).$$

It can be found using, e.g., the inverse differential operator

$$\frac{1}{\partial_x}\Phi(x) = \int_0^x d\xi\Phi(\xi).$$

We have

$$j(\varepsilon, x) = \frac{1}{\partial_x + eE(x)\partial_\varepsilon}\delta(x)j(\varepsilon)$$

$$= \exp[e\varphi(x)\partial_\varepsilon]\frac{1}{\partial_x}\exp[-e\varphi(x)\partial_\varepsilon]\delta(x)j(\varepsilon) = j(\mathscr{E}),$$

and $j(\varepsilon, x)$ takes nonzero values at a given x only if $\varepsilon > -U(x)$ ($\mathscr{E} \geq 0$). The total current through the sample is

$$J = A\int_0^\infty d\varepsilon j(\varepsilon, x)$$
$$= A\int_{-U(x)}^\infty d\varepsilon j[\varepsilon + U(x)] = A\int_0^\infty d\mathscr{E} j(\mathscr{E}). \tag{3.3}$$

From Eq. (2.9), we now obtain

$$f(\varepsilon, x) = -\frac{1}{\partial_x + eE(x)\partial_\varepsilon}\frac{j(\varepsilon, x)}{e\lambda(\varepsilon)} + f[\varepsilon + U(x)] \tag{3.4}$$

or

$$f(\varepsilon, x) = -j[\mathscr{E}]\int_0^x d\xi\frac{1}{e\lambda[\mathscr{E}-U(\xi)]} + f[\mathscr{E}], \tag{3.5}$$

where

$$\lambda(\varepsilon) \equiv \nu(\varepsilon)D(\varepsilon).$$

We have taken the boundary condition at the source into account. Equation (3.5) can be rewritten as

$$f[\mathscr{E}-U(x), x] = \left[f[\mathscr{E}-U(L)]\int_0^x d\xi\frac{1}{\lambda[\mathscr{E}-U(\xi)]}\right.$$
$$\left. + f(\mathscr{E})\int_x^L d\xi\frac{1}{\lambda[\mathscr{E}-U(\xi)]}\right]\left[\int_0^L d\xi\frac{1}{\lambda[\mathscr{E}-U(\xi)]}\right]^{-1}, \tag{3.6}$$

where $j(\varepsilon)$ is expressed through the difference of the distribution functions at $x = 0$ and $x = L$,

$$j(\mathscr{E})\int_0^L dx\frac{1}{e\lambda[\mathscr{E}-U(x)]} = f(\mathscr{E}) - f[\mathscr{E}-U(L)]. \tag{3.7}$$

An advantage of the form chosen for Eq. (3.6) is its physical transparency. The first term on the right-hand side gives the contribution of the right boundary, and the second term gives the contribution of the left boundary. The solution clearly demonstrates that the thermally excited carriers injected from the contact at $x = L$

make a negligible contribution to the distribution function $f[\mathscr{E} - U(x), x]$, because

$$f(\mathscr{E}) \gg f[\mathscr{E} - U(L)] \quad (\mathscr{E} \geq 0)$$

for the parameter $|U(L)|/k_BT$ is assumed to be large. Neglecting this term in our solution to Eq. (3.6), we arrive at the solution already obtained in [11] by assuming absorbing boundary conditions at the current drain.

4. THE FIELD DISTRIBUTION

We use the Poisson equation to determine the self-consistent electric field that can be expressed through the obtained distribution function. We consider the values of x such that $x > x_{\bar{\varepsilon}}$, where

$$-U(x_{\bar{\varepsilon}}) \gg \bar{\mathscr{E}} \sim k_BT,$$

$$\begin{aligned} -\frac{\kappa}{4\pi e^2}\frac{d^2U}{dx^2} &= \int_0^\infty d\mathscr{E}\,\nu[\mathscr{E} - U(x)]\,f[\mathscr{E} - U(x), x] \\ &= \int_0^\infty d\mathscr{E}\,\nu[\mathscr{E} - U(x)]\,j(\mathscr{E})\int_x^L \frac{d\xi}{e\lambda[\mathscr{E} - U(\xi)]} \\ &\approx \nu[-U(x)]\frac{J}{eA}\int_x^L \frac{d\xi}{\lambda[-U(\xi)]}. \end{aligned} \tag{4.1}$$

We finally obtain

$$-\frac{\kappa}{4\pi e^2}\frac{1}{\nu[-U(x)]}\frac{d^2U}{dx^2} = \frac{J}{eA}\int_x^L \frac{d\xi}{\lambda[-U(\xi)]}. \tag{4.2}$$

We now check that, for large x, this equation is consistent with the requirement of a uniform total current. Assuming

$$\nu(\varepsilon) = \nu_0\varepsilon^{d/2-1}$$

and

$$D(\varepsilon) = D_0\varepsilon^{s+1},$$

we integrate Eq. (2.9) over the transverse coordinates and energy, with the result

$$\begin{aligned} \frac{J}{A} &= -e\frac{d}{dx}\int_0^\infty d\varepsilon\,\nu(\varepsilon)D(\varepsilon)\,f(\varepsilon, x) \\ &+ \frac{eD_0\kappa(d+2s)}{16\pi}[-U(x)]^s\frac{d}{dx}E^2(x). \end{aligned} \tag{4.3}$$

We integrate the second term by parts and take into account that, at $x > x_{\bar{\mathscr{E}}}$, we can neglect $\mathscr{E}$ as compared to $|U(x)|$ and use Poisson equation (2.8). The first term in Eq. (4.3) can be simplified in the same way[1]

$$\begin{aligned} &\int_{-U(x)}^\infty d\varepsilon\,\nu(\varepsilon)D(\varepsilon)\,f(\varepsilon, x) \\ &= \int_0^\infty d\mathscr{E}\,\nu[\mathscr{E} - U(x)]D[\mathscr{E} - U(x)]\,f(\mathscr{E} - U(x), x) \\ &= D[-U(x)]\int_0^\infty d\mathscr{E}\,\nu[\mathscr{E} - U(x)]\,f(\mathscr{E} - U(x), x) \\ &= D[-U(x)]\frac{\kappa}{4\pi e}\frac{d}{dx}E. \end{aligned} \tag{4.4}$$

In the second equality, we used that $\mathscr{E} \ll |U(x)|$. Inserting Eq. (4.4) into Eq. (4.3), we obtain the simplified equation

$$\frac{4\pi|J|}{D_0\kappa A} = \frac{d}{dx}\left([-U]^{s+1}\frac{dE}{dx}\right) + |e|\frac{2s+d}{4}[-U]^s\frac{dE^2}{dx}. \tag{4.5}$$

It can be used to verify the self-consistency of our approach. Indeed, multiplying Eq. (4.2) by $U^{s+d/2}$ and taking the derivative, we arrive at Eq. (4.5) that was obtained from the equation for the current. A dimensionless version of Eq. (4.5) is

$$\chi^s\left(\frac{d-2}{2}\chi'\chi'' - \chi\chi'''\right) = 1, \tag{4.6}$$

where the dimensionless potential χ is related to φ by

$$\varphi = \left(\frac{4\pi|J|L^3}{D_0\kappa A|e|^{s+1}}\right)^{1/(s+2)}\chi(x/L). \tag{4.7}$$

5. THE CURRENT AND FIELD FLUCTUATIONS

In what follows, we consider the particular cases where

$$s = 0, \quad D(\varepsilon) = D_0\varepsilon;$$

$$s = -1/2, \quad D(\varepsilon) = D_0\varepsilon^{1/2};$$

and

$$s = 1/2, \quad D(\varepsilon) = D_0\varepsilon^{3/2}.$$

We begin by investigating the case of the energy-independent scattering time, $s = 0$. This case can be related to the scattering of electrons by neutral impurities, such as hydrogen-like shallow donor and acceptor states. The scattering is analogous to the scattering of electron by a hydrogen atom [13] (with the effective Bohr radius a_B). The scattering cross section turns out to be

[1] We note that, in view of Eq. (3.5), the distribution function $f(\varepsilon, x)$ takes nonzero values only for $\varepsilon > -U(x)$.

about $2\pi\hbar/pa_B$ times larger than the geometrical cross section πa_B^2 (that would result in an energy-independent scattering time).

In the case of defects with deep energy levels, we encounter a short-range scattering potential with the scattering length about the atomic length. The scattering cross section does not depend on the energy. As a result, the scattering rate is proportional to the electron density of states $\varepsilon^{1/2}$ and the diffusion coefficient $v^2\tau$ is proportional to $\varepsilon^{1/2}$, i.e., $s = -1/2$. (This is one of the main scattering mechanisms in metals because the scattering length is then determined by the screening radius, which is of the order of the interatomic distance.) The cases where $s = -1/2$ (which, in particular, describes elastic scattering by acoustic phonons) and $s = 1/2$ are discussed at the end of this section.

5.1. Energy-Independent Scattering Time

Integrating Eq. (2.11) over ε, we obtain

$$\frac{1}{A}(\delta J_\omega - G_\omega) = -e\frac{d}{dx}\int_{-U(x)}^{\infty} d\varepsilon \nu(\varepsilon) D(\varepsilon)\delta f_\omega(\varepsilon, x) + \frac{eD_0 d\kappa}{8\pi}\frac{d}{dx}E(x)\delta E_\omega(x). \quad (5.1)$$

We note that, because of Eq. (2.10), the Fourier transform of the current fluctuations δJ_ω is spatially homogeneous. Here, G_ω is the current fluctuation source integrated over the energy and transverse coordinates,

$$G_\omega(x) = \int_0^\infty d\varepsilon d\mathbf{r}_\perp g_\omega(\varepsilon, \mathbf{r}), \quad (5.2)$$

$$\langle G(x)G(x')\rangle_\omega = e^2\int_0^\infty d\varepsilon \int_0^\infty d\varepsilon' \sum_{\mathbf{p},\mathbf{p}'} \delta(\varepsilon - \varepsilon_\mathbf{p})\delta(\varepsilon' - \varepsilon_{\mathbf{p}'}) v_x v_x' \times \frac{1}{I_\mathbf{p}}\frac{1}{I_{\mathbf{p}'}}\int d\mathbf{r}_\perp d\mathbf{r}_\perp' \langle y_\mathbf{p} y_{\mathbf{p}'}\rangle_\omega. \quad (5.3)$$

The part of the distribution function that is odd with respect to $\mathbf{p} \longrightarrow -\mathbf{p}$ vanishes after inserting it into correlation function (2.14) of the Langevin forces and subsequently integrating over $\mathbf{p}$ and $\mathbf{p}'$. As a result, we are left with the integral of the even function

$$\langle G(x)G(x')\rangle_\omega = \delta_{xx'}\langle G^2(x)\rangle_\omega, \quad (5.4)$$

$$\langle G^2(x)\rangle_\omega = 2e^2 A\int_0^\infty d\varepsilon f(\varepsilon, x)\sum_\mathbf{p} \delta(\varepsilon - \varepsilon_\mathbf{p}) v_x \frac{1}{I_\mathbf{p}} v_x = 2e^2 A\int_0^\infty d\varepsilon \nu(\varepsilon) D(\varepsilon) f(\varepsilon, x). \quad (5.5)$$

The second term on the right-hand side of Eq. (5.1) can be simplified in the same way as Eq. (4.4),

$$\int_0^\infty d\varepsilon \nu(\varepsilon) D(\varepsilon)\delta f_\omega(\varepsilon, x) = D(-U(x))\frac{\kappa}{4\pi e}\frac{d}{dx}\delta E_\omega, \quad (5.6)$$

and we finally obtain the equation for δE_ω

$$\frac{d}{dx}\left(U(x)\frac{d}{dx}\delta E_\omega(x)\right) + e\frac{d}{2}\frac{d}{dx}E(x)\delta E_\omega(x) = \frac{4\pi}{AD_0\kappa}(\delta J_\omega - G_\omega). \quad (5.7)$$

In order to justify the simplification in Eq. (5.6), we now show that $\delta f_\omega(\varepsilon, x)$ is also a function taking nonzero values only at $\varepsilon > -U(x)$. Indeed, from Eq. (2.10) and Eq. (2.11), we can obtain the solutions

$$\delta j_\omega[\varepsilon - U(x), x] = \delta U_\omega(x)\frac{\partial}{\partial\varepsilon} j(\varepsilon) + \Delta j(\varepsilon)_\omega, \quad (5.8)$$

$$\delta f_\omega[\varepsilon - U(x), x] = \int_x^L d\xi e\delta E_\omega(\xi)\frac{\partial}{\partial\varepsilon} f[\varepsilon - U(\xi), \xi] - \int_x^L d\xi \frac{g_\omega[\varepsilon - U(\xi), \xi] - \delta j_\omega[\varepsilon - U(\xi), \xi]}{e\lambda[\varepsilon - U(\xi)]}, \quad (5.9)$$

which show that δf has the aforementioned property. Here, $\Delta j(\varepsilon)$ are the fluctuations of the current at the left boundary $x = 0$. The fluctuations of the distribution function $\Delta f(\varepsilon)$ at the right boundary are assumed to be zero. If we assume $\lambda(\varepsilon)$ to be a constant (independent of the energy), taking Eqs. (5.8) and (5.9) and the equation $\delta f_\omega(\varepsilon, 0) = 0$ into account, we immediately arrive at the result

$$\Delta J = \frac{1}{L}\int_0^L dx \int d\varepsilon g[\varepsilon - U(x), x] \quad (5.10)$$

obtained by Nagaev [12].

5.2. Comparison with the Approach in [4]

We now embark on setting forth the crucial point of the paper. Equation (5.7) does not coincide with the equation for the field fluctuations obtained in [11] by directly linearizing Eq. (4.5) for $s = 0$,

$$\frac{d}{dx}\left[\delta U_\omega(x)\frac{d}{dx}E(x)\right] + \frac{d}{dx}\left[U(x)\frac{d}{dx}\delta E_\omega(x)\right] + e\frac{d}{2}\frac{d}{dx}E(x)\delta E_\omega(x) = \frac{4\pi}{AD_0\kappa}(\delta J_\omega - G_\omega). \quad (5.11)$$

The origin of this discrepancy must be understood.

First, we temporarily adopt the scheme of [11] and reconsider Eq. (4.3) for the current

$$\frac{J}{A} = -e\frac{d}{dx}\int\limits_{-U(x)}^{\infty} d\varepsilon\nu(\varepsilon)D(\varepsilon)f(\varepsilon, x) + \frac{3}{2}D_0e^2E(x)\int\limits_{-U(x)}^{\infty} d\varepsilon\nu(\varepsilon)f(\varepsilon, x). \quad (5.12)$$

For the total current (the dc current plus fluctuations), the equation reads

$$\frac{J+\delta J-G}{A} = -e\frac{d}{dx}\int\limits_{-U(x)-\delta U(x)}^{\infty} d\varepsilon\nu(\varepsilon)D(\varepsilon) \times [f(\varepsilon, x)+\delta f(\varepsilon, x)] + \frac{3}{2}D_0e^2[E(x)+\delta E(x)] \times \int\limits_{-U(x)-\delta U(x)}^{\infty} d\varepsilon\nu(\varepsilon)[f(\varepsilon, x)+\delta f(\varepsilon, x)]. \quad (5.13)$$

Taking Eq. (5.12) into account, we obtain the linearized equation

$$\frac{\delta J-G}{A} = -e\frac{d}{dx}\int\limits_{-U(x)}^{\infty} d\varepsilon\nu(\varepsilon)D(\varepsilon)\delta f(\varepsilon, x) + \frac{3}{2}D_0e^2E(x)\left\{\int\limits_{-U(x)}^{\infty} d\varepsilon\nu(\varepsilon)\delta f(\varepsilon, x) + \delta U\frac{\delta}{\delta U(x)}\int\limits_{-U(x)}^{\infty} d\varepsilon\nu(\varepsilon)f(\varepsilon, x)\right\} + \frac{3}{2}D_0e^2\delta E(x)\int\limits_{-U(x)}^{\infty} d\varepsilon\nu(\varepsilon)f(\varepsilon, x) - e\frac{d}{dx}\delta U\frac{\delta}{\delta U(x)}\int\limits_{-U(x)}^{\infty} d\varepsilon\nu(\varepsilon)D(\varepsilon)f(\varepsilon, x). \quad (5.14)$$

If one linearized the Poisson equation in the spirit of [11], one would see that the term in the curly brackets in Eq. (5.14) would coincide with $(\kappa/4\pi e)(d\delta E/dx)$, and therefore,

$$\frac{\kappa}{4\pi e}\frac{d\delta E}{dx} = \int\limits_{-U(x)}^{\infty} d\varepsilon\nu(\varepsilon)\delta f(\varepsilon, x) + \delta U\frac{\delta}{\delta U(x)}\int\limits_{-U(x)}^{\infty} d\varepsilon\nu(\varepsilon)f(\varepsilon, x). \quad (5.15)$$

Simplifying the first, third, and fourth terms on the right-hand side of Eq. (5.14) with the help of Eq. (5.6) and inserting $(\kappa/4\pi e)(d\delta E/dx)$ instead of the term in the curly brackets, we arrive at

$$\frac{\delta J-G}{A} = e\frac{d}{dx}\left(D_0U(x)\frac{\kappa}{4\pi}\frac{d\delta E}{dx}\right) + \frac{3}{2}D_0e\frac{\kappa}{4\pi}\frac{d}{dx}E\delta E + eD_0\frac{\kappa}{4\pi}\frac{d}{dx}\delta U\frac{\delta}{\delta U}\left[U\frac{dE}{dx}\right]. \quad (5.16)$$

We can see that the last term on the right-hand side of this equation coincides with the first term on the left-hand side of Eq. (5.11). To avoid confusion, we note that we believe Eq. (5.15) to be also wrong. We have written it here only for the detailed comparison with the approach in [11]. We believe that the correct Poisson equation for the fluctuation field is

$$\frac{\kappa}{4\pi e}\frac{d\delta E}{dx} = \int\limits_{-U(x)}^{\infty} d\varepsilon\nu(\varepsilon)\delta f(\varepsilon, x). \quad (5.17)$$

In Eq. (4.3) for the dc current, we now add the terms that actually vanish because they are proportional to the integrals of the distribution function over ε with the upper limit $-U(x)$, whereas the distribution function $f(\varepsilon, x) = 0$ for $\varepsilon < -U(x)$. The point is that, when we calculate the fluctuations by the replacement

$$U(x) \longrightarrow U(x) + \delta U(x),$$

they give a nonvanishing result. We have

$$\frac{J}{A} = -e\frac{d}{dx}\int\limits_{0}^{-U(x)} d\varepsilon\nu(\varepsilon)D(\varepsilon)f(\varepsilon, x) - e\frac{d}{dx}\int\limits_{-U(x)}^{\infty} d\varepsilon\nu(\varepsilon)D(\varepsilon)f(\varepsilon, x) + \frac{3}{2}D_0e^2E(x)\int\limits_{0}^{-U(x)} d\varepsilon\nu(\varepsilon)f(\varepsilon, x) + \frac{3}{2}D_0e^2E(x)\int\limits_{-U(x)}^{\infty} d\varepsilon\nu(\varepsilon)f(\varepsilon, x). \quad (5.18)$$

Rewriting this equation for the total current, we obtain

$$\frac{J+\delta J-G}{A} = -e\frac{d}{dx}\int\limits_{0}^{-U(x)-\delta U} d\varepsilon\nu(\varepsilon)D(\varepsilon)[f(\varepsilon, x)+\delta f(\varepsilon, x)] - e\frac{d}{dx}\int\limits_{-U(x)-\delta U(x)}^{\infty} d\varepsilon\nu(\varepsilon)D(\varepsilon)[f(\varepsilon, x)+\delta f(\varepsilon, x)]$$

$$+\frac{3}{2}D_0e^2[E(x)+\delta E(x)] \tag{5.19}$$
$$\times\int_0^{-U(x)-\delta U(x)} d\varepsilon\nu(\varepsilon)[f(\varepsilon,x)+\delta f(\varepsilon,x)]$$
$$+\frac{3}{2}D_0e^2[E(x)+\delta E(x)]$$
$$\times\int_{-U(x)-\delta U(x)}^{\infty} d\varepsilon\nu(\varepsilon)[f(\varepsilon,x)+\delta f(\varepsilon,x)].$$

Linearizing this equation and using relations similar to

$$\delta U(x)\frac{\delta}{\delta U(x)}\int_{-U(x)}^{\infty} d\varepsilon\nu(\varepsilon)D(\varepsilon)f(\varepsilon,x) = -\int_{-U(x)}^{-U(x)-\delta U(x)} d\varepsilon\nu(\varepsilon)D(\varepsilon)f(\varepsilon,x), \tag{5.20}$$

we arrive at Eq. (5.1) that has been derived above. We see that the contributions to Eq. (5.19) that are linear in δU cancel because of the terms that vanish in the equation for the dc current but must be taken into account in considering fluctuations. This is why the linearization of Eq. (4.5) leads to Eq. (5.11) that we believe to be wrong because it does not take all the sources of fluctuations into account, or in other words, all the terms in Eq. (5.18) containing $U(x)$.

The solution to Eq. (5.7) with the boundary conditions

$$E(x)\delta E_\omega(x)\big|_{x\to 0}\longrightarrow 0,\qquad U(x)\frac{d}{dx}\delta E_\omega(x)\Big|_{x\to 0}\longrightarrow 0 \tag{5.21}$$

is given by

$$-\frac{AD_0\kappa}{4\pi}\delta E_\omega(x) = U^{d/2}(x)\times\left[C+\int_0^x\frac{d\xi}{U^{d/2+1}(\xi)}\int_0^\xi d\eta(\delta J_\omega - G(\eta)_\omega)\right], \tag{5.22}$$

where C is the integration constant. Requiring a nonfluctuating applied voltage

$$\int_0^L dx\delta E_\omega = 0,$$

we find from Eq. (5.22) that the constant is

$$C=\int_0^L dx\left(\frac{\psi(x)}{\psi(L)}-1\right)\frac{1}{U^{d/2+1}(x)}\int_0^x d\xi(\delta J_\omega - G_\omega(\xi)), \tag{5.23}$$

where

$$\psi(x) = \int_0^x d\xi U^{d/2}(\xi). \tag{5.24}$$

We now require

$$\frac{d}{dx}\delta E_\omega(x)\Big|_{x=L} = 0 \tag{5.25}$$

at the right boundary and obtain

$$\delta J_\omega = \frac{1}{Z}\int_0^L dx\Pi(x)G_\omega(x), \tag{5.26}$$

where

$$Z = L+\frac{dU'(L)U^{d/2}(L)}{2\psi(L)}\int_0^L dx\frac{x\psi(x)}{U^{d/2+1}(x)}, \tag{5.27}$$

$$\Pi(x) = 1+\frac{dU'(L)U^{d/2}(L)}{2\psi(L)}\int_x^L d\xi\frac{\psi(\xi)}{U^{d/2+1}(\xi)}. \tag{5.28}$$

The noise power P is then given by

$$P = \frac{2}{Z^2}\int_0^L dx\Pi^2(x)\langle G^2(x)\rangle_\omega. \tag{5.29}$$

In accordance with Eq. (5.5), we have

$$\langle G^2(x)\rangle_\omega = 2e^2A\int_0^\infty d\varepsilon\nu(\varepsilon)D(\varepsilon)f(\varepsilon,x) = 2e^2AD_0U(x)\frac{\kappa}{4\pi e}\frac{d^2U}{dx^2}. \tag{5.30}$$

We finally arrive at

$$P = \frac{4AD_0\kappa}{4\pi Z^2}\int_0^L dx\Pi^2(x)U(x)\frac{d^2U}{dx^2}. \tag{5.31}$$

The potential distribution can be found by following the method in [11], i.e., by solving Eq. (4.5) with boundary condition (4.2) at $x = L$. Using Eqs. (5.24), (5.27), (5.28), and (5.31), we calculate the suppression factor P/P_{Poisson}. For physically relevant different values of the dimensionality d, we obtain

$$\frac{P}{P_{\text{Poisson}}} = \begin{cases}0.3188 & \text{for } d=3,\\ 0.4512 & \text{for } d=2,\\ 0.682 & \text{for } d=1.\end{cases} \tag{5.32}$$

In this particular case, our results therefore differ from those calculated in [11] both analytically (which is of principal importance in our opinion) and numerically (although in this particular case the difference is

not great). Naturally, there is essentially no difference from the results calculated within an ensemble Monte Carlo scheme in [10].

5.3. Energy-Dependent Scattering Time

Here, we calculate the noise power for $s = \pm 1/2$ and $d = 3$. The equation for the fluctuations is

$$-\frac{4\pi}{\kappa D_0 A}(\delta J_\omega - G_\omega) = \frac{d}{dx}\left[(-U)^{s+1}\frac{d\delta E_\omega}{dx}\right] - e\frac{2s+d}{2}(-U)^s\frac{d}{dx}(E\delta E_\omega). \quad (5.33)$$

Introducing the dimensionless potential χ by Eq. (4.7) and the fluctuation of the field ΔE by

$$\delta E(x) = \frac{1}{L}\left(\frac{4\pi|J|L^3}{\kappa D_0 A|e|^{s+1}}\right)^{1/(s+2)}\Delta E\left(\frac{x}{L}\right), \quad (5.34)$$

we can rewrite Eq. (5.33) as

$$\Delta E'' + \left(1-\frac{d}{2}\right)\frac{\chi'}{\chi}\Delta E' - \left(s+\frac{d}{2}\right)\frac{\chi''}{\chi}\Delta E = \frac{1}{\chi^{s+1}}\frac{(G-\delta J)}{|J|}. \quad (5.35)$$

Setting $s = -1/2$ and $d = 3$, we obtain

$$\Delta E'' - \frac{1}{2}\frac{\chi'}{\chi}\Delta E' - \frac{\chi''}{\chi}\Delta E = \chi^{-1/2}\frac{(G-\delta J)}{|J|}. \quad (5.36)$$

This equation differs from that derived in [11], while the equation for the potential χ coincides with

$$\frac{1}{2\chi^{1/2}}\chi'\chi'' - \chi^{1/2}\chi''' = 1. \quad (5.37)$$

To calculate the Green's function of Eq. (5.36), we need the function $\psi_1(x)$ obeying the homogeneous equation

$$\psi_1'' - \frac{1}{2}\frac{\chi'}{\chi}\psi_1' - \frac{\chi''}{\chi}\psi_1 = 0 \quad (5.38)$$

and satisfying the boundary condition $\psi_1'\big|_{x=0} = 0$. The second function ψ_2 satisfying the boundary condition $\psi_2'\big|_{x=L} = 0$ can be expressed through the functions χ and ψ_1 as

$$\psi_2(x) = -\psi_1\left[\frac{\chi^{1/2}(1)}{\psi_1(1)\psi_1'(1)} + \int_x^1 d\xi\frac{\chi^{1/2}(\xi)}{\psi_1^2(\xi)}\right]. \quad (5.39)$$

The solution to Eq. (5.36) can be written using the Green's function

$$G(x, x') = \frac{1}{\chi^{1/2}(x')} \times [\theta(x-x')\psi_1(x')\psi_2(x) + \theta(x'-x)\psi_1(x)\psi_2(x')] \quad (5.40)$$

as

$$\Delta E = \int_0^1 dx' G(x, x')\frac{(G(x') - \delta J)}{\chi^{1/2}(x')|J|}. \quad (5.41)$$

Requiring a nonfluctuating applied voltage, we obtain

$$\delta J = \frac{1}{Z}\int_0^1 dx\frac{G(x)}{\chi(x)}\Pi(x), \quad (5.42)$$

where

$$\Pi(x) = \psi_1(x)\int_x^1 d\xi\psi_2(\xi) + \psi_2(x)\int_0^x d\xi\psi_1(\xi), \quad (5.43)$$

$$Z = \int_0^1 dx\frac{\Pi(x)}{\chi(x)}. \quad (5.44)$$

Expressing the correlation function $\langle G^2(x)\rangle$ through χ, we obtain the power supression factor

$$\frac{P}{P_{\text{Poisson}}} = \frac{2}{Z^2}\int_0^1 dx\frac{\chi''(x)}{\chi^{3/2}(x)}\Pi^2(x) \quad (5.45)$$

for the shot noise. We determine the potential χ following [11] and numerically find ψ_1 from Eq. (5.38). The functions ψ_2, Π and the constant Z can be found from Eqs. (5.39), (5.43), and (5.44). The suppression factor can be evaluated as

$$\frac{P}{P_{\text{Poisson}}} = 0.4257, \quad (5.46)$$

which is about 10% larger than the result obtained in [11]. The numerical simulation result in [10] for $s = -1/2$ is

$$\frac{P}{P_{\text{Poisson}}} = 0.42\text{–}0.44. \quad (5.47)$$

This interval is noticeably closer to the value given by Eq. (5.46) than the result in [11].

In the case where $s = 1/2$, the suppression factor can be evaluated as

$$\frac{P}{P_{\text{Poisson}}} = 0.1974, \quad (5.48)$$

which is slightly smaller than the result in [10].

6. CONCLUSION

In summary, we have developed an analytical theory of shot noise in a diffusive conductor under the space-charge-limited regime. We find that the present theory is different from the approach developed earlier and indicate a possible origin of the difference.

We now make several concluding remarks. The calculated nonequilibrium shot noise power in a nondegenerate diffusive semiconductor for two types of physically relevant elastic scattering mechanisms turned out to be very close to the ones obtained in numerical simulations by the authors of [10]. The computed noise suppression factor P/P_{Poisson} for the energy-independent scattering time is also rather close to the analytical results obtained earlier by Schomerus *et al.* [11]. However, for the energy-dependent scattering, the numerical difference between our results and those in [11] is considerable.

We clarify once more why the authors of [11] arrived at the equations that differ from ours. As an example, we take the Poisson equation. According to [11], one could write

$$n = \int_{-U(x)}^{\infty} d\varepsilon \nu(\varepsilon) f(\varepsilon, x), \tag{6.1}$$

where n and U are the exact total concentration and potential energy and f is the total distribution function (the mean value plus the fluctuating part). The linearization of this equation leads to the equations in [11]. The authors of [11] could have argued that, because the voltages in the reservoirs do not fluctuate and U is set to zero at the left boundary and because the total energy $\mathscr{E} = \varepsilon + U$ remains positive, the total distribution function is zero for $\varepsilon < -U$.

Our point is that Eq. (6.1) cannot be justifed for the total values of these variables including the stationary and fluctuating parts. This is readily seen from the fact that the fluctuating part of the distribution function itself implicitly depends on the mean value of the distribution function through the correlation function. One should bear in mind that an equation involving both the mean and the fluctuating quantities must be regarded symbolically. Indeed, such an equation is in fact equivalent to two equations, one for the mean values and the other for the fluctuating part. Regarded literally, it can lead to confusion. For example, analyzing the equation

$$\bar{n} + \delta n = \int_{-\bar{U}-\delta U} d\varepsilon \nu(\varepsilon)(\bar{f} + \delta f),$$

one can come to the wrong conclusion that the mean value $\bar{n}$ depends on such an average as $\overline{\delta U \delta f}$.

We add several words about the boundary conditions for the potential. The boundary conditions used here are not applicable within the length

$$R_V = \sqrt{\kappa V/4\pi e n(0)}$$

near the electrodes. Because the nonequilibrium noise power is a bulk property (we note, e.g., the integration over the coordinate in Eq. (5.45)), this approximation is justified since we assume that the sample length L is much greater than R_V.

Being interested in the analysis of the fluctuation phenomena in the simplest situation of the space-charge limited diffusive conduction regime, we have not taken the electron–electron collisions into account. These collisions can lead to an additional electron–electron correlation [7] that must be considered in analyzing a more general case.

ACKNOWLEDGMENTS

The authors are grateful to K.E. Nagaev for communicating to them his views concerning the role of boundary conditions at the contact between metal and semiconductor for the space-charge limited diffusive conduction.

This work was supported by the Russian Foundation for Basic Research (project no. 00-15-96748).

REFERENCES

1. E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon, Oxford, 1980).
2. V. L. Gurevich and R. Katilius, Zh. Éksp. Teor. Fiz. **49**, 1145 (1965) [Sov. Phys. JETP **22**, 796 (1966)].
3. Sh. M. Kogan and A. Ya. Shul'man, Fiz. Tverd. Tela (Leningrad) **9**, 2259 (1967) [Sov. Phys. Solid State **9**, 1771 (1968)].
4. S. V. Gantsevich, V. L. Gurevich, and R. Katilius, Fiz. Tverd. Tela (Leningrad) **11**, 308 (1969) [Sov. Phys. Solid State **11**, 247 (1969)].
5. Sh. M. Kogan and A. Ya. Shul'man, Zh. Éksp. Teor. Fiz. **56**, 862 (1969) [Sov. Phys. JETP **29**, 467 (1969)].
6. S. V. Gantsevich, V. L. Gurevich, and R. Katilius, Zh. Éksp. Teor. Fiz. **57**, 503 (1969) [Sov. Phys. JETP **30**, 276 (1970)].
7. S. V. Gantsevich, V. L. Gurevich, and R. Katilius, Riv. Nuovo Cimento **2**, 5 (1979).
8. M. A. Lampert and P. Mark, *Current Injection in Solids* (Academic, New York, 1970; Mir, Moscow, 1973).
9. N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals* (Clarendon, Oxford, 1940; Inostrannaya Literatura, Moscow, 1950).
10. T. González, J. González, J. Mateos, *et al.*, Phys. Rev. Lett. **80**, 2901 (1998).
11. R. Schomerus, E. G. Mishchenko, and C. W. E. Beenakker, Phys. Rev. B **60**, 5839 (1999).
12. K. E. Nagaev, Phys. Rev. Lett. **83**, 1267 (1999).
13. C. Erginsoy, Phys. Rev. **79**, 1013 (1950).

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 1035–1042.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1204–1212.
Original Russian Text Copyright © 2002 by Peschanskiĭ.

SOLIDS
Electronic Properties

Galvanomagnetic Phenomena in Organic Layered Conductors

V. G. Peschanskiĭ

Verkin Institute for Low Temperature Physics and Engineering, National Academy of Sciences of Ukraine, Kharkov, 61103 Ukraine
Karazin Kharkov National University, Kharkov, 61077 Ukraine
e-mail: vpeschansky@ilt.kharkov.ua

Abstract—Galvanomagnetic phenomena in organic conductors with a quasi-two-dimensional energy spectrum of an arbitrary form in the presence of several groups of charge carriers whose states belong to Fermi surface sheets with different topological structures are considered. The dependences of magnetoresistance, Shubnikov–de Haas oscillations, and Hall field on the intensity and orientation of a strong magnetic field with respect to the normal to layers **n** are analyzed for a Fermi surface consisting of a weakly corrugated cylinder and a plane weakly corrugated along the $p_z = \mathbf{pn}$ plane. *© 2002 MAIK "Nauka/Interperiodica".*

Interest in low-dimensional conductors of organic origin, which arose in the 1960s because of practical demands for new superconducting materials, does not seem to lessen. After the discovery of the superconducting state of fullerenes doped with various organic molecules (the transition to the superconducting state in these materials was observed at $T_c = 52$ K [1] and, more recently, at 120 K [2]), the preparation of organic high-T_c superconductors has become quite realistic. Alongside metal oxide superconductors, these materials will find wide applications in diverse fields of modern electronics. In addition, unusual behavior of organic compounds in the normal (not superconducting) state also attracts attention of researchers because of the presence of peculiar phase transitions in them and their very sensitive response to external magnetic fields. Reviews [3, 4] include more than three hundred publications concerned with studying electronic processes in organic conductors in the superconducting and normal states, especially in strong magnetic fields **H**. Under these conditions, experimental studies of magnetic susceptibility [5, 6] and transport phenomena can be used to solve the inverse problem of reproducing electronic energy spectra. This requires preparing fairly perfect samples with the free path of charge carriers l sufficiently long for frequency Ω of electron rotations in the magnetic field to be much higher than the frequency of its collisions $1/\tau$. This strong magnetic field condition ($\Omega\tau \gg 1$) was fulfilled in fields of the order of several dozen tesla units in radical-ion salts based on tetrathiafulvalene with layered structures. This is a likely cause of a considerable interest in studying electronic processes in organic conductors based on tetrathiafulvalene in strong magnetic fields, especially galvanomagnetic phenomena and quantum oscillation effects. These conductors are layered structures with sharp anisotropy of metal-type conduction, in which the electrical conductivity along layers is several orders of magnitude higher than that across layers.

In addition to quantum oscillation effects in tetrathiafulvalene salts observed as angle ϑ between the magnetic field and normal to layers **n** changed, periodic recurring narrow maxima were observed in the dependence of resistance to current across layers on $\tan\vartheta$ [7, 8]. This orientation effect arose because of the quasi-two-dimensional character of the energy spectrum of charge carriers and was absent in usual metals.

The sharp anisotropy of the conduction of layered conductors is likely to be related to a sharp anisotropy of the distribution of charge carrier velocities $\mathbf{v} = \partial\varepsilon(\mathbf{p})/\partial\mathbf{p}$ over the Fermi surface $\varepsilon(\mathbf{p}) = \varepsilon_F$. The energy of charge carriers,

$$\varepsilon(\mathbf{p}) = \sum_{n=0}^{\infty} \varepsilon_n(p_x, p_y)\cos\left\{\frac{anp_z}{\hbar} + \alpha_n(p_x, p_y)\right\},$$
$$\varepsilon_n(p_x, p_y) = \varepsilon_n(-p_x, -p_y), \quad (1)$$
$$\alpha_n(p_x, p_y) = -\alpha_n(-p_x, -p_y),$$

therefore weakly depends on the momentum projection $p_z = \mathbf{pn}$, and the $\max\{\varepsilon(\mathbf{p}) - \varepsilon_0(p_x, p_y)\}$ maximum function value equal to $\eta\varepsilon_F$ is much smaller than the ε_F Fermi energy.

The energy of an elementary excitation of a charge carrier in form (1) with arbitrary periodic functions $\varepsilon_n(p_x, p_y)$ and arbitrary phases $\alpha_n(p_x, p_y)$ satisfies translational symmetry and $\varepsilon(\mathbf{p})$ function parity, which follows from the Hermitian character of the Hamiltonian. Here, a is the distance between layers and $\hbar$ is the Planck's constant.

The quasi-two-dimensional character of the energy spectrum of charge carriers in organic layered conductors favors stronger manifestations of Shubnikov–de Haas [9] and de Haas–van Alphen [10] oscillation

effects in them because their formation involves a fairly large number of conduction electrons with Fermi energy ε_F. Shubnikov–de Haas oscillations of magnetoresistance in the $(BEDT\text{-}TTF)_2IBr_2$ and $(BEDT\text{-}TTF)_2I_3$ organic conductors discovered in 1988 [7, 8, 11–14] and then virtually in all tetrathiafulvalene salts and tetraselenotetracene halides [15] at various magnetic field orientations with respect to the layers are evidence that at least one Fermi surface sheet is a weakly corrugated cylinder with the openness direction along the p_z axis. All sections of this cylinder by the **pH** = const plane are closed if the ϑ angle between the **n** and **H** vectors equal to $(0, H\sin\vartheta, H\cos\vartheta)$ differs from $\pi/2$ and conduction electron energy levels are quantized. These energy levels should be determined by the Schrödinger equation

$$\hat{H}\left(\hat{\mathbf{P}} - \frac{e}{c}\mathbf{A}\right)\phi(x)\exp\left(\frac{i}{\hbar}yp_y + \frac{i}{\hbar}zp_z\right) = \varepsilon_N(p_y, p_z)\phi(x)\exp\left(\frac{i}{\hbar}yp_y + \frac{i}{\hbar}zp_z\right), \quad (2)$$

where e is the electron charge and c *i*s the velocity of light in the vacuum.

Here, we use the Landau calibration on the assumption that magnetic field vector-potential **A** only depends on coordinate x; that is, the $\hat{H}(\hat{\mathbf{P}} - e\mathbf{A}/c)$ Hamiltonian contains a single differential operator $\hat{p}_x$, and the p_y and p_z generalized momentum projections are good quantum numbers.

In the quasi-classical approximation, when the distance between quantized levels $\Delta\varepsilon_N = h\Omega$ is much smaller than $\eta\varepsilon_F$, the energy spectrum of charge carriers can be determined for an arbitrary form of Hamiltonian $\hat{H}$, which coincides with expression (1) for energy, if kinematic momentum **p** is replaced by $(\hat{\mathbf{P}} - e\mathbf{A}/c)$. If $\varepsilon_0(p_x, p_y)$ is a quadratic function of the conduction electron momentum, for instance,

$$\varepsilon_0(p_x, p_y) = \frac{p_x^2 + p_y^2}{2m},$$

and all the other $\varepsilon_n(p_x, p_y)$ functions with $n \geq 1$ on the Fermi surface equal constant values A_n, that is, do not depend on p_x and p_y, then the energy spectrum of charge carriers can easily be obtained at an arbitrary ratio between $\Delta\varepsilon_N$ and $\eta\varepsilon_F$. In the main approximation with respect to the a/r_H parameter, where r_H is the radius of curvature of the trajectory of charge carriers in the magnetic field, the $\phi(x)$ function is, as for free electrons, a Hermitian function with the argument shifted with respect to the center, and the energy of conduction electrons in the quantizing magnetic field has the form

$$\varepsilon_N(p_H) = \left(N + \frac{1}{2}\right)g\hbar\Omega + \sum_{n=1}^{\infty} A_n\cos\left\{\frac{anp_H}{\hbar\cos\vartheta}\right\} - \frac{1}{g^2\hbar\Omega} \times \sum_{n=1}^{\infty}(A_n n)^2\alpha\sin^2\left\{\frac{anp_H}{\hbar\cos\vartheta}\right\}, \quad (3)$$

where

$$\alpha = a^2\frac{eH\sin^2\vartheta}{\hbar c\cos\vartheta}, \quad \Omega = \frac{eH\cos\vartheta}{mc}, \quad N = 0, 1, 2, 3, \ldots,$$

and

$$g = \left\{1 - \frac{2}{\hbar\Omega}\sum_{n=1}^{\infty} n\alpha\cos\left(\frac{anp_H}{\hbar\cos\vartheta}\right)\right\}^{1/2}. \quad (4)$$

The cyclotron effective mass of conduction electrons with energy spectrum (3) has the form

$$m^* = \frac{m}{g\cos\vartheta} \quad (5)$$

and, because of the quasi-two-dimensional character of the electronic energy spectrum, is almost identical at all sections of the Fermi surface by the p_H = const plane. Generalized moments P_x and P_y, as in the quasi-classical approximation for an arbitrary spectrum of charge carriers, enter into the expression for energy only in the form of the integral of charge motion in magnetic field $P_z\cos\vartheta + P_y\sin\vartheta = p_H = \mathbf{pH}/H$.

If $h\Omega \ll \eta\varepsilon_F$ and the $\varepsilon_n(p_x, p_y)$ functions have an arbitrary form, the quantized energy spectrum of charge carriers is easy to determine with the use of the rule for quantizing areas [5, 6],

$$S(\varepsilon, p_H) = \frac{1}{\cos\vartheta}S_0(\varepsilon, p_H) = 2\pi\hbar\frac{eH}{c}\left(N + \frac{1}{2}\right), \quad (6)$$

where $S_0(\varepsilon, p_H) = \oint p_y\,dp_x$ is the projection onto the p_xp_y plane of the section of the isoenergy surface $S(\varepsilon, p_H)$ by the p_H = const plane.

Consider galvanomagnetic phenomena in a conductor in which the Fermi surface has the form of a weakly corrugated cylinder with an arbitrary cross section. The relation between current density and electric field E,

$$j_i = \mathrm{Sp}\{e\tilde{v}_i\hat{f}\} = \sigma_{ik}E_k, \quad (7)$$

can be found by solving the quantum kinetic equation for the $\hat{f} = \hat{f}_0 + \hat{f}_1$ statistical operator, where $\hat{f}_0$ is the statistical operator describing the equilibrium state of the system of electrons whose diagonal components coincide with the Fermi function of the distribution of charge carriers $\hat{f}_0^{NN} = f_0\{\varepsilon_N(P_y, P_z)\}$. The $\hat{f}_1$ operator describes the perturbation of the electronic system by

the electric field, and $\hat{\mathbf{v}}$ is the operator of the velocity of electrons.

In the linear approximation with respect to a weak electric field, the kinetic equation has the form [16]

$$\frac{i}{h}(\varepsilon_N - \varepsilon_{N'}) f_1^{NN'} + \hat{W}_{NN'}\{\hat{f}_1\} = e\mathbf{E}\mathbf{v}_{NN'}\frac{f_0(\varepsilon_N) - f_0(\varepsilon_{N'})}{\varepsilon_N - \varepsilon_{N'}}, \quad (8)$$

where $\hat{W}\{\hat{f}_1\}$ is the linear operator that describes scattering of conduction electrons on crystal defects and crystal lattice vibrations.

In a quantizing magnetic field, the density of conduction electron states has singularities periodically recurring as $1/H$ varies, which is the reason for oscillations of kinetic coefficients in a strong magnetic field. These singularities appreciably manifest themselves when the summation over conduction electrons is performed not only in Eq. (7) for current density, but also in the integral of collisions $\hat{W}$, which results in the appearance of terms oscillating with $1/H$ in the eigenvalues of the integral operator of conduction electron scattering. Taking them into account is very important in crossed fields $\mathbf{E} \perp \mathbf{H}$, when there are stationary conduction electron states in the collisionless limit ($\tau = \infty$). Ignoring the oscillating quantum addition to the relaxation time of charge carriers τ_{osc} then leads to a different, substantially underestimated amplitude of Shubnikov–de Haas oscillations [17, 18].

In a strong magnetic field, when not only $\Omega\tau \gg 1$, but also $\hbar\Omega/\varepsilon_F \geq \eta$, the periodic dependence of kinetic coefficients on $1/H$ is fairly complex. At $\hbar\Omega/\varepsilon_F \ll \eta$, this dependence, however, has a harmonic form and can easily be identified in the equation for current density with the use of the Poisson formula.

When current flows over layers, the Hall field at $\Omega\tau \gg 1$, as in a usual metal, substantially exceeds the electric field along the current if the magnetic field is appreciably deflected from the surface of layers. However, if current flows across layers, the galvanomagnetic characteristics (magnetoresistance and the Hall field) of a layered conductor show an essentially different behavior. At $\eta^2\Omega\tau \ll 1$, the Hall field is much lower than the electric field E_z along the normal to layers, and the electric field vector is therefore directed mainly along the electric current. The resistance to current across layers $\rho = \rho_{zz}$ is then determined fairly accurately by the single σ_{zz} component of the electrical conductivity tensor, and $\rho_{zz} = 1/\sigma_{zz}$, because the velocity of the drift of charge carriers along the z axis,

$$\bar{v}_z(p_H, \varepsilon_F) = \frac{\partial S/\partial p_H}{\partial S/\partial \varepsilon}\cos\vartheta, \quad (9)$$

is proportional to the η quasi-two-dimensionality parameter, and, at $(\pi/2 - \vartheta) \gg \eta$, the expansion in powers of η of the electrical conductivity tensor components σ_{ij} with at least one index coinciding with z begins with quadratic terms [19, 20].

In the quasi-classical approximation ($\hbar\Omega \ll \varepsilon_F\eta$), the σ_{zz} asymptote in a strong magnetic field at $\gamma = 1/\Omega\tau \ll 1$ can be calculated with the use of the solution to the kinetic equation in the τ approximation on the assumption that the $\hat{W}\{\hat{f}_1\}$ operator of collisions of electrons is an operator of multiplying $\hat{f}_1$ by the $\nu = 1/\tau + \nu_{osc}$ frequency of collisions. This results in loss of significance of low-importance numerical coefficients of the order of one, which does not change the amount of information about the form of the energy electronic spectrum contained in galvanomagnetic characteristics.

The Poisson formula yields the following asymptotic equation for σ_{zz} at $\gamma \ll 1$:

$$\sigma_{zz} = \frac{-2}{(2\pi\hbar)^3}\sum_{k=-\infty}^{\infty}\int d\varepsilon\frac{\partial f_0(\varepsilon)}{\partial\varepsilon} \times \int dp_H 2\pi m^* e^2 \bar{v}_z^2 \nu^{-1}\exp\{2\pi i k N(\varepsilon, p_H)\}, \quad (10)$$

where

$$m^* = \frac{1}{2\pi}\frac{\partial S(\varepsilon, p_H)}{\partial\varepsilon}$$

is the cyclotron effective mass.

The part of the electrical conductivity across layers that monotonically changes as the magnetic field varies,

$$\sigma_{zz}^{mon} = \frac{-2}{(2\pi\hbar)^3}\int d\varepsilon\frac{\partial f_0(\varepsilon)}{\partial\varepsilon}\int dp_H 2\pi m^* e^2 \bar{v}_z^2\tau, \quad (11)$$

substantially depends on the magnetic field orientation with respect to layers [20, 21] because, at some ϑ values, the velocity of conduction electrons along the normal to layers averaged over the $T = 2\pi/\Omega$ period,

$$\bar{v}_z(p_H, \varepsilon) = -\sum_{n=1}^{\infty}\frac{an}{\hbar}\sin\left(\frac{anp_H}{\hbar\cos\vartheta}\right)I_n\{\vartheta, p_H\}, \quad (12)$$

sharply decreases. Here,

$$I_n(\vartheta, p_H) = \frac{1}{T}\int_0^T dt\varepsilon_n\{p_x(t, p_H), p_y(t, p_H)\} \times \cos\left\{\frac{an}{\hbar}p_y(t, p_H)\tan\vartheta\right\}. \quad (13)$$

The ε = const and p_H = const electron orbits at $\eta \ll 1$ are almost indistinguishable, and the p_x and p_y functions and, therefore, $I_n\{\vartheta, p_H\}$ weakly depend on p_H, in accordance with the smallness of the quasi-two-dimensionality parameter η. It follows that the $\Delta I_n\{\vartheta, p_H\} = I_n\{\vartheta, p_H\} - I_n\{\vartheta\}$ difference should only be taken into

account in the vicinity of the ϑ_c angles at which $I_n\{\vartheta\}$ vanishes. If the $\varepsilon_n(p_x, p_y)$ functions fairly rapidly decrease as the n number increases, a minimum electrical conductivity value across layers should be expected at the $\vartheta = \vartheta_c$ values at which $I_1\{\vartheta_c\} = 0$. We must then take into account terms of higher orders in the expansion of σ_{zz}^{mon} in powers of γ and η,

$$\sigma_{zz}^{\text{mon}} = ae^2\tau m^* \frac{\cos\vartheta}{(2\pi\hbar)^4}\sum_{n=1}^{\infty} n^2 I_n^2(\vartheta) + \eta^2\sigma_0\{\eta^2\varphi_1(\vartheta) + \gamma^2\varphi_2(\vartheta)\}, \quad (14)$$

where the $\varphi_i(\vartheta)$ functions, which depend on the special form of the law of dispersion of charge carriers, are of the order of one and σ_0 coincides in order of magnitude with the electrical conductivity along layers in the absence of a magnetic field.

At $\tan\vartheta \gg 1$, rapidly oscillating functions in the integrand in (13) contribute to the integral mainly close to the neighborhood of the stationary phase points, where

$$\frac{\partial p_y}{\partial t} = \frac{eH}{c} v_x \cos\vartheta = 0.$$

The distance between these points is the electron orbit diameter D_p along the p_y axis. When ϑ varies, the $I_n\{\vartheta\}$ function, like $\tan\vartheta$, experiences periodic changes with the period

$$\Delta(\tan\vartheta) = \frac{2\pi\hbar}{anD_p}, \quad (15)$$

and experimental studies of the dependence of magnetoresistance at various strong magnetic field orientations make it possible to completely reproduce the shape of the Fermi surface [19–22].

Zeros of the function $I_n\{\vartheta\}$ with various n do not coincide, and the first term in Eq. (14) for σ_{zz}^{mon} therefore never vanishes. If we assume that, at $\vartheta = \vartheta_c$, when $I_1\{\vartheta_c\} = 0$, all the other terms in the sum over n in (14) are together proportional to η taken to a higher power than η^2, for instance, to $\eta^{2(1+q)}$, then, at $\eta^q \ll \gamma \ll 1$, magnetoresistance ρ_{zz} should increase proportionally to H^2, and the angular dependence of ρ_{zz} should be expected to contain fairly sharp peaks at $\vartheta = \vartheta_c$. In addition, $\rho_{\max}/\rho_{\min} \approx \gamma^{-2} \gg 1$. The experimentally observed magnetoresistance maxima of tetrathiafulvalene salts [7, 8] (also see [3, 4]), however, have the same order of magnitude as the magnetoresistance between the maxima. The small height of the maxima in the angular dependence of magnetoresistance is evidence that the $\varepsilon_n(p_x, p_y)$ functions decrease fairly slowly as n increases, and ignoring the $n \geq 2$ terms in the equation for the dispersion of charge carriers is incorrect.

Consider the equation for the part of the electrical conductivity across layers that oscillates with $1/H$

$$\sigma_{zz}^{\text{osc}} = \frac{2}{(2\pi\hbar)^3}\int d\varepsilon \frac{\partial f_0(\varepsilon)}{\partial\varepsilon} \times \int dp_H 2\pi m^* e^2 \bar{v}_z^2 \tau^2 \nu_{\text{osc}} - \frac{2}{(2\pi\hbar)^3} \times 2\text{Re}\sum_{k=1}^{\infty}\int d\varepsilon \frac{\partial f_0(\varepsilon)}{\partial\varepsilon}\int dp_H 2\pi m^* e^2 \bar{v}_z^2 \frac{\tau}{1+\tau\nu_{\text{osc}}} \times \exp\{2\pi i k N(\varepsilon, p_H)\}. \quad (16)$$

Here, not only $\bar{v}_z$ but also the $\exp\{2\pi i k N(\varepsilon, p_H)\}$ multipliers, which rapidly oscillate at $\hbar\Omega \ll \eta\varepsilon_F$, essentially depend on p_H. The first term on the right-hand side of (16) has the same order of magnitude as any terms in the sum over k with a small k index. As the significance of numerical coefficients of the order of one in the equation for σ_{zz} was lost in selecting the model of the collision integral, we will only analyze the sum over k in (16) ignoring the $\tau\nu_{\text{osc}}$ value, which is small compared with one, in the denominator.

The major contribution to σ_{zz}^{osc} is made by the integration of a small neighborhood of stationary phase points, which are found from the condition

$$dN(\varepsilon, p_H) = \left\{\frac{\partial S}{\partial\varepsilon}d\varepsilon + \frac{\partial S}{\partial p_H}dp_H\right\}\frac{c}{2\pi\hbar eH} = 0. \quad (17)$$

The stationarity condition when $S(\varepsilon, p_H)$ is almost constant is only fulfilled at

$$\frac{\partial\varepsilon}{\partial p_H} = -\frac{\partial S/\partial p_H}{\partial S/\partial\varepsilon}.$$

The conduction electron state density has singularities at $\partial\varepsilon/\partial p_H = 0$, that is, either on the extremal section of the isoenergy surface, when $\partial S/\partial p_H = 0$, or on a self-intersecting orbit, where

$$m^* = \frac{1}{2\pi}\frac{\partial S}{\partial\varepsilon} = \infty.$$

The Fermi surface has no orbits with self-intersections at $\eta\tan\vartheta \ll 1$, and the major contribution to σ_{zz}^{osc} is made by electron states from a small neighborhood of the extremal Fermi surface section by the $p_H = \text{const}$ plane.

Integrating (16) by parts yields σ_{zz}^{osc} in the form

$$\sigma_{zz}^{\text{osc}} = \frac{2}{(2\pi\hbar)^3}2\text{Re}\sum_{k=1}^{\infty}\int d\varepsilon \frac{\partial f_0(\varepsilon)}{\partial\varepsilon} e^2\tau \frac{eH\hbar}{ikc}(-1)^k$$

$$\times \int dp_H \exp\left\{\frac{ikcS(\varepsilon, p_H)}{eH\hbar}\right\} \tag{18}$$

$$\times \sum_{n=1}^{\infty}\left(\frac{an}{\hbar}\right)^2 \cos\left(\frac{anp_H}{\hbar\cos\vartheta}\right) I_n\{\vartheta\}.$$

Further calculations of σ_{zz}^{osc} by the stationary phase method pose no difficulties. As a result, the oscillating part of electrical conductivity across layers takes the form

$$\sigma_{zz}^{osc} = \sum_{k=1}^{\infty} e^2\tau\left(\frac{eH\hbar}{kc}\right)^{3/2}\frac{k\lambda}{\sinh(k\lambda)}$$

$$\times \frac{4\pi^{1/2}a^2(-1)^k}{(2\pi\hbar)^3\left\|\partial^2 S/\partial p_H^2\right\|^{1/2}\hbar^2}$$

$$\times\left\{\sum_{n=1}^{\infty}(2n-1)^2\sin\left\{\frac{\pi}{4}-\frac{kc(S_{max}-S_{min})}{2eH\hbar}\right\}\right. \tag{19}$$

$$\times\cos\left\{\frac{kc(S_{max}+S_{min})}{2eH\hbar}\right\} I_{2n-1}(\vartheta)$$

$$+\sum_{n=1}^{\infty}(2n)^2\cos\left\{\frac{\pi}{4}-\frac{kc(S_{max}-S_{min})}{2eH\hbar}\right\}$$

$$\left.\times\sin\left\{\frac{kc(S_{max}+S_{min})}{2eH\hbar}\right\} I_{2n}(\vartheta)\right\},$$

where $\lambda = 2\pi^2\theta/\hbar\Omega$ and θ is the temperature in energy units.

If the $\varepsilon_n(p_x, p_y)$ functions decrease fairly rapidly as n increases, a substantial decrease in σ_{zz}^{osc} is observed at the same $\vartheta = \vartheta_c$ angles as for σ_{zz}^{mon}, and the amplitude of Shubnikov–de Haas magnetoresistance ρ_{zz}^{osc} oscillations, which is proportional to $\sigma_{zz}^{osc}/(\sigma_{zz}^{mon})^2$, sharply increases at $\vartheta = \vartheta_c$. As a result, sharp maxima appear in the dependence of ρ_{zz}^{osc} on the angle ϑ between the **H** and **n** vectors. If $I_2(\vartheta)$ and $I_1(\vartheta)$ are values of approximately the same order, the maxima in the angular dependence of ρ_{zz}^{osc} are shifted with respect to the maxima of ρ_{zz}^{mon}. The shift value can be used to estimate the quasi-two-dimensionality parameter of the electronic energy spectrum.

Equations (10), (11) for σ_{zz}^{mon} and (16), (18), (19) for σ_{zz}^{osc} are valid at $\gamma \ll 1$, that is, when an electron manages to perform many revolutions in its orbit in the magnetic field during the mean free time and σ_{zz}^{osc} can be calculated ignoring quantum level broadening caused by collisions of charge carriers in the form of the so-called Dingle factor $\exp\{-\gamma\}$ [23].

The larger $\tan\vartheta$, the smaller the contribution to electrical conductivity of the functions in the integrand in (13) that rapidly oscillate at $\tan\vartheta \gg 1$, and the resistance of the sample grows proportionally to $\tan\vartheta$ as ϑ increases as long as $\gamma \ll 1$. Strongly prolate orbits, in which an electron fails to perform a complete revolution during the mean free time, appear as ϑ approaches $\pi/2$. At $\gamma = \gamma_0\cos\vartheta \geq 1$, σ_{zz} begins to grow as ϑ increases, and, at $\eta\tan\vartheta = 1$, when self-intersecting orbits appear, the magnetoresistance to current across layers reaches a minimum [24, 25]. The $\gamma_0 = 1/\Omega_0\tau$ parameter, where Ω_0 is the frequency of electron revolutions in the magnetic field at $\vartheta = 0$, can be assigned the meaning of the ratio of the time of electron movement in an orbit strongly prolate along the p_z axis by a distance of the order of $2\pi\hbar/a$. At $\eta\tan\vartheta > 1$, the resistance again increases as the ϑ angle widens and reaches a maximum at $\vartheta = \pi/2$. The ρ_{zz} maximum grows as the magnetic field increases first linearly at $\eta \ll \gamma_0 \ll 1$ and then proportionally to H^2 at $\eta \geq \gamma_0$ [21, 22]. The magnetoresistance along the normal to layers experiences unlimited growth at $\vartheta = \pi/2$ because $\overline{v}_z(p_H)$ vanishes in all Fermi surface sections, and we must retain expansion terms of higher order in small parameter γ_0 in the expression of σ_{zz}. The major contribution to σ_{zz}^{mon} is made by a small fraction of electrons with orbits close to the self-intersecting orbit, the period of revolution in which logarithmically diverges.

Azbel' showed that the self-intersecting orbit for σ_{zz}^{osc} stands out in the same sense as the extremal section of the Fermi surface [26, 27]. In some special situations, electrons from the vicinity of the self-intersecting orbit can form high-temperature oscillations [28]. Their contribution to σ_{zz}^{osc} is, however, negligibly small compared with the amplitude of oscillations for which electrons with the extremal Fermi surface section are responsible. A theoretical analysis of the fine effects predicted by Azbel' requires the collision integral in the quantum kinetic equation to be handled more accurately. For this reason, we do not pay due attention to these effects and interference oscillation effects, which arise when ν_{osc} is taken into account in the sum over k in (16) and which appear in the higher terms of the expansion in powers of the $\hbar\Omega/\eta\varepsilon_F$ small magnetic parameter. Although the amplitude of these oscillations is exceedingly small, the authors of [29] were able to observe them. Among the interference oscillations in a quasi-two-dimensional conductor, there are low-fre-

quency oscillations determined by the dimensionality of the maximum and minimum sections of the Fermi surface by the p_H = const plane. The amplitude of these magnetoresistance oscillations slowly decreases as the temperature increases as long as $\gamma \leq 1$ [29].

The Fermi surface of organic layered conductors containing more than a hundred atoms in the unit cell is fairly complex and can consist of elements with different topological structures. In addition to a corrugated cylinder, sheets in the form of planes weakly corrugated along the p_z axis are possible. For instance, the Fermi surface of $(BEDT\text{-}TTF)_2MHg(SCN)_4$ salts, where M is K, Rb, Tl, or NH_3, contains not only a weakly corrugated cylinder but also two quasi-one-dimensional sheets according to band calculations of the electronic energy spectrum [30]. These sheets are weakly corrugated planes, and their p_H = const sections are open at almost any magnetic field orientation. Although conduction electrons with a quasi-one-dimensional energy spectrum weakly respond to the presence of an external magnetic field, the existence of such a group of charge carriers can substantially change the magnetoresistance of a conductor. At magnetic field orientations at which open electronic trajectories appear in the momentum space, the magnetoresistance becomes sharply anisotropic even in the plane of layers. Experimental studies of the anisotropy of the transverse magnetoresistance ($\mathbf{j} \perp \mathbf{H}$) are a reliable method for determining the topology of the Fermi surface [31, 32] (also see the review by Novikov and Mal'tsev [33]).

Consider galvanomagnetic phenomena in an organic conductor whose Fermi surface consists of a weakly corrugated cylinder and sheets weakly corrugated along the p_z axis, and let the normal to the plane tangent to these sheets make angle ϕ with the p_x axis.

In the presence of several groups of charge carriers, all of them contribute to electric current. When charge carriers collide with the boundary of the sample, the transfer from one Fermi surface sheet to another is possible, which results in intermingling of electron states during the mean free time with respect to dissipative collisions in the bulk of the conductor. In thin conductors, such transfer processes substantially influence the magnetoresistance value [34]. In massive samples whose thickness is much larger than the mean free path of electrons, the total electrical conductivity can, however, fairly accurately be represented by the sum of the electrical conductivities of all groups of conduction electrons,

$$\sigma_{ik} = \sigma_{ik}^{(1)} + \sigma_{ik}^{(2)}, \tag{20}$$

where $\sigma_{ik}^{(1)}$ is the contribution of charge carriers whose state belongs to weakly corrugated plane Fermi surface sheets and $\sigma_{ik}^{(2)}$ is the contribution of conduction electrons that belong to the weakly corrugated cylinder.

Charge carriers involved in finite motion in the plane orthogonal to the magnetic field have a discrete-continuous energy spectrum. There is no quantization of the integral of motion p_H at ε = const for conduction electrons on open isoenergy surface sections. It follows that the major contribution to the oscillating part of electrical conductivity is made by charge carriers whose states belong to the Fermi surface sheet having the form of a weakly corrugated cylinder. It can, therefore, be taken that $\sigma_{zz}^{\mathrm{osc}} = \sigma_{zz}^{(2)\mathrm{osc}}$ to a fairly high accuracy.

If $\eta \tan\vartheta \ll 1$, there is no self-intersecting closed sections of the corrugated cylinder by the $p_H = \mathbf{p} \cdot \mathbf{H}/H$ = const plane, and the asymptotic equation for the $\sigma_{ik}^{\mathrm{mon}}$ matrix in a strong magnetic field takes a fairly simple form, namely,

$$\sigma_{ik}^{\mathrm{mon}} = \begin{pmatrix} \gamma^2 a_{xx} + \sigma_1 \cos^2\phi & \gamma a_{xy} + \sigma_1 \sin\phi\cos\phi & \gamma\eta^2 a_{xz} \\ \gamma a_{yx} + \sigma_1 \sin\phi\cos\phi & \sigma_1 \sin^2\phi + \gamma^2 a_{yy} + \sigma_{zz}\tan^2\vartheta & \gamma\eta^2 a_{yz} + \sigma_{zz}\tan\vartheta \\ \gamma\eta^2 a_{zx} & \gamma\eta^2 a_{zy} + \sigma_{zz}\tan\vartheta & \sigma_{zz} \end{pmatrix}, \tag{21}$$

where σ_1 is the largest contribution to electrical conductivity over layers at $H = 0$ made by charge carriers whose states belong to the open Fermi surface sheet in the form of a corrugated plane; the a_{ij} matrix components coincide with σ_0 in order of magnitude, that is, with the contribution to electrical conductivity over layers made by the other conduction electrons in the absence of a magnetic field.

The drift of charge carriers with open trajectories in the momentum space does not coincide with the magnetic field direction, and the fan of all possible drift directions covers the whole plane as p_H changes [31]. In the direction orthogonal to this plane, all charge carriers cannot, however, be displaced by a considerable distance. It is easy to see that, in the coordinate system in which one of the axes is directed along the magnetic field, electrical conductivity in the plane orthogonal to the magnetic field is sharply anisotropic at $\gamma \ll 1$, and one of the diagonal electrical conductivity tensor components is proportional to γ^2. As a result, the determinant of the $\sigma_{ik}^{\mathrm{mon}}$ matrix is also proportional to γ^2 no matter what the form of the energy spectrum of the

topological structure specified above. Simple calculations lead to the following asymptotic equations for the resistance over and across layers at $\eta \ll \cos\vartheta$ and $\gamma_0 \ll \cos\vartheta$:

$$\rho_{xx} = \frac{\sigma_1 \sin^2\phi \cos^2\vartheta + \gamma_0^2\sigma_0}{\gamma_0^2\sigma_0(\sigma_0 + \sigma_1)},$$
$$\rho_{yy} = \frac{\sigma_1 \cos^2\phi \cos^2\vartheta + \gamma_0^2\sigma_0}{\gamma_0^2\sigma_0(\sigma_0 + \sigma_1)}, \quad (22)$$

$$\rho_{zz} = \frac{1}{\sigma_{zz}} + \frac{\sigma_1 \cos^2\phi \sin^2\vartheta}{\gamma_0^2\sigma_0(\sigma_0 + \sigma_1)}. \quad (23)$$

Unimportant numerical coefficients of the order of one, which depend on the special form of the dispersion law for charge carriers, are, as previously, omitted in (22) and (23).

At $\eta \ll \gamma_0 \ll 1$, the electric field vector is, as previously, almost parallel to the current that flows along the normal to layers, and the $\rho_{zz}^{osc}/\rho_{zz}^{mon}$ ratio is of the same order of magnitude as when the Fermi surface is a single corrugated cylinder. At $\gamma_0 \leq \eta$, the presence of an additional Fermi surface sheet in the form of a corrugated surface, however, results in an unlimited increase in the resistance to current across layers, and the Hall field

$$E_x = \gamma_0^{-2}\frac{\sigma_1 \sin 2\phi \sin 2\vartheta}{\sigma_0(\sigma_0 + \sigma_1)} + \gamma_0^{-1}\frac{(\sigma_0 + \sigma_1 \sin^2\phi)\sin\vartheta}{\sigma_0(\sigma_0 + \sigma_1)}, \quad (24)$$

which is proportional to H^2 in the whole region of strong magnetic fields satisfying the $\gamma_0 \ll \cos\vartheta$ condition, is already comparable with E_z.

The existence of a group of charge carriers belonging to the Fermi surface sheet in the form of a corrugated plane is easy to determine from quadratic growth in magnetoresistance or the Hall field as H increases at various magnetic field orientations with respect to layers.

Equations (14)–(21) for the magnetoresistance and Hall field are valid for arbitrary Fermi surface corrugation in the $p_x p_y$ plane. The amplitude of corrugation or its absence in one of the Fermi surface sheets does not influence the character of the dependence of the galvanomagnetic characteristics of a two-dimensional conductor as long as $\eta \tan\theta < 1$. In the opposite limiting case of $\eta \tan\theta \gg 1$, the contribution of charge carriers with a quasi-one-dimensional energy spectrum to electrical conductivity across layers is, however, inversely proportional to H^2 in the whole region of strong magnetic fields $\gamma_0 \ll 1$, and the contribution of charge carriers with a quasi-two-dimensional spectrum is proportional to γ_0 at $\eta < \gamma_0 \ll 1$.

To summarize, the dependence of magnetoresistance and the Hall field on the strong magnetic field value at various field orientations with respect to layers allows us to answer the question to what extent the suggestion of the existence of a group of charge carriers with a quasi-one-dimensional energy spectrum in the organic charge transfer complexes $(BEDT\text{-}TTF)_2MHg(SCN)_4$ is justified.

ACKNOWLEDGMENTS

The author thanks the Foundation for International Cooperation (the Ukraine–NATO program, NATO grant PSF.CL G972846) for financial support.

REFERENCES

1. J. H. Schon, Ch. Kloe, and B. Batlogg, Nature **408**, 970 (2000).
2. J. H. Schon, Ch. Kloe, and B. Batlogg, Science **293**, 2432 (2001).
3. J. Wosnitza, *Fermi Surfaces of Low-Dimensional Organic Metals and Superconductors* (Springer-Verlag, Berlin, 1996).
4. J. Singleton, Rep. Prog. Phys. **63** (8), 1111 (2000).
5. L. Onsager, Philos. Mag. **43**, 1006 (1952).
6. I. M. Lifshits and A. M. Kosevich, Zh. Éksp. Teor. Fiz. **29**, 730 (1955) [Sov. Phys. JETP **2**, 636 (1956)].
7. M. V. Kartsovnik, V. N. Laukhin, V. I. Nizhankovskiĭ, and A. A. Ignat'ev, Pis'ma Zh. Éksp. Teor. Fiz. **47**, 302 (1988) [JETP Lett. **47**, 363 (1988)].
8. M. V. Kartsovnik, P. A. Kononovich, V. N. Laukhin, and I. F. Shchegolev, Pis'ma Zh. Éksp. Teor. Fiz. **48**, 498 (1988) [JETP Lett. **48**, 541 (1988)].
9. L. V. Schubnikov and W. J. de Haas, Leiden Comm. **207**, 210 (1930).
10. W. J. de Haas and P. M. van Alphen, Proc. R. Acad. Sci. Amsterdam **33**, 1106 (1930).
11. I. D. Parker, D. D. Pigram, R. H. Friend, *et al.*, Synth. Met. **27**, A387 (1988).
12. K. Oshima, T. Mori, H. Inokuchi, *et al.*, Phys. Rev. B **38**, 938 (1988).
13. N. Toyota, T. Sasaki, K. Murata, *et al.*, J. Phys. Soc. Jpn. **57**, 2616 (1988).
14. W. Kang, G. Montambaux, J. R. Cooper, *et al.*, Phys. Rev. Lett. **62**, 2559 (1989).
15. I. F. Shchegolev, P. A. Kononovich, M. V. Kartsovnik, *et al.*, Synth. Met. **39**, 357 (1990).
16. I. M. Lifshits, Zh. Éksp. Teor. Fiz. **32**, 1509 (1956) [Sov. Phys. JETP **5**, 1227 (1957)].
17. E. Adams and T. Holstein, J. Phys. Chem. Solids **10**, 254 (1959).
18. A. M. Kosevich and V. V. Andreev, Zh. Éksp. Teor. Fiz. **38**, 882 (1960) [Sov. Phys. JETP **11**, 637 (1960)].
19. V. G. Peschansky, J. A. Roldán López, and Toji Gnado Yao, J. Phys. I **1**, 1469 (1991).
20. V. G. Peschansky, Phys. Rep. **288**, 305 (1997).

21. V. G. Peschanskiĭ, Fiz. Nizk. Temp. **23**, 47 (1997) [Low Temp. Phys. **23**, 35 (1997)].
22. V. G. Peschanskiĭ, Zh. Éksp. Teor. Fiz. **112**, 618 (1997) [JETP **85**, 337 (1997)].
23. Yu. A. Bychkov, Zh. Éksp. Teor. Fiz. **39**, 1401 (1960) [Sov. Phys. JETP **12**, 971 (1961)].
24. V. G. Peschansky and M. V. Kartsovnik, Phys. Rev. B **60**, 11207 (1999).
25. V. G. Peschansky and M. V. Kartsovnik, J. Low Temp. Phys. **117**, 1717 (1999).
26. M. Ya. Azbel', Zh. Éksp. Teor. Fiz. **39**, 878 (1960) [Sov. Phys. JETP **12**, 608 (1961)].
27. M. Ya. Azbel', Zh. Éksp. Teor. Fiz. **39**, 1276 (1960) [Sov. Phys. JETP **12**, 891 (1961)].
28. M. Ya. Azbel', Zh. Éksp. Teor. Fiz. **45**, 2022 (1963) [Sov. Phys. JETP **18**, 1388 (1963)].
29. P. Grigoriev, M. Kartsovnik, W. Biberacher, and P. Wyder, cond-mat/0108352.
30. R. Rossenau, M. L. Doublet, E. Canadell, *et al.*, J. Phys. I **6** (12), 1527 (1996).
31. I. M. Lifshits and V. G. Peschanskiĭ, Zh. Éksp. Teor. Fiz. **35**, 1251 (1958) [Sov. Phys. JETP **8**, 875 (1959)].
32. I. M. Lifshits and V. G. Peschanskiĭ, Zh. Éksp. Teor. Fiz. **38**, 188 (1960) [Sov. Phys. JETP **11**, 137 (1960)].
33. S. P. Novikov and A. Ya. Mal'tsev, Usp. Fiz. Nauk **168**, 249 (1998) [Phys. Usp. **41**, 231 (1998)].
34. V. G. Peschansky, Sov. Sci. Rev. **16**, 1 (1992).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 853–861.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 995–1003.
Original English Text Copyright © 2002 by Huhtala, Volovik.

GRAVITATION, ASTROPHYSICS

Fermionic Microstates within the Painlevé–Gullstrand Black Hole¶

P. Huhtala[a] and G. E. Volovik[a, b, *]

[a]*Low Temperature Laboratory, Helsinki University of Technology FIN-02015 HUT, Finland*
[b]*Landau Institute for Theoretical Physics, Moscow, 117940 Russia*
**e-mail: volovik@boojum.hut.fi*

Received December 10, 2001

Abstract—We consider the quantum vacuum of a fermionic field in the presence of a black hole background as a possible candidate for the stabilized black hole. The stable vacuum state (as well as thermal equilibrium states at an arbitrary temperature) can exist if we use the Painlevé–Gullstrand description of the black hole and the superluminal dispersion of the particle spectrum at high energy, which is introduced in the free-falling frame. This choice is inspired by the analogy between the quantum vacuum and the ground state of quantum liquid, in which the event horizon for the low-energy fermionic quasiparticles can also arise. The quantum vacuum is characterized by the Fermi surface that appears behind the event horizon. We do not consider the back reaction, and therefore, there is no guarantee that the stable black hole exists. But if it does exist, the Fermi surface behind the horizon would be the necessary attribute of its vacuum state. We also consider the exact discrete spectrum of fermions inside the horizon, which allows us to discuss the problem of fermion zero modes. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In 1981, Unruh proposed to study black hole physics using its sonic analogue [1]. Originally suggested for classical liquids, this was later extended to quantum systems such as superfluids and Bose condensates [2–4]. The main advantage of the quantum liquids and gases is that, in many respects, they are similar to the quantum vacuum of fermionic and bosonic fields. This analogy forms a view on the quantum vacuum as a special type of condensed matter—the "ether"—where the physical laws that we have at present can arise emergently as the energy or temperature of the "ether" decreases [5]. A particular scenario of the emergent formation of the effective gravity together with gauge fields and chiral fermions can be found in the recent review paper [6].

According to the topology in the momentum space, there are three types (universality classes) of fermionic vacua. One of them has a trivial topology, and its fermionic excitations are therefore fully gapped (massive fermions). The other two have a nontrivial momentum-space topology characterized by certain topological invariants in the momentum space [6]. One of the two nontrivial universality classes contains systems with Fermi points; their excitations are chiral fermions, whose energy vanishes at points in the momentum space. Another class represents systems with a wider manifold of zeroes: their gapless fermionic excitations are concentrated in the vicinity of the 2D surface in momentum space, the Fermi surface. This class contains Fermi liquids.

Here, we discuss the properties of the quantum vacuum in the presence of the event horizon. We assume that, in the absence of the horizon, the fermionic vacuum belongs either to the trivial class (such as the Standard Model below the electroweak transition, where all fermions are massive) or to the class of Fermi points (such as the Standard Model above the electroweak transition, with its excitations being chiral massless fermions).

In the presence of a horizon, the region behind the horizon becomes the ergoregion: particles acquire negative energy there. In the true vacuum state, these negative-energy levels must be occupied, which means that the old vacuum must be reconstructed by filling these levels. We do not study the process of filling, which can be the smooth Hawking radiation process [7] or some other more violent process; we discuss the structure of the true vacuum state assuming that this state can be reached without destroying the horizon. In other words, we assume that the stable black hole can exist as a final ground state of the gravitational collapse. We find that, behind the horizon, the fermionic vacuum belongs to the class of the Fermi surface.

The main sources for the appearance of the Fermi surface originate in the following properties of the event horizon. First, the emergence of Planck physics in the vicinity of (and behind) the horizon. The event horizon serves as a magnifying glass through which the phenomena at the Planck length scale could be visual-

¶This article was submitted by the authors in English.

ized. At some scales, the Lorentz invariance—a property of low-energy physics—inevitably becomes invalid and deviations from the linear (relativistic) spectrum become important. This violation of the Lorentz invariance is now popular in the literature [1, 9–13]. It leads to either subluminal or superluminal propagation at high energy, e.g.,

$$E^2(p) = c^2p^2(1 \pm p^2/p_p^2),$$

where p_p is the Planck momentum. In accordance with the condensed matter analogy, we assume that high-energy (quasi)particles are superluminal, i.e., the sign is plus. Because of the superluminal dispersion, there is a bottom in the Dirac sea, and the process of filling the negative-energy levels is therefore limited. When all of these levels are occupied, we come to a global vacuum state (or the global thermodynamical equilibrium with a positive heat capacity, if the temperature is finite). Thus, the superluminal dispersion of the particle energy gives rise to the energetic stability of the vacuum in the presence of a black hole.

The second important consequence of the event horizon, due to which the vacuum belongs to the class of systems with the Fermi surface, is that the horizon violates the time reversal symmetry of the system: the incoming and outgoing particles have different trajectories. In condensed matter, the appearance of the Fermi surface due to the violation of the time reversal symmetry is a typical phenomenon (see, e.g., [8] and also Section 12.4 in [6]).

In [4, 14], a stable black hole that exhibits a finite positive heat capacity, an arbitrary temperature, and no Hawking radiation is also considered. However, it is assumed there that the time reversal symmetry is not broken in the final state (or is actually restored in the final state). The existence of such a stable black hole with unbroken time reversal symmetry is also supported by the condensed matter analogies [4, 15, 16], in which stable infinite-redshift surfaces arise. An example of the infinite-redshift surface with no time reversal symmetry breaking is also provided by the extremal black hole, whose condensed matter analogue is discussed in Section 12.6 of review [6]. In all these examples, the Fermi surface does not appear. The black hole ground states with time reversal symmetry are in some sense exceptional (in the same manner as the extremal black hole), and we do not discuss them here.

2. STATIONARY METRIC WITH EXPLICITLY VIOLATED TIME REVERSAL SYMMETRY

The vacuum can be well defined only if the metric is stationary. In general relativity, the stationary metric for the black hole is provided in the Painlevé–Gullstrand spacetime [17]. The line element of the Painlevé–Gullstrand metric is

$$\begin{aligned} ds^2 &= -c^2dt^2 + (d\mathbf{r} - \mathbf{v}dt)^2 \\ &= -(c^2 - v^2)dt^2 - 2\mathbf{v}d\mathbf{r}dt + d\mathbf{r}^2, \end{aligned} \tag{1}$$

where

$$\mathbf{v}(\mathbf{r}) = \pm\hat{\mathbf{r}}c\sqrt{\frac{r_h}{r}}, \quad r_h = \frac{2MG}{c^2}. \tag{2}$$

Here, M is the mass of the hole, r_h is the radius of the horizon, and G is the Newton gravitational constant; the minus sign in Eq. (2) gives the metric for the black hole, while the plus sign characterizes the white hole. The time reversal operation $t \longrightarrow -t$ transforms the black hole into the white whole. The stationary property of this metric and the fact that it describes the spacetime in both the exterior and the interior regions are very attractive features that were explored starting from [18] (see [19–21]; an extension of the Painlevé–Gullstrand spacetime to the rotating black hole can be found in [22]).

In the case of the black hole, the field $\mathbf{v}(\mathbf{r})$ has a simple interpretation: it is the velocity of the observer who freely falls along the radius towards the center of the black hole with zero initial velocity at infinity. The motion of the observer obeys the Newtonian laws all the way through the horizon,

$$\frac{d^2r}{dt^2} = -\frac{GM}{r^2}, \tag{3}$$

and his velocity is therefore given by

$$\mathbf{v}(\mathbf{r}) \equiv \frac{d\mathbf{r}}{dt} = -\hat{\mathbf{r}}\sqrt{\frac{2GM}{r}}. \tag{4}$$

The time coordinate t is the local proper time for the observer who drags the inertial coordinate frame with him.

As was first noticed by Unruh [1], the effective metric of type (1) is experienced by quasiparticles propagating in moving fluids. The field $\mathbf{v}(\mathbf{r})$ is then the velocity field of the liquid, and c is the "maximum attainable velocity" of quasiparticles in the low-energy limit, for example, the speed of sound in the case of phonons (see also [23–26, 6]). The horizon could be produced in liquids when the flow velocity becomes greater than c. The black hole and the white hole can be reproduced by the liquid flowing radially inward and outward, respectively. This is an explicit realization of the time reversal symmetry breaking by a flowing liquid: the time reversal operation reverses the direction of the flow of the "vacuum,"

$$\mathcal{T}\mathbf{v}(\mathbf{r}) = -\mathbf{v}(\mathbf{r}).$$

This Painlevé–Gullstrand spacetime, although not static, is stationary. That is why the energy $\tilde{E}$ of a

(quasi)particle in this spacetime is determined in both the exterior and the interior regions. It can be obtained as the solution to the equation

$$g^{\mu\nu}p_\mu p_\nu + m^2 = 0$$

with $p_0 = -\tilde{E}$, which gives

$$\tilde{E}(\mathbf{p}) = E(p) + \mathbf{p}\cdot\mathbf{v}(\mathbf{r}), \tag{5}$$

where $E(p)$ is the energy of the particle in the free-falling frame,

$$E^2(p) = p^2c^2 + m^2. \tag{6}$$

For the "sonic" black hole, it is the energy of the quasiparticle in the frame comoving with the superfluid vacuum.

We now consider a massless (quasi)particle moving in the radial direction from the black hole horizon to infinity; i.e., with a positive radial momentum p_r. Because the metric is stationary, the energy of a particle in the Painlevé–Gullstrand frame (or of a quasiparticle in the laboratory frame) is conserved and we have $\tilde{E} = \text{const}$. Its energy in the free-falling (superfluid comoving) frame is then given by

$$E(p) = cp_r = \frac{\tilde{E}}{1 + v(r)/c} = \frac{\tilde{E}}{1 - \sqrt{r_h/r}}. \tag{7}$$

This energy, which is very big near the horizon, decreases as the (quasi)particle moves away from the horizon. This is the gravitational red shift superimposed on the Doppler effect [27], because the emitter is freely falling with the velocity $v = v_s(r)$. The frequency of the spectral line measured by the observer at infinity is

$$\tilde{\omega} = \omega\sqrt{-g_{00}}\frac{\sqrt{1 - v^2/c^2}}{1 - v/c} = \omega\left(1 - \sqrt{\frac{r_h}{r}}\right), \tag{8}$$

where ω is the nominal frequency of this line. The surface $r = r_h$ is the infinite redshift surface, and the energy in Eq. (7) diverges there. This means that, if we observe particles coming to us from a very close vicinity of the horizon, these outgoing particles originally had a huge energy approaching the Planck energy scale. The event horizon can therefore serve as a magnifying glass that allows us to see what happens at the Planck length scale. At some point, the low-energy relativistic approximation inevitably becomes invalid and the Lorentz invariance is violated.

In quantum liquids, the nonlinear dispersion enters the velocity-independent energy $E(p)$ in the superfluid comoving frame. Taking the analogy with quantum liquids into account, we assume that, in our vacuum, the Planck physics also enters the energy in the free-falling frame. The energy spectrum of particles is therefore given by Eq. (5), where

$$E^2(p) = m^2 + p^2c^2\left(1 \pm \frac{p^2}{p_p^2}\right). \tag{9}$$

As for the incoming massless particle, its radial momentum $p_r < 0$, and its energy in the comoving frame is given by

$$E(p) = -cp_r = \frac{\tilde{E}}{1 - v(r)/c} = \frac{\tilde{E}}{1 + \sqrt{r_h/r}}. \tag{10}$$

It has no pathology at the horizon: the observer falling freely across the horizon sees no inconveniences when he crosses the horizon, and the Planck physics is therefore not evoked here.

The pathology reappears when one tries to construct the thermal global equilibrium state (or the vacuum state) in the presence of a horizon. In the global equilibrium, according to the Tolman law, the temperature measured by an observer in the comoving frame diverges at the horizon,

$$T(\mathbf{r}) = \frac{T_{\text{Tolman}}}{\sqrt{-g_{00}(\mathbf{r})}} = \frac{T_{\text{Tolman}}}{\sqrt{1 - v^2/c^2}}. \tag{11}$$

At some point, this temperature again becomes so high that the Planck physics becomes relevant. In the presence of a horizon, the global equilibrium is possible only for the superluminal dispersion, i.e., for the plus sign in Eq. (9). The reason is as follows. Behind the horizon, at $r < r_h$, the frame-dragging velocity exceeds the speed of light. In the relativistic domain, this implies that the radial coordinate r becomes timelike, because a (quasi)particle can move along the r coordinate in only one direction behind the horizon, towards the singularity. However, with the plus sign for the energy spectrum in Eq. (9), the (quasi)particles can go back and forth even behind the horizon. The spacelike nature of the r coordinate is therefore restored by the superluminal dispersion and the global equilibrium becomes possible.

Finally, the condensed matter analogue of the formation of quantum field theory as an emergent phenomenon at low energy suggests that our vacuum is fermionic, while all the bosonic degrees of freedom can be obtained as collective modes of the fermionic vacuum. It is the Pauli principle for fermions that allows us to construct a stable vacuum in the presence of a horizon. Thus, there are three main necessary conditions for the existence of a stable vacuum with the broken time reversal symmetry in the presence of a black hole: the vacuum is fermionic, its fermionic excitations have superluminal dispersion, and the black hole is described by the Painlevé–Gullstrand metric. All the three conditions are motivated by the quantum liquid similarities.

3. THE DIRAC EQUATION IN THE PAINLEVÉ–GULLSTRAND METRIC

In [28], fermions were considered in the semiclassical approximation. Here, we extend this analysis to the exact quantum-mechanical one. In the presence of a nontrivial gravitational background, fermions are described by the tetrad formalism. Here, we follow [29]. The metric $g_{\mu\nu}$ can be written in terms of e^a_μ as

$$g_{\mu\nu} = e^a_\mu e^b_\nu \eta_{ab}, \tag{12}$$

where $\eta^{ab} = \mathrm{diag}(-1, 1, 1, 1)$. The Dirac equation in a curved spacetime is

$$\begin{aligned} &(i\gamma^a E^\mu_a D_\mu - m)\Psi = 0, \\ &D_\mu = \partial_\mu + \frac{1}{4}\omega_{\mu;\,ab}\gamma^a\gamma^b, \end{aligned} \tag{13}$$

where the dual tetrad field E^ν_b obeys

$$g_{\mu\nu} = e^a_\mu e^b_\nu \eta_{ab}, \quad E^\mu_a e^a_\alpha = \delta^\mu_\nu, \quad E^\mu_a E^\nu_b \eta^{ab} = g^{\mu\nu}, \tag{14}$$

$$e^a_\mu = g_{\mu\nu}\eta^{ab}E^\nu_b, \quad e_{\nu b} = e^a_\nu \eta_{ab} = g_{\mu\nu}E^\mu_b, \tag{15}$$

and the torsion field is

$$\begin{aligned} \omega_{\mu;\,ab} &= E^\nu_a \eta_{bc}\nabla_\mu e^c_\nu = E^\nu_a \nabla_\mu (g_{\nu\alpha}E^\alpha_b) \\ &= E^\nu_a \nabla_\mu e_{\nu b} = E^\nu_a(\partial_\mu e_{\nu b} - \Gamma^\gamma_{\mu\nu} e_{\gamma b}). \end{aligned} \tag{16}$$

The vielbeins corresponding to the general "flow" metric in Eq. (1) are

$$e^a_\mu = \delta^a_\mu + \tilde{e}^a_\mu, \quad \tilde{e}^a_\mu = v^i \delta^a_i \delta^0_\mu. \tag{17}$$

The only nonzero correction to the tetrad field δ^a_μ for Minkowski spacetime is

$$\tilde{e}^i_0 = v^i \neq 0.$$

For the Painlevé–Gullstrand metric of the black hole in spherical coordinates, we have

$$\begin{aligned} e^0_\mu &= (1, 0, 0, 0), \quad e^1_\mu = (v, 1, 0, 0), \\ e^2_\mu &= (0, 0, r, 0), \quad e^3_\mu = (0, 0, 0, r\sin\theta), \end{aligned} \tag{18}$$

where $v(r) = -r^{-1/2}$, assuming that $c = r_h = 1$.

The violation of the Lorentz invariance at high energy can be introduced by adding a nonlinear γ_5 term that leads to the superluminal dispersion. As a result, we obtain the Dirac equation in the Painlevé–Gullstrand metric [22], which is now modified by a non-Lorentzian term,

$$i\partial_t \Psi = -ic\alpha^i \partial_i \Psi + m\gamma_0 \Psi + H_p \Psi + H_g \Psi. \tag{19}$$

Here, H_p and H_g are the respective Hamiltonians coming from the Planck physics and from the gravitational field,

$$H_p = -\frac{c}{p_p}\gamma_5 \partial_i^2, \quad H_g = ic\sqrt{\frac{r_h}{r}}\left(\frac{3}{4r} + \partial_r\right). \tag{20}$$

The γ matrices that we use are given by

$$\alpha^i = \begin{pmatrix} 0 & \sigma^i \\ \sigma^i & 0 \end{pmatrix}, \quad \gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{21}$$

and

$$\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{22}$$

After the multiplication by $r_h/\hbar c$, we obtain a dimensionless form and write $\hbar = c = r_h = 1$ and $p_0 = p_p r_h/\hbar \gg 1$.

4. EIGENSTATES OF FERMIONS IN THE PAINLEVÉ–GULLSTRAND BLACK HOLE

Because ∂_t is a timelike Killing vector in the Painlevé–Gullstrand black hole, the energy $\tilde{E}$ is a well-defined quantity and the variables t and $\mathbf{r}$ can be separated by writing

$$\Psi = \begin{pmatrix} \phi(\mathbf{r}) \\ \chi(\mathbf{r}) \end{pmatrix} e^{-i\tilde{E}t}. \tag{23}$$

The $\mathbf{r}$ equations are now given by

$$\begin{aligned} \tilde{E}\phi &= \sigma\cdot\mathbf{p}\chi + m\phi - i\frac{1}{p_0}p^2\chi + H_g\phi, \\ \tilde{E}\chi &= \sigma\cdot\mathbf{p}\phi - m\chi + i\frac{1}{p_0}p^2\phi + H_g\chi, \end{aligned} \tag{24}$$

where $p_i = -i\partial_i$. Using spherical symmetry, we introduce spherical harmonics in the standard way. These are eigenstates of the operators $\mathbf{J}^2$ and J_z, where $\mathbf{J}$ is the total angular momentum,

$$J_i = L_i + S_i = L_i + \frac{1}{2}\begin{pmatrix} \sigma_i & 0 \\ 0 & \sigma_i \end{pmatrix}, \tag{25}$$

and L_i is the orbital angular momentum operator in R^3. Because we are interested in the states with high momenta $J \sim p_0 \gg mr_h/\hbar$, we can neglect the mass term. We then obtain the ansatz

$$\begin{aligned} \phi_{J,J_3} &= \frac{1}{2r} \\ &\times ((f^+(r) + f^-(r))\Omega_l + (f^+(r) - f^-(r))\Omega_{l+1}), \end{aligned} \tag{26}$$

$$\chi_{J,J_3} = \frac{1}{2r} \times ((g^+(r) - f^+(r))\Omega_l + (g^+(r) + g^-(r))\Omega_{l+1}), \quad (27)$$

where the spherical harmonics are given by

$$\Omega_l = \begin{pmatrix} \sqrt{\frac{J+J_z}{2J}}Y_{l,J_z-1/2} \\ \sqrt{\frac{J-J_3}{2J}}Y_{l,J_3+1/2} \end{pmatrix},$$

$$\Omega_{l+1} = \begin{pmatrix} -\sqrt{\frac{J-J_3+1}{2J+2}}Y_{l+1,J_3-1/2} \\ \sqrt{\frac{J+J_3+1}{2J+2}}Y_{l+1,J_3+1/2} \end{pmatrix}, \quad (28)$$

with $l = J - 1/2$. The radial functions satisfy the equations

$$\tilde{E}\begin{pmatrix} f^+ \\ g^+ \end{pmatrix} = \left[i\partial_r \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + i\frac{l+1}{r}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \frac{l+1}{p_0 r^2} + \frac{1}{p_0}\left(-\partial_r^2 + \frac{(l+1)^2}{r^2}\right)\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} + i\sqrt{1/r}\left(\partial_r - \frac{1}{4r}\right)\right]\begin{pmatrix} f^+ \\ g^+ \end{pmatrix}, \quad (29)$$

$$\tilde{E}\begin{pmatrix} f^- \\ g^- \end{pmatrix} = \left[i\partial_r \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + i\frac{l+1}{r}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} - \frac{l+1}{p_0 r^2} + \frac{1}{p_0}\left(-\partial_r^2 + \frac{(l+1)^2}{r^2}\right)\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + i\sqrt{1/r}\left(\partial_r - \frac{1}{4r}\right)\right]\begin{pmatrix} f^- \\ g^- \end{pmatrix}. \quad (30)$$

Taking the complex conjugation of (29), we obtain Eq. (30) with the reversed sign of energy. This implies that the matrices cannot be diagonalized simultaneously unless $\tilde{E} = 0$, and therefore, either (f^+, g^+) or (f^-, g^-) is nonzero for the eigenstate with $\tilde{E} \neq 0$.

Equations (29) and (30) are the starting point for our analysis of the fermionic vacuum and excitations.

5. FERMIONS IN THE SEMICLASSICAL APPROXIMATION

In the classical limit, with $(f, g) \propto \exp(i\int p_r dr)$, we obtain the energy spectrum

$$\left(\tilde{E} + \frac{p_r}{\sqrt{r}}\right)^2 = p_r^2 + \frac{l^2}{r^2} + \frac{1}{p_0^2}\left(p_r^2 + \frac{l^2}{r^2}\right)^2, \quad (31)$$

where we neglected small terms of the relative order $1/p_0$. We are interested in the states with the lowest energy, because they give the main contribution to thermodynamics. For a given l, the energy of the fermion becomes zero at the following values of the radial momentum:

$$p_r^2(r, \tilde{E} = 0, l) = \frac{1}{2r}p_0^2(1-r) - \frac{l^2}{r^2} \pm \frac{1}{r}\sqrt{\frac{1}{4}p_0^4(1-r)^2 - \frac{p_0^2 l^2}{r}}. \quad (32)$$

This coincides with Eq. (13) in [28], where the quasiclassical approximation was used from the very beginning.

Within the completely classical analysis, with $p_\perp = l/r$ representing the transverse momentum of the fermion, Eq. (31) at $\tilde{E} = 0$ gives the closed 2D surface in the 3D momentum space. This surface, on which the energy of particles is zero, represents the Fermi surface; it exists only inside the horizon, i.e., at $r < r_h$ $(r < 1)$. Figure 1 demonstrates the Fermi surface $\tilde{E}(\mathbf{p}) = 0$ at two values of the radius r behind the horizon: $r = 2r_h/3$ and $r = r_h/3$. The area of the Fermi surface increases with decreasing r.

In the true ground state, all the levels inside the Fermi surface (i.e., those with $\tilde{E}(\mathbf{p}) < 0$) must be occupied. Of course, this reconstruction of the vacuum involving the Planck energy scale can have tremendous

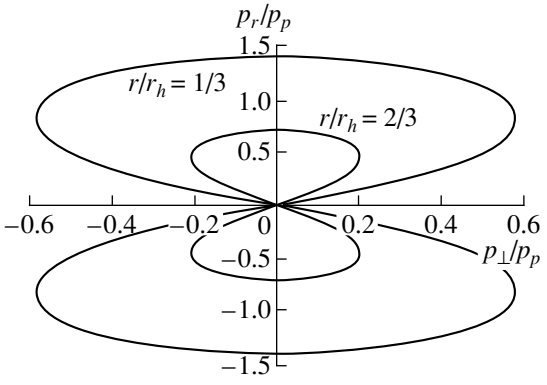

Fig. 1. Fermi surface $\tilde{E}(\mathbf{p}) = 0$ at two positions inside the black hole: $r = 2r_h/3$ and $r = r_h/3$.

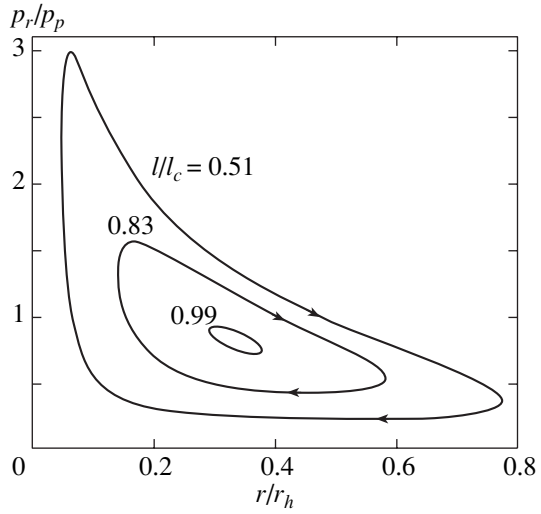

Fig. 2. Closed trajectories of the radial motion inside the black hole at zero energy $\tilde{E} = 0$ for different values of the angular momentum l.

consequences for the black hole itself. These cannot be described by phenomenological low-energy physics. Nevertheless, we can claim that, if the horizon survives the vacuum reconstruction, the Fermi surface also survives because of its topological robustness. In this case, the statistical physics of the black hole microstates is entirely determined by the fermionic states in the vicinity of the Fermi surface. In particular, the entropy and the heat capacity of the black hole are linear in the temperature T,

$$S = C = \frac{\pi^2}{3} N(0) T, \tag{33}$$

where $N(0)$ is the density of states at $\tilde{E} = 0$. From the general dimensionality arguments together with the fact that the density of states must be proportional to the volume of the Fermi liquid, we obtain

$$N(0) = \gamma N_F \frac{p_p^2 r_h^3}{\hbar^3 c}, \tag{34}$$

where N_F is the number of fermionic species and γ is a dimensionless constant of the order of unity. In our oversimplified model, $\gamma = 4/35\pi$ [28].

In the ulterior region, the equation of state is

$$p = \rho \propto T^2.$$

Incidentally, this coincides with the equation of state of the perfect fluid inside the horizon required to obtain the Bekenstein–Hawking entropy (see [30, 31] and [14]). In the Sakharov induced gravity [32], the Planck momentum and the gravitational constant are related by $N_F p_p^2 \sim \hbar c^3/G$. This actually implies that the microscopic parameters of the system, the fermion number N_F and the Planck momentum p_p, are combined to form the phenomenological parameter of the effective theory, the gravitational constant G. If we assume that only the thermal fermions are gravitating, we obtain

$$M \sim \int dV_\rho \sim T^2 M^3 G^2.$$

This gives estimates for the temperature and entropy of the black hole,

$$T \sim 1/GM, \quad S \sim GM^2,$$

which are in correspondence with the Hawking–Bekenstein entropy and the Hawking temperature. Only the phenomenological parameters G and c are involved here, while the microscopic parameters N_F and p_p drop out. This is in agreement with the observation made by Jacobson [33] that the black hole entropy and the gravitational constant are renormalized such that the relation between them is preserved. All this means that statistical properties of the black hole can be produced by the Fermi liquid in the interior of the black hole.

6. EXACT ENERGY LEVELS

Another problem that can be investigated using our scheme is that of the fermion zero modes: Are there fermionic modes that have exactly zero energy in the exact quantum mechanical problem? If yes, this would justify the conjectures that the black hole has a nonzero entropy even at $T = 0$ and also that the area of the black hole is a quantized quantity [34–36]. For this reason, we now proceed to solving eigenvalue equations (29) and (30).

It is impossible to solve these equations analytically, but one can choose the region of parameters where they can be solved using the perturbation theory expansion in the small parameter $1/p_0$. To find this region, we consider semiclassical trajectories of the radial motion $p_r(r)$ at $\tilde{E} = 0$ for different l, Eq. (32). These trajectories are shown in Fig. 2 (we used $p_0 = 10000$). If l is small compared to p_0, these trajectories are highly asymmetric: the incoming and outgoing particles experience essentially different motions. The conventional relativistic particles with a small momentum compared to the Planck momentum p_p can move only towards the singularity. However, when they acquire a large momentum, the nonlinear dispersion allows them to move away from the singularity. As a result, the trajectories of particles become closed. This asymmetry reflects the violation of the time reversal symmetry by the horizon.

However, as l increases, the trajectories become more and more symmetric. Near the maximum value

$$l^{(c)} = 3^{-3/2} p_0, \tag{35}$$

they become perfectly elliptic and increasingly more concentrated in the vicinity of the center point

$$r^{(c)} = \frac{1}{3}, \tag{36}$$

$$p^{(c)} = \pm\sqrt{\frac{2}{3}}p_0. \tag{37}$$

This implies that, in vicinity of $r^{(c)}$ and $p^{(c)}$, the Hamiltonian describing the radial motion becomes that of oscillators. We can therefore expand the equations in the vicinity of $p^{(c)}$ and $r^{(c)}$ using the small parameter $1/p_0$,

$$\begin{aligned} r &= r^{(c)} + x, \\ p_r &= p^{(c)} - i\partial_x. \end{aligned} \tag{38}$$

It can be seen that the regions where x and ∂_x are concentrated,

$$x \propto \frac{1}{\sqrt{p_0}} \ll r^{(c)}, \quad \partial_x \propto \sqrt{p_0} \ll \left|p^{(c)}\right|, \tag{39}$$

become really small compared to $r^{(c)}$ and $p^{(c)}$ as p_0 increases. As a result, after lengthy but straightforward expansion of Eq. (29) near the point with $p^{(c)} > 0$, we obtain (keeping the terms of the order of unity) the effective oscillator Hamiltonian

$$\begin{aligned} H_{\text{eff}} = &-3\sqrt{\frac{3}{2}}\delta l + \frac{13p_0}{2\sqrt{2}}x^2 + \frac{2\sqrt{2}}{3p_0}p^2 \\ &+ \frac{5}{2\sqrt{3}}(xp + px) + \frac{3\sqrt{3}}{4}, \end{aligned} \tag{40}$$

where

$$\delta l \equiv l^{(c)} - (l + 1). \tag{41}$$

Diagonalization gives the energy spectrum

$$\tilde{E}_1 = -3\sqrt{\frac{3}{2}}\delta l + 3n_r + \frac{3}{2} + \frac{3\sqrt{3}}{4}, \tag{42}$$

where $n_r = 0, 1, \ldots$ is the radial quantum number. Accordingly, the expansion near the point with $p^{(c)} < 0$ and the same procedure for Eq. (30) give the other three sets of energy levels,

$$\tilde{E}_2 = 3\sqrt{\frac{3}{2}}\delta l - 3n_r - \frac{3}{2} + \frac{3\sqrt{3}}{4}, \tag{43}$$

$$\tilde{E}_3 = -3\sqrt{\frac{3}{2}}\delta l + 3n_r + \frac{3}{2} - \frac{3\sqrt{3}}{4} = -\tilde{E}_2, \tag{44}$$

and

$$\tilde{E}_4 = 3\sqrt{\frac{3}{2}}\delta l - 3n_r - \frac{3}{2} - \frac{3\sqrt{3}}{4} = -\tilde{E}_1. \tag{45}$$

Finally, in dimensional units, we have the discrete levels of fermions in the vicinity of the Fermi surface,

$$\begin{aligned} \tilde{E}(J, n_r) &= \pm\frac{\hbar c}{r_h} \\ &\times\left(\frac{1}{\sqrt{2}}\frac{p_p r_h}{\hbar} - 3\sqrt{\frac{3}{2}}\left(J + \frac{1}{2}\right) - 3n_r - \frac{3}{2} \pm \frac{3\sqrt{3}}{4}\right), \end{aligned} \tag{46}$$

where all four signs must be taken into account. This equation is valid for J smaller than but close to the maximum value

$$J^{(c)} = p_p r_h/3\sqrt{3}\hbar$$

at which zero-energy states can still exist.

Equation (46) allows us to answer the question whether the true fermion zero modes exist in the presence of a black hole. For general values of $p_p r_h$ and, hence, for the general values of the black hole area $A = 4\pi r_h^2$, there are no states with exactly zero energy. A zero-energy eigenstate can be found for some special values of A. However, because of the incommensurability between the radial and orbital quantum numbers, the degeneracy of the $\tilde{E} = 0$ levels is small, and the fermion zero modes cannot therefore produce the entropy at $T = 0$ that is proportional to the area of the horizon. Accordingly, there are no microscopic reasons for the quantization of the area of the horizon.

There are no topological arguments ensuring the existence of the exact fermion zero modes. On the other hand, the momentum-space topology prescribes the existence of zero-energy fermion modes at the semiclassical level. These modes form a surface in the momentum space—the Fermi surface—in Fig. 1. The existence of the Fermi surface is a robust property of the fermionic vacuum; the Fermi surface survives when the back reaction is introduced (of course, if the horizon survives). It is the Fermi liquid whose thermal states give rise to the entropy proportional to the area, as was discussed in the previous section.

7. CONCLUSION

In deriving the fermionic microstates responsible for the statistical mechanics of the black hole, we used an analogy between quantum liquids and the quantum vacuum, the ether. We know that there are two preferred reference frames in superfluids. One of them is the "absolute" spacetime (x, t) of the laboratory frame, which can be Galilean as well as Minkowskian with c being the real speed of light. In the effective gravity experienced by the low-energy excitations in quantum liquids, the effective "acoustic" metric $g_{\mu\nu}^{\text{acoustic}}$ appears as a function of this "absolute" spacetime (x, t). The other preferred reference frame is the local frame, where the metric is Minkowskian in the acoustic sense,

i.e., with c being the maximum attainable speed of low-energy quasiparticles. This frame is comoving with the superfluid condensate. In this frame, the energy spectrum does not depend on the velocity $\mathbf{v}$ of the condensate and has the form given in Eq. (9). It is therefore in this frame that the Planck energy physics is properly introduced: if the energy becomes big in the superfluid comoving frame, the acoustic Lorentz symmetry is violated.

As for the quantum vacuum, the attainable energies are still so low that we cannot select the preferred reference frame. In particular, we cannot say in which reference frame the Planck energy physics must be introduced, and whether there is an absolute spacetime. The magnifying glass of the event horizon can serve as a possible source of spotting these reference frames.

In our low-energy corner, the Einstein action is covariant: it does not depend on the choice of the reference frame. That is why the Einstein equations can be solved in any coordinate system. However, in the presence of a horizon or ergoregion, some of the solutions are not defined in the entire spacetime of the quantum vacuum. In these cases, the discrimination between different solutions arises and one must choose between them. In quantum liquids, the choice is natural because the absolute coordinates are used from the very beginning. But in general relativity, the ambiguity in the presence of a horizon imposes the problem of properly choosing the solution. This problem cannot be solved within the effective theory, while the fundamental "microscopic" background is still not known, and one can only guess the proper solution of Einstein equations through which the vacuum state can be constructed.

It is clear that the Schwarzschild solution is not the proper choice, in particular, because the entire spacetime is not covered by the Schwarzschild coordinates. According to the quantum liquid analogy, the Painlevé–Gullstrand metric with inward frame dragging can be a reasonable choice. Its analogue can really be reproduced (at least in principle) in quantum liquids. The analogy also suggests that the Painlevé–Gullstrand spacetime can be considered as the absolute one in which the true vacuum must be determined. On the other hand, the local frame of the free-falling observer can be considered as an analogue of the superfluid comoving frame in which the Planck energy physics must be introduced. We again warn that this choice cannot be justified from the standpoint of the effective theory alone.

If the Planck physics is in addition superluminal, as is also suggested by the quantum liquid analogy, the stable quantum vacuum can even be constructed in the presence of a horizon. We argue that the main property of such a quantum vacuum, distinguishing it from the original vacuum of the Standard Model, is the existence of the Fermi surface inside the horizon. The statistical mechanics of the Fermi liquid formed inside the horizon is responsible for the thermodynamics of the black hole.

ACKNOWLEDGMENTS

G.E.V. thanks Jan Czerniawski and Pawel Mazur for fruitful discussions. This work was supported by the ESF COSLAB Programme. The work of G.E.V. was supported in part by the Russian Foundation for Basic Research.

REFERENCES

1. W. G. Unruh, Phys. Rev. Lett. **46**, 1351 (1981); Phys. Rev. D **51**, 2827 (1995).
2. T. Jacobson and G. E. Volovik, Phys. Rev. D **58**, 064021 (1998).
3. L. J. Garay, J. R. Anglin, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **85**, 4643 (2000); Phys. Rev. A **63**, 023611 (2001).
4. G. Chapline, E. Hohlfeld, R. B. Laughlin, and D. I. Santiago, Philos. Mag. B **81**, 235 (2001).
5. R. Laughlin and D. Pines, Proc. Natl. Acad. Sci. USA **97**, 28 (2000).
6. G. E. Volovik, Phys. Rep. **351**, 195 (2001).
7. S. W. Hawking, Nature **248**, 30 (1974).
8. G. E. Volovik, Phys. Lett. A **142**, 282 (1989).
9. S. Corley and T. Jacobson, Phys. Rev. D **54**, 1568 (1996).
10. S. Corley, Phys. Rev. D **57**, 6280 (1998).
11. S. Corley and T. Jacobson, Phys. Rev. D **59**, 124011 (1999).
12. A. Starobinsky, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 415 (2001) [JETP Lett. **73**, 371 (2001)].
13. T. Jacobson, gr-qc/0110079; T. Jacobson and D. Mattingly, Phys. Rev. D **63**, 041502 (2001); gr-qc/0007031.
14. P. O. Mazur and E. Mottola, gr-qc/0109035.
15. M. Mohazzab, J. Low Temp. Phys. **121**, 659 (2000).
16. G. E. Volovik, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 717 (1999) [JETP Lett. **70**, 711 (1999)].
17. P. Painlevé and C.R. Hebd, Acad. Sci., Paris **173**, 677 (1921); A. Gullstrand, Ark. Mat., Astron. Fys. **16**, 1 (1922).
18. P. Kraus and F. Wilczek, Mod. Phys. Lett. A **9**, 3713 (1994).
19. K. Martel and E. Poisson, Am. J. Phys. **69**, 476 (2001).
20. R. Schützhold, Phys. Rev. D **64**, 024029 (2001).
21. M. K. Parikh and F. Wilczek, Phys. Rev. Lett. **85**, 5042 (2000).
22. C. Doran, Phys. Rev. D **61**, 067503 (2000).
23. M. Visser, Class. Quantum Grav. **15**, 1767 (1998).
24. S. Liberati, S. Sonego, and M. Visser, Class. Quantum Grav. **17**, 2903 (2000).

25. M. Stone, cond-mat/0012316.

26. M. Sakagami and A. Ohashi, gr-qc/0108072.

27. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 2: *The Classical Theory of Fields* (Nauka, Moscow, 1973; Pergamon, Oxford, 1975).

28. G. E. Volovik, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 721 (2001) [JETP Lett. **73**, 637 (2001)].

29. S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, 1972).

30. W. H. Zurek and Don N. Page, Phys. Rev. D **29**, 628 (1984).

31. G. 't Hooft, Nucl. Phys. (Proc. Suppl.) **68**, 174 (1998).

32. A. D. Sakharov, Dokl. Akad. Nauk SSSR **177**, 70 (1967) [Sov. Phys. Dokl. **12**, 1014 (1968)].

33. T. Jacobson, gr-qc/9404039.

34. J. D. Bekenstein, in *Proceedings of the 8th Marcel Grossman Meeting*, Ed. by Tsvi Piran (World Sci., Singapore, 1999); gr-qc/9710076.

35. H. A. Kastrup, Phys. Lett. B **413**, 267 (1997).

36. V. F. Mukhanov, Pis'ma Zh. Éksp. Teor. Fiz. **44**, 78 (1986) [JETP Lett. **44**, 63 (1986)].

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 862–868.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1004–1011.
Original Russian Text Copyright © 2002 by Andreev, Van'kov, Platonov, Rozhdestvenskiĭ, Chizhov, Yashin.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Determination of the Radiation Cross Sections of Low-Energy Transitions of Isomeric Nuclei from Observation of Laser-Induced γ Fluorescence

A. A. Andreev, A. K. Van'kov, K. Yu. Platonov, Yu. V. Rozhdestvenskiĭ*, S. P. Chizhov, and V. E. Yashin

Institute of Laser Physics, St. Petersburg, 199034 Russia
**e-mail: RozdYu@soi.spb.su, Rozd-Yu@mail.ru*

Received December 14, 2001

Abstract—Theoretical estimates are given along with the first experimental results on the observation of resonance fluorescence in nuclei of rubidium isomer under conditions of laser plasma X-ray pumping of the contiguous transition with an energy of 3.4 keV. The laser plasma is prepared by irradiating a silver target by a powerful radiation of a Nd laser with a pulse duration of 600 ps. It is demonstrated how one can use the recorded number of emitted γ quanta to determine the probability of low-energy nuclear transition excited by laser plasma X rays. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The acquisition of spectroscopic information about low-energy transitions (1 to 100 keV) in isomeric nuclei is one of the urgent problems of nuclear spectroscopy. For example, the confirmation (or refutation) of the Axel–Brink theory, according to which the cross sections of low-energy transitions in nuclei depend on the energy difference alone rather than on the absolute values of level energies [1], requires direct measurements of the probabilities of low-energy transitions. In addition, a knowledge of cross sections of X-ray absorption by nuclei is essential from the standpoint of practical realization of both coherent (laser) and incoherent sources of the γ range when using two-stage pumping [2].

At present, one tries to obtain such spectroscopic information by conventional methods of nuclear spectroscopy: the possibilities of the latter decrease with the energy of nuclear transition employed. As a result, due to special features of the detection techniques used in nuclear spectroscopy [3], such conventional methods fail to produce adequate information even at transition energies of the order of 100 keV. On the other hand, the development of intense sources of thermal X-radiation enables one to detect such low-energy transitions using methods that are characteristic of laser, rather than nuclear, spectroscopy.

For example, an attempt was made to observe γ fluorescence in an isomer of ^{178}Hf, induced by X-radiation of a medical source [4]. However, in spite of the fact that the hafnium isomer selected for experiment features a number of indisputable advantages over other isomers (the lifetime of 31 years and a significant cross section of radiative absorption of X-ray quanta), the obtained results cannot be regarded as reliable, primarily because of the weak intensity of the X-ray source employed (see [4]). In [5], observations were performed of the excitation of nuclei of the stable isotope ^{181}Ta, which has a low-lying nuclear level with an energy of 6.238 keV [5]. In this case, the nuclei were excited by the intrinsic X-radiation of a laser plasma, which was obtained as a result of irradiation of a tantalum target with subpicosecond optical pulses having an intensity of 10^{16} W/cm^2. The choice of stable isotope in this case necessitated the use of complex techniques of time selection in recording the signal from excited nuclei. The point is that, because in [5] the X-ray pumping of a low-lying level of tantalum from the ground state was performed, the energy of spontaneously emitted quanta coincided with that of pumping quanta. Therefore, special techniques were required for separation in time of the useful signal from excited nuclei and the signal from the intrinsic radiation of a dense plasma. As a result, as was pointed out in [5], most promising in this situation is the use of unstable nuclear isomers, when the energy of X-ray pumping is much lower than the energy of the γ quantum being recorded. In this case, all of the advantages offered by the use of a laser plasma as a source of X-radiation are retained and, at the same time, the system of recording spontaneously emitted quanta is simplified considerably.

In this paper, we present the first experimental results and theoretical estimates of the potential of the new method of determining radiative low-energy γ transitions in isomeric nuclei. The method suggested by us is based on the observation of γ fluorescence of strong nuclear transition under conditions of X-ray pumping of a low-strength contiguous transition in isomeric

nuclei. In order to validate this method, we used rubidium nuclei in the isomeric state $|0\rangle$ with a lifetime of 20.26 min (Fig. 1). Such isomeric nuclei were obtained as a result of irradiation of a bromine target in the cyclotron at the St. Petersburg State University. Then, a target activated by isomeric nuclei of rubidium was subjected to the thermal X-radiation of a laser plasma produced as a result of irradiation of a silver plate with the optical radiation of a neodymium laser with a duration of 600 ps. The intensity of laser radiation was selected to be such that the intensity of X-ray pumping reached its maximal value at an energy close to that of the low-energy $|0\rangle$–$|1\rangle$ transition in rubidium isomer (3.4 keV), see Fig. 1.

We emphasize that the information about such a transition with an energy of 3.4 keV was previously obtained indirectly in analyzing the spectral distribution of conversion electrons. It was our objective to obtain the spectroscopic characteristics of this transition by way of direct observation of γ fluorescence of isomeric nuclei on the contiguous transition $|1\rangle$–$|2\rangle$ with an energy of the order of 200 keV. Because, as a result of X-ray pumping of the $|0\rangle$–$|1\rangle$ transition, a portion of the isomeric nuclei must change to the upper excited state with a lifetime of several nanoseconds with subsequent spontaneous decay to the $|2\rangle$ state, we would have to record γ quanta with an energy of the order of 200 keV, which exceeds considerably the energy of X-ray pumping. Note that, because of the significant difference between the energies of absorbed and emitted quanta, there was no need for us to use the techniques for time selection of emitted γ quanta, as was done in [5] in the case of X-ray pumping of thallium nuclei. By measuring the number of spontaneously emitted γ quanta with an energy of 200 keV, we can determine the interaction cross section of the $|0\rangle$–$|1\rangle$ transition if we allow for the dependence of the number of emitted quanta on the interaction cross section, the intensity of X-ray pumping, and the geometry of the observation scheme.

This paper is organized as follows. In Section 2, the procedure of determining the cross section of radiative transitions in isomeric nuclei by the number of spontaneously emitted quanta is analyzed, and the possibilities of increasing the sensitivity of the suggested method are examined. In Section 3, the experimental facility is described; in Section 4, the obtained results are discussed. The Conclusion gives the inferences made as a result of our investigations.

2. PROCEDURE FOR DETERMINING THE CROSS SECTIONS OF RADIATIVE TRANSITIONS IN ISOMERIC NUCLEI

We will estimate the sensitivity of the method of determining the cross section of low-energy nuclear transitions by way of observation of resonance γ fluorescence induced by X-radiation. In order to validate this method, we will treat the excitation of a nuclear

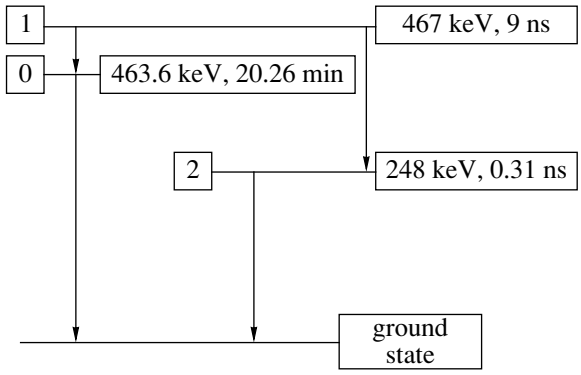

Fig. 1. A diagram of the levels of $^{84}_{37}Rb_{47}$ isomer used in our experiment. The irradiation of a bromine target in a cyclotron produces a target activated by nuclei of this isomer in the long-lived state 0. Laser plasma X rays are effective in the 0–1 transition. A portion of nuclei change to the state 1 and decay spontaneously with the relative probability of 1/2 to the states 0 and 2. In the latter case, γ quanta with an energy of 219 keV must be observed.

isomer by laser plasma X rays; the diagram of the isomer levels is given in Fig. 1. The pumping of the $|0\rangle$–$|1\rangle$ low-energy transition of a nuclear isomer is accomplished by thermal X-radiation of laser plasma. As a result, during the laser plasma de-excitation time τ, a portion of the isomeric nuclei go to the $|1\rangle$ state with subsequent spontaneous decay both back to the $|1\rangle$ state and to the low-lying $|2\rangle$ state. In the latter case, hard γ quanta with an energy of the order of 100 to 500 keV are obtained, the number of which will define the radiation cross section Σ_x of the weak $|0\rangle$–$|1\rangle$ transition of nuclear isomer.

It must be emphasized that this method enables us to measure the radiation cross section Σ_x of the low-energy $|0\rangle$–$|1\rangle$ transition of nuclear isomer rather than the total cross section Σ_{tot}. The total absorption cross section in the $|0\rangle$–$|1\rangle$ transition may be written as

$$\Sigma_{tot} = \Sigma_x + \Sigma_{01},$$

where Σ_{01} is the cross section for nonradiative transitions due to internal conversion for the $|0\rangle$–$|1\rangle$ transition. Note that, for transitions in the energy range from 1 to 50 keV, $\Sigma_x \ll \Sigma_{01}$ [1].

The total number of γ quanta obtained during the laser plasma de-excitation time τ may be defined as

$$N_\gamma = 2N_{is}\omega_x\tau(S_{sp}/S_{is})\Sigma_x\mu/lE_x\lambda_x^2, \quad (1)$$

where N_{is} is the total number of isomeric nuclei in an activated target irradiated by X rays with the energy E_x; ω_x and λ_x denote the X-radiation frequency and wavelength, respectively; S_{sp} is the area of the spot of optical radiation on a metal target; S_{is} is the area of a target activated by isomeric nuclei; μ is the mean absorption path of the quanta of X-ray pumping in an activated target;

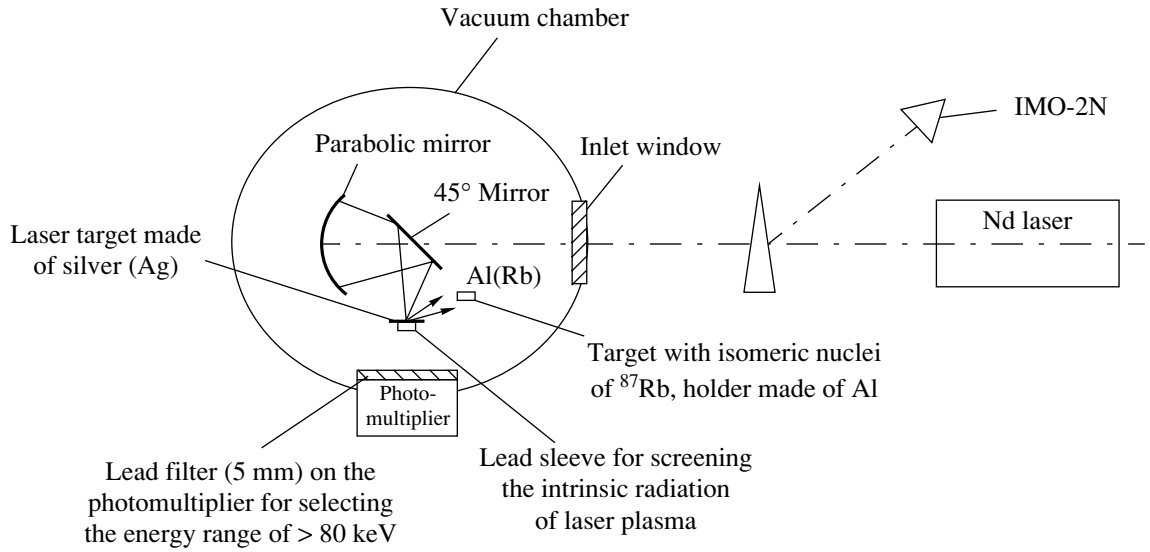

Fig. 2. A diagrammatic view of the experimental facility.

and l is the depth of alloying the target with isomeric nuclei (see Fig. 2). In deriving Eq. (1), we assumed that the laser plasma radiation is described by the Planck formula for blackbody radiation, and the energy of X-ray quantum corresponds to the plasma temperature T_{cold},

$$h\omega_x \approx 2\pi \times 2.822 k_B T_{\text{cold}}.$$

Therefore, the radiation absorption cross section Σ_x is related to the number N_γ of emitted γ quanta by the laser radiation parameters and activated target parameters,

$$\Sigma_x = \alpha(E_x, \tau, S_{\text{sp}}; S_{\text{is}}, \mu, l, N_{\text{is}})N_\gamma, \qquad (2)$$

where α is a dimensional numerical factor.

For the typical values of the parameters involved in our experiment with isomeric nuclei of ^{87}Rb (Fig. 2), namely, $N_{\text{is}} = 10^{11}$, $S_{\text{sp}} = 4 \times 10^{-6}$ cm^2, $S_{\text{is}} = 0.25$ cm^2, $\mu \approx (1\text{–}2) \times 10^{-3}$ cm, and the energy of the $|0\rangle$–$|1\rangle$ transition of 3.4 keV, assuming the plasma de-excitation time to be equal to the laser pulse duration of 600 ps, we derive the following expression for the radiation absorption cross section:

$$\Sigma_x \approx 10^{-26} \text{ cm}^2 \text{ eV } N_\gamma. \qquad (3)$$

From this, after measuring the number of γ quanta, one can determine the value of Σ_x and compare it with the respective value from [6],

$$\Sigma_x' = 10^{-26} \text{ cm}^2 \text{ eV}.$$

Therefore, if a single "shot" of optical radiation produces several γ quanta, the value of the radiation absorption cross section of the $|0\rangle$–$|1\rangle$ transition in an isomer of ^{87}Rb will coincide with that measured previously,

$$\Sigma_x \approx \Sigma_x'.$$

Note further that the numerical values of transition cross sections obtained in [6] call for further verification, because these quantities are given with a question mark. Therefore, our method may be used to either refine these data or confirm that the true values of the respective quantities do not exceed the values given in [6]. This is true of all nuclear isomers featuring low-energy transitions.

We will now discuss the increase in the sensitivity of the suggested method of measuring the cross sections of low-energy transitions in isomeric nuclei. As was demonstrated above, the increase in the sensitivity in this case is directly related to the increase in the number N_γ of γ quanta emitted in the $|1\rangle$–$|2\rangle$ transition, which may be accomplished by increasing the total amount of isomeric nuclei or by increasing the number of X-ray quanta in the $|0\rangle$–$|1\rangle$ transition.

The number of isomeric nuclei may be increased, for example, by using radiochemical methods in preparing the target. This enables one to increase the number of isomeric nuclei in a target approximately a hundred times. However, the use of radiochemical methods (in view of the relatively slow rate of chemical reactions) is justified only in cases in which we deal with a long-lived isomer; for the rubidium isomer employed by us, the number of isomeric nuclei may hardly be increased more than ten times.

On the other hand, it is impossible to accomplish a significant increase in the number of quanta of X-ray pumping by simply raising the laser pulse intensity because, to raise the pumping yield in the desired

energy range a hundred times, it is necessary to raise laser radiation intensity by a factor of at least 10^4. Therefore, we believe that, instead of increasing the total number of X-ray quanta, it appears promising to increase only the number of pumping quanta with an energy of the order of 3.4 keV in the case of the ^{87}Rb isomer. This, in principle, may be accomplished when the ion emission lines in a laser plasma are sufficiently close to the desired nuclear transition [7]. In this case, the number of quanta of X-ray pumping increases sharply, because, along with the thermal radiation of laser plasma, powerful line radiation of plasma ions is present in the desired energy range.

The radiation intensity of plasma ions (in W/cm^2) may be defined as [8]

$$I_i \approx 4 \times 10^8 \Delta E_i E_L Z_n^5 J_z^{-1/2} d^{-2}, \quad (4)$$

where E_i is the ion transition energy (in keV), $E_L = 0.1$ J is the laser pulse energy, $Z_n = 10$ is the nuclear charge, J_z is the atom ionization energy (in keV), and $d = 20$ μm is the focal waist diameter of optical radiation. Compare expression (4) with the expression used to estimate the intensity of blackbody radiation in the same range of quantum energy,

$$I_T \approx 2.5 \times 10^{11} T^4 (\Gamma_i/T)(\Delta E_i/T)^3 \times [\exp(\Delta E_i/T) - 1]^{-1} d^{-2}, \quad (5)$$

where T is the temperature (in keV), $\Gamma_i = \Delta E_i (v_i/c)$ is the line width, and $d = 20$ μm.

For example, for $T = 1$ keV, $E_L = 1$ J, $\Delta E_i = 3.4$ keV, $Z_n = 16$, $J_z = 50$, and $A = 32$, we have

$$I_i/I_T = 2.4 \times 10^2.$$

Therefore, we have found that, as a result of radiation of the plasma ions, the intensity of X-ray pumping is two orders of magnitude higher than the intensity of thermal radiation pumping. We will select sulfur ($Z_n = 16$, $A = 32$) as the material for the laser target. Then, the lower state energy estimated by the hydrogen-like spectrum is 3.482 keV. This value is very close to the value of energy of the $|0\rangle$–$|1\rangle$ transition in the nucleus of the ^{87}Rb isomer, which is 3.4 keV. In addition, because the sulfur atoms in a laser plasma are partly ionized, they possess a spectrum of free crowding high-lying states with level energies

$$\Delta E_i \approx 3.482(1 - 1/n^2).$$

By way of exhaustion of the number n of the high-lying state, one can select the transition energy of sulfur ions so as to be close to the energy of nuclear level. Note that the Doppler line width for pumping amounts to several electronvolts, and this also contributes to the fact that the transition energies in the sulfur atom and in the nuclear isomer of rubidium coincide. So, the number of quanta of X-ray pumping may be increased more than a hundred times.

As a result, one can use radiochemical methods and excite the lines of respective ions in the plasma to raise the yield of γ quanta in the $|1\rangle$–$|0\rangle$ transition to 10^3. In this case, even with the existing recording of γ quanta, we will have about 10^2 quanta with the given aperture of the photomultiplier, which will result in a considerable increase in the sensitivity of the suggested method.

3. EXPERIMENTAL FACILITY FOR THE OBSERVATION OF γ RADIATION OF ISOMERIC NUCLEI

Figure 2 is a diagrammatic view of the experimental facility for the observation of resonance γ fluorescence of isomeric nuclei of rubidium, induced by X-radiation of a laser plasma. The optical radiation of an Nd laser with a duration of 600 ps and total energy of 1–2 J is introduced via a window into a steel chamber 10 cm in size in which a vacuum is maintained. A parabolic 10-cm mirror focuses the optical radiation onto the polished surface of a silver plate (we used silver to provide for a higher yield of thermal X-radiation) by means of a rotating flat mirror. As a result of such focusing, a high-temperature laser plasma is formed on the metal surface; the radiation spectrum of this plasma has the form of a blackbody distribution with a maximum corresponding to an energy of the order of 3 keV. The photomultiplier was screened from X-radiation of the laser plasma by a layer of lead 15 mm thick.

The thus obtained X-ray quanta fall on a target activated by isomeric nuclei of rubidium. The distance between the targets is approximately 1 cm. The activated target was prepared by irradiation of bromine nuclei by a flow of fast ions in the cyclotron at the St. Petersburg State University. The target diameter was 0.5 cm, with the thickness selected from the condition of total absorption of X-radiation with an energy of 3.4 keV in the activation region. The total number of isomeric nuclei was monitored by the intensity of natural decay and did not exceed the standard medical norms for work with radioactive substances in laboratories without special equipment. The photomultiplier with a lead-plate filter 6 mm thick made it possible to record X-radiation only above 80 keV (see below); the inlet window for the photomultiplier was 10 cm in diameter. The photomultiplier signal indicated the total number of quanta which came to the recording system at some moment of time.

The experimental procedure was as follows. First, an activated target was prepared in a cyclotron and the total number of isomeric nuclei was estimated, and then the target was placed in chamber *1* (the time between preparing the target and placing it into the chamber was approximately equal to the decay period of 0 to

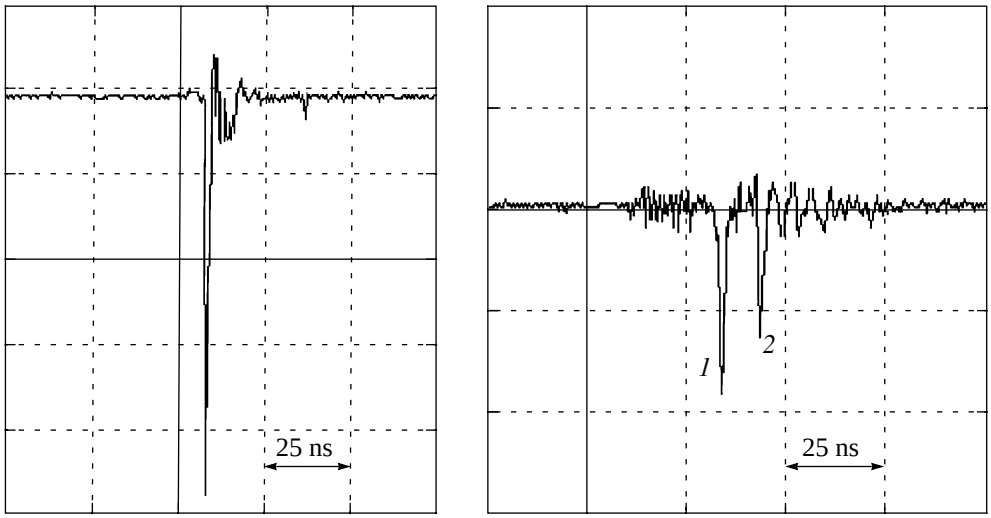

Fig. 3. The time dependence of the signal from the photomultiplier (a) in the absence of activated target and (b) in the presence of a target activated by $^{84}_{37}Rb_{47}$ isomer. The target for laser radiation is made of silver. Curve *1* indicates the signal of induced γ fluorescence of excited nuclei, and curves *2* correspond to the signal of X-radiation of plasma ions with a quantum energy above 80 keV.

20 min); after producing a vacuum in chamber *1* (the evacuation time was equal to two decay periods), a "shot" of optical radiation was produced, after which a metal plate *4* was shifted, and one more "shot" was made. Five targets with isomeric nuclei of rubidium were investigated.

4. DISCUSSION OF THE EXPERIMENTAL RESULTS

We will first treat the results of measurements without a rubidium target. Figure 3a gives the time dependence of the photomultiplier photocurrent under conditions of irradiation of a silver target with optical radiation. One can see from the result of measurement of the photocurrent that an X-ray pulse exists with an energy above 80 keV (see above). Also observed is a weak dependence of the intensity of such radiation on the material of the holder securing the activated target.

The quanta of such energy (above 80 keV) may form in two ways, namely, as a result of de-excitation of nuclei and as a result of bremsstrahlung of fast particles, both electrons and ions. Here, the ions themselves do not emit because of the large mass. However, when an ion hits the target, recoil electrons arise with an energy of up to the energy of the ion itself. It is these electrons that generate bremsstrahlung.

In order to determine the source of generation of hard quanta, an experiment was performed without a holder and a working target. In this case, the photomultiplier signal disappeared completely. This means that the bremsstrahlung in the hot laser plasma and bremsstrahlung from the chamber walls do not produce a working signal. The former is screened from the photomultiplier by the layer of lead, and the latter is low due to the remoteness of the walls (because the intensities of the particle flows are relatively low). Bremsstrahlung always arises in the case of interaction between the ions and fasteners (the holder, etc.), because such radiation is recorded; in the absence of isomeric nuclei, it is this radiation that presents a spurious signal. We will estimate quantitatively the intensity of the spurious signal and the possibility of recording the desired signal from excited isomeric nuclei against the background of bremsstrahlung. With the selected intensity of laser radiation $I = 10^{15}$ W/cm^2, the temperature of hot electrons, according to [9], may be

$$T_{\text{hot}} = 14T_{\text{cold}}[I/10^{16}\ \text{W/cm}^2]^{1/3}\ \text{keV},$$

which amounts to 6.5 keV for $T_{\text{cold}} \approx 1$ keV. This value of T_{cold} follows from the scaling formulas for estimating the laser plasma parameters given in [8].

Note that this temperature of hot electrons is obviously insufficient for the generation of quanta in the range above 80 keV. However, the hot electrons escaping from a laser target entrain ions due to the emerging electrostatic field. The energy of such ion is given by [10]

$$\varepsilon_i = 2Z^*T_{\text{hot}}\ln(0.9\Omega_{\text{pi}}\tau_L), \qquad (6)$$

where Z^* and Ω_{pi} denote the effective charge and the plasma frequency of escaping ions, respectively, and τ_L is the laser pulse duration. Therefore, multiply charged ions with $Z^* \geq 10$ possess an energy sufficient for the generation of hard X rays.

We will estimate the number of such ions. The effective degree of ionization of silver Z_{eff} at a given temperature T_{cold} is defined as [8]

$$Z_{\text{eff}} = 16(T_{\text{cold}}/1\ \text{keV})^{1/3}$$

and has a value of the order of 16. It is more difficult to estimate the fraction of multiply charged ions relative to the total number of ions. The thing is that, in addition to silver ions, protons escape intensely from the target (hydrogen is contained in the target as a part of water vapors adsorbed by the surface and of traces of polishing pastes). By virtue of quasineutrality, the total charge of all ions is approximately equal to the charge of fast electrons leaving the target. For estimating the number of multiply charged ions, we will use the experimental data of [11], which were obtained for similar parameters of laser pulse for a copper target. The number of copper ions with a charge of more than eight was approximately 10^3 per shot.

The number of hard radiation quanta with an energy of 0.3 to 1.0 of the ion energy which arise during deceleration of a single ion is estimated by the formula [12]

$$N_\gamma \approx \frac{16}{3}\frac{n_e Z^{*2} e^6 \tau_L}{m_e^2 c^4 \hbar},$$

where n_e is the electron concentration in the holder material (Al). For $Z^* \approx 10$, the quantum yield is approximately 0.04 quantum per ion. Therefore, silver ions in an amount of the order of 10^3 develop a signal corresponding to tens of hard quanta, which is the lower limit of sensitivity of our detector. The desired signal from the de-excitation of nuclei is of the same or somewhat lower level. We can isolate this signal against the background of bremsstrahlung only owing to its time characteristics. The desired signal arises after $t \sim 10^{-9}$ s (the lifetime of the excited state of nuclei), while bremsstrahlung arises after tens of nanoseconds (the transit time of ions from the laser target to the working one). Therefore, in the case of nuclear decay, we must see a characteristic double-peak structure of the signal. Note that, if there existed a superfast electron component that would also generate bremsstrahlung of requisite energies, the photomultiplier signal would always have a double-peak form (electrons with an energy equal to that of ions fly tens of times faster). In our case, in the absence of excited nuclei in all experiments (24 shots), we observed a single-peak form of signal.

Figure 3b gives the results of measurements in the presence of a second target activated by isomeric nuclei of ^{87}Rb. One can see that the measured signal has in this case a clearly defined double-peak structure. The origin of these peaks may be explained as follows: the X-radiation of laser plasma transforms the isomeric nuclei of rubidium to the upper excited state, and, as a result of spontaneous decay, X-ray quanta are formed with an energy of about 200 keV; we observe the signals from these quanta as the first peak. Then, after the transit time necessary for the laser plasma ions to reach the target fastenings, the second peak is observed in the photomultiplier signal; the origin of this peak may be attributed to the bremsstrahlung of ions (see above). The time between two peaks in the photomultiplier signal coincides with the transit time of plasma ions, and the probability of recording the X-radiation of the nuclei in our experiment is about 20%. Finally, the sensitivity threshold of the photomultiplier amounts to several X-ray quanta. All this leads one to conclude that we have indeed observed the resonance fluorescence of isomeric nuclei and obtained the value of the radiative cross section of the $|0\rangle$–$|1\rangle$ transition of the order of

$$\Sigma_x \approx 10^{-26}\ \text{cm}^2\ \text{eV},$$

which agrees with the previously obtained data.

5. CONCLUSION

We will formulate the basic conclusions reached as a result of this study.

(i) It has been demonstrated that the X-ray pumping of low-energy transitions in isomeric nuclei with subsequent recording of hard γ quanta in the adjacent nuclear transition may be used as a new method of nuclear spectroscopy.

(ii) The potential possibilities have been determined of increasing the sensitivity of the suggested method using the K_α line of radiation of matrix ions.

(iii) The spurious signal of X-radiation from fast ions has been calculated.

(iv) The first qualitative experiments involving the observation of induced γ fluorescence have been performed.

(v) It has been found that the measured total cross section Σ_x of the low-energy transition $|0\rangle$–$|1\rangle$ of the rubidium isomer does not exceed 10^{-26} cm^2 eV.

ACKNOWLEDGMENTS

We are deeply grateful to E.V. Antropov, A.P. Dubetskiĭ, and V.A. Lazarev for the preparation of targets of rubidium isotope and for their permanent and profound interest in our work.

REFERENCES

1. A. Bohr and B. R. Mottelson, *Nuclear Structure* (Benjamin, New York, 1969, 1975; Mir, Moscow, 1971, 1977), Vols. 1, 2.
2. C. C. Balwin and J. C. Solem, Rev. Mod. Phys. **69**, 1085 (1997).
3. L. V. Groshev and I. S. Shapiro, *Spectroscopy of Atomic Nuclei* (Nauka, Moscow, 1952).
4. C. B. Collins, F. Davanloo, M. C. Iosif, *et al.*, Phys. Rev. Lett. **82**, 695 (1999).

5. A. V. Andreev, R. V. Volkov, V. M. Gordienko, *et al.*, Zh. Éksp. Teor. Fiz. **118**, 1343 (2000) [JETP **91**, 1163 (2000)].
6. *Nuclear Data Sheets*, Ed. by M. J. Martin (Academic, New York, 1989–1998).
7. V. S. Letokhov, Kvantovaya Élektron. (Moscow) **1**, 125 (1973).
8. A. A. Andreev and A. N. Semakhin, Probl. Nauchn. Priborostr. **3**, 35 (1993).
9. F. N. Beg, A. R. Bell, *et al.*, Phys. Plasmas **4**, 447 (1997).
10. A. V. Gurevich and A. P. Meshcherkin, Zh. Éksp. Teor. Fiz. **80**, 1810 (1981) [Sov. Phys. JETP **53**, 937 (1981)].
11. E. Woryna, P. Parys, *et al.*, Laser Part. Beams **14**, 293 (1996).
12. V. B. Berestetskii, E. M. Lifshitz, and L. P Pitaevskii, *Course of Theoretical Physics*, Vol. 4: *Quantum Electrodynamics* (Nauka, Moscow, 1980; Pergamon, New York, 1982).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 869–881.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1012–1027.
Original Russian Text Copyright © 2002 by Zabolotskiĭ.

ATOMS, SPECTRA, RADIATION

Amplification of Extremely Short Pulses in Optical Media

A. A. Zabolotskiĭ

Institute of Automatics and Electrometry, Siberian Division, Russian Academy of Sciences, Universitetskiĭ pr. 1, Novosibirsk, 630090 Russia

e-mail: zabolotskii@iae.nsk.su

Received September 12, 2001

Abstract—The dynamics of pulses with durations comparable to the inverse transition frequency that propagate in an optical medium is studied in terms of two integrable systems of Maxwell–Bloch equations. The first model describes the field interaction with a nondegenerate medium with a permanent dipole moment and permanent external pumping. A general formula is derived for the *N*-soliton solution. Particular solutions are used as examples to investigate the effect of permanent dipole moment and pumping on the soliton dynamics. The second model describes the interaction between two-component electric-field pulses and a two-level degenerate medium with permanent upper-level pumping. For different initial magnetic-sublevel populations, soliton solutions are used as examples to show that pumping causes a change in polarization dynamics. A two-soliton solution is used to analyze the interaction of solitons in a two-level medium with external pumping. *© 2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The dynamics and amplification of femtosecond optical pulses attract constant attention because of their applications in various fields of physics. In many cases, these pulses are associated with soliton solutions of the integrable systems of Maxwell–Bloch equations (see, e.g., [1, 2] for a review). In general, the application of integrable models [3] requires imposing a number of physical constraints. In practice, however, the evolution of ultrashort pulses in nonlinear media can be described in detail in terms of these models. One of the integrable models, the system of Maxwell–Bloch equations describing the dynamics of quasi-monochromatic electric-field pulses in a two-level degenerate media, was studied extensively [4–6]. Related models of three-level media [7, 8] are also noteworthy. The assumption of slow (compared to $\pi\omega_0^{-1}$, where ω_0 is the frequency of the energy transition between electronic levels of the medium) variations in the amplitudes and phases of the field components and the rotating-wave approximation for the nondiagonal elements of the density matrix [1, 9] were used to derive these equations.

The multisoliton periodic solutions and quasi-self-similar solutions describing the decay of an initial unstable state of the medium are known for the Maxwell–Bloch equations that describe the dynamics of quasi-monochromatic fields [1, 2, 10, 11].

Recently, interest in electromagnetic pulses with characteristic durations close to π/ω_0 has increased. In the literature, these are called ultrashort pulses [12–14], electromagnetic bubbles, video pulses [16], and extremely short pulses [17, 18]. The domain of parameters for such pulses is restricted by the condition that there is no photoionization; i.e., the field amplitude is no larger than $\sim 10^8$–10^9 W cm^{-1}, and the lower limit for the soliton duration is $\sim 10^{-15}$–10^{-16} s [15]. However, the parameters of light pulses with durations $\tau_p \sim \pi/\omega_0$ make the conditions for applicability of the two-level model for the medium to describing the actual field interaction with the medium very stringent. The levels with energy $\hbar\omega_k$ that are disregarded when constructing a model must be at the distance

$$|\omega_k - \omega_0| \gg \pi/\tau_p.$$

Thus, we must use an isolated energy transition with a relatively low frequency ω_0 as the laser transition and restrict the range of pulse parameters to

$$\tau_p \gg \pi|\omega_k - \omega_0|^{-1}.$$

The conditions for applicability of the two-level model are improved if the coefficient of nonlinear susceptibility or the dipole moment corresponding to the chosen laser transition is much larger than the same coefficients for adjacent transitions [19]. Effects related to the generation of such pulses are observable within the scope of the currently available laser technology. The generation of pulses with durations that are longer than the oscillation half-period by only several times was described, for example, in [19–22].

When the femtosecond range is investigated, the approximation of slowly varying amplitudes and phases of light fields in the models used for this purpose is inapplicable. The corresponding Maxwell–Bloch equations prove to be too difficult to analyze. Therefore, the assumption of pulse duration $\tau_p \gg \pi/\omega_0$ is used in many theoretical studies [12–18, 23], which simpli-

fies the problem. For soliton solutions, this approximation is equivalent to the strong-field approximation

$$\mu_{12}E/\hbar \gg \omega_0, \quad (1)$$

where μ_{12} and E are the transition dipole moment and the field amplitude, respectively. When condition (1) is satisfied, some of the Maxwell–Bloch equations for a nondegenerate two-level medium can be reduced to the sine-Gordon equation with well-known soliton and other solutions [3, 9]. At the same time, the application of the two-level model approximation for a resonant atomic or molecular medium in the optical range is unfeasible under this condition [16].

Gibbon *et al.* [24] did not use condition (1) to derive the reduced Maxwell–Bloch equations (RMBE). They showed that the RMBE could be solved for a nondegenerate two-level medium by applying the inverse scattering transform. The low-density active-medium approximation is used instead of the slow-envelope approximation to derive the RMBE. As a result, the problem can be reduced to studying unidirectional wave propagation [24–29].

In [24–29], the inverse scattering transform was used to find soliton solutions to the RMBE and to study their properties for a nondegenerate two-level medium. Light pulses with durations of the order of π/ω_0^{-1}, at which the applicability conditions for the two-level model hold, can be analyzed in terms of these integrable models. Here, our goal is to study the dynamics of such pulses in media with pumping. Following [2], we use the term "extremely short pulse" for these pulses.

It is well known that, apart from the dipole moment μ_{12}, where

$$\mu_{ij} \propto \int \psi_i^* z \psi_j dz,$$

$\psi_{1,2}$ being the wave functions for the states of levels 1 and 2, a nonzero permanent dipole moment (PDM) is possible in asymmetric media. It is defined by the difference

$$\Delta\mu_{12} = \mu_{11} - \mu_{22} \neq 0.$$

A nonzero PDM arises in polar molecules, in asymmetric semiconductor quantum wells, and in other media. The PDM contribution to nonlinear absorption and dispersion was studied, for example, in [30–33], where it was found that PDM could play a significant role in nonlinear multifrequency processes. Kŏcinac *et al.* [33] pointed out that the PDM for quantum wells increases in importance with wavelength during second-harmonic generation. These authors gave parameters of the semiconducting medium for which the $\Delta\mu_{ij}/\mu_{ij}$ ratio varies over the range 0.15–7.1 for various transition frequencies.

Investigation of the PDM effect on the shape and evolution of pulses with durations comparable to the oscillation period appears to be restricted to a recent paper by Agrotis *et al.* [34]. These authors showed that the integrable RMBE for a nondegenerate two-level transition could be generalized to a nonzero PDM. They derived a general expression for the multisoliton solution by using the Backlund transformation and gave one- and two-soliton solutions in explicit form.

During the evolution of an extremely short pulse with two polarization components in a nonlinear medium, a number of new nonlinear effects related to the mutual influence of these components can arise. The polarization effects in nonlinear media were first studied in terms of an integrable model by Manakov [35], who analyzed a two-component nonlinear Schrödinger equation. The results of this author are of current interest and have been extensively used up until now. In a dipole medium with a degenerate two-level transition, polarized radiation can produce transitions with a change in magnetic moment by one. For example, the $J = 0 \longleftrightarrow J = 1$ transition interacting with a light field with two polarization components belongs to such systems. For quasi-monochromatic fields, this scheme was investigated in [5, 6, 28] (see also [1] for a review) for the integrable Maxwell–Bloch equations. These equations are mathematically equivalent to the system of equations that describes the field interaction with a three-level medium with two allowed transitions [7] arranged in the form of Λ and V schemes. Note that the soliton solutions for these Maxwell–Bloch equations often coincide in form and have a number of properties similar to the properties of the soliton solutions for the two-component nonlinear Schrödinger equation.

Sazonov [17] studied the rotation of the field polarization plane for extremely short pulses for which inequality (1) holds. However, as was pointed out by the author himself, the approximation of extremely short pulse duration and, accordingly, extremely high field intensity that he used is more of methodological importance, because this model is inapplicable to the actual situation in the optical range. Parkhomenko and Sazonov [18] investigated the self-induced transparency for an extremely short pulse by using an approximate model of a multilevel medium with restrictions on possible transitions. In these papers, the Maxwell–Bloch equations were reduced to a single-component sine-Gordon equation for strong fields and to a modified nonlinear Schrödinger equation for weak fields.

A similar interaction with a degenerate transition is possible for an extremely short pulse of duration $\sim\pi/\omega_0$. In general, polarization effects must be taken into account to construct a full picture of the evolution of solitons and other solutions. A number of such effects can be related, for example, to a change in pulse polarization at small distances. For instance, based on the Maxwell–Bloch model describing the dynamics of quasi-monochromatic fields, Bol'shov *et al.* [7] showed for a three-level medium that a difference between the initial level populations could result in effective Raman

scattering of solitons. A similar result was obtained for a more complex interaction scheme in [8].

When light pulses propagate in very long optical fibers, pulse amplification on the segments with implanted erbium atoms periodically arranged along the fiber length is used to compensate for any losses. The amplification efficiency depends on parameters of the nonlinear processes within such a segment. Two amplification schemes are most commonly used in a resonant two-level medium. In the first scheme, it is assumed that the two-level medium is initially inverted and that the pulse absorbs the stored energy during its evolution [9]. In the second scheme, the pulse is amplified through permanent upper-level pumping [36, 37]. The system of Maxwell–Bloch evolution equations with external pumping is used to model the processes in gaseous media, solids, and dyes [36, 37]. Three- or four-level transition schemes are used in such laser media. These schemes can be reduced to an effective system of Maxwell–Bloch equations for a two-level medium with an additional term in the Bloch equations for the population difference if a number of conditions are satisfied [36]. Note that the scheme of a two-level laser transition with external pumping [36] is more commonly used in nonlinear optics to describe the generation in lasers than the above scheme of a single-pass laser amplifier.

In optical fibers, the amplification is nonadiabatic [38]. Therefore, to describe it requires using appropriate models that describe the dynamics of the nonlinear stage outside the scope of the approximation theory. Such models include the integrable Maxwell–Bloch equations for a two-level medium with permanent upper-level pumping derived in the slow-envelope approximation. These equations belong to the so-called integrable deformations [39–42], i.e., equations integrable by the inverse scattering transform with a variable spectral parameter.

Burtsev and Gabitov [41] showed the RMBE that describe a two-level laser with permanent level pumping in terms of the inverse scattering transform with a variable spectral parameter to be integrable. Pumping causes the upper level to be populated, and laser generation begins when an inverse population arises. For small seed pulses, the laser generation is asymptotically described by the (quasi-)radiative solution associated with the real continuum of the Zakharov–Shabat problem, just as for a long laser amplifier [10]. In [44–46], this solution was shown to consist of nonlinear oscillations with an amplitude monotonically increasing with distance z.

At the same time, a two-level laser medium with pumping can be used to amplify both quasi-monochromatic soliton pulses with a carrier frequency close to the transition frequency and solitons with durations of the order of π/ω_0. Particular soliton and linear solutions to the integrable Maxwell–Bloch equations for quasi-monochromatic waves with pumping were obtained in [39–44]. In [40, 41], the dynamics of the soliton associated with an isolated eigenvalue of the Zakharov–Shabat spectral problem with a positive imaginary part under pumping was reduced to a change in its parameters. At the same time, the effect of the radiative solution, which is determined by the real continuum of this problem, on the soliton dynamics must generally be taken into account.

Studying the dynamics and change in the shape of optical solitons with durations $\sim\pi/\omega_0$ during their amplification is evidently of practical interest. However, as far as we know, the soliton dynamics under permanent pumping have not yet been investigated in terms of the integrable RMBE in nondegenerate media with PDM and polarization effects in degenerate two-level media.

Here, we study the amplification of an extremely short pulse in a nondegenerate two-level medium with PDM and in a degenerate two-level medium with pumping in terms of the integrable RMBE deformations. Using these integrable models allows us to abandon condition (1) and to investigate the soliton amplification dynamics outside the scope of perturbation theory and the adiabatic approximation. In the former case, we use a generalization of the integrable system of Maxwell–Bloch equations with PDM derived by Agrotis *et al.* [34]. Here, we additionally take into account the level pumping and examine the role of PDM in the soliton amplification. Subsequently, we construct an effective integrable RMBE system for a medium with a degenerate (in magnetic-moment projections) transition. This system allows us to analyze the nonlinear polarization effects related to the magnetic-sublevel population difference and the soliton interaction and conversion, as well as a number of other effects. We use one- and two-soliton solutions as examples to study the soliton dynamics in the presence of pumping. As far as we know, no similar studies of the monochromatic Maxwell–Bloch system for a degenerate transition with pumping have been carried out.

The paper has the following structure. In the next section, we generalize the integrable RMBE to include pumping and PDM and find soliton solutions. The integrable Maxwell–Bloch equations that describe the interaction of a two-component light field with a degenerate medium for two interaction schemes are derived in Section 3. In Section 4, we find one- and two-soliton solutions of this model and analyze their properties. Our results are discussed in the final section.

2. THE MAXWELL–BLOCH EQUATIONS WITH PUMPING

The pumping of the upper energy level in a laser medium is phenomenologically taken into account by adding an extra term to the right-hand side of the Bloch equation for the population of this level (see, e.g., Eqs. (9.106) in [36]). This effective two-level scheme

can be obtained by reducing some models of a three- or four-level medium interacting with monochromatic fields that are resonant for these transitions [42]. It is assumed that there is a pumping mechanism under which electron transfer from the ground level 0 to level 3 lying above the upper level 1 of the laser transition takes place in these schemes. For an ion laser, this population results from the passage a strong electric current through gas. Subsequently, because of the rapid electron transfer from level 3 to level 1, the population difference between level 1 and level 2 with a lower energy increases. In such a system, a population inversion of levels 1 and 2 and the generation of a field with amplitude E_{12} on this transition with frequency ω_{12} can be obtained. However, to observe the soliton dynamics, a situation with a large initial population of level 2 is preferred. In this case, the generation of the nonsoliton part of the solution, which is associated with the real continuum of the corresponding spectral problem for an integrable model [43], is suppressed. Therefore, a three-level scheme with the coincident levels 2 and 0 is preferred for observing the soliton dynamics.

As was pointed out above, the RMBE for a nondegenerate two-level medium were derived in [23, 24] in the approximation of a low effective density of the medium. Agrotis *et al.* [34] generalized these equations for a nonzero permanent dipole moment, i.e., for

$$\mu_{11} - \mu_{22} \neq 0.$$

In this approximation, the pulse group velocity is close to the speed of light in the medium and the reduced Maxwell equations reduce to one equation,

$$\left(\frac{\partial}{\partial t} + \frac{c}{\sqrt{\epsilon_\infty}}\frac{\partial}{\partial z}\right)E = i\frac{N_0\omega\mu_{12}}{2\epsilon_0\epsilon_\infty}(\rho_{12} - \rho_{21}). \qquad (2)$$

Here, N_0 is the density of the active atoms, and ϵ_0 and ϵ_∞ are the permittivities of the medium [34]. The nondiagonal values of the dipole moment were reduced to real form by a simple phase shift of the density matrix elements ρ_{12} and ρ_{21} with $\mu_{12} = \mu_{21}$.

The Bloch equations for the density matrix with components ρ_{ij} $(i, j = 1, 2)$ of a nondegenerate two-level medium with PDM are [34]

$$\partial_t\rho_{12} = i\left(\frac{\Delta\mu_{12}}{\hbar}E - \omega\right)\rho_{12} - i(\rho_{11} - \rho_{22})\frac{\mu_{12}}{\hbar}E + \frac{b}{2}, \quad (3)$$

$$\partial_t\rho_{11} = i\frac{\mu_{12}}{\hbar}E(\rho_{21} - \rho_{12}) + c_1, \qquad (4)$$

$$\partial_t\rho_{22} = i\frac{\mu_{12}}{\hbar}E(\rho_{12} - \rho_{21}) + c_2. \qquad (5)$$

Here,

$$\Delta\mu_{12} = (\mu_{11} - \mu_{22}),$$

and μ_{ij} are the dipole moment components of the medium. Equations (3)–(5) include the level pumping ($\propto c_1$, c_2) and the constant external force ($\propto b/2$) produced by an additional field or fields on this transition.

Let us rewrite system (2)–(5) as

$$\begin{aligned} \partial_z E &= iR^-, \\ \partial_\tau R^- &= i(aE - \omega)R^+ - 2iN\mu E, \\ \partial_\tau R^+ &= i(aE - \omega)R^- + b, \\ \partial_\tau N &= -2i\mu E R^- + c, \end{aligned} \qquad (6)$$

where

$$R^\pm = \rho_{12} \pm \rho_{21}, \quad N = \rho_{11} - \rho_{22},$$

$$a = \Delta\mu_{12}/\hbar, \quad z = x\mu_{12}\omega N_0(2c\epsilon_0\sqrt{\epsilon_\infty})^{-1},$$

$$\tau = t - x\sqrt{\epsilon_\infty}c^{-1}, \quad c = c_1 - c_2, \quad \mu = \mu_{12}/\hbar.$$

The Lax representation of system (6) was found for $b = c = 0$ in [34]. In our notation, this representation is

$$\frac{\partial}{\partial\tau}\Psi = \begin{bmatrix} -i\lambda & -a\omega D^{-1} + DE \\ a\omega D^{-1} - DE & i\lambda \end{bmatrix}\Psi \equiv \mathbf{L}_1\Psi, \qquad (7)$$

$$\frac{\partial}{\partial z}\Psi = \frac{D}{4(\lambda^2D^2 - \mu^2\omega^2)} \times \begin{bmatrix} iD(2\mu\Lambda N - a\lambda R^+) & 2A_{12} \\ 2A_{21} & -iD(2\mu\lambda N - a\lambda R^+) \end{bmatrix}\Psi \equiv \mathbf{A}_1\Psi; \qquad (8)$$

here,

$$D^2 = 4\mu^2 + a^2,$$

$$A_{12} = \mu a\omega N + 2\mu^2\omega R^+ + \lambda D^2 R^-,$$

$$A_{21} = -\mu a\omega N - 2\mu^2\omega R^+ + \lambda D^2 R^-.$$

Including the arbitrary functions $c = c(z)$ and $b = -ac(z)/2\mu$ gives rise to a dependence of the spectral parameter λ on variable z. This dependence is described by the equation

$$\frac{\partial}{\partial z}\lambda = \frac{1}{4}\frac{\lambda D^2}{\lambda^2D^2 - \mu^2\omega^2}[b(z)a - 2c(z)\mu]. \qquad (9)$$

Note that for $c = b = 0$ and for large amplitudes at which inequality (1) holds, system (2)–(5) reduces to the sine-Gordon equation after discarding the term $i\omega\rho_{12}$ in Eq. (3):

$$\partial_T\partial_z\theta = \sin\theta, \qquad (10)$$

where

$$T = D\tau, \quad \partial_T\theta = E, \quad iR^- = \sin\theta,$$

$$2\mu N - aR^+ = D\cos\theta.$$

Estimation of the field amplitudes indicates that they are within the scope of the currently available laser technology [22]. The solutions to the sine-Gordon equation for various initial–boundary conditions have been studied extensively. At the same time, as was pointed out above, the conditions under which the actual medium can be modeled by a two-level medium in the visible range are too stringent for such field amplitudes. Therefore, we must consider the more general model (6) for such field amplitudes that

$$\Delta\mu_{ij}E\hbar^{-1}, \mu_{ij}E\hbar^{-1} \sim \omega.$$

It is easy to show that system (6) at $c = b = 0$ has the following constant, linear stable solution:

$$E_v^{(0)} = a\omega D^{-2}, \quad R_v^{-(0)} = 0, \qquad 2\mu R_v^{+(0)} = -aN_v. \tag{11}$$

This solution is a generalization of the standard vacuum solution to the RMBE for $a = 0$ that corresponds to the Bloch vector

$$\mathbf{B} \equiv (R^+, iR^-, N) = (0, 0, -1)N_0,$$

where

$$(R^+)^2 + (iR^-)^2 + N^2 = N_0^2.$$

For $a \neq 0$, the following Bloch vector corresponds to the vacuum solution (11):

$$\mathbf{B} = (a, 0, -2\mu)N_0/D.$$

Pumping leads to a dependence of the vacuum solution on variables. It must be taken into account when constructing the soliton solutions. For $b, c \neq 0$, system (6) has the vacuum solution

$$E_v^{(1)} = \frac{\omega}{D^2}\left[a + 2\mu\tau\frac{2\mu b + ac}{DN_0 + \tau(ab - 2\mu c)}\right], \tag{12}$$

$$R_v^{-1} = 0, \quad R_v^{+(1)} = \frac{aN_0}{D} + b\tau, \qquad N_v^{(1)} = -\frac{2\mu N_0}{D} + c\tau. \tag{13}$$

The Bloch vector for this solution has a variable length.

For a special choice of the pumping constants, such that

$$2\mu b = -ac, \tag{14}$$

the vacuum solution (12) for the field amplitude does not depend on τ. In this case,

$$E_v^{(1)} \equiv a\omega D^{-2}.$$

Physically, this means that the pumping energy does not enter the radiative part of the solution for the field. At the same time, as we show below, the soliton amplitude can increase under pumping if condition (14) is satisfied. Solution (12) for the field amplitude $E_v^{(1)}$ at $ab - 2\mu c < 0$ is singular; i.e., it becomes infinite in a finite time at point N_0/γ_s if condition (14) is not satisfied. As can be easily shown using (9), the soliton amplitude increases for $ab - 2\mu c < 0$. In this paper, we consider the parameters and time intervals for which this singularity may be disregarded. With this goal in mind, we restrict our analysis to condition (14) or to the parameters at which

$$DN_0 \gg \tau(2\mu c - ab), 2\mu\tau(2\mu b + ac). \tag{15}$$

N-soliton solutions to RMBE (6) for $c = b = \Delta = 0$ were found in [24]. To construct similar solutions for system (6), we must fix the vacuum solution with the solitons propagating against its background. In the linear approximation, it is easy to show that solution (12) may be chosen as the vacuum solution if the approximate (15) or exact (14) condition is satisfied.

To construct N-soliton solutions for model (6), it will suffice to modify the corresponding solutions found in [24]. Indeed, the form of the Marchenko integral equations used to construct soliton solutions is determined by the form of the spectral problem and by the asymptotic behavior of the Jost function ($\tau \longrightarrow \pm\infty$). The spectral problem (7) and the Jost function asymptotics for the vacuum solution (12) and condition (15) or (14) match the corresponding spectral problem and the asymptotics in [24]. The solution is also determined by the dependence of the scattering data on variable z, which can be derived by using system (8) for $\tau \longrightarrow \pm\infty$. This dependence is significantly simplified if the functions in the expression for matrix $\mathbf{A}_1$ obey Eqs. (12) and (14). In this case, the nondiagonal elements of matrix $\mathbf{A}_1$ become zero and the z dependence of the spectral data manifests itself only in the appearance of an exponential factor with the exponent

$$-\int_0^z 2[\mathbf{A}_1]_{11}(z'; \lambda(z'))dz'.$$

In the presence of pumping, we must also take into account the z dependence of λ, which is given by Eq. (9). Repeating the construction of soliton solutions in [24], we then find the N-soliton solution to system (6) with PDM and pumping

$$\left|DE - \frac{a\omega}{D}\right|^2 = \frac{d^2}{d\tau^2}\ln\det\|\mathbf{I} + \mathbf{B}\mathbf{B}^*\|, \tag{16}$$

where $\mathbf{I}$ is a unit $(N \times N)$ matrix. Matrix $\mathbf{B}$ has the following elements:

$$B_{nm} = \frac{\sqrt{\beta_n\beta_m}}{\zeta_n - \zeta_m^*}\exp\{i[\zeta_n(z) - \zeta_m^*(z)]\tau\}, \tag{17}$$

where $\zeta_n(z)$ is the solution to the equation

$$\frac{\partial}{\partial z}\zeta_n = -\frac{i}{8\mu}\frac{\zeta_n D^4 c}{\zeta_n^2 D^2 - \mu^2\omega^2} \tag{18}$$

with the boundary condition

$$\zeta_n(0) = \lambda_n.$$

Condition (14) was used to derive Eq. (18). We emphasize that, since β_n also depend on τ in the problem under consideration, it should be remembered when formally using formula (16) that the differentiation is performed only with respect to τ in the exponent on the right-hand side of (17).

In (17), λ_n are either purely imaginary or enter in pairs in the form

$$\lambda_n = -\lambda_m^*,$$

and, using (18), we obtain

$$\beta_n(\tau, z) = \beta_n(0, 0)\exp\left\{i\frac{(2\mu N_0 - Dc\tau)}{Dc}[\zeta_n(z) - \lambda_n]\right\}.$$

In the former case, the following soliton solution corresponds to the only eigenvalue $\lambda_1 = i\eta_1$ and $\beta_1 = \beta_1^*$:

$$DE_s = \frac{a\omega}{D} + \frac{4\nu_1(z)}{\cosh[2\eta_1\tau - 2(\nu_1(z) - \eta_1)\mu N_0(Dc)^{-1} - \phi_1]}; \tag{19}$$

here,

$$\mathrm{Im}\zeta_1(z) = \nu_1(z), \quad \nu_1(0) = \eta_1, \quad \mathrm{Re}\zeta_1(z) \equiv 0,$$

$$\phi_1 = \ln\left|\beta_1(0)[2\zeta_1(z)]^{-1}\right|.$$

The solution describes the increase in soliton amplitude proportional to $\sqrt{z}$, $z \longrightarrow \infty$. The soliton duration does not change during this increase, which distinguishes this solution from a similar soliton solution found in [40] for the Maxwell–Bloch equations with pumping for quasi-monochromatic waves. The mechanism of the increase in soliton amplitude attributable to pumping for $c > 0$ and $b = 0$ is known (see [39, 41]). At the same time, it follows from (9) that the soliton is also amplified for a medium with a nonzero PDM and $ab < 0$, $c = 0$. Note that the coefficient a in actual media can be positive or negative [32]. The possibility of the soliton amplification attributable to a nonzero PDM appears to be pointed out for the first time. Quasi-monochromatic fields with the carrier frequency equal to the transition frequency may be used as pumping proportional to b.

The following breatherlike solution corresponds to the pair of boundary ($z = 0$) eigenvalues $\lambda_{2,3} = \pm\xi + i\eta$ and $\beta_2 = \beta_3^*$:

$$DE_b = \frac{a\omega}{D} + 8\mathrm{Im}\zeta_2\mathrm{Re}\zeta_2 \times \frac{\mathrm{Re}\zeta_2\cos\theta_1\cosh\theta_2 - \mathrm{Im}\zeta_2\sin\theta_1\sinh\theta_2}{|\zeta_2|^2 - (\mathrm{Im}\zeta_2)^2\cos(2\theta_1) + (\mathrm{Re}\zeta_2)^2\cosh(2\theta_2)}, \tag{20}$$

where

$$\theta_1(z, \tau) = 2\xi\tau - 4\mu N_0\frac{\mathrm{Re}\zeta_2(z) - \xi}{Dc} + \arg\left[\frac{\beta(0)}{2\zeta_2(z)}\right],$$

$$\theta_2(z, \tau) = 2\eta\tau - 4\mu N_0\frac{\mathrm{Im}\zeta_2(z) - \eta}{Dc} + \ln\left[\frac{\mathrm{Re}\zeta_2(z)|\beta(0)|}{2\mathrm{Im}\zeta_2|\zeta_2(z)|}\right],$$

$$\zeta_2(0) = \lambda_2.$$

The numerical solution of Eq. (18) under the condition

$$2\mu c(z) - b(z)a = \mathrm{const} > 0$$

indicates that the imaginary part of $\zeta_n(z)$ monotonically increases with z. Depending on the boundary values, the real part can initially increase (for $|\zeta_2| < \omega|\mu|$), reach a maximum at $\sim\omega\mu$, and then asymptotically tend to zero for all boundary values. The phase shift, i.e., the last term in the expression for θ_1, also tends to zero as z increases. In particular, this implies that the breather that corresponds to the pair of eigenvalues $\lambda_1 = -\lambda_2^*$ at the boundary ($z = 0$) is transformed during the amplification into a soliton–antisoliton pair with amplitudes increasing proportionally to $\mathrm{Im}\zeta_1(z)$ and with an increasing separation between them. Figure 1 shows solution (20) at equal intervals of the physical time t for such pumping rates that

$$(2c\mu - ba)D^2 = 1, \quad D\mathrm{Im}\zeta(0) = D\mathrm{Re}\zeta(0) = 1.$$

In this section, we obtained the soliton solutions for specially chosen values of constants (14). At the same time, these solutions also remain approximately valid if this condition is violated for the finite time τ for which the pumping contribution is small. Inequalities (15) imply that the number of electrons transferred to the upper level during the entire process is small compared to their total number in the active medium. This condition can be satisfied if the lower level of a two-level transition coincides with the ground level. As we see from our solution, the soliton amplitude can also increase if conditions (15) are satisfied.

Our soliton solutions are peculiar in that we may pass to the formal limits $a = 0$ or $\mu = 0$ in them. In this case, the solutions do not lose their soliton properties.

This suggests that both mechanisms, one related to electron transfer between the levels ($\mu \neq 0$ in the Bloch equations) and the other related to frequency modulation ($a \neq 0$), independently produce solitons or breathers. At the same time, the joint action of these mechanisms leads to new, nonsoliton solutions of form (12) and to new properties of the system.

3. THE INTEGRABLE MAXWELL–BLOCH EQUATIONS FOR A DEGENERATE TRANSITION WITH PUMPING

Let us consider a Λ scheme in which the magnetic moments $J = 0$ and $J = 1$ correspond to the upper (denoted by 1 in Fig. 2) and lower levels, respectively. The $1 \longleftrightarrow 2, 3$ transitions occur with a change in magnetic-moment projection by one, $\Delta M = -1, 1$. These transitions interact with the mutually orthogonal polarization projections $E_{2,3}$ of the electric field propagating along the x axis. The Maxwell equations in a low-density approximation for active molecules similar to that used above are

$$\left(\frac{\partial}{\partial t} + \frac{c}{\sqrt{\epsilon_\infty}}\frac{\partial}{\partial x}\right)E_j = i\frac{N_0\omega_j}{2c\epsilon_\infty\epsilon_0}(\mu_{1j}^*\rho_{1j} - \mu_{1j}\rho_{j1}), \quad j = 2, 3. \tag{21}$$

The Bloch equations for the density matrix with components ρ_{ij} ($i, j = 1, 2, 3$) of a degenerate two-level medium interacting with this field are

$$\begin{aligned}
-i\partial_t\hbar\rho_{12} &= \mu_{13}\rho_{32}E_3 + \Delta_2 E_2\rho_{12} \\
&\quad -(\rho_{11} - \rho_{22})\mu_{12}E_2 - \hbar\omega_2\rho_{12}, \\
-i\hbar\partial_t\rho_{13} &= \mu_{12}\rho_{23}E_2 + \Delta_3 E_3\rho_{13} \\
&\quad -(\rho_{11} - \rho_{33})\mu_{13}E_3 - \hbar\omega_3\rho_{13}, \\
-i\hbar\partial_t\rho_{23} &= (\mu_{21}\rho_{13}E_2 - \mu_{13}\rho_{21}E_3), \\
-i\hbar\partial_t\rho_{11} &= E_2(\mu_{12}\rho_{21} - \mu_{21}\rho_{12}) \\
&\quad + E_3(\mu_{13}\rho_{31} - \mu_{31}\rho_{13}) + c, \\
-i\hbar\partial_t\rho_{kk} &= E_k(\mu_{k1}\rho_{1k} - \mu_{1k}\rho_{k1}) + c_1, \\
&\quad k = 2, 3.
\end{aligned} \tag{22}$$

Here, $\Delta_j = (\mu_{11} - \mu_{jj})/\hbar$, μ_{ij} are the dipole moment components for the medium.

The Lax representation for the system of equations (21) and (22) can be constructed only if $\Delta_j = 0$ and $\mu_{12} = \mu_{13}$. As above, the coefficients μ_{1k} can be reduced to real form by a simple phase shift of the nondiagonal elements in the density matrix by a constant. Let us pass to the laboratory frame of reference and to the renormalized variables

$$z = x\omega_2\mu_{12}^2 N_0(2\hbar c\epsilon_\infty\epsilon_0)^{-1}, \quad \tau = t - \sqrt{\epsilon_\infty}x/c.$$

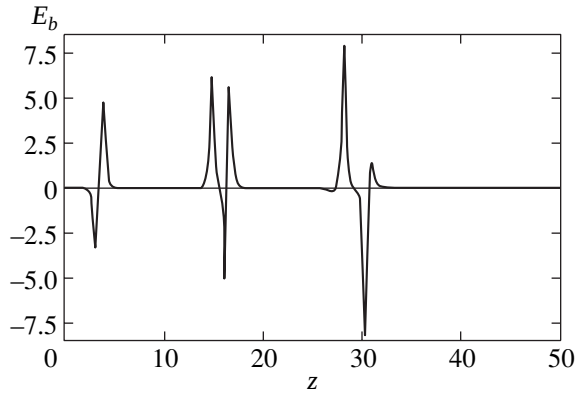

Fig. 1. The dynamics a growing breather at equal time intervals t. The units are arbitrary.

Denote

$$\mathscr{E}_j = E_j\mu_{1j}\hbar^{-1}, \quad R_j^\pm = \rho_{1j} \pm \rho_{j1}, \quad j = 2, 3,$$

$$R_1^\pm = \rho_{23} \pm \rho_{32}, \quad \omega_2 = \omega_3 = \omega$$

and rewrite the Maxwell–Bloch equations for $\Delta_j = 0$ and $\mu_{12} = \mu_{13}$ as

$$\begin{aligned}
\partial_\tau R_k^+ &= i(-1)^{k+1}\mathscr{E}_k R_1^- - i\omega R_k^-, \\
\partial_\tau R_k^- &= i(-1)^{k+1}\mathscr{E}_k R_1^+ - i\omega R_k^+ - 2i(\rho_{11} - \rho_{jj})\mathscr{E}_k, \\
\partial_\tau R_1^+ &= iR_3^-\mathscr{E}_2 + iR_2^-\mathscr{E}_3, \\
\partial_\tau R_1^- &= iR_3^+\mathscr{E}_2 - iR_2^+\mathscr{E}_3, \\
\partial_\tau\rho_{11} &= -i\mathscr{E}_2 R_2^- - i\mathscr{E}_2 R_3^- + c, \\
\partial_\tau\rho_{kk} &= i\mathscr{E}_k R_k^- + c_1, \\
\partial_z\mathscr{E}_k &= iR_k^-;
\end{aligned} \tag{23}$$

here, $k = 2, 3$.

The Lax representation of system (23) is

$$\frac{\partial}{\partial\tau}\Psi = \begin{bmatrix} -2i\lambda & \mathscr{E}_2 & \mathscr{E}_3 \\ -\mathscr{E}_2 & 0 & 0 \\ -\mathscr{E}_3 & 0 & 0 \end{bmatrix}\Psi \equiv \mathbf{L}_2\Psi, \tag{24}$$

$$\begin{aligned}
\frac{\partial}{\partial z}\Psi &= \frac{1}{4\lambda^2 - \omega^2} \\
&\times \begin{bmatrix} 4i\lambda\rho_{11} & \omega R_2^+ + 2\lambda R_2^- & \omega R_3^+ + 2\lambda R_3^- \\ -\omega R_2^+ + 2\lambda R_2^- & 4i\lambda\rho_{22} & -2i\lambda R_1^+ + i\omega R_1^- \\ -\omega R_3^+ + 2\lambda R_3^- & -2i\lambda R_1^+ - i\omega R_1^- & 4i\lambda\rho_{33} \end{bmatrix} \\
&\times \Psi \equiv \mathbf{A}_2\Psi.
\end{aligned} \tag{25}$$

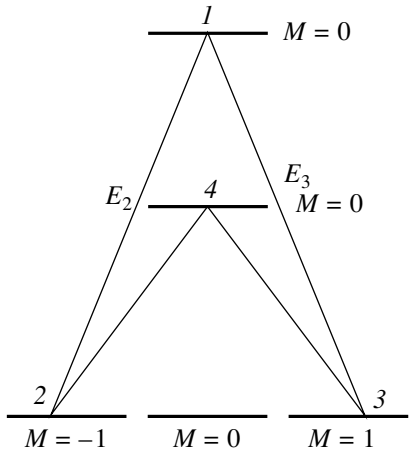

Fig. 2. The transition scheme with a change in magnetic-moment projection by one.

System (23) incorporates the pumping of the upper ($\propto c$) and lower ($\propto c_1$) levels. Only the upper-level pumping may be included in the Lax representation with matrix $\mathbf{L}_2$ that we chose in form (24). This pumping is taken into account by an additional dependence of the spectral parameter on variable z:

$$\frac{\partial}{\partial z}\lambda = -\frac{2c(z)\lambda}{4\lambda^2 - \omega^2}. \tag{26}$$

Including the pumping of the lower level $\propto c_1$ in this Lax representation involves no difficulty. This requires adding a matrix proportional to the unit matrix

$$-i\lambda\frac{2c_1}{c - c_1}\mathbf{I}$$

to $\mathbf{L}_2$. However, since this leads to obvious changes in the results, we set $c_1 = 0$ for simplicity.

The interaction between an extremely short pulse of duration less than π/ω and a multilevel medium differs qualitatively from a similar interaction of quasi-monochromatic waves. The nonresonant nature of the interaction between extremely short pulses manifests itself in the fact that the corresponding evolution equations are more universal and simpler than those for quasi-monochromatic waves. As a physical example that confirms this universality, let us consider a four-level medium with two pairs of transitions (a double Λ scheme) for which the magnetic moment changes by one when passing from level 1 or 4 to levels 2 and 3 during the interaction with a two-component electric field (see Fig. 2).

Below, we show that the RMBE describing this interaction can be reduced to the integrable model (23). Let us write the equations for the wave functions ψ_k of levels $k = 1$–4 as

$$-i\hbar\partial_t\psi_1 = \mu_{12}\psi_2 E_2 + \mu_{13}\psi_3 E_3 + \hbar\omega_1\psi_1, \tag{27}$$

$$-i\hbar\partial_t\psi_2 = \mu_{12}\psi_1 E_2 + \mu_{24}\psi_4 E_2 + \hbar\omega_2\psi_2, \tag{28}$$

$$-i\hbar\partial_t\psi_3 = \mu_{13}\psi_1 E_3 + \mu_{34}\psi_4 E_3 + \hbar\omega_3\psi_3, \tag{29}$$

$$-i\hbar\partial_t\psi_4 = \mu_{42}\psi_2 E_2 + \mu_{43}\psi_3 E_3 + \hbar\omega_4\psi_4. \tag{30}$$

The Maxwell equations in the same low-density approximation as that used above are

$$\left(\frac{\partial}{\partial t} + \frac{c}{\sqrt{\epsilon_\infty}}\frac{\partial}{\partial z}\right)E_j = i\frac{N_0\omega_j}{2c\epsilon_\infty\epsilon_0} \times [(\mu_{1j}\psi_1 + \mu_{4j}\psi_4)\psi_j^* - (\mu_{1j}\psi_1 + \mu_{4j}\psi_4)^*\psi_j], \tag{31}$$

where $j = 2, 3$. Assuming that $\omega_2 = \omega_3$, $\mu_{12} = \mu_{13}$, and $\mu_{42} = \mu_{43}$ and that these coefficients are real, we rewrite system (27)–(31) as

$$-i\partial_t\phi = \mu\hbar^{-1}(\psi_2 E_2 + \psi_3 E_3) + \omega_+\psi + \omega_-\frac{\mu_{12}\psi_1 - \mu_{24}\psi_4}{\mu}, \tag{32}$$

$$-i\partial_t\psi_k = \mu\hbar^{-1}E_k\phi + \omega_2\psi_k, \tag{33}$$

$$\left(\frac{\partial}{\partial t} + \frac{c}{\sqrt{\epsilon_\infty}}\frac{\partial}{\partial z}\right)E_j = i\frac{\mu N_0\omega_j}{2c\epsilon_\infty\epsilon_0}(\phi\psi_j^* - \phi^*\psi_j), \quad j = 1, 2, \tag{34}$$

where

$$\omega_\pm = \frac{\omega_1 \pm \omega_4}{2}, \quad \phi = \frac{\mu_{12}\psi_1 + \mu_{42}\psi_4}{\mu}, \quad \mu = (\mu_{12}^2 + \mu_{42}^2)^{1/2}.$$

For this system to be equivalent to system (23), the last term on the right-hand side of Eq. (32) must be removed. This can be done by assuming that $\omega_+ \gg \omega_-$. On the other hand, system (32)–(34) has the solution

$$\mu_{12}\psi_1 \equiv \mu_{42}\psi_4, \quad \forall x, t,$$

which also reduces this system to the integrable system (23) for all $\omega_\pm$.

4. THE DYNAMICS OF SOLITON POLARIZATION IN A DEGENERATE MEDIUM WITH PUMPING

As was pointed out above, to construct the soliton solutions of (23) requires fixing a stable vacuum solution of the model. We assume the populations of the lower-level sublevels to be larger than the upper-level population during the entire process. The populations of the lower-level magnetic sublevels are initially unequal. For simplicity, we assume that the initial level populations

$$N_k = \rho_{kk}(z, 0), \quad k = 1, 2, 3,$$

do not depend on z. Note that the upper-level population in an actual system can be produced by a thermal mechanism.

We choose a vacuum solution with the solitons propagating against its background in the form

$$E_{2,3}(z,\tau) = 0, \quad R^{\pm}_{2,3}(z,\tau) = 0,$$
$$\rho_{11}(z,\tau) = N_1 + c\tau, \quad \rho_{kk}(z,\tau) = N_k, \qquad (35)$$
$$k = 2,3.$$

We restrict our analysis to the time interval

$$c\tau \ll N_0 - N_1, \quad N_0 = \rho_{11} + \rho_{22} + \rho_{33}.$$

In this case, solution (35) is invariable before and after the soliton passage. A radiative solution is generated for $c\tau \geq N_0 - N_1$. This solution consists of a packet of nonlinear pulsations oscillating about the mean that increases proportionally to $\sqrt{cz}$, $z \longrightarrow \infty$.

The analytic apparatus for our model is basically similar to the apparatus developed by Manakov for the two-component nonlinear Schrödinger equation [10]. The results of this author after a modification related to the evolution of the scattering data and to the z dependence of λ can be applied to our model.

Problem (24), (25) has two sets of solutions: the Jost functions $\varphi_i(\tau, z; \lambda)$ and $\Psi_i(\tau, z, \lambda)$ with the asymptotics

$$(\varphi_i)_k = \delta_{ik}\exp[-2I_k\lambda(z)\tau], \quad \tau \longrightarrow -\infty,$$
$$(\Psi_i)_k = \delta_{ik}\exp[-2I_k\lambda(z)\tau],$$
$$\tau \longrightarrow \infty, \quad I_1 = i, \quad I_m = 0, \quad m = 2,3.$$

The scattering matrix $\mathbf{S} = \{\alpha_{ij}\}$ is given by the relation

$$\varphi_i(\tau, z; \lambda) = \sum_{j=1}^{3} \alpha_{ij}(z; \lambda)\Psi_j(\tau, z; \lambda).$$

The scattering data include the continuum for real λ

$$R_m(z; \lambda) = \alpha_{1m}(z; \lambda)(\alpha_{11}(z; \lambda))^{-1}, \quad \mathrm{Im}\lambda = 0, \quad (36)$$

and bound states composed of zeros ζ_k for $\mathrm{Im}\lambda > 0$. The following coefficients correspond to each such zero:

$$C_m(z; \zeta_k(z)) = \frac{\alpha_{1m}(z; \zeta_k(z))}{\alpha'_{11}(z; \zeta_k(z))}, \quad m = 2,3. \qquad (37)$$

Here,

$$\alpha'_{11} = \frac{d\alpha_{11}}{d\zeta}, \quad \zeta = \zeta_k.$$

The z dependence of these scattering data can be derived from the formula

$$\partial_z \mathbf{S} = \lim_{\tau\to+\infty} \mathbf{A}_2\mathbf{S} - \mathbf{S}\lim_{\tau\to-\infty} \mathbf{A}_2. \qquad (38)$$

Formula (38) is valid for the asymptotic behavior described by the vacuum solution (35), because we chose the conditions under which pumping and injected pulses do not change these asymptotics. If condition (14) is satisfied, matrix $\mathbf{A}_2$ for this solution (35) is diagonal for $t \longrightarrow \pm\infty$. In this case, the z dependence of $C_m(z; \zeta_k)$ derived from (38) is

$$C_m(z; \zeta_k) = C_m(0; \zeta_k)$$
$$\times \exp\left\{\int_0^z \frac{4i\zeta_k(z)[\rho_{11}(z,0) - \rho_{mm}(z,0)]}{4\zeta_k^2(z) - \omega^2}dz\right\}. \qquad (39)$$

We do not give a similar expression for $R_m(z; \lambda)$, because the contribution from the continuum is disregarded here.

For the functions

$$\chi(\tau, z) = [\Psi_1(\tau, z)\exp(2i\zeta(z)\tau), \Psi_2(\tau, z), \Psi_3(\tau, z)]^T,$$

we have the triangular representation

$$\chi_1(\tau, z; \xi) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \frac{1}{2\pi i}$$
$$\times \int_{\mathscr{C}} \sum_m \frac{R_m(z; \zeta)\chi_m(\tau, z; \zeta)\exp(2i\zeta(z)\tau)}{\zeta(z) - \xi(z)}d\zeta, \qquad (40)$$

$$\chi_m(\tau, z; \xi) = \begin{pmatrix} 0 \\ \delta_{m2} \\ \delta_{m3} \end{pmatrix} + \frac{1}{2\pi i}$$
$$\times \int_{\tilde{\mathscr{C}}} \frac{R^*_m(z; \zeta)\chi_1(\tau, z; \zeta)\exp(-2i\zeta(z)\tau)}{\zeta(z) - \xi(z)}d\zeta. \qquad (41)$$

Here, δ_{mk} is the delta function, $m = 2, 3$. The integration is performed along contour $\mathscr{C}$, which includes the real axis and passes above all poles in the upper half-plane and along the mirror reflection $\tilde{\mathscr{C}}$ of this contour.

For a soliton spectrum composed of such N poles ζ_k that either ζ_k are purely imaginary or they enter in pairs $\zeta_n = -\zeta_n^*$, Eqs. (40) and (41) reduce to the algebraic system

$$\chi_1(\tau, z; \xi) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$
$$- \sum_{n=1}^{N}\sum_{m=2,3} \frac{C_{mn}(z; \zeta_n)\chi_m(\tau, z; \zeta_n)\exp(2i\zeta_n\tau)}{\zeta_n(z) - \zeta(z)}, \qquad (42)$$

$$\chi_m(\tau, z; \xi) = \begin{pmatrix} 0 \\ \delta_{m2} \\ \delta_{m3} \end{pmatrix} + \sum_{n=1}^{N} \frac{C^*_{mn}(z; \zeta^*_n)\chi_1(\tau, z; \zeta^*_n)\exp(-2i\zeta^*_n\tau)}{\zeta^*_n(z) - \zeta(z)}. \quad (43)$$

The potential $\mathscr{E}_m(z, t)$ can be reconstructed using the formula

$$\mathscr{E}_m(z, \tau) = \lim_{\zeta \to \infty} 2i\zeta\chi_m(\tau, z)(1, 0, 0). \quad (44)$$

Solving the algebraic system of equations (42) and (43) for a single pole, $\lambda_1 = \zeta_1(0) = i\eta$, we find that the one-soliton solution is described by the projector

$$\mathscr{P}_{ij} = \frac{m_i m^*_j}{\sum_{k=1}^{3} |m_k|^2}. \quad (45)$$

Here, m_k are the vectors that can be determined by taking into account the solution to Eq. (26) for $\lambda = i\eta$, $c(z) = \text{const}$, and $\rho_{kk}(z) = N_k = \text{const}$:

$$m_1 = \alpha_1 \exp\left\{2\eta(0)\tau + N_1\frac{2}{c}[\eta(z) - \eta(0)]\right\},$$
$$m_k = \alpha_k \exp\left\{N_k\frac{2}{c}[\eta(z) - \eta(0)]\right\}, \quad k = 2, 3, \quad (46)$$

where $\alpha_1 = \alpha^*_k$ and $\eta(z)$ is the solution to Eq. (26)

The field components $\mathscr{E}_k$ are described by the projector

$$\mathscr{E}_k = -i(\lambda - \lambda^*)\mathscr{P}_{1k}. \quad (47)$$

For the one-soliton solution (47), we finally obtain

$$\mathscr{E}_2 = \frac{2\eta(z)}{\cosh[\tilde{\theta}_2 + \phi_2] + |\alpha_3^2/\alpha_1\alpha_2|\exp[-\tilde{\theta}_2 + 2\xi(N_3 - N_2)]}, \quad (48)$$

$$\mathscr{E}_3 = \frac{2\eta(z)}{\cosh[\tilde{\theta}_3 + \phi_3] + |\alpha_2^2/\alpha_1\alpha_3|\exp[-\tilde{\theta}_3 - 2\xi(N_3 - N_2)]}, \quad (49)$$

where

$$\tilde{\theta}_2 = 2\eta(0)\tau + (N_1 - N_2)\xi,$$
$$\tilde{\theta}_3 = 2\eta(0)\tau + (N_1 - N_3)\xi,$$
$$\xi = \frac{2}{c}[\eta(z) - \eta(0)], \quad \phi_k = \ln\left|\frac{\alpha_1}{\alpha_k}\right|, \quad k = 2, 3.$$

Let us analyze the asymptotic behavior of solutions (48) and (49). Let $\xi(N_3 - N_2) > 0$. We then have

$$\mathscr{E}_3 \longrightarrow 2\eta(z)\operatorname{sech}[\tilde{\theta}_3 + \phi_3], \quad \mathscr{E}_2 \longrightarrow 0, \quad z \longrightarrow \infty.$$

The derived solutions (48) and (49) describe the transformation of the soliton pair $\mathscr{E}_{2,3}(z = 0)$ into soliton $\mathscr{E}_3(z \longrightarrow \infty)$. The energy transfer takes place at distances

$$z_0 \sim c(N_2 - N_3)^{-2}$$

for large z. The estimate is valid for $z_0 c \gg \eta(0), \omega$. This effect is similar to the Raman transformation of solitons found in [7, 8] in terms of a model for the interaction of quasi-monochromatic waves with a three-level medium. However, in contrast to the results of these studies, the conversion here is determined by the pumping rate constant c. The conversion efficiency was found to decrease with increasing c. As can be seen from the soliton solution, this is because the soliton group velocities and amplitudes level off during the amplification. It can be shown that, in the inverse limit $c \longrightarrow 0$, the distance at which a complete soliton conversion takes place is proportional to $|N_3 - N_2|\eta^{-1}(0)$.

Let us investigate the dynamics of the solution that corresponds to two eigenvalues λ_j. To this end, we simplify the problem by setting $N_2 = N_3$. Below, we give a formal general solution to the algebraic system (42) and (43) for two arbitrary eigenvalues $\lambda_{1,2}$. However, this solution for the field amplitudes is generally complex. Therefore, in order to obtain solutions to model (23), constraints on the form of λ_k and the coefficients must be imposed to ensure that the potential $\mathscr{E}_{2,3}$ be real. Recall that for this problem, $\lambda_{1,2}$ either must be purely imaginary or enter as $\lambda_1 = -\lambda^*_2$.

Let $\lambda_k = i\nu_k$, where ν_k are arbitrary complex numbers. The general two-soliton solution is then

$$\begin{pmatrix} \mathscr{E}_2 \\ \mathscr{E}_3 \end{pmatrix} = \left[\begin{pmatrix} a_1 \\ b_1 \end{pmatrix}e^{\theta_1} + \begin{pmatrix} a_2 \\ b_2 \end{pmatrix}e^{\theta_2} + \begin{pmatrix} d_1 \\ g_1 \end{pmatrix}e^{\theta_1 + \theta_1^* + \theta_2} + \begin{pmatrix} d_2 \\ g_2 \end{pmatrix}e^{\theta_1 + \theta_2 + \theta_2^*}\right]D^{-1}, \quad (50)$$

where

$$D = 1 + p_1 e^{\theta_1 + \theta_1^*} + p_2 e^{\theta_1 + \theta_2^*} + p_2^* e^{\theta_2 + \theta_1^*} + p_3 e^{\theta_2 + \theta_2^*} + p_4 e^{\theta_1 + \theta_1^* + \theta_2 + \theta_2^*},$$

$$p_1 = \frac{q_{11}}{\nu_1 + \nu_1^*}, \quad p_2 = \frac{q_{12}}{\nu_1 + \nu_2^*}, \quad p_3 = \frac{q_{22}}{\nu_2 + \nu_2^*},$$

$$d_1 = \frac{(\nu_1 - \nu_2)(a_1 q_{21} - a_2 q_{11})}{(\nu_1 + \nu_1^*)(\nu_2 + \nu_1^*)},$$

$$d_2 = \frac{(\nu_2 - \nu_1)(a_2 q_{12} - a_1 q_{22})}{(\nu_2 + \nu_2^*)(\nu_1 + \nu_2^*)},$$

$$g_1 = \frac{(\nu_1 - \nu_2)(b_1 q_{21} - b_2 q_{11})}{(\nu_1 + \nu_1^*)(\nu_2 + \nu_1^*)},$$

$$g_2 = \frac{(\nu_2 - \nu_1)(b_2 q_{12} - b_1 q_{22})}{(\nu_2 + \nu_2^*)(\nu_1 + \nu_2^*)},$$

$$p_4 = \frac{|\nu_1 - \nu_2|^2 (q_{11} q_{22} - q_{12} q_{21})}{(\nu_1 + \nu_1^*)(\nu_2 + \nu_2^*)|\nu_2^* + \nu_1|^2},$$

$$q_{ij} = \frac{a_i a_j^* + b_i b_j^*}{\nu_i + \nu_j^*}.$$

The two-soliton solution of (23) corresponds to two real values of $\nu_{1,2}$. The breatherlike solution is given by the relations $\nu_{1,2} = \eta \mp i\xi$, $\eta > 0$. The exponents for the vacuum solution (35) are

$$\theta_{2,3} = 2\nu_{2,3}(0)\tau + \frac{2N_{12}}{c}[\nu_{2,3}(z) - \nu_{2,3}(0)].$$

We numerically found that the form of the breather solution for one component is qualitatively the same as that shown in Fig. 1.

Let us analyze the asymptotics of solution (50) for the two-soliton solution with $\nu_{1,2}(0) = \eta_{1,2}$. Let $\eta_3 > \eta_2$ and let there be no pumping ($c = 0$). In that case, solution (50) for large z splits into two solitons, with the soliton of field $\mathscr{E}_2$ with a smaller amplitude having a higher velocity. The separation between the solitons linearly increases with z. Pumping causes the dynamics of the soliton pair to change. We initially restrict our analysis to one polarization component and choose the condition

$$\nu_1 = 1, \quad \nu_2 = 1.1, \quad a_1 = a_2 = 1, \quad b_1 = 0, \quad b_2 = 0.$$

Since the initial populations of the lower-level magnetic sublevels are equal, no polarization rotation takes place. We see from Fig. 3 that the amplitudes of the growing solitons asymptotically level off and that the separation between them increases much more slowly than in the absence of pumping.

The presence of a second polarization component causes the structure of the soliton pair to change. Figures 4a and 4b show, respectively, the field components $\mathscr{E}_2$ and $\mathscr{E}_3$ in the presence of homogeneous pumping, $c(z) = \text{const}$. The boundary and initial values of the parameters for $\mathscr{E}_2$ are the same as those for the case shown in Fig. 3, except that $b_1 = 0.5$ and $b_2 = 2$. We see from this figure that the dynamics of the soliton pairs if there is amplification differs significantly from the pair dynamics in the absence of pumping. In the former case, a semblance of a quasi-bound soliton state arises. However, this state is not accompanied by a decrease in energy and is attributable solely to the asymptotic leveling of the soliton velocities for arbitrary initial data.

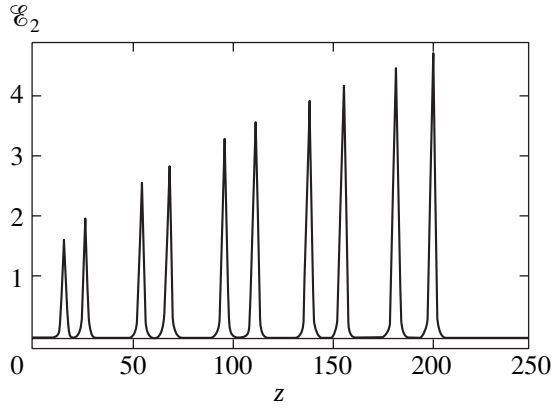

Fig. 3. The dynamics of a soliton pair for one field component $\mathscr{E}_2$ at equal intervals t. The units are arbitrary.

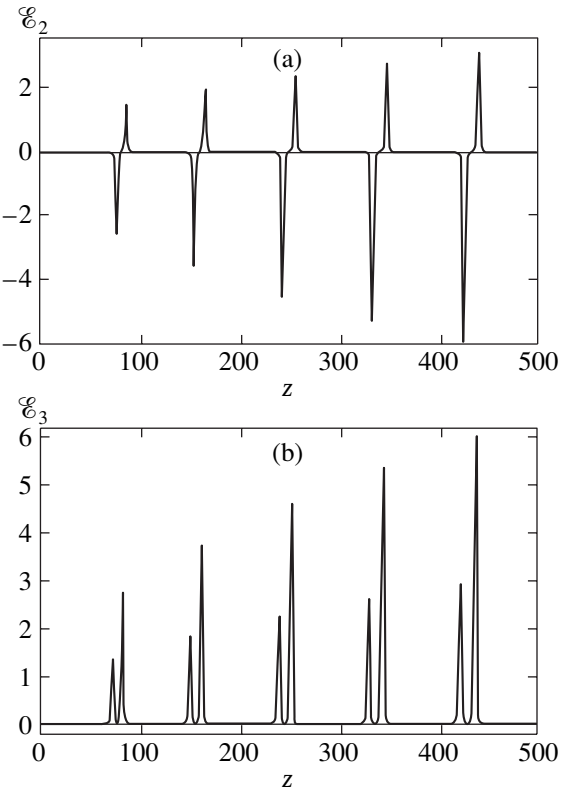

Fig. 4. The two-soliton solution of system (23) for homogeneous pumping at equal intervals t. The amplitudes of the field (a) $\mathscr{E}_2$ and (b) $\mathscr{E}_3$. The units are arbitrary.

5. CONCLUSION

We have investigated the dynamics of pulses with durations close to the oscillation period in a two-level

medium with pumping. The soliton solutions were used to study the dynamics of extremely short pulses. In particular, we used the soliton solutions to the RMBE obtained in Section 2 as examples to show new possibilities for soliton amplification in media with PDM. It follows from these solutions that solitons can be generated in a two-level medium virtually without electron transitions between levels 1 and 2, in contrast to the standard generation mechanism of 2π pulses [9]. Formally, this implies that the dipole moment of the medium in the Bloch equations (3)–(5) becomes zero. In this case, the density of the medium must be high enough. These conditions correspond to the limit

$$\mu \longrightarrow 0, \quad N_0\mu \sim \mathbb{O}(1). \tag{51}$$

If, in addition, the PDM contribution is of the order of unity, then this approximation leads to the simplified integrable system of equations (6). The soliton solutions associated with self-induced transparency also exist in this limit. Thus, we note that the integrable RMBE can be derived not in the low-density approximation but in the approximation of an extremely small dipole moment μ_{12}.

The self-induced transparency solitons related to PDM alone can exist only outside the scope of the quasi-monochromatic approximation in a time interval when the medium polarization instantly tracks the change in level state. Using the Lax representation (7), (8) after passing to the limit (51), we can show that the necessary initial condition for the generation of traveling solitons is a partial but not complete inversion of the medium, i.e., $R^+(0, z) \neq 0$. An ultrashort field pulse with a duration of the order of π/ω gives rise to a coherent coupling between the levels. At the same time, the electric field produces nonlinear oscillations of the upper-level electrons, which results in the generation of stable solitons.

A nonzero PDM also allows a new (compared to the standard) soliton amplification mechanism to be used. In the example considered in Section 2, the additional field that gives rise to a term proportional to b in the equation for the nondiagonal part of the density matrix (3) can cause both an increase and a decrease in soliton amplitude, depending on the sign. This property can be used to modulate a pulse packet. It is of interest to use this amplification mechanism of extremely short pulses to compensate for any losses of the light pulses that propagate in long optical fibers.

The PDM varies over a wide range for different media. The quantum confinement of the carriers in semiconductors gives rise to discrete bands with large oscillator strengths related to band-to-band transitions. A nonzero PDM is possible in asymmetric quantum wells. The dipole moments for $GaAs/Al_{0.14}Ga_{0.86}$ with $Al_{0.3}Ga_{0.3}As$ barriers between the wells are given in [33]. The dipole moments in a three-band scheme are

$$\mu_{12} = 17.6\mu_0, \quad \mu_{13} = 11.1\mu_0, \quad \mu_{23} = 33.4\mu_0,$$

and the corresponding permanent dipole moments are

$$\Delta\mu_{12} = -77.8\mu_0, \quad \Delta\mu_{13} = -72.9\mu_0,$$
$$\Delta\mu_{23} = 4.9\mu_0,$$

where $\mu_0 = 1.6 \times 10^{-29}$ C m. It was noted that the currently available technologies allow the PDM to be increased relative to the transition dipole moment.

In this paper, we did not consider the contribution of the real continuum of problem (7) to the radiative solution. At the same time, the apparatus developed here allows the phenomena related to the generation of the corresponding fields to be studied in terms of the RMBE. In general, for example, when the lower level of the medium does not coincide with the ground level and when initial population is small, the contribution of the radiative solution must be taken into account. This solution for some initial-value–boundary-value problems reduces to the Painlevé V transcendent [44].

At the same time, we used the conditions under which the contribution of the radiative solution may be disregarded. Physically, these conditions correspond to the case where the number of electrons transferred to the upper level through pumping is much smaller than the total number of electrons. Note that some forms of weak nonlinearities lead to the relative suppression of the radiative part of the solution. These can be taken into account in terms of our perturbation theory for almost integrable systems of equations to which the inverse scattering transform with a variable spectral parameter is applicable [42]. A threshold in pumping rate below which no generation is observed is known to exist in experimental laser physics [36]. In terms of the models studied above, this generation corresponds to the radiative solution.

As we showed here, the system of Maxwell–Bloch equations for a degenerate transition with pumping is also integrable in terms of the inverse scattering transform and has multisoliton solutions. The polarization effects in a two-level amplifier can cause a qualitative change in the dynamics and shape of pulses in a nonlinear regime. Under pumping conditions, the velocity and interaction of the solitons that correspond to different polarization components also change. Our solutions demonstrate an effective transformation of the soliton polarization components for different initial magnetic-sublevel populations. The population difference can be produced by a magnetic field, which leads to an electron redistribution over magnetic sublevels. Thus, a soliton counterpart of the Faraday effect arises. It can be analyzed in terms of the Maxwell–Bloch equations. Sazonov [17] investigated this effect for solitons in the low-frequency approximation, which allowed the problem to be reduced to a differentiated nonlinear Schrödinger equation. However, for such a reduction of the system of equations, part of the important information related to the medium is lost. Using the RMBE allows the influence of the initial state of the medium on the evolution of an extremely short pulse in the medium to be studied. As our results show, this influence can be significant.

Equations (6) do not admit a direct integrable generalization of form (23) to a degenerate transition for a nonzero PDM. On the other hand, in the limit of very small μ but for a nonzero polarization of the medium on the $1 \longleftrightarrow 2, 3$ transitions, system (22) breaks down into two individually integrable independent systems. Each of them admits soliton solutions.

In conclusion, note that some of the Maxwell–Bloch equations considered here can also be used to analyze the dynamics of an extremely short pulse with duration $\gg \pi/\omega_0$. However, a number of effects predicted above are possible only for pulse durations $\gtrsim \pi/\omega_0$. The self-induced transparency attributable to a nonzero PDM described above belongs to such effects.

ACKNOWLEDGMENTS

I am grateful to S.V. Manakov for several valuable remarks.

REFERENCES

1. A. I. Maimistov and A. M. Basharov, *Nonlinear Optical Waves* (Kluwer, Dordrecht, 1999).
2. A. I. Maĭmistov, Kvantovaya Élektron. (Moscow) **30**, 287 (2000).
3. S. Novikov, S. V. Manakov, L. P. Pitaevskiĭ, and V. E. Zakharov, *Theory of Solitons: the Inverse Scattering Method* (Nauka, Moscow, 1980; Consultants Bureau, New York, 1984).
4. A. I. Maimistov, Kvantovaya Élektron. (Moscow) **11**, 567 (1984).
5. A. M. Basharov and A. I. Maĭmistov, Zh. Éksp. Teor. Fiz. **87**, 1594 (1984) [Sov. Phys. JETP **60**, 913 (1984)].
6. A. M. Basharov and A. I. Maĭmistov, Zh. Éksp. Teor. Fiz. **94** (12), 61 (1988) [Sov. Phys. JETP **67**, 2426 (1988)].
7. L. A. Bol'shov, N. N. Elkin, V. V. Likhanskiĭ, and M. I. Persiantsev, Zh. Éksp. Teor. Fiz. **88**, 47 (1985) [Sov. Phys. JETP **61**, 27 (1985)].
8. A. A. Zabolotskiĭ, Zh. Éksp. Teor. Fiz. **92**, 46 (1987) [Sov. Phys. JETP **65**, 25 (1987)].
9. G. L. Lamb, Rev. Mod. Phys. **43**, 99 (1971).
10. S. V. Manakov, Zh. Éksp. Teor. Fiz. **83**, 68 (1982) [Sov. Phys. JETP **56**, 37 (1982)].
11. A. A. Zabolotskiĭ, Zh. Éksp. Teor. Fiz. **115**, 1168 (1999) [JETP **88**, 642 (1999)].
12. E. M. Belenov, P. G. Kryukov, A. V. Nazarkin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **47**, 442 (1988) [JETP Lett. **47**, 523 (1988)].
13. E. M. Belenov and A. V. Nazarkin, Pis'ma Zh. Éksp. Teor. Fiz. **51**, 252 (1990) [JETP Lett. **51**, 288 (1990)].
14. E. M. Belenov, A. V. Nazarkin, and V. A. Ushchapovskiĭ, Zh. Éksp. Teor. Fiz. **100**, 762 (1991) [Sov. Phys. JETP **73**, 422 (1991)].
15. A. E. Kaplan and P. L. Schkolnikov, Phys. Rev. Lett. **75**, 2316 (1995).
16. S. V. Sazonov and E. V. Trifonov, J. Phys. A **27**, L7 (1994).
17. S. V. Sazonov, Zh. Éksp. Teor. Fiz. **107**, 20 (1995) [JETP **80**, 10 (1995)].
18. A. Yu. Parkhomenko and S. V. Sazonov, Zh. Éksp. Teor. Fiz. **114**, 1393 (1998) [JETP **87**, 864 (1998)].
19. M. Wittmann, A. Nazarkin, and G. Korn, Opt. Lett. **26**, 298 (2001).
20. B. I. Greene, J. F. Federici, D. R. Dykaar, *et al.*, Appl. Phys. Lett. **59**, 893 (1991).
21. D. You, D. R. Dykaar, and P. H. Bucksbaum, Opt. Lett. **18**, 290 (1993).
22. A. Nazarkin and B. Korn, Phys. Rev. A **58**, R61 (1998).
23. R. K. Bullough and F. Akhmad, Phys. Rev. Lett. **27**, 330 (1971).
24. J. D. Gibbon, P. J. Coudrey, J. K. Eilbeck, and R. K. Bullough, J. Phys. A **6**, 1237 (1973).
25. P. J. Coudrey, J. K. Eilbeck, J. D. Gibbon, and R. K. Bullough, J. Phys. A **6**, L53 (1973).
26. J. K. Eilbeck, P. J. Coudrey, J. D. Gibbon, and R. K. Bullough, Phys. Rev. Lett. **30**, 237 (1973).
27. P. J. Coudrey, J. K. Eilbeck, J. D. Gibbon, and R. K. Bullough, Lett. Nuovo Cimento **8**, 775 (1973).
28. A. I. Maĭmistov and S. O. Elyutin, Opt. Spektrosk. **70**, 101 (1991) [Opt. Spectrosc. **70**, 57 (1991)].
29. A. I. Maimistov and S. O. Elyutin, Chaos, Solitons and Fractals **80**, 369 (1997).
30. R. Bavli and Y. B. Band, Phys. Rev. A **43**, 5039 (1991).
31. J. P. Lavoine, C. Hoerner, and A. A. Villaeys, Phys. Rev. A **44**, 5947 (1991).
32. G. B. Hadjichristov, M. D. Stamova, and P. P. Kircheva, J. Phys. B **28**, 3441 (1995).
33. S. Kŏcinac, Z. Ikonić, and V. Milanović, Opt. Commun. **140**, 89 (1997).
34. M. Agrotis, N. M. Ercolani, S. A. Glasgow, and J. V. Moloney, Physica D (Amsterdam) **138**, 134 (2000).
35. S. V. Manakov, Zh. Éksp. Teor. Fiz. **65**, 505 (1973) [Sov. Phys. JETP **38**, 248 (1973)].
36. O. Svelto, *Principles of Lasers* (Plenum, New York, 1976; Mir, Moscow, 1979).
37. A. A. Apolonsky, A. A. Zabolotskii, V. P. Drachev, and E. I. Zinin, Proc. SPIE **2041**, 385 (1993).
38. A. Hasegawa and Y. Kodama, *Solitons in Optical Communications* (Oxford Univ. Press, Oxford, 1995).
39. S. P. Buptsev, A. V. Mikhaĭlov, and V. E. Zakharov, Teor. Mat. Fiz. **70**, 323 (1987).
40. A. V. Rybin, J. Phys. A **24**, 5235 (1991).
41. S. P. Burtsev and I. R. Gabitov, Phys. Rev. A **49**, 2065 (1994).
42. A. A. Zabolotskiĭ, Zh. Éksp. Teor. Fiz. **120**, 740 (2001) [JETP **93**, 649 (2001)].
43. S. P. Burtsev, I. R. Gabitov, and V. E. Zakharov, *Plasma Theory and Nonlinear and Turbulent Processes in Physics* (World Sci., Singapore, 1988), p. 897.
44. *Painlevé Transcendents. Their Asymptotics and Physical Applications*, Ed. by P. Winternitz and D. Levi (Plenum, New York, 1992), NATO ASI Ser., Ser. B: Phys., Vol. 278.
45. S. P. Burtsev, Phys. Lett. A **177**, 341 (1993).
46. A. V. Kitaev, A. V. Rybin, and J. Timonen, J. Phys. A **26**, 3583 (1993).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 882–891.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1028–1040.
Original Russian Text Copyright © 2002 by Gadomsky, Voronov.

ATOMS, SPECTRA, RADIATION

On a Physical Implementation of Logical Operators NOT and CNOT in a Two-Qubit Quantum Computer Controlled by Ultrashort Optical Pulses

O. N. Gadomsky[a, *] and Yu. Yu. Voronov[b]

[a]*Ul'yanovsk State University, Ul'yanovsk, 432700 Russia*
[b]*Ul'yanovsk Branch, Institute of Radio Engineering and Electronics, Russian Academy of Sciences, Ul'yanovsk, 432700 Russia*
**e-mail: qed_group@mail.ru*

Received October 16, 2001

Abstract—It is shown that it is preferable to perform quantum computations on a system of two-level atoms with metastable states using optical dipole transitions that occur under the effect of ultrashort light pulses. It is suggested to measure the quantum information that is passed to qubits using Bloch, rather than pure, quantum states of two-level atoms. Moreover, the inversion of atoms can be used as the measure of quantum information. In order to describe the logical operators NOT and CNOT in the system of interacting two-level atoms (qubits), modified optical equations for the Bloch vectors of individual qubits are derived. These equations are solved in combination with field equations, without using the slowly varying amplitude approximation, for a small two-qubit system in the field of ultrashort intense optical pulses of arbitrary shape. A numerical analysis of the solution shows that it is possible to control the recording of information on individual qubits in a small quantum system of a dimension much smaller than the length of the optical wave by smoothly varying the irradiation conditions of qubits. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is known that a quantum computer can be constructed from two logical elements—an arbitrary single-qubit unitary transformation and a two-qubit element realizing the controlled NOT (CNOT) [1]. It is shown in [1] that, for the implementation of the NOT element, one should act upon a quantum particle (qubit) by an external intense pulse that takes the qubit from the ground state to an excited state and, conversely, from the excited to the ground state. The CNOT element is implemented by applying an excitation pulse to two interacting qubits. In the process, one qubit controls the evolution of the other one by means of the interaction. In [1], it is suggested to implement the NOT and CNOT operations on the basis of electron and nuclear spins using magnetic resonance spectroscopy [2]. In this paper, we show that it is preferable to design a two-qubit quantum computer on the basis of two-level atoms with metastable states using optical dipole–dipole transitions occurring under the effect of ultrashort light pulses.

By now, suggestions for the implementation of quantum computers on the basis of ions and molecules in laser traps [3], on the basis of nuclear spins of ^{31}P in crystalline silicon [4], on electron spins at quantum points created in a two-dimensional electron gas in GaAs heterostructures [5], and on Josephson junctions [6] have been put forward. Simulation experiments related to quantum computers were carried out using a pulse nuclear magnetic spectrometer [7–10] on the basis of two spins of ^{13}C nuclei and a single proton spin in the trichloroethylene molecule. However, in these experiments, a ensemble quantum computer was studied. The output signals were combined from signals of a large number of molecules in a liquid solution. In this paper, we suggest an implementation of a two-qubit quantum computer on the basis of an arsenic atom dimer on a pure GaAs surface; those dimers can be obtained using well-known techniques [11, 12]. It was shown in [13] that the lines at 3 eV and 4.5 eV, which were observed in [11, 12] in the spectrum of the anisotropic reflection (100) from GaAs surfaces stabilized by arsenic, can be explained on the basis of optical dimensional resonances that were theoretically predicted in [14]. It was also noted that the dispersion properties of the GaAs surface significantly affect the properties of dimensional resonances. Therefore, it is reasonable to implement quantum computers on the basis of two-level atoms (identical or different) by implanting those atoms into surfaces with frequency-independent optical properties.

The concept of quantum information is based on the quantum mechanical superposition principle, which is applied to two-level quantum particles [15–17]. In this paper, we suggest measuring the quantum information using Bloch states, which are well known in resonance spectroscopy [18, 19]. In the process, we have to solve some important problems concerning writing and read-

1063-7761/02/9405-0882$22.00 © 2002 MAIK "Nauka/Interperiodica"

ing quantum information and transferring it from one two-level atom to another. We believe that these problems can be solved using polarizing fields that occur in a system of interacting atoms in an external radiation field.

The concept of polarizing field was suggested in [20–22] on the basis of third-order quantum electrodynamical effects; it takes into account intermediate states with a positive and negative energy in the spectrum of interacting atoms lying at an arbitrary distance from each other. In accordance with the intermediate state type, we distinguish between electron and positron polarizing fields. The scheme of quantum transitions corresponding to the process of inducing the polarizing field is as follows. Let two hydrogen-like atoms be in the ground state with the energy W_0. As a result of exchanging virtual photons, one of the atoms (the polarizing atom) goes to an intermediate state with the energy W_l and then returns to the original state with the energy W_0; in the process, the other atom (the observer) goes to the excited state with the energy W_1. This means that one real photon was absorbed by the two-atom system. It was shown in [20–22] that this process is equivalent to inducing a polarizing field with the corresponding vector potential, which is different from the vector potential of the free photon field at the location of the observer atom. In the electric dipole approximation for classical fields, the electron polarizing field is an electric dipole field induced by the polarization atom at the location of the observer atom. Then, quantum-mechanical considerations make it possible to reveal the nature of the electric dipole field, which is considered as a third-order electrodynamical effect. In [21], a comparison of the physical nature of this field considered as field of virtual photons with another field, which also is a third-order electrodynamical effect but depends only on real photons, was made. In contrast to the polarizing field, as was noted in [21], the transfer of energy from one atom to the other through the real photon field occurs with the characteristic time $t_p = R/c$, where R is interatomic distance and c is the speed of light in vacuum. The polarizing field is not associated with the transfer of energy; therefore, the characteristic time of inducing this field on one atom of the system when the other atom is excited by an external field is different. In this paper, we consider the process of inducing polarizing fields in a strong resonance optical field; in particular, this is of interest for revealing the physical nature of the logical CNOT operator.

A full-scale quantum computer includes about 10^3 qubits [1]; however, the physical foundations of its operation can be clarified on the basis of a two-atom quantum computer that is theoretically studied in this paper. We formulate several key points that provide a basis for the operation of a quantum computer.

1. Before the computer starts to operate, all qubits must be in the ground state. The system of N qubits in the ground state constitutes a memory register [1]. It can store data and perform quantum computations. Evidently, this state is most easily implemented for optical transitions. In this case, two energy levels, W_0 and W_1, exist in the spectrum of two-level atoms such that $W_1 - W_0 \ll kT$, where T is the absolute temperature and k is the Boltzmann constant.

2. A method for controlling the selective action of a pulse upon individual qubits should be devised. One such technique was suggested in [19], whereby two interacting atoms in a nanostructure take different inversion values under the effect of intense stationary radiation that acts upon the nanostructure at various angles. In this paper, we consider a technique for selectively acting upon individual qubits using ultrashort light pulses.

3. The CNOT operation can be implemented in a system consisting of N qubits by exploiting the interaction between them. In this paper, we consider this transformation using the electric dipole–dipole interaction of two-level atoms in the field of pulse radiation in small systems when one of the qubits is excited.

4. In the process of quantum computations, the computer's qubits are affected by relaxation processes that disturb the computation process. The computation process acquires random features during the decoherence time. To increase this time, metastable states can be used that have the lifetime of about 1 s for optical transitions. During this time interval, about 10^{14} elementary unitary transformations can be performed if femtosecond pulses are used.

5. Qubit states must be measured in the process of computations and after their completion. A measurement technique based on the trial pulse radiation and an analysis of interference pattern formed by qubits in the wave zone was suggested in [23].

In this paper, we focus on the physical implementation of the NOT and CNOT operators under the effect of powerful light pulses.

2. MEASURING QUANTUM INFORMATION ON THE BASIS OF BLOCH STATES

The classical theory [24, 25] considers information as a negative contribution to entropy. This way of defining information is appropriate for macroscopic physical systems. It is evident that we need another way for measuring information in quantum systems consisting of a small number of qubits.

Classical computers operate with bistable transistor circuits that demonstrate nonlinear dependence between the input and output voltage [1]. In a quantum computer, the corresponding element is a two-level atom that can be described using the concept of effective spin [18]. We assign the logical zero "0" $\equiv |0\rangle$ to the state W_0 with the wave function $|\Psi_0\rangle$ and the logical unity "1" $\equiv |1\rangle$ to the state with the wave function $|\Psi_1\rangle$ and the energy W_1 ($W_1 > W_0$). Transitions $|0\rangle \longrightarrow |1\rangle$ in

the quantum system correspond to transitions "0" ⟶ "1" in the bistable transistor circuit. It was noted in [1] that the quantum bistable element (qubit) has a new property compared with the classical element. This is the superposition property: a qubit can be in any superposition state $|\Psi\rangle = a|0\rangle + b|1\rangle$, where a and b are complex numbers such that $|a|^2 + |b|^2 = 1$.

Consider a two-level atom with electric dipole transitions from one state to the other. The Hamiltonian of such an atom can be represented in terms of the Pauli matrices σ_1, σ_2, and σ_3 [18]. In the absence of the radiation field, we have

$$H_A = \frac{1}{2}(W_1 + W_0)\hat{I} + \frac{1}{2}(W_1 - W_0)\sigma_3,$$

where $\hat{I}$ is an identity 2-by-2 matrix. The average values of the operators σ_1, σ_2, and σ_3 calculated with the help of the wave functions Ψ of superposition states are written as

$$\begin{aligned} s_1 &= \langle\Psi|\sigma_1|\Psi\rangle = ab^* e^{i\omega_0 t} + a^* b e^{-i\omega_0 t}, \\ s_2 &= \langle\Psi|\sigma_2|\Psi\rangle = -i(ab^* e^{i\omega_0 t} - a^* b e^{-i\omega_0 t}), \\ s_3 &= \langle\Psi|\sigma_3|\Psi\rangle = |b|^2 - |a|^2, \end{aligned} \quad (1)$$

where $W_1 - W_0 = \hbar\omega_0$. It is easily seen that the conservation law

$$s_1^2 + s_2^2 + s_3^2 = 1 \quad (2)$$

holds. The quantity $s_3(W_1 - W_0)/2$, which is the atomic energy, differs from the average energy $(W_1 + W_0)/2$, and s_3 is the inversion of the two-level atom [18]. s_1 and s_2 determine the induced dipole moments of the atom. Thus, quantum information can be measured, for example, using the inversion and taking into account the fact that Eqs. (1) hold if there are two quantum states that form a superposition. For a quantum system to store a bit of quantum information, it is sufficient that the system have two states and satisfy the superposition principle. It is possible to establish a unique correspondence between the Bloch states s_1, s_2, s_3 and the quantum superposition states.

Taking into account relation (2) between the inversion and the induced dipole moment of the atom, we may conclude that the measure of quantum information is determined in terms of physically observable quantities. The pseudospin vector $\mathbf{s} = (s_1, s_2, s_3)$ in the energy space traces curves on the unit sphere for various values of the superposition coefficients a and b, which vary under the effect of the radiation field. For individual atoms that interact with the radiation field at the frequency ω, it is convenient to change from the variables s_1, s_2, and s_3 to the variables u, v, and w using the transformations

$$\begin{aligned} s_1 &= u\cos(\omega t) - v\sin(\omega t), \\ s_2 &= u\sin(\omega t) + v\cos(\omega t), \quad s_3 = w. \end{aligned} \quad (3)$$

Here, $u^2 + v^2 + w^2 = 1$ and u, v, and w satisfy the Bloch optical equations [18]

$$\begin{aligned} \dot{u} &= -(\omega_0 - \omega)v, \quad \dot{v} = (\omega_0 - \omega)u + \kappa E_0 v, \\ \dot{w} &= -\kappa E_0 v, \end{aligned} \quad (4)$$

where $\kappa = 2d_0/\hbar$, E_0 is the envelope of the optical pulse acting on the atom, and d_0 is the transition dipole moment. Thus, the behavior of an individual qubit in the process of computations can be represented in terms of the observable quantities u, v, and w, which describe the atomic local dipole moments and its inversion and, therefore, the quantum information encoded in the atom.

All available transformations, such as the Hadamard transformation [1]

$$H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad H|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle), \quad (5)$$

that contain the variables a and b in the quantum superposition of states can be replaced by the corresponding unitary transformations that transform the observable variables u, v, and w. Comparing (1) with (3) for $\omega = \omega_0$, we obtain the following correspondence between the variables:

$$\begin{aligned} u = 1, v = 0, w = 0 &\longleftrightarrow a = \frac{1}{\sqrt{2}}, b = \frac{1}{\sqrt{2}}, \\ u = -1, v = 0, w = 0 &\longleftrightarrow a = \frac{1}{\sqrt{2}}, b = -\frac{1}{\sqrt{2}}, \\ u = 0, v = 1, w = 0 &\longleftrightarrow a = \frac{1}{\sqrt{2}}, b = -\frac{i}{\sqrt{2}}, \\ u = 0, v = -1, w = 0 &\longleftrightarrow a = -\frac{i}{\sqrt{2}}, b = \frac{1}{\sqrt{2}}, \\ u = 0, v = 0, w = -1 &\longleftrightarrow a = 1, b = 0, \\ u = 0, v = 0, w = 1 &\longleftrightarrow a = 0, b = 1. \end{aligned} \quad (6)$$

Below, we derive modified optical Bloch equations that apply to two-level systems with regard for the interatomic interaction. These new equations will replace Eqs. (4).

3. EQUATIONS OF MOTION OF A TWO-ATOMIC SYSTEM IN A RADIATION FIELD IN TERMS OF ATOMIC AND FIELD VARIABLES

In this section, we analyze properties of a single-qubit NOT element and a two-qubit CNOT element in the field of optical radiation. To this end, we consider optical properties of a quantum system consisting of

two (generally different) atoms in the field of an intense wave. The Hamiltonian of such a system has the form

$$H = \frac{1}{2}\hbar\sum_{j=1}^{2}\omega_{0j}\sigma_{3j} - \frac{1}{2}\sum_{j=1}^{2}\sigma_{+j}(\mathbf{d}'_{0j} + i\mathbf{d}''_{0j})\cdot\mathbf{E}_j - \frac{1}{2}\sum_{j=1}^{2}\sigma_{-j}(\mathbf{d}'_{0j} - i\mathbf{d}''_{0j})\cdot\mathbf{E}_j. \quad (7)$$

Here $\mathbf{E}_1$ and $\mathbf{E}_2$ are the electric field strengths at the location points of the first and second atoms, respectively; ω_{01} and ω_{02} are the frequencies of the atoms' transitions; the dipole moment operators

$$\mathbf{d}_j = \frac{1}{2}\sigma_{+j}(\mathbf{d}'_{0j} + i\mathbf{d}''_{0j}) + \frac{1}{2}\sigma_{-j}(\mathbf{d}'_{0j} - i\mathbf{d}''_{0j})$$

are determined in terms of the real, $\mathbf{d}'_{0j}$, and imaginary, $\mathbf{d}''_{0j}$, parts of the dipole moments of the first and second atom transitions; and $\sigma_{\pm j} = \sigma_{1j} \pm i\sigma_{2j}$ and σ_{aj} are the operators of effective spin of the jth atom ($\alpha = 1, 2, 3$ and $j = 1, 2$). We have

$$[\sigma_{1j}, \sigma_{2k}] = 2i\sigma_{3j}\delta_{jk}, \quad [\sigma_{2j}, \sigma_{3k}] = 2i\sigma_{1j}\delta_{jk},$$
$$[\sigma_{3j}, \sigma_{1k}] = 2i\sigma_{2j}\delta_{jk}, \quad [\sigma_{+j}, \sigma_{-k}] = 4\sigma_{3j}\delta_{jk}, \quad (8)$$
$$[\sigma_{+j}, \sigma_{3k}] = -2\sigma_{+j}\delta_{jk}, \quad [\sigma_{-j}, \sigma_{3k}] = 2\sigma_{-j}\delta_{jk},$$

where δ_{jk} is the Kronecker delta.

Let us derive Heisenberg's equations of motion using the commutation relations (8) and Hamiltonian (7) for the operators

$$\sigma_{+j} = \sigma_{+j}(\mathbf{d}'_{0j} + i\mathbf{d}''_{0j}), \quad \sigma_{-j} = \sigma_{-j}(\mathbf{d}'_{0j} - i\mathbf{d}''_{0j}).$$

Calculations yield the following system of equations:

$$\dot\sigma_{+j} = i\sigma_{+j}\omega_{0j} + \frac{2i}{\hbar}(\mathbf{d}'_{0j} + i\mathbf{d}''_{0j})\sigma_{3j}((\mathbf{d}'_{0j} - i\mathbf{d}''_{0j})\mathbf{E}_j),$$
$$\dot\sigma_{-j} = -i\sigma_{-j}\omega_{0j} - \frac{2i}{\hbar}(\mathbf{d}'_{0j} - i\mathbf{d}''_{0j})\sigma_{3j}((\mathbf{d}'_{0j} + i\mathbf{d}''_{0j})\mathbf{E}_j), \quad (9)$$
$$\dot\sigma_{3j} = \frac{i}{\hbar}\sigma_{+j}\mathbf{E}_j - \frac{i}{\hbar}\sigma_{-j}\mathbf{E}_j.$$

We represent the field that acts on the atoms in the quantum system as

$$\mathbf{E}_j = \mathbf{E}_{0j}e^{-i\omega t}, \quad \mathbf{E}_{0j} = \mathbf{E}'_j - i\mathbf{E}''_j,$$

where $\mathbf{E}'_j$ and $\mathbf{E}''_j$ are the real and imaginary parts of the field at the location point of the jth atom. If $\mathbf{E}_j$ are not operators, we pass to the mean values in Eqs. (9) in the usual fashion. For classical fields, the form of Eqs. (9) does not change when passing to mean values. Defining the transformations

$$\sigma_{1j} = u_j\cos(\omega t) - v_j\sin(\omega t),$$
$$\sigma_{2j} = u_j\sin(\omega t) + v_j\cos(\omega t), \quad \sigma_{3j} = w_j,$$
$$\sigma_{\pm j} = (u_j \pm v_j)e^{\pm i\omega t},$$
$$\sigma_{\pm j} = (\mathbf{d}'_{0j} \pm i\mathbf{d}''_{0j})(u_j \pm iv_j)e^{\pm i\omega t}$$

and the notation

$$(u_j + iv_j)\mathbf{d}^*_{0j} = \mathbf{X}^*_j, \quad (u_j - iv_j)\mathbf{d}_{0j} = \mathbf{X}_j, \quad \mathbf{d}_{0j} = \mathbf{d}'_{0j} - i\mathbf{d}''_{0j}, \quad (10)$$

we obtain the equations

$$\dot{\mathbf{X}}_j = -i\mathbf{X}_j(\omega_{0j} - \omega) - \frac{2i}{\hbar}w_j|\mathbf{d}_{0j}|^2\mathbf{E}_{0j},$$
$$\dot{\mathbf{X}}^*_j = i\mathbf{X}^*_j(\omega_{0j} - \omega) - \frac{2i}{\hbar}w_j|\mathbf{d}_{0j}|^2\mathbf{E}^*_{0j}, \quad (11)$$
$$\dot w_j = \frac{i}{\hbar}\mathbf{X}^*_j\cdot\mathbf{E}_{0j} - \frac{i}{\hbar}\mathbf{X}_j\mathbf{E}^*_{0j}.$$

Here, we took into account that

$$\mathbf{d}_{0j}(\mathbf{d}^*_{0j}\cdot\mathbf{E}_{0j}) = |\mathbf{d}_{0j}|^2\mathbf{E}_{0j}.$$

This equality is proved for $\mathbf{d}_{0j} \parallel \mathbf{E}_{0j}$.

The changeover from Eqs. (9) to Eqs. (11) corresponds to the changeover to the rotating reference frame that is well known in resonance optical spectroscopy [18]. Equations (11) obey the following conservation laws:

$$\frac{d}{dt}(|\mathbf{X}_j|^2 + w_j^2|\mathbf{d}_{0j}|^2) = 0;$$

therefore,

$$u_j^2 + v_j^2 + w_j^2 = 1. \quad (12)$$

We will call Eqs. (11) the modified Bloch equations. These equations are set up with regard for the fact that the fields and dipole moments are complex-valued. Below, we show that this allows us to describe the mutual influence of dipoles in the two-atom quantum system.

In the general case, Eqs. (11) should be supplemented by relaxation terms by phenomenologically introducing the phase relaxation, T'_{2j}, and the energy relaxation, T_{1j}, of the jth atom of the system. Then, instead of Eqs. (11), we obtain the equations

$$\dot{\mathbf{X}}_j = i\mathbf{X}_j\Delta_j + \frac{2i}{\hbar}w_j|\mathbf{d}_{0j}|^2\mathbf{E}^*_{0j} - \frac{\mathbf{X}_j}{T'_{2j}},$$
$$\dot{\mathbf{X}}^*_j = -i\mathbf{X}^*_j\Delta_j - \frac{2i}{\hbar}w_j|\mathbf{d}_{0j}|^2\mathbf{E}_{0j} - \frac{\mathbf{X}^*_j}{T'_{2j}},$$
$$\dot w_j = \frac{i}{\hbar}\mathbf{X}^*_j\cdot\mathbf{E}_{0j} - \frac{i}{\hbar}\mathbf{X}_j\cdot\mathbf{E}^*_{0j} - \frac{w_j - w_{0j}}{T_{1j}},$$

where $\Delta_j = \omega_{0j} - \omega$ is the detuning from the jth atom resonance and w_{0j} is the initial value of the jth atom inversion. We consider the interaction of atoms with a field of short light pulses with duration much less than the relaxation times T'_{2j} and T_{1j}.

The electric field strength obeys the equation

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_I(\mathbf{r}, t) + \sum_{j=1}^{2} \operatorname{curl}\operatorname{curl}\frac{\mathbf{p}_j(t - R_j/c)}{R_j}, \quad (13)$$

where $\mathbf{r}$ is an arbitrary observation point (it may coincide with the location of the atoms), $\mathbf{E}_I(\mathbf{r}, t)$ is the external field that affects the system,

$$\mathbf{p}_j = \frac{1}{2}\mathbf{X}_j e^{-i\omega t} + \text{c.c.} \quad (14)$$

is the induced dipole moment of the jth atom, c is the speed of light in vacuum, and $R_j = |\mathbf{r} - \mathbf{r}_j|$. The combination of Eq. (13) with Eqs. (11) forms a closed system of equations; we will use it to describe various optical properties of the two-atom quantum system.

Using Eq. (13), we explicitly represent the fields $\mathbf{E}_{01}$ and $\mathbf{E}_{02}$ in Eqs. (11). We single out the negative frequency parts of those fields, which are proportional to $\exp(-i\omega t)$. Let the axis of the two-atom system be directed along the axis y. Then, atom 1 has the radius vector $\mathbf{r}_1(0, 0, 0)$, and atom 2 has the radius vector $\mathbf{r}_2(0, R, 0)$. The field at the location of atom 2 has the form

$$\begin{aligned}
\mathbf{E}_2 = (\mathbf{E}'_{02} - i\mathbf{E}''_{02})\exp(-i\omega t) = \mathbf{y}_0\Big(&E_{0I}\exp(i\mathbf{k}_0\cdot\mathbf{R})e_{Iy} \\
&+ 2G\exp(i\mathbf{k}_0\cdot\mathbf{R})[p^y_{01}] + \frac{2\exp(i\mathbf{k}_0\cdot\mathbf{R})}{cR^2}[\dot{p}^y_{01}]\Big) \\
&\times\exp(-i\omega t) + \mathbf{x}_0\Big(E_{0I}\exp(i\mathbf{k}_0\cdot\mathbf{R})e_{Ix} \\
&- F\exp(i\mathbf{k}_0\cdot\mathbf{R})[p^x_{01}] - \frac{\exp(i\mathbf{k}_0\cdot\mathbf{R})}{cR^2}[\dot{p}^x_{01}] \\
&+ \frac{2ik_0\exp(\exp(i\mathbf{k}_0\cdot\mathbf{R}))}{cR}[\dot{p}^x_{01}]\Big)\exp(-i\omega t) \\
&+ \mathbf{z}_0\Big(E_{0I}\exp(i\mathbf{k}_0\cdot\mathbf{R})e_{Iz} - F\exp(i\mathbf{k}_0\cdot\mathbf{R})[p^z_{01}] \\
&- \frac{\exp(i\mathbf{k}_0\cdot\mathbf{R})}{cR^2}[\dot{p}^z_{01}] \\
&+ \frac{2ik_0\exp(\exp(i\mathbf{k}_0\cdot\mathbf{R}))}{cR}[\dot{p}^z_{01}]\Big)\exp(-i\omega t).
\end{aligned} \quad (15)$$

Here $\mathbf{x}_0$, $\mathbf{y}_0$, and $\mathbf{z}_0$ are the coordinate unit vectors; E_{0I} is the real amplitude of the external wave with the polarization vector $\mathbf{e}_I = \mathbf{e}'_I + i\mathbf{e}''_I$ and the wave vector $\mathbf{k}_0$;

$$G = \frac{1}{R^3} - i\frac{k_0}{R^2}, \quad F = G - \frac{k_0^2}{R}$$

are the parameters of the dipole–dipole interaction of the atoms; $k_0 = \omega/c$; $\mathbf{p}_{0j} = \mathbf{X}_j/2$; and the symbol $[\ldots]$ denotes that the corresponding quantity is determined at the time instant $t - R/c$. A similar expression can be written for the field $\mathbf{E}_1$ at the location of atom 1; generally, this field is different from $\mathbf{E}_2$.

3.1. A Small Two-Atom System

Expressions for $\mathbf{E}_1$ and $\mathbf{E}_2$ become considerably simpler when $k_0R \longrightarrow 0$, i.e., when the system dimensions are much less than the external radiation wave length. In this case, we may neglect the lag in the dipole–dipole interaction of the atoms, which yields $G = F = 1/R^3$; we also may neglect the terms proportional to $\dot{\mathbf{p}}_j$ in expression (15) for $\mathbf{E}_j$. Then, for the fields at the location points of atoms 1 and 2, we have

$$\mathbf{E}_{01} = E_{0I}\mathbf{e}_{0I} + \frac{1}{2}\hat{G}\mathbf{X}_2, \quad \mathbf{E}_{02} = E_{0I}\mathbf{e}_I + \frac{1}{2}\hat{G}\mathbf{X}_1, \quad (16)$$

where the tensor $\hat{G}$ is

$$\begin{bmatrix} -F & 0 & 0 \\ 0 & 2G & 0 \\ 0 & 0 & -F \end{bmatrix}. \quad (17)$$

Substitute (16) into Eqs. (11) to obtain a closed system of equations in the unknowns $\mathbf{X}_j$ for the given value of the external field in the form of a short pulse. Moreover, Eqs. (16) and (11) for the field and atomic variables hold for optical pulses of an arbitrary shape and duration, including femtosecond pulses.

4. NOT OPERATOR IN A SYSTEM OF INDEPENDENT ATOMS

Consider the case of independent atoms when, according to (16), the field at the location of the atoms (for example, of atom 1) is equal to the external field. This means that, under certain irradiation conditions of atom 2, $\mathbf{X}_2 = 0$.

Let us write Eqs. (11) in matrix form. For atom 1, we have

$$\frac{d}{dt}\begin{bmatrix} X_{1y} \\ X^*_{1y} \\ w_1|\mathbf{d}_{01}| \end{bmatrix} = \hat{M}_{1y}\begin{bmatrix} X_{1y} \\ X^*_{1y} \\ w_1|\mathbf{d}_{01}| \end{bmatrix}, \quad (18)$$

where

$$\hat{M}_{1y} = \begin{bmatrix} -i\Delta_1 & 0 & -\frac{2i}{\hbar}|\mathbf{d}_{01}|E_{01y} \\ 0 & i\Delta_1 & \frac{2i}{\hbar}|\mathbf{d}_{01}|E^*_{01y} \\ -\frac{i}{\hbar}|\mathbf{d}_{01}|E^*_{01y} & \frac{i}{\hbar}|\mathbf{d}_{01}|E_{01y} & 0 \end{bmatrix}. \quad (19)$$

Similar equations can be obtained for the other components of the induced dipole moments of atom 1. In the case of independent atoms, we have $\mathbf{E}_{01} = \mathbf{E}_{0I}$ in (19).

Consider a rectangular pulse acting on atom 1. Then, the solution to Eq. (18) is written in the form

$$\begin{bmatrix} X_{1y} \\ X^*_{1y} \\ w_1|\mathbf{d}_{01}| \end{bmatrix} = \exp(\hat{M}_{1y}t)\begin{bmatrix} X_{1y} \\ X^*_{1y} \\ w_1|\mathbf{d}_{01}| \end{bmatrix}_{t_0}, \quad (20)$$

where t_0 is the initial instant of time. For a rectangular pulse, the unitary operator $\exp(\hat{M}_{1y}t)$ can be represented as

$$\exp(\hat{M}_{1y}t) = \hat{M}^2_{1y}\frac{1}{\Omega^2_{1y}}(1-\cos(\Omega_{1y}t)) + \hat{M}_{1y}\frac{1}{\Omega_{1y}}\sin(\Omega_{1y}t) + \hat{I}, \quad (21)$$

where

$$\Omega_{1y} = \sqrt{\Delta_1^2 + \frac{4}{\hbar^2}|\mathbf{d}_{01}|^2|E_{01y}|^2}.$$

Taking into account (20) and (21), we determine the variation of the atom inversion in time. For $t_0 = 0$, we have

$$w_1(t) = \frac{1}{\Omega^2_{1y}}(1-\cos(\Omega_{1y}t))\left[-\Delta_1\frac{1}{\hbar}E^*_{01y}X_{1y}(0) - \Delta_1\frac{1}{\hbar}E_{01y}X^*_{1y}(0) - \frac{4}{\hbar^2}|\mathbf{d}_{01}|^2|E_{01y}|^2 w_1(0)\right] + \frac{1}{\Omega_{1y}} \times \sin(\Omega_{1y}t)\left[-\frac{i}{\hbar}E^*_{01y}X_{1y}(0) + \frac{i}{\hbar}E_{01y}X^*_{1y}(0)\right] + w_1(0). \quad (22)$$

We point out some properties of solution (22). If d''_{0j} and E''_{0j} are zero, this solution coincides with Rabi's solution [18]. The change in polarization of the external field does not affect the atom inversion if $|E_{01y}| = |E_{01\beta}|$, where $\beta = x, z$. In the case of the exact resonance, the inversion of the two-level atom varies in the range from –1 through 1 for the corresponding values of the field.

In order to obtain a complete description of the atomic Bloch state, one should, in addition to the inversion w, determine the quantities u and v, which determine the pseudospin or Bloch vector. Setting $\omega = \omega_{01}$, we find from solution (20) that

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 1-2\sin^2\varphi\sin^2\left(\frac{\Omega}{2}t\right) & -\sin 2\varphi\sin^2\left(\frac{\Omega}{2}t\right) & \sin\varphi\sin(\Omega t) \\ -\sin 2\varphi\sin^2\left(\frac{\Omega}{2}t\right) & 1-2\cos^2\varphi\sin\left(\frac{\Omega}{2}t\right) & \cos\varphi\sin(\Omega t) \\ -\sin\varphi\sin(\Omega t) & -\cos\varphi\sin(\Omega t) & \cos(\Omega t) \end{bmatrix}\begin{bmatrix} u_0 \\ v_0 \\ w_0 \end{bmatrix}, \quad (23)$$

where

$$\Omega = \frac{2|\mathbf{d}_{01}||E_{01}|}{\hbar}, \quad e^{i\varphi} = \frac{E_{01}}{|E_{01}|},$$

and u_0, v_0, and w_0 are the values of the corresponding quantities at $t = 0$. For convenience, we omit the subscripts on the variables u, v, and w. At $\varphi = 0$, solution (23) to Eq. (20) turns into Rabi's solution [18] of the Bloch optical equations for an isolated rectangular pulse in the case of the exact resonance.

Assume that the initial state of the atom is $u = v = 0$, $w = w_0$, where $w_0 = \pm 1$. When the atom is affected by a rectangular pulse of duration $\tau = \pi/\Omega$, it goes to the state $u = 0$, $v = 0$, $w = -w_0$, as is seen from (23). Thus, using the so-called 180° pulse, we can change the atom state to the inverse one or, in other words, perform the logical NOT operator.

In the general case, operator (23) corresponds to the rotation of the Bloch vector of the initial state by the angle Ωt about an axis that lies in the plane uv and is determined by the rectangular pulse phase φ. If $w = \pm 1$ at the initial state, the axis of rotation makes no difference when the 180° turn is made. If we make a 90° turn, i.e., use rectangular pulses of duration $\tau = \pi/2\Omega$, the situation is different.

As in the models of quantum computers based on the NMR [1], the single-qubit Hadamard operator (5) is very useful. It can be realized by 90° rotations about the axes u and v. For such rotations, pulses with the phase 0 and $\pi/2$, respectively, should be used.

Assume that the atom is in the state $u = v = 0$, $w = -1$ or $|0\rangle$. Let it be affected by a sequence of two pulses. The first of them makes a rotation about the axis u by 180°, and the second one makes the rotation about the axis v by 90°. As a result, the atom goes to the state $u = 1$, $v = 0$, $w = 0$. According to (6), this state is associated with the wave function

$$|\Psi\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle.$$

If the atom is initially in the state $|1\rangle$, then the same sequence of pulses takes it to the state

$$|\Psi\rangle = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle.$$

5. CNOT OPERATOR IN A SYSTEM OF INTERACTING ATOMS

Consider the case of a small quantum system in which the field at the location of atoms 1 and 2 is determined by relations (16). In this case, we write the solution to Eqs. (11) as

$$\begin{bmatrix} X_{j\gamma}(t) \\ X^*_{j\gamma}(t) \\ |\mathbf{d}_{0j}| w_j(t) \end{bmatrix} = \exp\left(\int_{t_0}^{t} \hat{M}_{j\gamma}(t')dt'\right) \begin{bmatrix} X_{j\gamma}(t_0) \\ X^*_{j\gamma}(t_0) \\ |\mathbf{d}_{0j}| w_j(t_0) \end{bmatrix}, \quad (24)$$

where $\gamma = x, y, z$. Depending on the subscript γ in (24), the corresponding components of tensor (17) should be used when determining the field that acts on the jth atom of the quantum system.

We represent the unitary operator in (24) in the form

$$\exp\left(\int_{t_0}^{t} \hat{M}_{j\gamma}(t')dt'\right) = \prod_{k=1}^{N_t} \exp\{\hat{M}^{(k)}_{j\gamma}(t_k - t_{k-1})\}, \quad (25)$$

where N_t is the number of short time intervals between the instant t_0 of switching on the external field and the current instant t. We assume that the field that affects atoms 1 and 2 at every kth interval is a rectangular pulse. Since these intervals are short, we have

$$\exp(\hat{M}^{(k)}_{j\gamma}\Delta t_k) = (\hat{M}^{(k)}_{j\gamma})^2 \frac{1}{2}(\Delta t_k)^2 + \hat{M}^{(k)}_{j\gamma}\Delta t_k + \hat{I}, \quad (26)$$

where the matrices $\hat{M}^{(k)}_{j\gamma}$ have the form (19) with the corresponding numerical values of the involved quantities for the kth time intervals. Making the number of time intervals sufficiently large, we can solve Eqs. (11) with any desired accuracy. Below, we use this algorithm for solving Eqs. (11) to analyze the influence of various factors on the interaction of atoms in the radiation field and on the evolution of the quantum system state with time. This will give us a good idea of how the logical CNOT operator can be implemented.

5.1. Factors That Determine the Properties of Atom Interaction

Assume that the qubit atoms of the small system are identical, i.e., $|\mathbf{d}_{01}| = |\mathbf{d}_{02}|$, $\omega_{01} = \omega_{02}$, and the frequency of the external field is such that resonance condition is fulfilled, i.e., $\Delta_1 = \Delta_2 = 0$ in Eqs. (11) and in the matrices $\hat{M}_{j\gamma}$. Nevertheless, the atomic inversions can be different.

According to (16), the field acting on atoms 1 and 2 is composed of the external and internal fields. Moreover, the internal field is self-consistent. This implies that the field at the location of atom 1 depends on the field at the location of atom 2 and inversely. This is described by the nonlinear terms in Eqs. (11) for inversion and in the evolution operators (25) and (26).

Let us define two time intervals describing the behavior of the two-atom quantum system. This is $T_{act} = 2\pi/\Omega$, where $\Omega = 2|\mathbf{d}_{01}||E_{0I}|/\hbar$, which corresponds to the action of the external field, and $T_{int} = 2\pi/\Omega_b$, which corresponds to the interaction of atoms in the small quantum system. T_{act} and T_{int} must be considerably less than the phase and relaxation times $T_{11} = T_{12} = T_1$ and $T'_{21} = T'_{22} = T'_2$ of the atoms in the system under consideration.

Let $\mathbf{d}_{01}$ and $\mathbf{d}_{02}$ be directed along the line passing through each atom and have the same length d. In this case, the maximum value of the internal field is $E_b = d/R^3$. It was shown above that the action of any field on the atom causes the change in its state with the Rabi frequency. In our case, this frequency is $\Omega_b = 2dE_b/\hbar$.

It is important for quantum computations that the interaction interval is much less than the relaxation time. Therefore, the atoms and the distance between them should be chosen so that $T_{int} = 2\pi/\Omega_b \ll T_{1j}, T'_{2j}$. For example, if we take atoms for which $d = 10^{-18}$ esu and place them at the distance $R = 10$ nm, then the interaction interval $T_{int} \sim 10^{-9}$ s. For the case $R = 1$ nm, we have $T_{int} \sim 10^{-12}$ s.

If $\mathbf{d}_{01}$ and $\mathbf{d}_{02}$ are directed perpendicularly to the line connecting the atoms, then $E_b = d/2R^3$. In this case, Ω_b is half as large and the interaction interval is twice as long.

Note that the interaction interval T_{int} and the interval of the external field action T_{act} can be substantially different. Indeed,

$$\frac{T_{int}}{T_{act}} = \frac{E_{0I}R^3}{d};$$

hence, if the external field $E_{0I} > d/R^3$, then the interval of the external field action is less than the interaction interval.

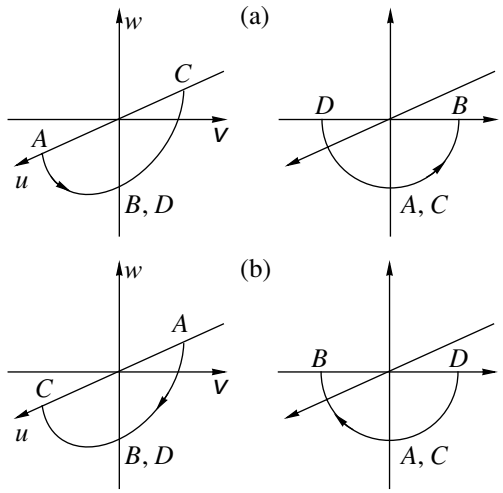

Fig. 1. Evolution of the quantum state of the system in the absence of the external field in the case when the second qubit atom is initially in the ground state and the first atom is in the state $H|0\rangle$ (a) and $H|1\rangle$ (b).

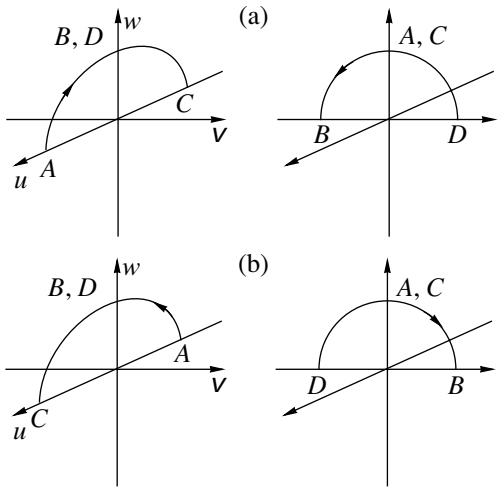

Fig. 2. Evolution of the quantum state of the system in the absence of the external field in the case when the second qubit atom is initially in an excited state and the first atom is in the state $H|0\rangle$ (a) and $H|1\rangle$ (b).

5.2. The Exchange of Information between Qubits under Various Initial Conditions

In order to determine the response of a two-atom system to an action, it is sufficient to know the response of four basic states $|00\rangle$, $|10\rangle$, $|01\rangle$, and $|11\rangle$ to this action. For the first state, the inversion of both atoms is –1; for the second state, the inversion of the first atom is 1 and of the second one is –1; and so on. Accordingly, the effect of the CNOT operator on the system is as follows:

$$\begin{aligned} |00\rangle &\longrightarrow |10\rangle, \quad |10\rangle \longrightarrow |00\rangle, \\ |01\rangle &\longrightarrow |01\rangle, \quad |11\rangle \longrightarrow |11\rangle. \end{aligned} \tag{27}$$

Although CNOT is one of the basic logical operations, it can be implemented by a set of elementary unitary transformations in each of which either one qubit atom or both are involved. Below, we analyze transformations of the second type assuming that $R = 10$ nm and $d = 10^{-18}$ esu.

Assume that no external field affects any of the atoms. In this case, there exists a set of initial conditions under which the quantum system does not evolve. For example, let $|w_1| = |w_2| = 1$. Then, first of all, we have from (12) that $u_1 + iv_1 = 0$ and $u_2 + iv_2 = 0$. Second, (16) implies that $E_1 = E_2 = 0$. Taking this fact into account, we obtain from Eqs. (11) $\dot{\mathbf{X}}_j = 0$ and the value of $\dot{w}_j$. It is seen that this state of the system does not vary with time. Such a situation occurs if the value of u_1, v_1, or w_1 for one of the atoms is ±1 and, for the other atom, $u_2 = \pm u_1$, $v_2 = \pm v_1$, and $w_2 = \pm w_1$.

We see that in order to use this type of interatomic interaction for the implementation of the CNOT operator, one of the atoms should first be put in a superposition state. Figures 1 and 2 show the evolution of the basic states in time when the first qubit atom is affected by the Hadamard operator. The point determining the Bloch state of the atom performs a periodic motion along an arc of unit circle while successively taking the states *A, B, C, D,* and *A*. The time required to complete a 90° turn about a certain axis is $T_{\text{int}}/4 = 7.5 \times 10^{-10}$ s. It is easily seen that the evolution proceeds differently in each case. In particular, the evolution of the first qubit-atom state depends on the state of the second qubit atom at the initial instant of time.

Assume that the second atom is known to be in the ground state and is affected by an external field about 1000 times less strong than the internal field. Figure 3 illustrates the evolution of the Bloch state of the system in two cases—when the first atom is in the ground and in the excited state. If both atoms were initially in the ground state (Fig. 3b), then they will be in the same state after a while if small amplitude oscillations are neglected. If the first atom was excited, then it goes to the ground state after a while, and the second atom goes to the excited state. In between, there are rather long time intervals (delays) when nothing happens. These delays are the longer the less the external field strength.

5.3. Switching between the NOT and CNOT Operators

It was mentioned above that the operation of a quantum computer requires that individual qubits could be affected selectively. The assumption on the localization of the external field in the region of an individual qubit does not suggest that the action is selective, since the qubits interact. We show that, under certain excitation conditions, the CNOT operator can be transformed into the NOT operator.

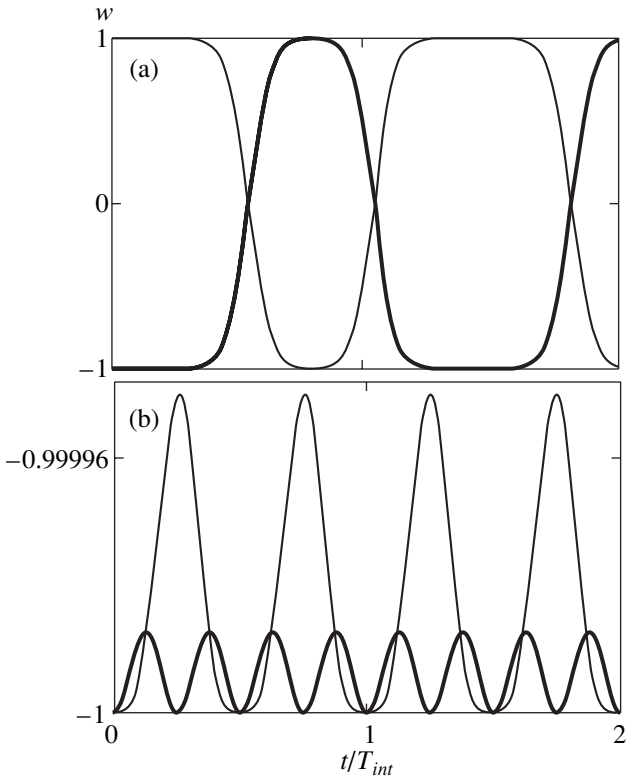

Fig. 3. The dependence of inversion on time in the case when the external field at the location of the second qubit atom is $E_{0I} = 0.005$ esu. The bold curve corresponds to the inversion of the second qubit atom and the normal line corresponds to that of the first qubit atom. (a) Atom 1 is initially in the excited state, and atom 2 is in the ground state; (b) both atoms are initially in the ground state.

Let there be two identical atoms the first of which is affected by an external field with the strength greater than that of the internal field that affects both the first and the second atom. Figure 4 illustrates the evolution of the Bloch state of the system for various strengths of the external field. In case (a), when the internal field $E_b = 1$ esu is ten times as strong as the external one, there is practically no evolution. In cases (b) and (c), the states of both atoms vary substantially. In case (d), when the external field is five times as strong as the internal one, the state of the second atom varies substantially, while the first atom state undergoes only insignificant variations.

6. CONCLUSION

In this paper, a mode of operation of a quantum computer is suggested. It is based on a small system of two-level atoms and makes use of optical quantum transitions and the dipole–dipole interaction of atoms in the radiation field. Basic problems that must be solved when designing a quantum computer are formulated, and it is shown that the mode of operation suggested in the paper has certain advantages over other types of quantum computers.

In the theory of quantum information and quantum computations, both individual quantum states of qubits and entangled states are used. We suggest using Bloch

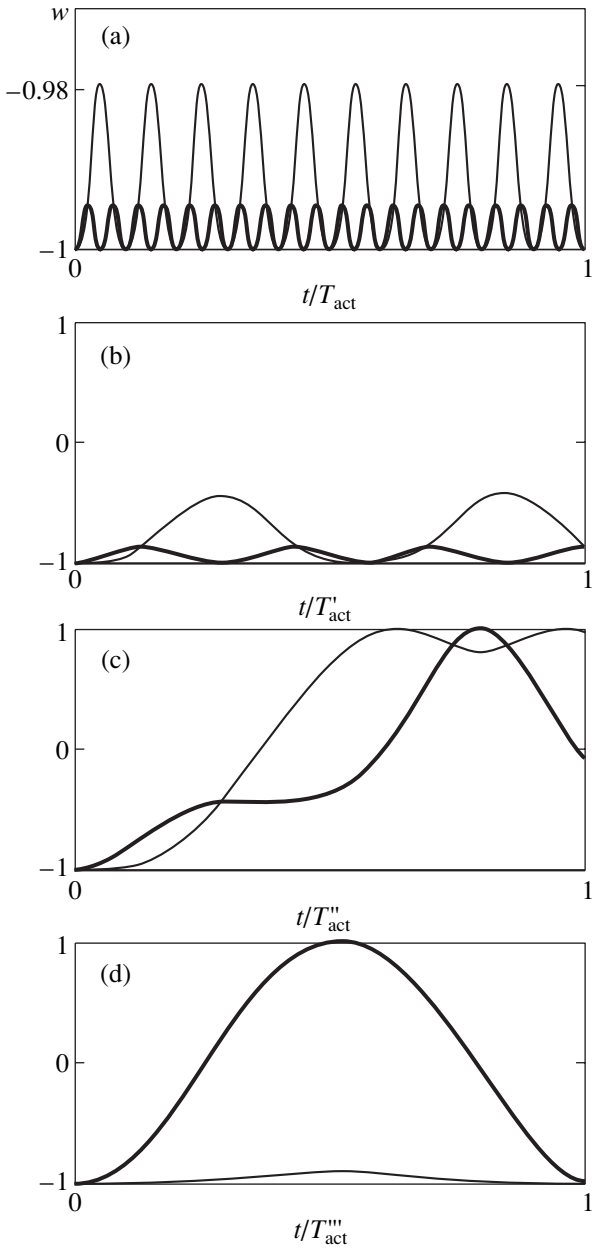

Fig. 4. The dependence of inversion on time for various strengths of the external field at the location of the second atom: E_{0I} = (a) 0.1, (b) 0.5, (c) 1, (d) 5 esu. The bold curve corresponds to the inversion of the second qubit atom, and the normal line corresponds to that of the first qubit atom. The scale of time is different for each figure; 3×10^{-8} s = $T_{act} > T'_{act} > T''_{act} > T'''_{act} = 6 \times 10^{-10}$ s; $T_{act} = 2\pi/\Omega$, where $\Omega = 2|\mathbf{d}_{01}||E_{0I}|/\hbar$. The intervals T'_{act}, T''_{act}, and T'''_{act} are different from T_{act} in the strength of the external field E_{0I}.

states instead. Bloch states are related to the observable quantities through such quantities as induced dipole moments and inversions of qubit atoms. We show that all operators used in the theory of quantum information can be implemented on the basis of Bloch states, and entangled states are implemented with regard for the dipole–dipole interaction of atoms.

In order to describe the mode of operation of a quantum computer, we derive modified optical Bloch equations for the vector operators of effective spin of two-level atoms and the atom inversion that take into account the amplitude and phase properties of all physical quantities determining the mutual influence of qubit atoms in the process of executing the NOT and CNOT logical operators.

A solution to the modified optical Bloch equations for ultrashort optical pulses with duration much less than the time of the phase and energy relaxation of atoms including femtosecond pulses is obtained.

A numerical analysis of this solution to the consistent system of equations in atomic and field variables for the small two-atomic quantum system consisting of two identical atoms in the case of the exact resonance is carried out. We assume that individual qubit atoms can be selectively affected by an external field so that the evolution of the Bloch states of individual qubits and internal polarizing fields can be traced.

In this paper, we suggest controlling the evolution of atoms using short and ultrashort light pulses with duration much less than the time of the phase and energy relaxation. The external field is considered as a classical one, which corresponds to powerful pulses that can substantially change the inversion and local dipole moments of qubits.

Properties of the NOT and CNOT logical operators under various excitation conditions and initial inversion of the atoms are analyzed. It is shown that, when one of the atoms is excited by an ultrashort external light pulse, a polarizing field at the location of the other atom occurs with a certain time delay. The delay is determined by the strength of the internal and external fields rather than by the photon transit time of the interatomic distance. The possibility of switching the NOT and CNOT operators using a smooth change of the strength of the external radiation field is shown.

The analysis of the solution to the equations of motion shows the importance of the results for the physical implementation of a quantum computer. A further analysis of this solution requires a separate consideration.

ACKNOWLEDGMENTS

The work was supported by the Ministry of Education of the Russian Federation (project no. E00-3.2-136); the "Integration" Federal Program (project no. A-0066); and the "Universities of Russia" Program.

REFERENCES

1. K. A. Valiev, Usp. Fiz. Nauk **169**, 691 (1999).
2. A. Abragam, *The Principles of Nuclear Magnetism* (Clarendon, Oxford, 1961; Inostrannaya Literatura, Moscow, 1963).
3. J. I. Cirac and P. Zoller, Phys. Rev. Lett. **74**, 4091 (1995).
4. B. E. Kane, Nature (London) **393**, 133 (1998).
5. D. Loss and D. P. DiVincenzo, Phys. Rev. A **57**, 120 (1998).
6. A. Shnirman, G. Schoen, and Z. Hermon, Phys. Rev. Lett. **79**, 2371 (1997).
7. I. L. Chuang, Lieven M. K. Vandersypen, D. W. Leung, *et al.*, Nature (London) **393**, 143 (1998).
8. I. L. Chuang, N. Gershenfeld, M. G. Kubinec, and D. W. Leung, Proc. R. Soc. London, Ser. A **454**, 447 (1998).
9. D. G. Cory, A. F. Farmy, and T. F. Havel, Proc. Natl. Acad. Sci. USA **94**, 1634 (1997).
10. J. A. Jones, M. Mosca, and R. H. Hansen, Nature (London) **393**, 344 (1998).
11. V. L. Berkovits and D. Paget, Thin Solid Films **233**, 9 (1993).
12. V. L. Berkovits and D. Paget, Appl. Phys. Lett. **61**, 1835 (1992).
13. O. N. Gadomsky and K. Yu. Moiseev, Opt. Spektrosk. **92**, 613 (2002) [Opt. Spectrosc. **92**, 560 (2002)].
14. O. N. Gadomsky and A. S. Kunitsyn, Zh. Prikl. Spektrosk. **67**, 777 (2000).
15. R. P. Feynman, Int. J. Theor. Phys. **21**, 467 (1982).
16. R. P. Feynman, Opt. News **11**, 11 (1985).
17. R. P. Feynman, Found. Phys. **16** (6), 507 (1986).
18. L. Allen and J. H. Eberly, *Optical Resonance and Two-Level Atoms* (Wiley, New York, 1975; Mir, Moscow, 1978).
19. O. N. Gadomsky and Yu. Yu. Voronov, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 750 (1999) [JETP Lett. **69**, 804 (1999)].
20. O. N. Gadomsky and K. V. Krutitsky, J. Eur. Opt. Soc. B **9**, 343 (1997).
21. O. N. Gadomsky and K. K. Altunin, Zh. Éksp. Teor. Fiz. **114**, 1555 (1998) [JETP **87**, 842 (1998)].
22. O. N. Gadomsky, Usp. Fiz. Nauk **170**, 1145 (2000).
23. O. N. Gadomsky and Yu. V. Abramov, Opt. Spektrosk. **90**, 727 (2001) [Opt. Spectrosc. **90**, 649 (2001)].
24. L. Brillouin, *Science and Information Theory* (Academic, New York, 1956; Fizmatgiz, Moscow, 1960).
25. L. Brillouin, *Scientific Uncertainty and Information* (Academic, New York, 1964; Mir, Moscow, 1966), p. 28.

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 892–900.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1040–1050.
Original English Text Copyright © 2002 by Medvedev, Shapiro, Fedoruk, Turitsyna.

ATOMS, SPECTRA, RADIATION

The Theory of Optical Communication Lines with a Short-Scale Dispersion Management¶

S. B. Medvedev*, E. G. Shapiro, M. P. Fedoruk, and E. G. Turitsyna

Institute of Computational Technologies, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia
Institute of Automation and Electrometry, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia
**e-mail: mife@ict.nsc.ru*

Received November 9, 2001

Abstract—We investigate, theoretically and numerically, properties of dispersion-managed (DM) solitons in fiber lines with the dispersion compensation period L much shorter than the amplification distance Z_a. We present the path-averaged theory of DM transmission lines with a short-scale management in the case of asymmetric maps. Applying a quasi-identical transformation, we demonstrate that the path-averaged dynamics in such systems can be described by an integrable model in some limits. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Realization of soliton-based optical data transmission has clearly demonstrated how results of fundamental soliton theory (see, e.g., [1–12]) can be successfully used in very important practical applications. The dispersion management technique proposed recently provides for an increase in the bit rate per channel and leads to suppression of the interchannel interaction in WDM systems in comparison with traditional soliton transmission [13]. The dispersion-managed (DM) soliton is a novel type of an optical information carrier with many attractive properties (see, e.g., [15–57] and references therein) combining features of the traditional fundamental soliton and the dispersion-managed non-return-to-zero transmission. The power of the DM soliton is enhanced [19] compared to the corresponding fundamental soliton. This increases the signal-to-noise ratio, reduces the Gordon–Haus jitter, and therefore improves the transmission system performance. However, in the systems (transmission regimes) limited by nonlinear pulse interactions rather than by noise, the enhanced soliton power can become a less attractive feature. For instance, data transmission with high bit rates of 40 Gb/s per channel and more requires a dense pulse packing and, consequently, short soliton widths. The DM soliton energy increases with the decrease in the pulse width (or, in other words, with the increase in the map strength). The average power of the traditional soliton signal increases with the increase in the bit rate (assuming the soliton width to be a fraction of the time slot) as the square of the bit rate. For the DM soliton, this growth is even more drastic, and for short pulses, the DM soliton power can therefore become too high to be realized in practice [55]. Additionally, soliton interaction becomes an important issue as the signal power increases [55]. The energy control by the corresponding reduction of the average dispersion is limited by fluctuations of the dispersion along the fiber and by higher order dispersive effects. Therefore, in designing soliton-based (and also general return-to-zero signal) transmission systems, the soliton power must be kept sufficiently large for the signal-to-noise ratio requirement and suppressed jitter and, at the same time, not too large to avoid strong soliton interaction and to meet the telecommunication standards on signal power. One way to find such an optimum for a high-bit-rate DM transmission is to use a chirped-return-to-zero signal [55, 56] with less power than the DM soliton power in the corresponding system. Even though such carriers are not stable in a rigorous mathematical sense and emit radiation as they propagate, they can be successfully used in practical systems. A challenge for the soliton theory, however, is to find high-bit-rate (≥ 40 Gb/s per channel) transmission regimes with a truly periodic soliton-like signal propagation. Short-scale dispersion management is a means of controlling the DM soliton energy while keeping the average dispersion not too small and taking advantage of the four-wave-mixing (FWM) suppression in the WDM transmission by a high local dispersion.

The traditional dispersion management for long-haul transmission assumes the amplification distance to be much shorter than the dispersion compensation period (see, e.g., [14]). Another important application is the implementation of dispersion-compensating schemes in the existing terrestrial fiber links based on standard monomode fibers, which typically requires rather close spacing of the dispersion compensating fibers because of the high dispersion of standard monomode fibers at 1.55 μm. In this case, the amplification distance is typically of the order of the compensation period. The existing technologies make it possible to

¶This article was submitted by the authors in English.

manufacture fibers with the continuous alternation of positive and negative dispersion sections of a few kilometers long without any splicing [27]. The fundamental properties of the optical signal transmission in this regime are less studied compared to other regimes. In this paper, we investigate the optical pulse transmission in DM fiber systems with a compensation length that is much shorter than the amplification distance [41]. We examine the case of an asymmetric dispersion map. Compared to lossless models, systems with different periods of the amplification (Z_a) and dispersion compensation (L) possess an important new degree of freedom, the parameter L/Z_a. A short-scale dispersion compensation ($L \ll Z_a$) leads to a reduction of the DM soliton power if we fix all system parameters and the pulse width and vary only L/Z_a. Below, we show that short-scale management can be considered as a possibility of an advantageous practical realization of the weak-map regime.

2. THE BASIC MODEL

We first recall the basic equations and the notation. The optical pulse propagation in a cascaded transmission system with varying dispersion is governed by

$$i\frac{\partial E}{\partial z} + \frac{\lambda_0^2 D(z)}{4\pi c_l}\frac{\partial^2 E}{\partial t^2} + \frac{2\pi n_2}{\lambda_0 A_{\text{eff}}}|E|^2 E = i\left[-\gamma(z) + r_k\sum_{k=1}^{N}\delta(z-z_k)\right]E = iG(z)E, \quad (1)$$

where z is the propagation distance in km, t is the retarded time in ps, $|E|^2 = P$ is the optical power in W, and $D(z)$ is the group velocity dispersion measured in ps/(nm km). We assume a periodic dispersion management with the period L, $D(z + L) = D(z)$; z_k are the amplifier locations. We consider a periodic amplification with the period Z_a. If $\gamma = \gamma_k$ is constant between two adjacent amplifiers, then $r_k = [\exp(\gamma_k Z_a) - 1]$ is the amplification coefficient after the fiber span between the kth and $(k-1)$th amplifier, n_2 is the nonlinear refractive index, A_{eff} is the effective fiber area, $\gamma = 0.05\ln 10\alpha$ (with α measured in dB/km) is the fiber loss of the corresponding fiber, c_l is the speed of light, and $\lambda_0 = 1.55$ μm is the carrier wavelength. We consider the general case where L and Z_a are rational and commensurable, namely, $nZ_a = mL = Z_0$ with integer n and m. In this paper, we focus on systems with short-scale management with $n = 1$, $m > 1$, and $Z_0 = Z_a = mL$. It is customary to pass from the original optical field $E(z, t)$ to

$$A(z, t) = E(z, t)\exp\left[\int_0^z G(z')dz'\right].$$

The evolution of the scaled envelope A is then given by the nonlinear Schrödinger (NLS) equation with periodic coefficients

$$iA_z + d(z)A_{tt} + \epsilon c(z)|A|^2 A = 0, \quad (2)$$

where

$$\epsilon c(z) = \frac{2\pi n_2}{\lambda_0 A_{\text{eff}}}\exp\left[2\int_0^z G(z')dz'\right], \qquad d(z) = \frac{\lambda_0^2 D(z)}{4\pi c_l}. \quad (3)$$

3. THE PATH-AVERAGED MODEL

In this section, we briefly recall the derivation of the path-average model [28, 42] describing the change in the signal waveform over one compensation period. Equation (3) governing the z *e*volution of an optical pulse can be written in the Hamiltonian form

$$i\frac{\partial A}{\partial z} = \frac{\delta H}{\delta A^*} = -d(z)A_{tt} - \epsilon c(z)|A|^2 A \quad (4)$$

with the Hamiltonian

$$H = \int\left\{d(z)|A_t|^2 - \frac{\epsilon c(z)}{2}|A|^4\right\}dt. \quad (5)$$

The true breathing soliton is a solution to Eq. (3) of the form

$$A(z, t) = \exp(ikz)F(z, t)$$

with a periodic function $F(z + Z_0, t) = F(z, t)$. It is interesting to find a systematic way to describe a family of periodic solutions F with different quasi-momenta k. The basic idea suggested in [28] is to use the small parameter ϵ to derive a path-averaged model that gives a regular description of the breathing soliton in the leading order in ϵ. Averaging cannot be performed directly in Eq. (1) in the case of large variations

$$\tilde{d} \gg \langle d\rangle,$$

where

$$d(z) = \tilde{d} + \langle d\rangle \text{ with } \langle\tilde{d}\rangle = 0.$$

However, a path-averaged propagation equation can be obtained in the frequency domain [28]. We show that, in some important limits, the averaged equation for the periodic breathing pulse can be transformed to the integrable NLS equation.

First, to eliminate the periodic dependence of the linear part, we follow [28] in applying the so-called Floquet–Lyapunov transformation

$$A_\omega = \phi_\omega\exp\{-i\omega^2 R(z)\}, \quad \frac{dR(z)}{dz} = d(z) - \langle d\rangle, \quad (6)$$

where $A_\omega = A(z, \omega)$ is the Fourier transform of

$$A(z, t) = \int A_\omega \exp[-i\omega t] d\omega.$$

An important observation used in what follows is that, for a fixed amplitude of d, the amplitude of the variation of R decreases as $m = Z_a/L$ increases. It can easily be found that

$$\max[R(z)] \propto 1/m.$$

In the new variables, the equation becomes

$$i\frac{\partial \phi_\omega}{\partial z} - \langle d \rangle \omega^2 \phi_\omega + \epsilon \int G_{\omega 123}(z) \delta(\omega + \omega_1 - \omega_2 - \omega_3) \times \phi_1^* \phi_2 \phi_3 d\omega_1 d\omega_2 d\omega_3 = 0, \quad (7)$$

where

$$G_{\omega 123}(z) = c(z) \exp\{i\Delta\Omega R(z)\}$$

is Z_a-periodic and

$$\Delta\Omega = \omega^2 + \omega_1^2 - \omega_2^2 - \omega_3^2.$$

We note that $G_{\omega 123}$ depends only on the specific combination of the frequencies given by the resonance surface $\Delta\Omega$. Both the Fourier transform and Floquet–Lyapunov transform (6) are canonical, and the transformed Hamiltonian H is given by

$$H = \langle d \rangle \int \omega^2 |\phi_\omega|^2 d\omega - \varepsilon \int \frac{G_{\omega 123}}{2} \delta(\omega + \omega_1 - \omega_2 - \omega_3) \times \phi_\omega^* \phi_1^* \phi_2 \phi_3 d\omega d\omega_1 d\omega_2 d\omega_3. \quad (8)$$

It is important that ε and $\langle d \rangle$ are small, and Eq. (7) therefore has the so-called Bogolyubov standard form and the averaging procedure can then be applied. We now apply the Hamiltonian averaging [50, 51]. We change the variables as

$$\phi_\omega = \varphi_\omega + \epsilon \int V_{\omega 123} \delta(\omega + \omega_1 - \omega_2 - \omega_3) \times \varphi_1^* \varphi_2 \varphi_3 d\omega_1 d\omega_2 d\omega_3 + \ldots,$$

where

$$V_{\omega 123}(z) = i\int [G_{\omega 123}(\tau) - T_{\omega 123}] d\tau + iV_{\omega 123}(0),$$

$$\langle V_{\omega 123} \rangle = 0$$

with

$$T_{\omega 123} = \langle G_{\omega 123} \rangle = \int_0^1 G_{\omega 123}(z) dz = \int_0^1 c(z) \exp\{i\Delta\Omega R(z)\} dz. \quad (9)$$

In the leading order in ϵ, the path-averaged evolution of signal in the DM line is governed by the Gabitov–Turitsyn model [28]

$$i\frac{\partial \varphi_\omega}{\partial z} - \langle d \rangle \omega^2 \varphi_\omega + \epsilon \int T_{\omega 123} \delta(\omega + \omega_1 - \omega_2 - \omega_3) \times \varphi_1^* \varphi_2 \varphi_3 d\omega_1 d\omega_2 d\omega_3 = 0. \quad (10)$$

The Hamiltonian averaging introduced here represents a regular way to calculate the next-order corrections to the averaged model. We note that Eq. (10) possesses a remarkable property. The matrix element $T_{\omega 123} = T(\Delta\Omega)$ is a function of $\Delta\Omega$, and on the resonant surface given by

$$\omega + \omega_1 - \omega_2 - \omega_3 = 0,$$

$$\Delta\Omega = \omega^2 + \omega_1^2 - \omega_2^2 - \omega_3^2 = 0,$$

both $T_{\omega 123}$ and its derivative with respect to $\Delta\Omega$ are regular. This observation allows us to make a quasi-identical transformation that eliminates the variable part of the matrix element $T_{\omega 123}$

$$\varphi_\omega = a_\omega - \frac{\epsilon}{\langle d \rangle} \int \frac{T_0 - T_{\omega 123}}{\Delta\Omega} \times a_1^* a_2 a_3 \delta(\omega + \omega_1 - \omega_2 - \omega_3) d\omega_1 d\omega_2 d\omega_3, \quad (11)$$

where $T_0 = T(0)$. This transformation has no singularities. If the integral part in this transformation is small compared to a_ω, then, in the leading order, we obtain

$$i\frac{\partial a_\omega}{\partial z} - \langle d \rangle \omega^2 a_\omega + \epsilon \int T_0 \delta(\omega + \omega_1 - \omega_2 - \omega_3) \times a_1^* a_2 a_3 d\omega_1 d\omega_2 d\omega_3 = 0. \quad (12)$$

This is nothing else but the integrable nonlinear Schrödinger equation written in the frequency domain. Obviously, this transformation is quasi-identical only if the integral in Eq. (11) is small compared to a_ω. This is not true in the general case, and that is why the path-averaged DM soliton given by the solution to Eq. (10) then has a form different from the cosh-shaped NLS equation soliton [28, 43, 49]. A comprehensive analysis of the DM soliton solutions to the Gabitov–Turitsyn equation has been published in [46–48]. The first high-precision numerical solution of the Gabitov–Turitsyn equation was presented in [48]. We note that, if the kernel function in Eq. (11) is small,

$$|S(\Delta\Omega)| = \left| \frac{T_0 - T_{\omega 123}(\Delta\Omega)}{\Delta\Omega} \right| \ll 1, \quad (13)$$

then the averaged model can be reduced to the NLS equation. In other words, this is a condition on the functions $c(z)$ and $d(z)$ that makes the quasi-identical transformation possible. The path-averaged DM soliton propagation in systems satisfying requirement (13) is close to the dynamics of the traditional soliton and at

the same time preserves all the advantages of the suppression of FWM by a high local dispersion.

4. SYSTEMS WITH A SHORT-SCALE MANAGEMENT

In this section, we calculate the matrix element $T_{\omega 123}$ for systems with a short-scale management ($L \ll Z_a$) and demonstrate that a path-averaged propagation (even with large variations of the dispersion) can be described by the integrable NLS equation in this regime. The matrix element T plays an important role in the description of the FWM [52]. To be specific, we consider a two-step dispersion map with the amplification distance $Z_a = Z_0$ ($n = 1$) and dispersion compensation period $L = Z_a/m$ km. The dispersion is

$$d(z) = d + \langle d \rangle$$

if

$$\frac{k}{m} < \frac{z}{Z_a} < \frac{k+a}{m}$$

and

$$d(z) = \frac{da}{a-1} + \langle d \rangle$$

if

$$\frac{k+a}{m} < \frac{z}{Z_a} < \frac{k+1}{m},$$

where $k = 0, 1, 2, \ldots, m-1$ and the parameter $a \in (0, 1)$ describes the position of the step. The mean-free function R defined above can be found as

$$R(z) = d(z - Z_a k/m - aZ_a/(2m))$$

if

$$\frac{k}{m} < \frac{z}{Z_a} < \frac{k+a}{m}$$

and

$$R(z) = \frac{da}{a-1}\left[z - \frac{Z_a k}{m} - \frac{(a+1)Z_a}{2m}\right]$$

if

$$\frac{k+a}{m} < \frac{z}{Z_a} < \frac{k+1}{m}.$$

Straightforward calculations show that, in this system, the matrix element $T_{\omega 123}$ is

$$T_{\omega 123} = \frac{\exp(2\gamma Z_a) - 1}{2\gamma Z_a \exp(2\gamma Z_a)}\left[1 + \frac{id\Delta\Omega}{2\gamma - id\Delta\Omega} \times \left[1 - \frac{2\gamma Z_a/m}{\exp(2\gamma Z_a/m) - 1} \times \frac{\exp[(2(1-a)\gamma + iad\Delta\Omega)Z_a/m] - 1}{(2(1-a)\gamma + iad\Delta\Omega)Z_a/m}\right]\right] \times \exp\left[-iad\Delta\Omega\frac{Z_a}{2m}\right]. \quad (14)$$

To show a self-similar structure of this matrix element, we rewrite $T_{\omega 123}$ as

$$T_{\omega 123} = B(G)F(a, X, Y), \quad (15)$$

$$B(G) = \frac{G-1}{G\ln G},$$

$$F(a, X, Y) = \left[1 + \frac{iY}{X - iY}\left[1 - \frac{X}{e^X - 1} \times \frac{\exp[(1-a)X + iaY] - 1}{(1-a)X + iaY}\right]\right]\exp\left(-\frac{iaY}{2}\right), \quad (16)$$

where the amplitude B is a function of only $G = \exp(2\gamma Z_a)$ and is independent of m. The shape $F(a, X, Y)$ is a function of the parameter a and specific combinations of $X = 2\gamma Z_a/m$ and $Y = d\Delta\Omega Z_a/m$. The real part (solid curve), the imaginary part (dashed curve), and the absolute value (dotted curve) of $F(a, X, Y)$ are plotted in Fig. 1. Here, $a = 0.21$ dB/km, $Z_a = 60$ km, $m = 2$, and $a = 0.5$. Minima of the function $|F|$ correspond to operation regimes with the suppressed FWM [52]. In the $d = 0$ limit, we obviously recover results of the traditional path-averaged (guiding-center) soliton theory [24–26].

In Fig. 2, the function $|F(a, X, Y)||$ is plotted versus Y for different a with the same parameters as in Fig. 1. We now estimate the matrix element of the quasi-identical transformation

$$|S(\Delta\Omega)| \le \left|\frac{1}{Z_a}\int_0^{Z_a}\frac{c(z)[\exp(i\Delta\Omega R(z)) - 1]}{\Delta\Omega}dz\right| \le \max(R)\langle c\rangle = \frac{ad}{2m}\langle c\rangle.$$

It can be seen that as m increases (with the other parameters fixed), the path-averaged model (10) governing the DM soliton propagation converges to the integrable NLS equation with

$$T(0) = \frac{G-1}{G\ln G}.$$

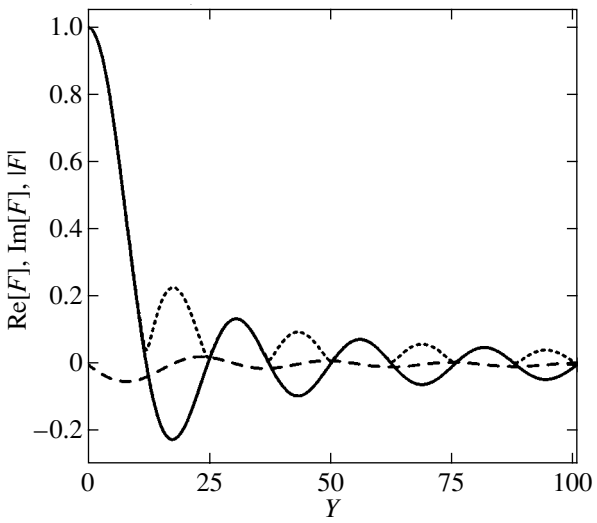

Fig. 1. Real (solid curve) and imaginary (dashed curve) parts and the absolute value (dotted curve) of the function $|F(a, X, Y)|$ are plotted for $a = 0.5$ and $X = 0.63$ ln(10)dB.

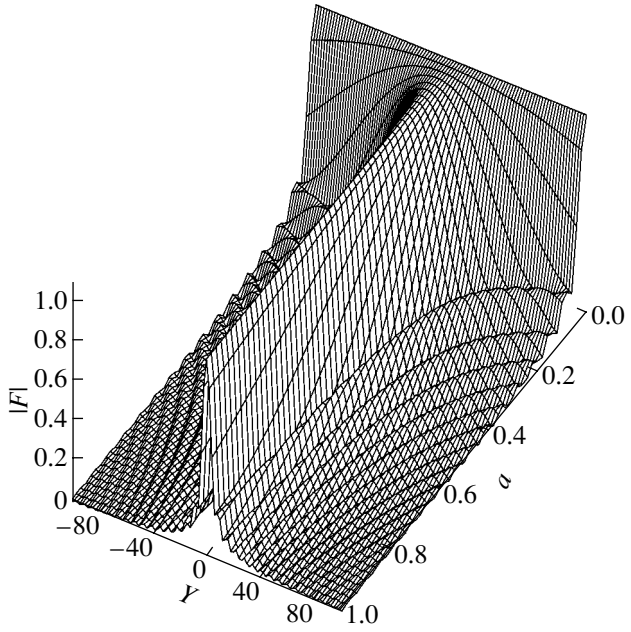

Fig. 2. The function $|F(a, X, Y)|$ versus Y for the system with different a.

It is obvious that, in the limit of a very weak loss (small γ), we again obtain the lossless model approximation for T,

$$T_{\omega 123} = \frac{\sin(aY)}{aY}.$$

However, the increase in m (decrease in L) under the fixed characteristic bandwidth of the signal makes the oscillatory structure of the kernel insignificant. This implies that, if $T(\Delta\Omega)$ is practically concentrated in some region, then the corresponding region in $\Delta\Omega$ is larger for large m than for small m. For pulses with the same spectral width, this means that T is much flatter for large m: as a matter of fact, the function T can be well approximated by the value $T(0)$ for large m (small L). As a result, the NLS equation model works rather well in this limit and the solution (of the path-averaged model!) should be close to the cosh-like soliton of the NLS equation. We note that, although it is known that the DM soliton shape is close to cosh for the lossless model in the so-called weak-map ($S < 1$) limit [19, 28, 36, 34], this is not so obvious for a system with loss and different periods of the amplification and dispersion variations. In such a system, DM solitons therefore possess the dual advantages of being chirped (which is important for the suppression of the four-wave mixing in WDM systems) and of having integrable path-averaged dynamics, which allows the use of well-developed mathematical tools in studying practical perturbations. This additionally implies that all the control techniques developed for the improvement of the traditional soliton transmission can be directly used in these systems.

5. A SINGLE-PULSE PROPAGATION

In this section, we consider numerical simulation results for a single pulse propagation in systems with a short-scale management. In contrast to the lossless model, the evolution of soliton parameters over one period is asymmetric here because of the loss. Rapid variations of the pulse width, peak power, and chirp are accompanied by the exponential decay of the power due to the loss. Nevertheless, numerical simulations have revealed that there exists a true periodic solution that reproduces itself at the end of the compensation cell (in this case, at the end of the amplification period). For the DM soliton with the map strength $S = 2$, the evolutions of its peak power (right top), chirp (left bottom), and full width at half maximum (right bottom) along one section are shown in Fig. 3 for a transmission system with the short-scale dispersion map (left top). The amplification distance is 40 km, and the dispersion compensation length is 4 km. The following parameters were used in the simulations: the dispersion in the two-step map ±16 + 0.1 ps/(nm km) (see Fig. 3), the nonlinear coefficient $\sigma = 2\pi n_2/\lambda_0 A_{\text{eff}} = 2.43\ \text{W}^{-1}\ \text{km}^{-1}$, and the fiber loss $a = 0.21$ dB/km.

The observed DM soliton is very stable and propagates without radiation as seen in Fig. 4 (where system parameters are the same as in Fig. 3). Figure 4 illustrates the chirp of the DM soliton versus the width. The left and right figures show this dependence for the first and the 140th sections, respectively.

An important feature of solitons in systems with a short-scale dispersion management is the reduced power. The DM soliton identified here has a reduced power compared to the previously studied DM soliton regimes ($L \geq Z_a$) for the same width propagating in a

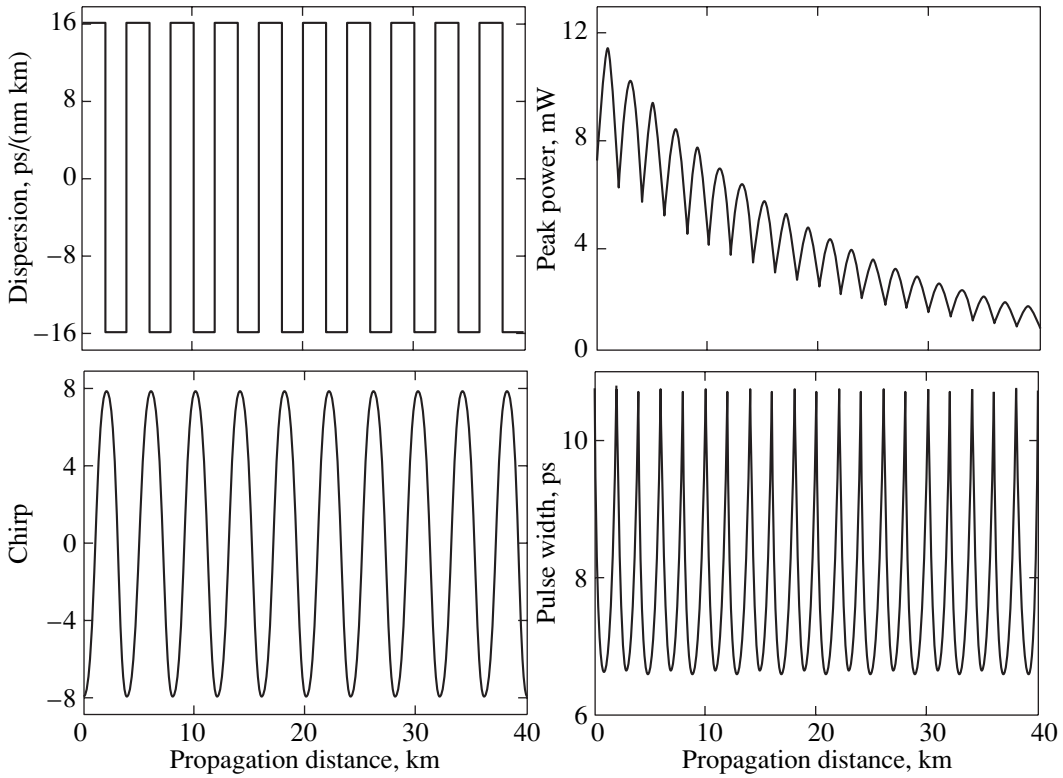

Fig. 3. Evolution of the soliton peak power (right top), chirp (left bottom), and full width at half maximum (right bottom) along one section is shown for the transmission system with the short-scale dispersion map (left top). Here, $S = 2$, the amplification distance is 40 km, and the dispersion compensation period is 4 km.

fiber system with the same average dispersion (with the same parameters except the L/Z_a ratio). This observation is illustrated by Fig. 5, where we present results of the modeling based on the zero-mode Gaussian approximation of the DM soliton (in the expansion using a complete basis of the chirped Gauss–Hermite functions—see [57] for details). Using this approach, we have built the evolution of the DM soliton peak power dependence on the pulse width; the dispersion compensation length was changed, but the average dispersion and the amplification distance were kept the same. In Fig. 5, the dependence of the DM soliton peak power on the pulse width at the beginning of the compensation section $z = 0$ is shown for different ratios of the dispersion period $L = Z_a/m$ to the amplification distance Z_a (40 km here): $m = 10$ (solid curve), 1 (long-dashed curve), 0.5 (dashed curve), 0.2 (dotted curve), 0.1 (dash-dotted curve). For control, we also show the peak power dependence for the true DM soliton found numerically (in the full model) in the case where $m = 10$ (squares) and $m = 0.2$ (rhombuses).

We also note that the energy of the short-scale DM soliton is very close to that of the conventional soliton (although the pulse is chirped and experiences breathing oscillations of the width and chirp during propagation). This is because the effective map strength is small here due to small L. It is seen from Fig. 5 that the short-scale dispersion management ($m = 10$) indeed provides a reduced power of the DM soliton for the same pulse width (and the same average dispersion and the same other parameters except the ratio L/Z_a). Because the soliton power grows very rapidly with the reduction of the pulse width (after the curves in Fig. 5 pass some critical turning points; for instance, for m = 1 such a point is around 16 ps), this effect can be very important for high-bit-rate transmissions using short pulses.

6. SOLITON INTERACTION

The nonlinear pulse-to-pulse interaction is one the main limiting factors in high-bit-rate optical data transmission. In this section, we present results on the soliton interaction in systems with a short-scale management with the amplification period $Z_a = 60$ km and the dispersion compensation period $L = 4$ km ($m = 15$), $L = 6$ km ($m = 10$), and $L = 12$ km ($m = 5$). Numerical simulations in this section include the third-order dispersion and Raman effects. An important advantage of operating close to the integrable limit (weak maps) discussed above is that the well-developed techniques to suppress soliton interaction can be applied. Figures 6 and 7 show the effect of the initial phase alternation of neighboring solitons. Figure 6 shows the propagation of two in-phase solitons initially separated by 10 ps (100 Gb/s). The solitons collapse after approximately

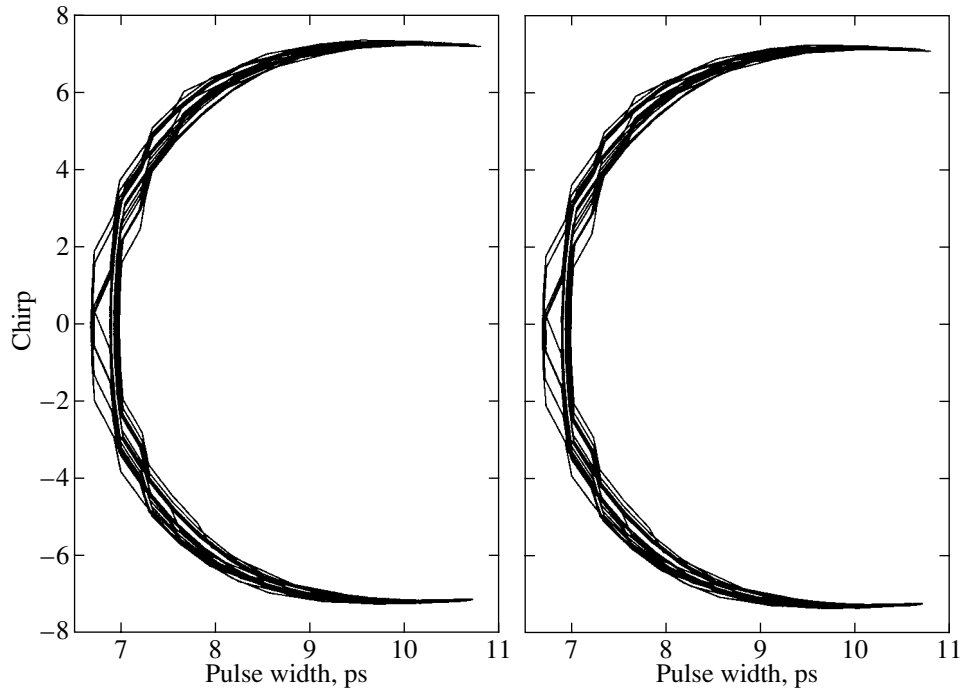

Fig. 4. Chirp versus width of the DM soliton for the first (left) and the 140th (right) sections.

500 km. In contrast, DM solitons with the initial phase shift π can propagate over 5000 km without fusion. Here, $D = \pm 2.4 + 0.0785$ ps/(nm km), $Z_a = 60$ km, $m = 15$, the peak power of the single soliton is 5.44 mW, and the pulse width is 2.93 ps at the chirp-free point (0.56 km from the end of the map). We recall that the interaction of DM solitons with larger S is independent of the initial phase shift [33].

Figure 8 shows the normalized distance between the Gaussian pulses for different initial phase shifts along the total distance $z = 1018.5$ km. The initial distance is $z = 12.5$ ps (80 Gb/s), and the maps are $D = \pm 1.6 + 0.04$ ps/(nm km), and $D = \pm 2.4 + 0.04$ ps/(nm km), and $D = \pm 3.2 + 0.04$ ps/(nm km), with the respective strengths $S = 1.06$, 1.58, and 2.12. Figure 9 shows an improvement of the system performance resulting from

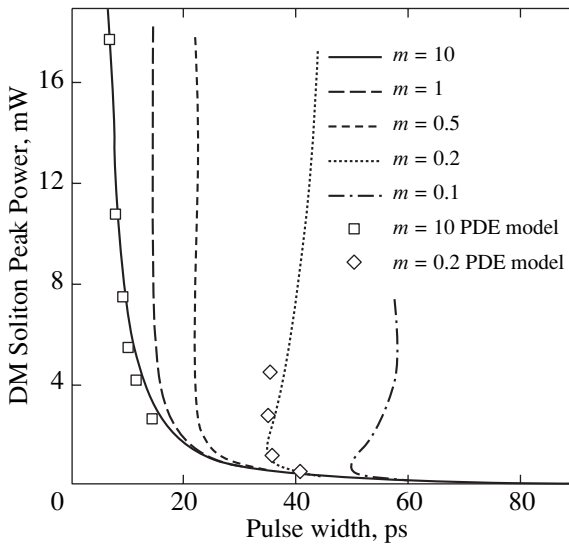

Fig. 5. The dependence of the DM soliton peak power on the pulse width at the beginning of the compensation section $z = 0$ for different ratios of the dispersion period $L = Z_a/m$ to the amplification distance Z_a: $m = 10$ (solid curve), 1 (long-dashed curve), 0.5 (dashed curve), 0.2 (dotted curve), and 0.1 (dash-dotted curve). The same dependences for the true DM soliton found numerically (in the full model) are shown for $m = 10$ (squares) and $m = 0.2$ (rhombuses).

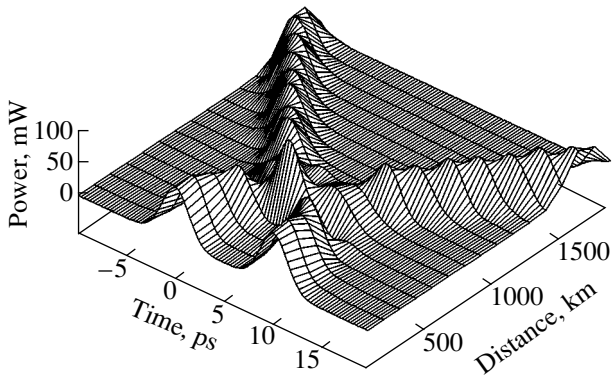

Fig. 6. The interaction of two in-phase DM solitons at 100 Gb/s. Here, $D = \pm 2.4 + 0.0785$ ps/(nm km), $Z_a = 60$ km, and $m = 15$; solitons with a peak power of 5.44 mW and pulse width of 2.93 ps are launched at a chirp-free point located 0.56 km before the end of the section.

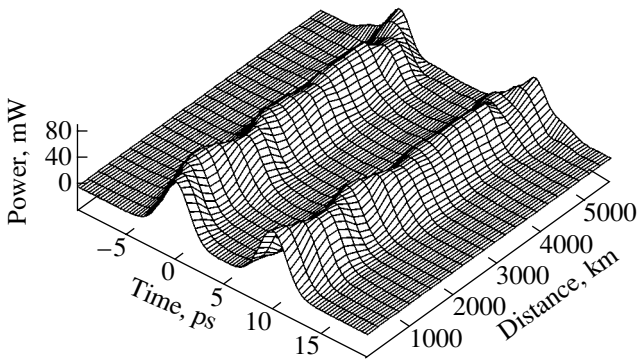

Fig. 7. The same as in Fig. 6 but with the initial phase alternation (out-of-phase solitons).

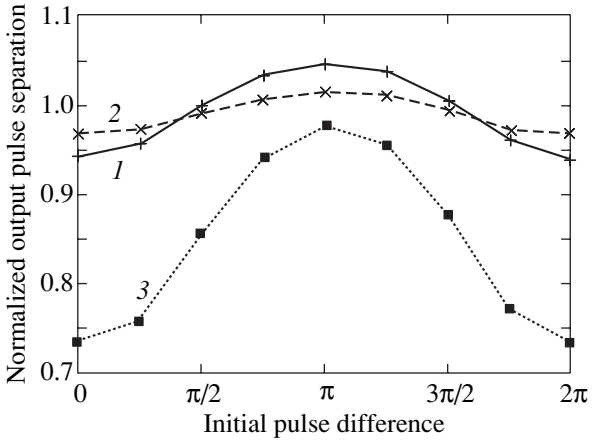

Fig. 8. The normalized pulse separation versus the initial phase shift (the total distance z = 1018.5 km) for different map strengths: S = 1.06 (solid curve *1*), S = 1.58 (long-dashed curve *2*), and S = 2.12 (dashed curve *3*). Here, Z_a = 60 km and L = 6 km.

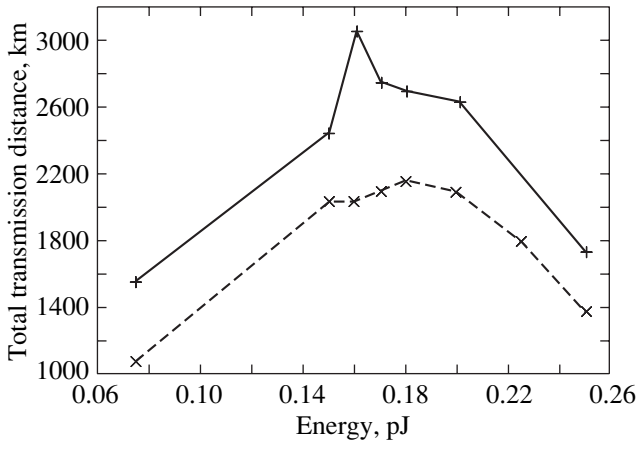

Fig. 9. The transmission distance at 80 Gb/s with (solid curve) and without (dashed curve) the initial phase alternation versus the DM soliton energy. Here, $D = \pm 2.4 + 0.04$ ps/(nm km), L = 6 km, and Z_a = 60 km.

the initial phase alternation. We plot the total transmission distance versus the DM soliton energy (at the beginning of the section) at 80 Gb/s. Here, the dispersion is $D = \pm 2.4 + 0.04$ ps/(nm km), L = 6 km, and Z_a = 60 km. The total transmission distance has been defined as the distance at which the Q factor becomes less than 6 for two test random 128-bit patterns. The solid lines are for the initial signals with a phase alternation, and dashed lines are for the in-phase input pulses. It can be seen that short-scale dispersion-managed systems are quite attractive candidates for the transmission of optical data at ultrahigh-bit rates. Optimization of such lines will lead to a further improvement of the system performance.

7. CONCLUSION

We have identified a stable optical pulse propagation regime in fiber systems with short-scale dispersion management when the compensation period is much shorter than the amplification distance. In systems with a short-scale management, the DM soliton has a reduced power compared to the usual DM soliton ($L > Z_a$) of the same width (and the same amplification distance and average dispersion). Short-scale management is a means of controlling the strength of the map (and, consequently, pulse energy, interactions, etc.) while keeping the average dispersion finite and taking advantage of the FWM suppression in WDM by a high local dispersion. We show that the path-averaged dynamics of chirped DM solitons in systems with a short-scale management for weak maps is close to that in the integrable model. Therefore, DM solitons in such systems possess the dual advantages of being chirped (which is important for the suppression of the four-wave mixing in WDM systems) and of possessing integrable path-averaged dynamics, which allows the use of well-developed mathematical tools for studying practical perturbations.

ACKNOWLEDGMENTS

We would like to thank S.K. Turitsyn for useful discussions. The support of EPRSC and RFBR (grant no. 99-02-16688) is acknowledged.

REFERENCES

1. A. Hasegawa and F. D. Tappert, Appl. Phys. Lett. **23**, 142 (1973).
2. V. E. Zakharov and A. B. Shabat, Zh. Éksp. Teor. Fiz. **60**, 136 (1971) [Sov. Phys. JETP **33**, 77 (1971)].
3. S. V. Manakov, Zh. Éksp. Teor. Fiz. **65**, 505 (1973) [Sov. Phys. JETP **38**, 248 (1973)].
4. L. F. Mollenauer, R. H. Stolen, and J. P. Gordon, Phys. Rev. Lett. **45**, 1095 (1980).
5. L. F. Mollenauer and K. Smith, Opt. Lett. **13**, 675 (1988).
6. S. P. Novikov, S. V. Manakov, L. P. Pitaevskii, and V. E. Zakharov, *Theory of Solitons: the Inverse Scatter-*

ing Method (Nauka, Moscow, 1980; Consultants Bureau, New York, 1984).
7. A. C. Newell and J. V. Moloney, *Nonlinear Optics* (Addison-Wesley, Redwood City, 1992).
8. M. J. Ablowitz and P. A. Clarkson, *Solitons, Nonlinear Evolution Equations, and Inverse Scattering* (Cambridge Univ. Press, Cambridge, 1991).
9. A. Hasegawa and Y. Kodama, *Solitons in Optical Communications* (Clarendon, Oxford, 1995).
10. R. K. Dodd, J. C. Eilbeck, J. D. Gibbon, and H. C. Morris, *Solitons and Nonlinear Wave Equations* (Academic, New York, 1984).
11. M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 294 (1974).
12. E. A. Kuznetsov, A. M. Rubenchik, and V. E. Zakharov, Phys. Rep. **142**, 103 (1986).
13. L. F. Mollenauer, P. V. Mamyshev, and M. J. Neubelt, in *Proceedings of the Conference on Optical Fiber Communication, OFC'96, San Jose, 1996*, Post Deadline Presentation, PD22-1.
14. *Optical Fiber Telecommunications*, Ed. by I. P. Kaminow and T. L. Koch (Academic, San Diego, 1997), Vol. IIIA.
15. F. M. Knox, W. Forysiak, and N. J. Doran, IEEE J. Lightwave Technol. **13**, 1955 (1995).
16. M. Suzuki, I. Morita, N. Edagawa, *et al.*, Electron. Lett. **31**, 2027 (1995).
17. M. Nakazawa and H. Kubota, Electron. Lett. **31**, 216 (1995).
18. H. A. Haus, K. Tamura, L. E. Nelson, and E. P. Ippen, IEEE J. Quantum Electron. **31**, 591 (1995).
19. N. Smith, F. M. Knox, N. J. Doran, *et al.*, Electron. Lett. **32**, 54 (1996).
20. M. Nakazawa, H. Kubota, A. Sahara, and K. Tamura, IEEE Photonics Technol. Lett. **8**, 452 (1996).
21. M. Nakazawa, K. Suzuki, H. Kubota, and E. Yamada, Electron. Lett. **32**, 1686 (1996).
22. M. Suzuki, I. Morita, K. Tanaka, *et al.*, in *ECOC'97: Proceedings of the European Conference on Optical Communication, Edinburgh, 1997*, Vol. 3, p. 99.
23. D. Le Guen, F. Favre, M. L. Moulinard, *et al.*, in *ECOC'97: Proceedings of the European Conference on Optical Communication, Edinburgh, 1997*, Postdeadline Paper, Vol. 5, p. 25.
24. L. F. Mollenauer, S. G. Evangelides, Jr., and H. A. Haus, IEEE J. Lightwave Technol. **9**, 194 (1991).
25. A. Hasegawa and Y. Kodama, Opt. Lett. **15**, 1443 (1990); Phys. Rev. Lett. **66**, 161 (1991).
26. K. J. Blow and N. J. Doran, IEEE Photonics Technol. Lett. **3**, 369 (1991).
27. A. F. Evans, in *Optical Fiber Communication Conference, 1998* (OSA, Washington, DC, 1998), OSA Technical Digest Series, Vol. 2, p. 22.
28. I. Gabitov and S. K. Turitsyn, Opt. Lett. **21**, 327 (1996).
29. N. J. Smith, N. J. Doran, F. M. Knox, and W. Forysiak, Opt. Lett. **21**, 1981 (1996).
30. T. Georges and B. Charbonnier, IEEE Photonics Technol. Lett. **9**, 127 (1997).
31. J. M. Jacob, E. A. Golovchenko, A. N. Pilipetskii, *et al.*, IEEE Photonics Technol. Lett. **9**, 130 (1997).
32. M. Matsumoto and H. A. Haus, IEEE Photonics Technol. Lett. **9**, 785 (1997).
33. E. A. Golovchenko, A. N. Pilipetskii, and C. R. Menyuk, Opt. Lett. **22**, 793 (1997).
34. T. S. Yang and W. L. Kath, Opt. Lett. **22**, 985 (1997).
35. D. Breuer, F. Kueppers, A. Mattheus, *et al.*, Opt. Lett. **22**, 546 (1997).
36. J. H. B. Nijhof, N. J. Doran, W. Forysiak, and F. M. Knox, Electron. Lett. **33**, 1726 (1997).
37. V. S. Grigoryan, T. Yu, E. A. Golovchenko, C. R. Menyuk, and A. N. Pilipetskii, Opt. Lett. **22**, 1609 (1997).
38. A. Sahara, H. Kubota, and M. Nakazawa, IEEE Photonics Technol. Lett. **9**, 1179 (1997).
39. S. Kumar and A. Hasegawa, Opt. Lett. **22**, 372 (1997).
40. A. Hasegawa, Y. Kodama, and A. Maruta, Opt. Fiber Technol.: Mater., Devices Syst. **3**, 197 (1997).
41. S. K. Turitsyn, M. Fedoruk, and A. Gornakova, Opt. Lett. **24**, 869 (1999).
42. S. B. Medvedev and S. K. Turitsyn, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 465 (1999) [JETP Lett. **69**, 499 (1999)].
43. M. J. Ablowitz and G. Biondini, Opt. Lett. **23**, 1668 (1998).
44. S. K. Turitsyn and V. K. Mezentsev, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 791 (1998) [JETP Lett. **68**, 830 (1998)].
45. V. E. Zakharov and S. V. Manakov, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 573 (1999) [JETP Lett. **70**, 578 (1999)].
46. P. M. Lushnikov, Opt. Lett. **25**, 1144 (2000).
47. D. E. Pelinovsky, Phys. Rev. E **62**, 4283 (2000).
48. P. M. Lushnikov, Opt. Lett. **26**, 1535 (2001).
49. V. E. Zakharov, in *Optical Solitons. Theoretical Challenges and Industrial Perspectives*, Ed. by V. E. Zakharov and S. Wabnitz (Springer-Verlag, Berlin, 1999).
50. V. I. Arnold, *Geometrical Methods in the Theory of Ordinary Differential Equations* (Nauka, Moscow, 1978; Springer-Verlag, New York, 1988).
51. V. E. Zakharov, V. S. L'vov, and G. Falkovich, *Kolmogorov Spectra of Turbulence* (Springer-Verlag, Berlin, 1992).
52. S. K. Burtsev and I. Gabitov, in *Proceedings of the II International Symposium on Physics and Applications of Optical Solitons in Fibers, Kyoto, 1997* (Kluwer, Dordrecht, 1998), p. 261.
53. J. N. Kutz, P. Holmes, S. G. Evangelides, Jr., and J. P. Gordon, J. Opt. Soc. Am. B **15**, 87 (1998).
54. S. K. Turitsyn and E. G. Shapiro, Opt. Fiber Technol.: Mater., Devices Syst. **4**, 151 (1998).
55. D. S. Govan, W. Forysiak, and N. J. Doran, Opt. Lett. **23**, 1523 (1998).
56. N. S. Bergano, C. Davidson, M. Ma, *et al.*, in *Proceedings of the Conference on Optical Fiber Communication, OFC'98, San Jose, 1998*, Post Deadline Presentation, PD12-1.
57. S. K. Turitsyn, E. G. Shapiro, and V. K. Mezentsev, Opt. Fiber Technol.: Mater., Devices Syst. **4**, 384 (1998).

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 901–915.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1051–1067.
Original Russian Text Copyright © 2002 by Drachev, Perminov, Rautian, Safonov, Khaliullin.

ATOMS, SPECTRA, RADIATION

Polarization Effects in Nanoaggregates of Silver Caused by Local and Nonlocal Nonlinear-Optical Responses

V. P. Drachev[a, §], S. V. Perminov[a, *], S. G. Rautian[b], V. P. Safonov[b], and É. N. Khaliullin[a, §]

[a]*Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia*

**e-mail: serge@isp.nsc.ru*

[b]*Institute of Automatics and Electrometry, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia*

Received December 27, 2001

Abstract—A theoretical and experimental study is made into the combined manifestation of local and nonlocal optical responses in a cubic nonlinear isotropic medium such as an aggregated colloidal silver solution. The phenomenological treatment of polarization effects is performed for the general case with due regard for the frequency dispersion of both local and nonlocal nonlinearities and for the noncollinear propagation of pump and probe light waves. The inverse Faraday effect, the optical Kerr effect, and the self-rotation of the polarization ellipse in a fractal-disordered nonlinear medium are observed for the first time. The tensor components of the local and nonlocal cubic nonlinearities of colloidal silver solutions are measured for different degrees of aggregation. It is demonstrated that, as the size of silver aggregate increases, the nonlocal nonlinear response increases much more strongly than the local one. An inference is made that the mechanical motion of metal nanoparticles because of their dynamic interaction with the light wave field can contribute to the nonlinear polarization effects. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The nonlinear-optical properties of aggregated metal nanocomposites have been subjected to intensive studies for well over ten years (see the review [1]). Interest in such media is primarily due to the special character of interaction between a light wave and an ensemble of highly polarizable nanoparticles which form a disordered structure referred to as aggregate or cluster. Large clusters, which form, in particular, during aggregation in colloidal solutions, consist of hundreds and thousands of nanoparticles and have a fractal structure [2]. The interaction between the induced dipole moments of particles results in the formation of collective modes of dipole excitation in clusters, and it is these modes that define the optical properties of nanocomposites. The spatial configurations of collective modes, characterized by the electric field distribution, may be highly diversified; in particular, a localization of optical excitation is possible in a region which is small compared with the aggregate size [3, 4]. This implies a considerable increase in the local electric field in this cluster region. The latter fact proves to be of special significance from the standpoint of nonlinear-optical processes depending on the square, cube, and higher powers of the electric field. Butenko *et al.* [5] predicted a considerable increase in the nonlinear-optical responses of aggregates compared with isolated nanoparticles. In subsequent experiments involving colloidal silver solutions, a giant amplification was observed of degenerate four-photon scattering [6], as well as of nonlinear refraction and nonlinear absorption [7] during aggregation of silver particles into clusters.

Polarization effects in the case of nonlinear self-action of a light wave in an aggregated colloidal silver solution were first observed and experimentally investigated by us in [8], where a nonlinear optical activity of fractal aggregates of silver was observed, which was due to spatial dispersion of nonlinear response of the third order (in other words, to the nonlocality of interaction between the medium and the light wave field). It is well known, however, that the polarization of a wave in an isotropic medium may also change due to purely local cubic nonlinearity. If the radiation polarization is other than strictly linear or circular, polarization self-action is observed, which consists in the rotation of the polarization ellipse [9]. Nonlinear polarization effects may be studied using the probe field method as well. The interaction between two waves (pump and probe ones) in a nonlinear medium with a local response brings about a variation of the polarization of the probe wave (except for the situation in which both waves are linearly polarized in a single plane). Such phenomena include, in particular, the inverse Faraday effect (IFE), i.e., the rotation of the plane of polarization of the probe field under the effect of circular-polarized pumping [10], and the optical Kerr effect (OKE) [11], i.e., the

[§]Present address: Purdue University, School of Electrical and Computer Engineering, West Lafayette, IN 47907 USA.

induction of birefringence in an isotropic medium by a linearly polarized pump wave.

The nonlinear polarization of an isotropic medium is characterized by three independent constants [12] (and, if the frequency dispersion of nonlocal response is taken into account, by four constants; see below). As a rule, the above-identified effects of nonlinear optical activity, inverse Faraday effect, and optical Kerr effect depend on the combination of these constants.

It is known that the ratio between nonlinear coefficients are important parameters of the theory. In particular, it follows from the theory of optical activity that the local and nonlocal parts of polarizability correlate by the order of magnitude as the wavelength and some effective dimension characterizing an optically active medium [12]. In the case of small molecules, this effective dimension is provided by the distance between at least two groups with anisotropic polarizability. For spiral macromolecules, the pitch of the helix serves as such an effective geometric parameter. In exciton transitions of crystals, the exciton radius [13] (10 nm) serves as the parameter of nonlocality. It is of interest to clarify the physical meaning of the nonlocality parameter for fractal clusters of silver and to directly measure this parameter.

The relative magnitude of two nonlinear constants corresponding to linear response depends on the nature of nonlinearity [9]. For example, for the orientation mechanism, this ratio is six; it is equal to unity for a nonresonant electron response and tends to zero for electrostriction [14]. No measurements were previously performed of the totality of the components of the tensor of local cubic susceptibility of colloidal aggregates of silver.

This paper deals with the investigation of polarization nonlinearities of colloidal aggregates of silver. The experiments were performed with samples of different degrees of aggregation of nanoparticles in order to reveal the dependence of the magnitude of cubic susceptibility on the geometric structure of the aggregate. The measurements of nonlinear optical activity, inverse Faraday effect, optical Kerr effect, and self-rotation of ellipse were performed with identical samples and with light beams of identical spatial and time properties, which enabled us to find nonlinear coefficients of the medium for local and nonlocal responses.

Our measurements produced fairly high magnitudes of the IFE and OKE for nanoaggregates of silver. The angle of rotation of the polarization plane of the probe field in the process of IFE was 0.66 deg/cm MW ($\alpha_{IFE} \approx$ 23 deg/cm with the strong field intensity of 35 MW/cm^2), and the induced birefringence (OKE) was $\Delta n \approx 3 \times 10^{-6}$ with the intensity of 8 MW/cm^2. No saturation of the intensity dependence of the rotation angle α_{IFE} and Δn was observed.

The next, second, section gives the relations between the polarization and the field with due regard for the frequency dispersion of nonlocal nonlinear response and noncollinear propagation of interacting waves. The third section contains a description of the procedures of measurements and preparation of samples. The results of measurements of polarization nonlinearities are given in Section 4. In Section 5, the obtained intensity dependences of induced birefringence and nonlinear angle of rotation are discussed. The Appendix deals with the results of qualitative analysis of the motion of nanoparticles of silver in the aggregate due to the light-induced interaction of monomers.

2. POLARIZATION SELF-ACTION AND INTERACTION OF LIGHT WAVES IN A CUBIC NONLINEAR MEDIUM WITH DUE REGARD FOR SPATIAL DISPERSION: PHENOMENOLOGY

2.1. Basic Relations

We will write the electric field of a light wave in the form

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{A}(\mathbf{r})\exp(-i\omega t) + \text{c.c.} \quad (1)$$

The cubic nonlinear polarization of the medium in the general case contains two terms,

$$\tilde{\mathbf{P}}^{\text{NL}}(\mathbf{r}, t) = \hat{\chi}^{(3)}\mathbf{EEE} + \hat{\Gamma}^{(3)}\mathbf{EE}\nabla\mathbf{E}. \quad (2)$$

The first term in the right-hand part of relation (2) is responsible for the third-order local nonlinear response of the medium. Corresponding to this term is the nonlinear polarization of the medium, whose Fourier component on the frequency ω has the form

$$(\tilde{P}_{\text{loc}})_i(\mathbf{r}, \omega) = 3\chi^{(3)}_{ijkl}(\omega, \omega, -\omega)A_jA_kA_l^* + \text{c.c.} \quad (3)$$

In the case of an isotropic medium, the tensor $\hat{\chi}^{(3)}(\omega, \omega, -\omega)$ may be represented as [14]

$$\chi^{(3)}_{ijkl}(\omega + \omega - \omega) = \chi_{1122}(\omega + \omega - \omega)(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl}) + \chi_{1221}(\omega + \omega - \omega)\delta_{il}\delta_{jk}, \quad (4)$$

which enables one to write Eq. (3) as

$$\tilde{\mathbf{P}}_{\text{loc}}(\mathbf{r}, \omega) = \chi_1\mathbf{A}((\mathbf{AA})^*) + \chi_2\mathbf{A}^*(\mathbf{AA}) + \text{c.c.}, \quad (5)$$

$$\chi_1 = 6\chi^{(3)}_{iijj} = 6\chi^{(3)}_{ijij}; \quad \chi_2 = 3\chi^{(3)}_{ijji};$$

$$i, j = 1, 2, \quad i \neq j.$$

The second term in Eq. (2) describes the spatial dispersion of nonlinear polarization of the third order. We will derive the expression for the nonlinear polarization $\tilde{\mathbf{P}}_{\text{nonloc}}$ related to spatial dispersion,

$$(\tilde{P}_{\text{nonloc}})_i(\mathbf{r}, t) = \int_0^\infty dt_1(A_je^{-i\omega(t-t_1)} + A_j^*e^{i\omega(t-t_1)})$$

$$\times \int_0^\infty dt_2 (A_k e^{-i\omega(t-t_2)} + A_k^* e^{i\omega(t-t_2)})$$

$$\times \left[\int_0^\infty \Gamma_{ijklm}(t_1, t_2, t_3) \nabla_m (A_l) e^{-i\omega(t-t_3)} dt_3 \right.$$

$$\left. + \int_0^\infty \Gamma_{ijklm}(t_1, t_2, t_3) \nabla_m (A_l^*) e^{i\omega(t-t_3)} dt_3 \right]$$

$$= \int_0^\infty dt_1 (\ldots) \int_0^\infty dt_2 (\ldots) [\nabla_m (A_l) e^{-i\omega t} \Gamma_{ijklm}(t_1, t_2, \omega) + \nabla_m (A_l^*) e^{i\omega t} \Gamma_{ijklm}(t_1, t_2, -\omega)] \quad (6)$$

$$= \int_0^\infty dt_1 (\ldots) [A_k \nabla_m A_l e^{-2i\omega t} \Gamma_{ijklm}(t, \omega, \omega) + A_k^* \nabla_m A_l \Gamma_{ijklm}(t_1, -\omega, \omega) + A_k \nabla_m A_l^* \Gamma_{ijklm}(t_1, \omega, -\omega) + A_k^* \nabla_m A_l^* e^{2i\omega t} \Gamma_{ijklm}(t_1, -\omega, -\omega)].$$

We will further perform integration with respect to t_1 and leave only the terms proportional to $\exp(\pm i\omega t)$ to derive the following expression for the Fourier harmonic of nonlocal response on the frequency ω (the subscripts of the tensor Γ_{ijklm} are omitted for simplicity):

$$\begin{aligned}(\tilde{P}_{\text{nonloc}})_i(\mathbf{r}, \omega) &= A_j A_k^* \nabla_m A_l \Gamma(\omega, -\omega, \omega) \\ &+ A_j A_k \nabla_m A_l^* \Gamma(\omega, \omega, -\omega) \\ &+ A_j^* A_k \nabla_m A_l \Gamma(-\omega, \omega, \omega) \\ &+ A_j A_k^* \nabla_m A_l^* \Gamma(\omega, -\omega, -\omega) \\ &+ A_j^* A_k^* \nabla_m A_l \Gamma(-\omega, -\omega, \omega) \\ &+ A_j^* A_k \nabla_m A_l^* \Gamma(-\omega, \omega, -\omega).\end{aligned} \quad (7)$$

We use the frequency-commutation relations for the tensor $\hat{\Gamma}^{(3)}$ [14]

$$\Gamma_{ijklm}(\omega_1, \omega_2, \omega_3) = \Gamma_{ikjlm}(\omega_2, \omega_1, \omega_3), \quad (8)$$

as well as the equality

$$\Gamma_{ijklm}(\omega_1, \omega_2, \omega_3) = \Gamma^*_{ijklm}(-\omega_1, -\omega_2, -\omega_3), \quad (9)$$

to derive

$$\begin{aligned}(\tilde{P}_{\text{nonloc}})_i(\mathbf{r}, \omega) &= 2\Gamma_{ijklm}(-\omega, \omega, \omega) A_j^* A_k \nabla_m A_l \\ &+ \Gamma_{ijklm}(\omega, \omega, -\omega) A_j A_k \nabla_m A_l^* + \text{c.c.}\end{aligned} \quad (10)$$

In an isotropic medium, the tensor $\hat{\Gamma}(-\omega, \omega, \omega)$ has four nonzero components (of which only one is independent). We will direct the unit vector $\mathbf{e}_3$ along the vector $\nabla \times \mathbf{A}$. Then, the relations

$$\begin{aligned}\Gamma_{11123}(-\omega, \omega, \omega) &= \Gamma_{12223}(-\omega, \omega, \omega) \\ &= -\Gamma_{21113}(-\omega, \omega, \omega) = -\Gamma_{22213}(-\omega, \omega, \omega) \equiv g_1\end{aligned} \quad (11a)$$

are valid between the components $\hat{\Gamma}(-\omega, \omega, \omega)$. Analogous relations are valid for the components of the tensor $\hat{\Gamma}(\omega, \omega, -\omega)$ as well,

$$\begin{aligned}\Gamma_{11123}(\omega, \omega, -\omega) &= \Gamma_{12223}(\omega, \omega, -\omega) \\ &= -\Gamma_{21113}(\omega, \omega, -\omega) = -\Gamma_{22213}(\omega, \omega, -\omega) \equiv g_2.\end{aligned} \quad (11b)$$

Therefore, in an isotropic medium, the nonlocal nonlinear response in view of frequency dispersion is described by two constants.

With due regard for the introduced notation, we derive the following expression for nonlinear polarization associated with the nonlinear response of the medium:

$$\begin{aligned}\tilde{\mathbf{P}}_{\text{nonloc}}(\mathbf{r}, \omega) &= -2g_1(\mathbf{A}^* \cdot \mathbf{A}) \nabla \times \mathbf{A} \\ &- g_2(\mathbf{A} \cdot \mathbf{A}) \nabla \times \mathbf{A}^* + \text{c.c.}\end{aligned} \quad (12)$$

2.2. Self-Action of Elliptically Polarized Radiation

On substituting nonlinear polarizations given by Eqs. (5) and (12) into the wave equation, one can derive the equation (analogously with that obtained in [12]) for the slow amplitude of a wave propagating in a nonlinear medium,

$$\begin{aligned}\frac{dA_\pm}{dz} &= -\delta A_\pm + i\frac{2\pi\omega^2}{kc^2} \\ &\times \{(\chi_1 |A|^2 + 2\chi_2 |A_\mp|^2) \pm 2k(g_1 |A|^2 - g_2 |A_\mp|^2)\} A_\pm,\end{aligned} \quad (13)$$

where $\delta = (\omega^2/2kc^2)\text{Im}\varepsilon_0$ is the linear absorption coefficient, $k = \omega n_0/c$, ε_0 is the linear permittivity of the medium, n_0 is the refractive index, c is the velocity of light, and $A_\pm$ denotes circular components of the complex amplitude

$$\mathbf{A} = \hat{\sigma}_+ A_+ + \hat{\sigma}_- A_-, \quad (14)$$

which are related to the Cartesian components $A_{1,2}$ as

$$A_\pm = \frac{A_1 \pm iA_2}{\sqrt{2}}. \quad (15)$$

The unit vectors $\hat{\sigma}_+$ and $\hat{\sigma}_-$ correspond to right-hand and left-hand circular polarization, respectively, and satisfy the relations

$$\hat{\sigma}_\pm^* = \hat{\sigma}_\mp; \quad \hat{\sigma}_\pm\hat{\sigma}_\pm = 0;$$
$$\mathbf{A}^*\cdot\mathbf{A} = A_+^*A_+ + A_-^*A_- = |A_+|^2 + |A_-|^2, \tag{16}$$
$$\mathbf{A}\cdot\mathbf{A} = A_+A_- + A_-A_+.$$

Equation (13) may be used to derive the equations for the azimuth of polarization ellipse $\alpha = \arg(A_+A_-^*)/2$ under conditions of wave propagation in the medium,

$$\frac{d\alpha}{dz} = \frac{2\pi\omega}{n_0c}\mathrm{Re}\chi_2(|A_-|^2 - |A_+|^2) + \frac{2\pi\omega^2}{c^2}|A|^2\mathrm{Re}(2g_1 - g_2). \tag{17}$$

The first term on the right-hand side corresponds to the known effect of rotation of polarization ellipse [9] in an isotropic medium with local cubic nonlinearity. The second term describes the nonlinear optical activity caused by the spatial dispersion of nonlinear response. One can see that the magnitude of specific rotation is proportional to the combination of nonlinear constants $2g_1 - g_2$. If the frequency dispersion of the nonlocal nonlinear response of the medium is ignored (i.e., if it is assumed that $g_1 = g_2$), Eq. (17) agrees with that derived in [12] within the term corresponding to the natural (linear) activity, which is not treated in this paper.

2.3. The Effect of an Pump Wave on a Probe Wave under Conditions of Noncollinear Propagation

We will assume that the electric field is the sum of two quasi-plane waves with identical frequencies but with different wave vectors. Then, the complex amplitude introduced in Eq. (1) may be represented as

$$\mathbf{A}(\mathbf{r}) = \mathbf{F}(\mathbf{r})\exp(i\mathbf{K}\cdot\mathbf{r}) + \mathbf{S}(\mathbf{r})\exp(i\mathbf{k}\cdot\mathbf{r}), \tag{18}$$

where **F**, **K** and **S**, **k** are the complex amplitudes and wave vectors of pump and probe waves, respectively. The dependence of **F** and **S** on **r** will be assumed to be slow compared with the exponential factor.

We will substitute expression (18) into relation (5) and, assuming the probe wave to be weak, retain only the terms linear with respect to **S**. As a result, we will derive the following expression for the complex amplitude of nonlinear polarization caused by the local response of the medium (from here on, we omit the complex conjugate terms on the right-hand side, assuming that **P** (without the tilde) implies the complex amplitude):

$$\mathbf{P}_{\mathrm{loc}}(\mathbf{r},\omega) = \chi_1[(\mathbf{F}^*\cdot\mathbf{F})\mathbf{S}\exp(i\mathbf{k}\cdot\mathbf{r}) + (\mathbf{F}\cdot\mathbf{S}^*)\mathbf{F}\exp[i(2\mathbf{K}-\mathbf{k})\cdot\mathbf{r}] + (\mathbf{F}^*\cdot\mathbf{S})\mathbf{F}\exp(i\mathbf{k}\cdot\mathbf{r})] + \chi_2[2(\mathbf{F}\cdot\mathbf{S})\mathbf{F}^*\exp(i\mathbf{k}\cdot\mathbf{r}) + (\mathbf{F}\cdot\mathbf{F})\mathbf{S}^*\exp[i(2\mathbf{K}-\mathbf{k})\cdot\mathbf{r}]]. \tag{19}$$

Similarly, we use Eq. (12) to find the contribution to polarization due to nonlocal interaction,

$$\mathbf{P}_{\mathrm{nonloc}}(\mathbf{r},\omega) = -2g_1\{(\mathbf{F}^*\cdot\mathbf{F})\nabla\times\mathbf{S}\exp(i\mathbf{k}\cdot\mathbf{r}) + (\mathbf{F}^*\cdot\mathbf{S})\exp[i(\mathbf{k}-\mathbf{K})\cdot\mathbf{r}]\nabla\times\mathbf{F}\exp(i\mathbf{K}\cdot\mathbf{r}) + (\mathbf{S}^*\cdot\mathbf{F})\times\exp[i(\mathbf{K}-\mathbf{k})\cdot\mathbf{r}]\nabla\times\mathbf{F}\exp(i\mathbf{K}\cdot\mathbf{r})\} \tag{20}$$
$$- g_2\{(\mathbf{F}\cdot\mathbf{F})\exp(2i\mathbf{K}\cdot\mathbf{r})\nabla\times\mathbf{S}^*\exp(-i\mathbf{k}\cdot\mathbf{r}) + 2(\mathbf{F}\cdot\mathbf{S})\exp[i(\mathbf{k}+\mathbf{K})\cdot\mathbf{r}]\nabla\times\mathbf{F}^*\exp(-i\mathbf{K}\cdot\mathbf{r})\}.$$

We will be interested in the part of $\mathbf{P}^{\mathrm{NL}} = \mathbf{P}_{\mathrm{loc}} + \mathbf{P}_{\mathrm{nonloc}}$ that propagates in the form of a plane wave with the wave vector **k** and, therefore, is under conditions of wave synchronism with the probe wave. We retain in Eqs. (19) and (20) the terms proportional to $\exp(i\mathbf{kr})$ to derive the expressions for the Fourier amplitudes of nonlinear polarization,

$$\mathbf{P}_{\mathrm{loc}}(\mathbf{k},\omega) = \chi_1[(\mathbf{F}^*\cdot\mathbf{F})\mathbf{S} + (\mathbf{F}^*\cdot\mathbf{S})\mathbf{F}] + 2\chi_2(\mathbf{F}\cdot\mathbf{S})\mathbf{F}^*,$$
$$\mathbf{P}_{\mathrm{nonloc}}(\mathbf{k},\omega) = 2ig_1\{(\mathbf{F}^*\cdot\mathbf{F})\mathbf{S}\times\mathbf{k} + (\mathbf{F}^*\cdot\mathbf{S})\mathbf{F}\times\mathbf{K}\} - 2ig_2(\mathbf{F}\cdot\mathbf{S})\mathbf{F}^*\times\mathbf{K}. \tag{21}$$

We will further substitute the electric field in the form

$$\mathbf{E} = \mathbf{S}\exp[-i(\omega t - kz)]$$

and nonlinear polarization

$$\tilde{\mathbf{P}}^{\mathrm{NL}} = (\mathbf{P}_{\mathrm{loc}} + \mathbf{P}_{\mathrm{nonloc}})\exp[-i(\omega t - kz)] + \mathrm{c.c.}$$

into the wave equation

$$\nabla^2\mathbf{E}(z,t) = \frac{\varepsilon_0\partial^2\mathbf{E}(z,t)}{c^2\partial t^2} + \frac{4\pi}{c^2}\frac{\partial^2}{\partial t^2}\tilde{\mathbf{P}}^{\mathrm{NL}}(z,t). \tag{22}$$

The Laplacian will be approximately written as

$$\nabla^2\mathbf{E}(z,t) \approx \frac{\partial^2\mathbf{E}(z,t)}{\partial z^2} \approx \left(2ik\frac{\partial\mathbf{A}}{\partial z} - k^2\mathbf{A}\right)\exp[-i(\omega t - kz)] + \mathrm{c.c.}, \tag{23}$$

ignoring the transverse derivatives and the second derivative with respect to *z*. In this approximation, we derive the following equation for the complex amplitude of the probe field:

$$\frac{d\mathbf{S}}{dz} = -\delta\mathbf{S} + i\frac{2\pi\omega^2}{kc^2}(\mathbf{P}_{\mathrm{loc}} + \mathbf{P}_{\mathrm{nonloc}}). \tag{24}$$

We will direct the unit vector $\mathbf{e}_3$ along the vector $\mathbf{k}$ and will assume that the angle between the wave vectors $\mathbf{K}$ and $\mathbf{k}$ to be small enough for the components F_3 and K_1, K_2 to be ignored. We will use relations (16) and rewrite Eqs. (21) in the circular components,

$$\begin{aligned}(P_{\text{loc}})_{\pm} &= \chi_1(|F|^2 S_{\pm} + |F_{\pm}|^2 S_{\pm} + F_{\pm}F_{\mp}^* S_{\mp}) \\ &+ 2\chi_2(|F_{\mp}|^2 S_{\pm} + F_{\mp}^* F_{\pm} S_{\mp}),\end{aligned} \tag{25}$$

$$\begin{aligned}(P_{\text{nonloc}})_1 &= 2ig_1\{(\mathbf{F}^* \cdot \mathbf{F})S_2 k_3 + (\mathbf{F}^* \cdot \mathbf{S})F_2 K_3\} \\ &- 2ig_2(\mathbf{F} \cdot \mathbf{S})F_2^* K_3 = \sqrt{2}g_1\{|F|^2(S_+ - S_-)k_3 \\ &+ K_3(|F_+|^2 S_+ - |F_-|^2 S_-)\} + 2g_2 K_3(|F_+|^2 S_- - |F_-|^2 S_+), \\ (P_{\text{nonloc}})_2 &= -2ig_1\{(\mathbf{F}^* \cdot \mathbf{F})S_1 k_3 + (\mathbf{F}^* \cdot \mathbf{S})F_1 K_3\} \\ &+ 2ig_2(\mathbf{F} \cdot \mathbf{S})F_1^* K_3 = -i\sqrt{2}g_1\{|F|^2(S_+ + S_-)k_3 \\ &+ K_3(|F_+|^2 S_+ + |F_-|^2 S_-)\} \\ &+ 2ig_2 K_3(|F_+|^2 S_- + |F_-|^2 S_+), \\ (P_{\text{nonloc}})_+ &= 2g_1\{|F|^2 S_+ k_3 + |F_+|^2 K_3 S_+\} \\ &- 2g_2 K_3 |F_-|^2 S_+, \\ (P_{\text{nonloc}})_- &= -2g_1\{|F|^2 S_- k_3 + |F_-|^2 K_3 S_-\} \\ &+ 2g_2 K_3 |F_+|^2 S_-.\end{aligned} \tag{26}$$

In the case of arbitrary polarization of strong and probe fields, the slow amplitude $\mathbf{S}(z)$ satisfies the equation

$$\begin{aligned}\frac{d}{dz}\begin{pmatrix} S_+ \\ S_- \end{pmatrix} &= -\delta\begin{pmatrix} S_+ \\ S_- \end{pmatrix} + i\frac{2\pi\omega^2}{k_3 c^2} \\ &\times \begin{pmatrix} M_{++} & F_-^* F_+(\chi_1 + 2\chi_2) \\ F_- F_+^*(\chi_1 + 2\chi_2) & M_{--} \end{pmatrix}\begin{pmatrix} S_+ \\ S_- \end{pmatrix},\end{aligned} \tag{27}$$

$$\begin{aligned}M_{++} &= \chi_1(|F|^2 + |F_+|^2) + 2\chi_2|F_-|^2 \\ &+ 2g_1 k_3|F|^2 + 2K_3(g_1|F_+|^2 - g_2|F_-|^2), \\ M_{--} &= \chi_1(|F|^2 + |F_-|^2) + 2\chi_2|F_+|^2 \\ &- 2g_1 k_3|F|^2 - 2K_3(g_1|F_-|^2 - g_2|F_+|^2).\end{aligned}$$

In the general case, the effect of a strong field on the polarization of a probe field may result in two effects, namely, the rotation of the polarization plane (or, to be more precise, the rotation of the polarization ellipse) and the variation of ellipticity, i.e., of the ratio between the ellipse semiaxes. We will analyze Eq. (27) as applied to these effects.

We will treat in more detail two particular cases of the ratio between the polarizations of strong and probe fields.

Inverse Faraday Effect

Let an pump wave have a circular polarization (right-hand, for definiteness) and a probe wave be linearly polarized. We will be interested in the rotation of the polarization plane of the probe wave. This nonlinear effect (ignoring the spatial dispersion of nonlinearity) is known in the literature as the inverse Faraday effect.

We assume in Eq. (27) that $F_- = 0$ to derive

$$\begin{aligned}\frac{d}{dz}\begin{pmatrix} S_+ \\ S_- \end{pmatrix} &= -\delta\begin{pmatrix} S_+ \\ S_- \end{pmatrix} + i\frac{2\pi\omega^2}{k_3 c^2}|F_+|^2 \\ &\times \begin{pmatrix} 2\chi_1 + 2g_1(k_3 + K_3) & 0 \\ 0 & \chi_1 + 2\chi_2 + 2g_2 K_3 \end{pmatrix}\begin{pmatrix} S_+ \\ S_- \end{pmatrix},\end{aligned} \tag{28}$$

whence we derive the following equation for the angle of rotation of the polarization plane:

$$\begin{aligned}\frac{d\alpha}{dz} &= \frac{\pi\omega}{cn_0}|F_+|^2 \\ &\times \text{Re}[\chi_1 - 2\chi_2 + 2g_1(k_3 + K_3) - 2g_2 K_3].\end{aligned} \tag{29}$$

Therefore, the angle of rotation in the general case depends on all four constants of cubic susceptibility of the medium.

Optical Kerr Effect

Let both waves have a linear polarization, with the pump wave along the axis x, $F_2 = 0$, and the polarization of the probe wave forming an angle of 45° with respect to the pump wave, so that $S_1 = S_2$. It is known that a strong field of linear polarization induces in the medium an anisotropy of the refractive index, i.e., the medium becomes birefringent. As a result, a phase shift arises between the probe field components S_1 and S_2, which leads to a variation of the ellipticity of polarization.

In order to find the expression for induced birefringence, we will rewrite Eq. (27) in Cartesian components and assume that $F_2 = 0$,

$$\begin{aligned}\frac{d}{dz}\begin{pmatrix} S_1 \\ S_2 \end{pmatrix} &= -\delta\begin{pmatrix} S_1 \\ S_2 \end{pmatrix} + i\frac{2\pi\omega^2}{c^2 k_3}|F_1|^2 \\ &\times \begin{pmatrix} 2(\chi_1 + \chi_2) & i[2g_1 k_3 + K_3(g_1 - g_2)] \\ -i[2g_1 k_3 + K_3(g_1 - g_2)] & \chi_1 \end{pmatrix} \\ &\times \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}.\end{aligned} \tag{30}$$

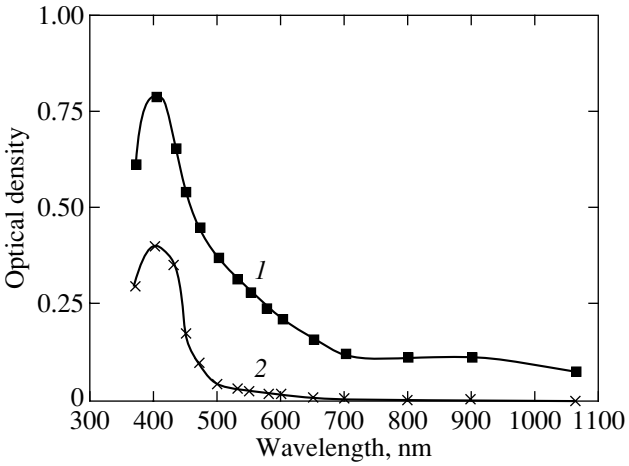

Fig. 1. Absorption spectra of colloidal solutions of silver.

On turning to equations for the real and imaginary parts $S_{1,2} \equiv s_{1,2}\exp(i\varphi_{1,2})$, we derive the following set of equations:

$$\frac{ds_1}{dz} = -\delta s_1 - \beta[2\mathrm{Im}(\chi_1 + \chi_2)s_1 + s_2 \mathrm{Re}G\cos(\varphi_1 - \varphi_2)],$$

$$\frac{ds_2}{dz} = -\delta s_2 + \beta[s_1 \mathrm{Re}G\cos(\varphi_1 - \varphi_2) - s_2 \mathrm{Im}\chi_1], \quad (31)$$

$$\frac{d(\varphi_1 - \varphi_2)}{dz} = \beta\left[\mathrm{Re}\chi_1 + 2\mathrm{Re}\chi_2 - \mathrm{Im}G\left(\frac{s_1}{s_2}\sin(\varphi_1 - \varphi_2) - \frac{s_2}{s_1}\sin(\varphi_1 - \varphi_2)\right)\right],$$

where the notation

$$\beta \equiv \frac{2\pi\omega^2}{c^2 k_3}|F_1|^2, \quad G \equiv 2g_1 k_3 + K_3(g_1 - g_2)$$

is used for brevity.

The main contribution in the right-hand part of the latter equation is made by the first two terms, because the third term describes a nonlinear effect of a higher order, the difference of the ratio s_1/s_2 from unity *per se* being a consequence of nonlinearity of the medium. In view of the foregoing, the nonlinear phase shift between the x and y components $\Delta\varphi^{\mathrm{NL}} = \varphi_1 - \varphi_2$ is described with a good accuracy by the following equation:

$$\frac{d\Delta\varphi^{\mathrm{NL}}}{dz} = \frac{2\pi\omega}{cn_0}|F_1|^2 \mathrm{Re}(\chi_1 + 2\chi_2). \quad (32)$$

The nonlocal response of the medium (constants g_1 and g_2) in the adopted approximation makes no contribution to the phase difference between the Cartesian components.

Thus, we have clarified the correlation between the nonlinear polarization effects, i.e., OKE and IFE, and the coefficients of polarizability of the medium of the third order.

3. EXPERIMENT

3.1. Preparation of Aggregated Colloids

Colloidal solutions of metals may be obtained in different solvents such as water, ethyl alcohol, and others. Several methods exist for the preparation of metal nanoparticles in reducing reactions in salt solutions and further stabilization of colloidal solution. In our experiments, we investigated an alcoholic colloidal solution of silver prepared by reduction of silver from $AgNO_3$ in a solution containing NaOH and polyvinylpyrrolidone with a mean molecular weight of 360 000. This procedure was described by Hirai [15]; such a colloid will be designated below as Ag(PVP). In accordance with the results of an electron-microscopic study, a typical diameter of nanoparticles in a colloidal solution prepared by this formula is approximately 14 nm.

We had at our disposal two colloid samples with different degrees of aggregation, namely, a less aggregated sample no. 2 and more aggregated sample no. 1. Figure 1 gives spectra of linear (i.e., in a weak field) absorption of these samples. A powerful long-wave wing of the absorption band of sample no. 1 is indicative of aggregation of nanoparticles [1]. According to the data of electron micrography, the spectrum of the more aggregated solution no. 1 corresponds to the presence in the colloid of large clusters consisting of hundreds of particles; such aggregates have a fractal structure with the fractal dimension $D_f \approx 1.8$. However, most of the nanoparticles are contained in small (of the order of ten particles) aggregates. The less aggregated solution (no. 2) contains clusters consisting of several particles and, apparently, a significant number of isolated nanoparticles.

For experimental purposes, colloids of both types were diluted with alcohol in one and the same proportion (approximately 1 : 5) so that the transmission of colloid no. 1 on the wavelength of 532 nm (at a low intensity of light) would be approximately 50%.

3.2. Measurement Procedure

The second harmonic of a YAG:Nd pulsed laser with the pulse duration $\tau_p \approx 10$ ns and wavelength of 532 nm was used in the experiments. After the frequency doubler, the radiation passed through a polarizer, i.e., a Glan prism, which provides for residual ellipticity on a level of the order of 10^{-5} with respect to intensity. Then, the radiation was separated into two beams, pump and probe ones, which were focused to a cuvette of fused

quartz 3 mm thick containing the silver solution being investigated. In so doing, the waist diameter of both beams (with respect to the 1/*e* level) was 0.5–0.6 mm, and the power ratio between the pump and probe beams was approximately 13. The signals were recorded by photodiodes with a response time of 10 μs.

In experiments with a probe field (IFE, OKE), the parameters of the probe beam that passed through a polarimeter were measured. In experiments involving the investigation of self-action effects (nonlinear optical activity, self-rotation of the polarization ellipse), the probe beam was shut off, and the parameters of the pump field were measured.

Self-Action Effects

In developing the measuring method, the following circumstances were taken into account. In a medium that is nongyrotropic in a linear approximation, the nonlinear rotation of the plane of polarization of a linearly polarized wave ($|A_+|^2 = |A_-|^2$ is caused by the nonlocality of nonlinear response alone. In Eq. (17), the term containing $(2g_1 - g_2)$ corresponds to this effect. In the case of elliptic polarization, an additional rotation of the polarization plane occurs, which is associated with the local nonlinear response (the term with $\mathrm{Re}\chi_2$). Because, under the experimental conditions, the polarization of radiation is always "weakly elliptic," both terms must be taken into account. In order to separate the contributions by the local and nonlocal responses, use was made of the method suggested by us in [8]. This method is essentially as follows. Note that, in Eq. (17),

$$|A_+|^2 - |A_-|^2 = 2|A_x||A_y|\sin(\phi_x - \phi_y),$$

where $|A_{x,y}|$ and $\phi_{x,y}$ denote the amplitudes and phases of linearly polarized field components. By varying the phase difference $\phi_x - \phi_y$, one can vary the local response contribution, while the nonlocal response contribution will remain constant.

The scheme of measurements of self-action effects is given in Fig. 2. The radiation passed successively through a polarizer *1* (Glan prism), a phase element *2*, a cuvette *3* of thickness $l = 3$ mm with the colloid being investigated, and an analyzer *4* (calcite wedge) and was registered by two silicon photodiodes *5* and *6*. For measuring the intensity, a part of the radiation was directed by a beam-splitting plate to a photodiode (not shown in the scheme). The radiation that passed through the polarizer had a weak ellipticity with the semiaxis ratio

$$|A_y|^2/|A_x|^2 \approx 5 \times 10^{-5}$$

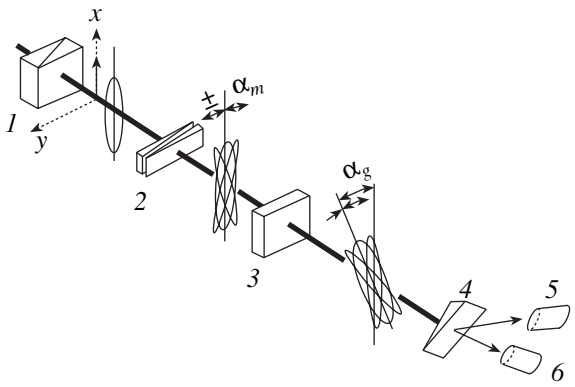

Fig. 2. The scheme of measurements of rotation of the polarization plane in the self-action configuration. The transformation of the state of polarization of a wave passing through optical elements including the medium being investigated is demonstrated.

(the *x* axis is directed along the polarizer axis, and the *y* axis is directed normally to the *x* axis and to the wave vector). The components of the complex amplitude of radiation that passed through the phase element assume an additional phase shift,

$$\begin{aligned} A_x(0) &= |A_x(0)|e^{i\phi_x(0)}, \\ A_y(0) &= |A_y(0)|e^{i\phi_y(0)}, \\ \Delta\phi_0 &= \phi_x(0) - \phi_y(0). \end{aligned} \tag{33}$$

So, the effect of the phase element brings about a variation of the azimuth and ratio of the semiaxes of the radiation polarization ellipse at the inlet to the medium. The ellipse azimuth $\alpha(0)$ relative to the polarizer in view of the smallness of $|A_y|/|A_x| \ll 1$ is given by

$$\alpha(0) = \frac{|A_y(0)|}{|A_x(0)|}\cos(\Delta\phi_0). \tag{34}$$

The radiation that passed through the medium being investigated falls on the analyzer, whose axis is oriented at an angle of 45° to the polarizer. The orthogonally polarized components isolated by the analyzer hit the photodiodes *5* and *6*. The difference of the signals from the photodiodes was recorded,

$$\Delta I = I_2 - I_1 \approx |A_x|^2 2\alpha, \tag{35}$$

where α is the azimuth of the polarization ellipse at the outlet from the medium; here, we use the fact of $\alpha \ll 1$, as well as the smallness of ellipticity.

The $\Delta I(\Delta\phi_0)$ dependence for different radiation intensities at the inlet to the medium was measured in the experiment. We will assume that

$$\Delta\phi(z) = \Delta\phi_0 + \Delta\phi_{\mathrm{NL}}(z)$$

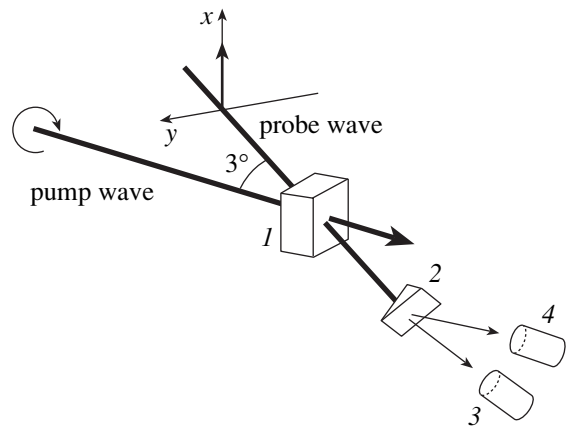

Fig. 3. The scheme of the experimental facility for the investigation of the inverse Faraday effect: (*1*) 3-mm cuvette with a colloidal silver solution, (*2*) analyzer, (*3*, *4*) photodetectors.

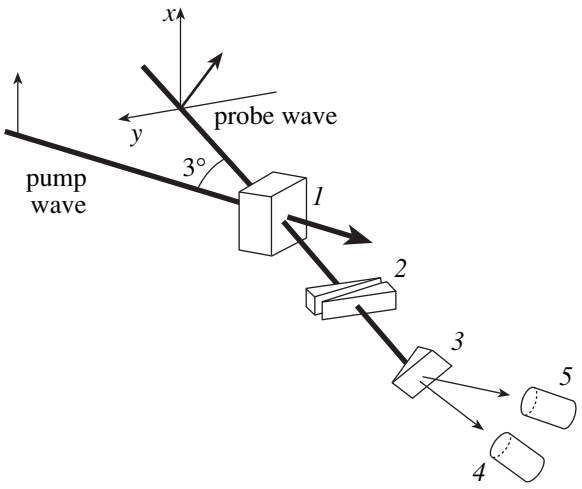

Fig. 4. The scheme of the facility for measuring the optical Kerr effect: (*1*) 3-mm cuvette with a colloidal silver solution, (*2*) phase element, (*3*) analyzer, (*4*, *5*) photodetectors.

and, in view of the smallness of $\Delta\phi_{NL}(z) \ll 1$, derive, from Eqs. (17) and (35),

$$\Delta I(\Delta\phi_0) = a + b\cos(\Delta\phi_0 + c),$$
$$a = 2\rho_1'|A_x(l)|^2\int_0^l |A_x|^2 dz, \quad b = 2|A_x(l)|^2\frac{|A_y(0)|}{|A_x(0)|}, \tag{36}$$
$$c = -2\sigma_2'\frac{|A_y(0)|}{|A_x(0)|}\int_0^l |A_x||A_y| dz,$$

where

$$\sigma_2' = \frac{2\pi\omega}{n_0 c}\mathrm{Re}\chi_2, \quad \rho_1' = \frac{2\pi\omega^2}{c^2}\mathrm{Re}(2g_1 - g_2).$$

One can see that the local (σ_2) and nonlocal (ρ_1) nonlinearities manifest themselves in $\Delta I(\Delta\phi_0)$ differently as the inlet intensity varies: the term with σ_2' may cause some phase shift, while the term with ρ_1' causes a shift of the $\Delta I(\Delta\phi_0)$ curve as a whole on the ordinate. By the sign of vertical shift, one can determine the direction of rotation, considering that the plus sign corresponds to counterclockwise rotation if one looks toward the beam.

Inverse Faraday Effect

The scheme of the experiment in recording the IFE is given in Fig. 3. In this case, a strong beam was imparted clockwise circular polarization with the aid of a phase element.

The experiment involved measurements of the rotation angle of the polarization plane of the probe beam. For this purpose, the beam was directed to the analyzer (calcite wedge), whose axes were oriented at an angle of 45° to the *x* axis. The analyzer divided the probe field into two orthogonally polarized beams, which then fell on photodetectors. The signals from the detectors are expressed in terms of the angle α between the *x* axis and the polarization plane of the probe field,

$$I_1 = |S_x|^2\cos^2(45° - \alpha),$$
$$I_2 = |S_x|^2\sin^2(45° - \alpha). \tag{37}$$

For small values of α,

$$\alpha \approx \frac{I_1 - I_2}{2|S_x|^2} = \frac{\Delta I}{2|S_x|^2} \tag{38}$$

is valid.

Optical Kerr Effect

The scheme given in Fig. 4 was used for measuring the OKE. The probe beam polarization was directed at an angle of 45° to the pump beam polarization. A phase element, which makes possible the introduction of the phase difference between the *x* and *y* components, was placed in the way of the probe beam after the cuvette. Then, the probe beam passed through the analyzer oriented as in the previous scheme (the axis at 45° to the *x* axis). The beams isolated by the analyzer fell on photodetectors, the signals from which may be written as follows:

$$I_1 = \frac{1}{2}(|S_x|^2 + |S_y|^2 - 2|S_x||S_y|\cos\Delta\varphi),$$
$$I_2 = \frac{1}{2}(|S_x|^2 + |S_y|^2 + 2|S_x||S_y|\cos\Delta\varphi), \tag{39}$$

where $\Delta\varphi$ is the phase difference between the *x* and *y* components; $\Delta\varphi$ contains the nonlinear phase shift

$\Delta\varphi^{NL}$ accumulated in the medium and the phase difference introduced by the phase element $\Delta\varphi_0$,

$$\Delta\varphi = \Delta\varphi_0 + \Delta\varphi^{NL}. \quad (40)$$

In the experiment, the difference signal $I_2 - I_1$ was measured; with a small nonlinear phase shift, this signal may be written as

$$\Delta I \equiv I_2 - I_1 \approx 2|S_x||S_y|(\cos\Delta\varphi_0 - \Delta\varphi^{NL}\sin\Delta\varphi_0). \quad (41)$$

The phase element set the shift $\Delta\varphi_0 = -\pi/2$; in so doing, the nonlinear shift is expressed as follows in terms of the quantities being measured (in view of the fact that $|S_x| = |S_y|$):

$$\Delta\varphi^{NL} = \frac{\Delta I}{2|S_x|^2}. \quad (42)$$

3.3. Inclusion of Time and Space Averagings

In a real experiment, the pumping field intensity is always characterized by some distribution in time and space. Nevertheless, the phenomenological treatment performed above for plane monochromatic waves may be applied directly if the nonlinearity relaxation time (both local and nonlocal) is short compared with the pulse duration and the characteristic size of nonlocality of interaction is small compared with the beam diameter.

We used photodetectors with the reaction time $T \gg \tau_p$ (T = 10 μs), so that time-average quantities were actually measured. The size of the photodiode pad exceeded the light beam diameter; therefore, cross-section-averaged signals were recorded during measurements. On assuming the Gaussian distribution of intensity in time and in the cross section, one can readily find that, in processing the experimental data, the values obtained from Eqs. (38) and (42) must be multiplied by $2\sqrt{2}$.

The reduction of the pump field during the propagation along the medium due to linear absorption is taken into account by introducing the effective intensity

$$I_{eff} = \frac{\int_0^l I(z)dz}{l}.$$

4. MEASUREMENT RESULTS

The results of measurement of the angle of nonlinear rotation of the polarization plane, which will be designated as α_g, are given in Fig. 5. In the $I_0 < 2$ MW/cm^2 range, the value of α_g for sample no. 1 depends approximately linearly on I_0 (I_0 is the intensity before the cuvette on the beam axis at the pulse maximum). For sample no. 2, the linear dependence $\alpha_g(I_0)$ persists until

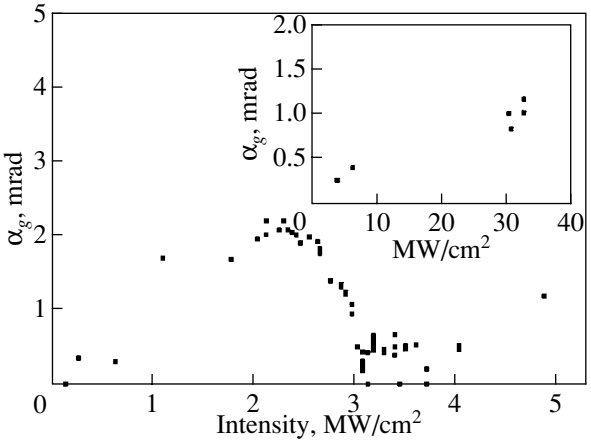

Fig. 5. The angle of rotation of the polarization plane due to nonlinear optical activity as a function of the intensity of incident radiation. The main graph, colloid no. 1; the inset, colloid no. 2.

$I_0 \approx 30$ MW/cm^2. If we reduce the measurement results to the same concentration of silver, the value of

$$\alpha_g/I_{eff} \propto \mathrm{Re}\Gamma^{(3)}$$

in this linear range for a more aggregated colloid no. 1 is approximately 80 times that for colloid no. 2. Note the decrease in the nonlinear constant at the intensity $I_0 > 2$–3 MW/cm^2 approximately corresponding to the threshold of photomodification of fractal clusters of silver, i.e., to the variation of their structure under the light pulse effect [16]. Previously, a similar manifestation of photomodification was observed for nonlinear refraction of $\mathrm{Re}\chi^{(3)}$ measured by dispersion interferometry [17]. A fivefold decrease in $|\chi^{(3)}|$ was observed using the method of degenerate four-photon scattering [7]. The absence of any special features in $\alpha_g(I_0)$ for colloid no. 2 in the $I_0 \approx 2$ MW/cm^2 range supports this assumption indirectly: nonaggregated particles of silver in colloid no. 2 do not experience the photomodification that is characteristic of clusters in sample no. 1.

The combination of components of the tensor $\Gamma^{(3)}$ may be found from the relation

$$\alpha_g = \frac{16\pi^4 l}{n_0 c\lambda^2} \times 10^{13}\mathrm{Re}(2g_1 - g_2)I_{eff}\ [\mathrm{MW/cm^2}], \quad (43)$$

where all quantities, except for the intensity, are expressed in units of the CGS electrostatic system (CGSE) and n_0 is the refractive index of colloid. We take into account the difference between the concentrations of silver in colloids and perform the averaging with respect to time and beam cross section, which produces the factor $2\sqrt{2}$, to derive

$\mathrm{Re}(2g_1 - g_2) \approx 0.9 \times 10^{-16}$ SGSE units for colloid no. 1,

$\mathrm{Re}(2g_1 - g_2) \approx 1.1 \times 10^{-18}$ SGSE units for colloid no. 2.

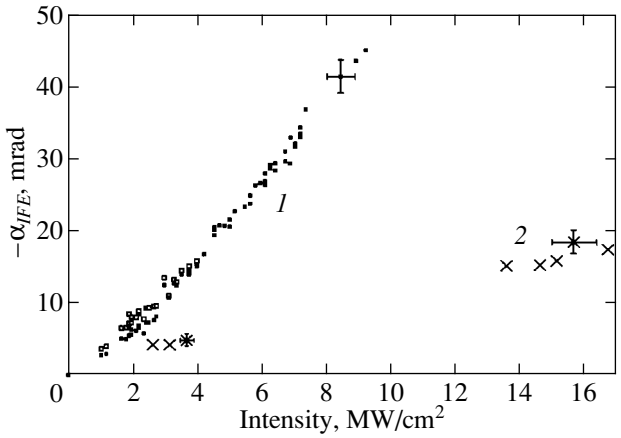

Fig. 6. The rotation angle of the polarization plane of probe field as a function of the intensity of pump field with circular polarization. Solid squares, colloid no. 1; crosses, colloid no. 2. Hollow squares indicate the results of subtraction of the angle of rotation due to nonlinear optical activity, according to the data of Fig. 5 (for colloid no. 1).

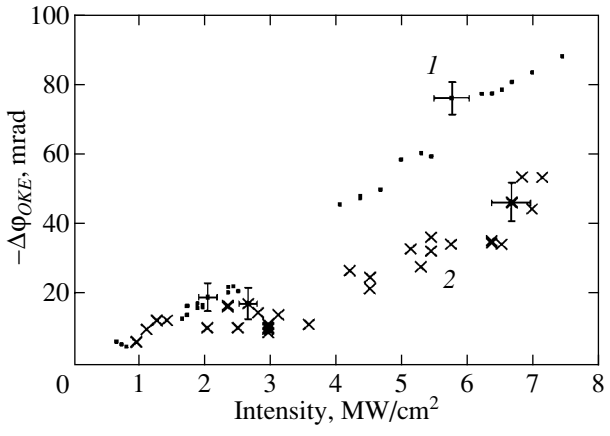

Fig. 7. The nonlinear phase shift between the *x* and *y* components of probe field as a function of the intensity of pump field. Solid squares, colloid no. 1; crosses, colloid no. 2.

In both cases, the medium turned out to be levorotatory.

In order to find the constant χ_2, we performed an experiment involving the measurement of the rotation angle of polarization ellipse of pump field through a nonlinear medium. The measurements were performed for colloidal solution no. 1. With the intensity before the cuvette varying from ≈0.7 to ≈5 MW/cm^2, the ellipse rotation was $\alpha_{SR} \approx -1.8 \pm 0.2$ mrad for right-hand polarization and $\alpha_{SR} \approx 2.1 \pm 0.2$ mrad for left-hand polarization, with the ellipticity $|F_y|^2/|F_x|^2 \approx 1:6$. One can see that the rotation angles for left-hand and right-hand polarizations differ somewhat in magnitude; this is apparently a manifestation of nonlinear gyrotropy in accordance with Eq. (17) and approximately corresponds to nonlinear optical activity at 5 MW/cm^2. By taking a half-difference of these values, one can eliminate the effect of nonlinear optical activity and use formula (17) to calculate χ_2,

$$\mathrm{Re}\chi_2 = (1.3 \pm 0.2) \times 10^{-11} \text{ CGSE units}. \qquad (44)$$

These experiments further involved measurements of the nonlinear absorption of the light wave in colloidal solution no. 1, which enabled one to determine the imaginary part of the sum of the components χ_1 and χ_2,

$$\mathrm{Im}(\chi_1 + \chi_2) = 2\sqrt{2}\frac{\delta(e^{-2\delta l} - T)}{T(1 - e^{-2\delta l})} \times \frac{cn_0\lambda \text{ [CGSE units]}}{16\pi^3 \times 10^{13} I_0 \text{ [MW/cm}^2\text{]}} \approx -1.1 \times 10^{-10} \text{ CGSE units}, \qquad (45)$$

where T is the transmission coefficient (with respect to intensity), $\delta = 1.15$ cm^{-1}, and $l = 3$ mm. In the experiment, the transmission was $T \approx 0.65$ with $I_0 \approx 6$ MW/cm^2.

Figure 6 gives the rotation angle of the polarization plane of probe beam, α_{IFE}, in the IFE scheme as a function of the pump beam intensity for two colloidal solutions with different degrees of aggregation. The value of the angle α was found by formula (38) with due regard for the averaging factor $2\sqrt{2}$. The results of measurements of the OKE for the same samples are given in Fig. 7. Given in the graph is the value of the nonlinear phase shift between the *x* and *y* components of the probe beam as a function of the pumping intensity, determined from the experimental data according to Eq. (42).

The values of errors given in Figs. 6 and 7 are reflective of the estimates for the random error of measurement at each point. However, the error in the values of the nonlinear constants χ_1 and χ_2 given below is largely preassigned by the systematic error of determination of the radiation intensity at the inlet to the medium; according to our estimates, this latter error is ±10%.

Note that the experiments described by us failed to reveal any saturation of either the IFE or the OKE with increasing intensity of incident radiation up to 35 MW/cm^2. The possible reason for this is discussed below (see Section 5).

It follows from Eq. (29) that both the local and nonlocal responses of the nonlinear medium contribute to the rotation of the plane of polarization of the probe wave, IFE. In our case, the rotation due to nonlinear optical activity proceeds in the opposite direction compared with the IFE; this brings about a deviation of the dependence of the IFE (curve *1* in Fig. 6) from linear with the intensity of up to 3 MW/cm^2. The result of subtraction (with due regard for the sign) of the effect of

nonlinear optical activity is also given in Fig. 6 (hollow squares). In the case of high intensity, the contribution by nonlocal nonlinearity becomes insignificant, and the slope of the dependence is in fact defined by the local part alone, which enables one to find the following quantity:

$$\mathrm{Re}(\chi_1 - 2\chi_2) = 2\sqrt{2}\frac{\lambda\alpha_{\mathrm{IFE}}}{l} \times \frac{n_0^2 c\ [\mathrm{CGSE\ units}]}{4\pi^3 \times 10^{13} I_{\mathrm{eff}}\ [\mathrm{MW/cm^2}]}\ \mathrm{CGSE\ units}. \quad (46)$$

The absorption in colloid no. 1 with the cuvette thickness $l = 3$ mm was 50%, which corresponds to $I_{\mathrm{eff}} \approx 0.72 I_0$. We substitute numerical values to derive

$$\mathrm{Re}(\chi_1 - 2\chi_2) = -(1.8 \pm 0.23) \times 10^{-10}\ \mathrm{CGSE\ units}$$
for colloid no. 1,

$$\mathrm{Re}(\chi_1 - 2\chi_2) = -(4.0 \pm 0.55) \times 10^{-11}\ \mathrm{CGSE\ units}$$
for colloid no. 2.

This error includes also the systematic error caused by possible nonoptimal overlapping of beams of pump and probe fields in the cuvette with colloid.

The magnitude of birefringence due to the OKE is proportional to the sum of two nonlinear coefficients χ_1 and $2\chi_2$. We approximate the data in Fig. 7 by a linear dependence to find

$$\mathrm{Re}(\chi_1 + 2\chi_2)\ [\mathrm{CGSE\ units}] = 2\sqrt{2}\frac{\lambda\Delta\varphi^{\mathrm{NL}}}{l} \times \frac{n_0^2 c\ [\mathrm{CGSE\ units}]}{8\pi^3 \times 10^{13} I_{\mathrm{eff}}\ [\mathrm{MW/cm^2}]} \quad (47)$$

and the following numerical values:

$$\mathrm{Re}(\chi_1 + 2\chi_2) = -(2.1 \pm 0.35) \times 10^{-10}\ \mathrm{CGSE\ units}$$
for colloid no. 1,

$$\mathrm{Re}(\chi_1 + 2\chi_2) = -(1.0 \pm 0.42) \times 10^{-10}\ \mathrm{CGSE\ units}$$
for colloid no. 2.

One can find χ_1 by combining the results of measurements of the IFE and OKE,

$$\mathrm{Re}\chi_1 = -(1.9 \pm 0.21) \times 10^{-10}\ \mathrm{CGSE\ units}$$
for colloid no. 1,

$$\mathrm{Re}\chi_1 = -(7 \pm 2.1) \times 10^{-11}\ \mathrm{CGSE\ units} \quad (48)$$
for colloid no. 2.

The error of our measurements prevents us from finding χ_2 using these data.

5. DISCUSSION

The polarization measurements of $\mathrm{Re}(\chi_1 + \chi_2)$ and the values of $\mathrm{Im}(\chi_1 + \chi_2)$ (it will be recalled that $\chi_{1111} = (\chi_1 + \chi_2)/3$) produce the value of $|\chi_{1111}| \approx 6.7 \times 10^{-11}$ CGSE units, which is close to the susceptibility of the most aggregated aqueous colloid ($|\chi_{1111}| \sim 10^{-10}$ CGSE units), measured by the efficiency of degenerate four-photon scattering [18] on the wavelength of 532 nm, and considerably exceeds the nonlinear susceptibility of monomers $|\chi_{1111}| \sim 10^{-13}$ CGSE units [18]. According to the results of our measurements of the OKE and IFE, the local nonlinear susceptibility depends little on the cluster size.[1] Note that, according to the data of [6] obtained for aqueous borohydride colloid using the method of degenerate four-photon scattering, the cubic susceptibility increases most rapidly in the early stages of aggregation, when monomers unite into small aggregates. In the case of comparison, the colloids were taken to have the same degree of aggregation if the contours of the linear absorption band were similar. Therefore, a significant increase in $\chi^{(3)}$ compared with monomer occurs even for a weakly aggregated colloid, in which, according to the data of electron microscopy, monomers form small groups of two–three particles. A further increase of clusters to 100–1000 particles causes a weaker growth of $\chi^{(3)}$. This means that it is of key importance from the standpoint of amplification of nonlinear responses to provide for the conditions of collective resonance on the frequency of incident radiation, which may well be accomplished even in the case of two fairly closely arranged monomers owing to the interaction between induced dipoles (or multipoles). Indeed, an increase in a local field with respect to an incident one for resonant modes in a binary approximation produces

$$E_{i\,\mathrm{res}}/E_0 \approx \varepsilon_i^2/3\varepsilon_h\varepsilon_2 = 18 \quad \text{for} \quad \lambda = 532\ \mathrm{nm},$$

where $\varepsilon = \varepsilon_1 + i\varepsilon_2$ and ε_h denote the permittivity of metal particle and surrounding medium, respectively, which is comparable with the estimate for large aggregates.

A singular feature of the results of our measurements of local nonlinear response in Ag(PVP) is that $\mathrm{Re}\chi^{(3)}$ (OKE and IFE) and $\mathrm{Im}\chi^{(3)}$ (nonlinear absorption) almost do not vary with intensity up to 30 MW/cm^2 (the corresponding energy density, 300 mJ/cm^2), which exceeds the known thresholds of photomodification of aggregates of silver in colloids [19]. In previous studies, a significant variation of nonlinear responses was observed for the energy density exceeding the threshold value, namely, a reduction of $|\chi^{(3)}|$ in experiments with 30-ps pulses for an aggregated colloid of silver with denaturated proteins [18] and a reduction, in experi-

[1] It must be borne in mind that strongly and weakly aggregated solutions usually differ by the concentration of metallic silver as well; naturally, this reflects on the magnitude of the nonlinear effect. The data in Figs. 6 and 7 correspond to the concentration of nanoparticles in colloid no. 1 exceeding that in colloid no. 2 by a factor of approximately 2.4; the inclusion of this difference will bring about the convergence of the curves for colloids with different degrees of aggregation.

ments with 15-ns pulses, of $\mathrm{Im}\chi^{(3)}$ in a borohydride colloid [7] and of $\mathrm{Re}\chi^{(3)}$ in a borohydride colloid and in Ag(PVP) [17] (use was made of PVP with a molecular weight of 40 000).[2] The photomodification thresholds for colloids of different types are in the limits from 5 to 10 mJ/cm^2 and are approximately the same for pico- and nanosecond ranges. At the same time, in some colloids containing polymer molecules (for example, in collargol [16] and in AG(PVP)), higher values of the threshold energy of photomodification were observed. Presumably, the reason for photomodification is the integration of resonant monomers as a result of their heating by laser radiation and subsequent motion of dislocations, which arises in heated particles upon their contact owing to a high shear stress [4]. Such a mechanism may be realized directly in coalescent aggregates consisting of nanoparticles in contact with one another. Colloidal silver aggregates prepared by the borohydride technique, as well as under conditions of reduction of $AgNO_3$ in a mixture of EDTA and NaOH, often have a coalescent structure [20]. However, in the case of coagulation aggregates in which the monomer boundaries are separated by some distance, this mechanism of photomodification presumes a preliminary mutual approach of particles. Characteristic distances between particles are as follows [20]: in collargol, 2 to 2.5 nm (with the average diameter $d_{av} \approx 14$ nm); in Ag(PVP), 1–2 nm ($d_{av} \approx 14$ nm); in silver colloid prepared by the Carey–Lee method, 1 nm (with $d_{av} \approx 12 \pm 3$ nm); and, in aggregated gold colloid prepared by the citrate technique, 0.7 nm ($d_{av} \approx 19$ to 22 nm). Therefore, Ag(PVP) colloid belongs to a fairly frequently occurring coagulation type.

The motion of monomers may occur owing to the particle interaction induced by the field [21, 22]. Estimates (see the Appendix) indicate that the particle displacement in such colloids during the time of a nanosecond pulse may be of the order of 1 nm.

The results of analysis performed in the Appendix lead one to three important conclusions.

(i) The interaction of particles in the field of laser radiation may lead to appreciable shifts (1 nm at 20 mJ/cm^2; the time of relaxation of the process to steady state, 100 ps) which enable the particles to pass the potential barrier and converge to a distance at which the dispersion forces of attraction are effective. Given the same energy, the pulse duration (pico- or nanoseconds) is of no importance.

(ii) The result of mutual approach of particles (whether or not it leads to adhesion required for photomodification) depends strongly on the potential curve of particle interaction in the absence of a field, i.e., on the double-layer parameters. Note that, in some colloids, the so-called steric force is significant, which arises as a result of overlapping of the adsorption layers of surfactants and polymers. The steric force has an additional stabilizing effect. Therefore, in polymer colloids, the threshold of integration of particles may be much higher.

(iii) The particle displacement must bring about nonlinear optical effects, i.e., contribute to $\chi^{(3)}$, because the range of distances between particles, in which a pair is in resonance with low-frequency radiation, is of the order of 1 nm, which is comparable with the estimate for induced shift of particles.

The particle shift occurs with a characteristic relaxation time of 100 ps (see the Appendix). This inertia may serve as an explanation of the following result of our measurements: according to Eqs. (44) and (48), we have

$$\mathrm{Re}(\chi_2/\chi_1) \approx -0.07,$$

which is an order of magnitude less than the typical value (approximately 0.5) characteristic of the nonresonance electron nonlinearity of solids [23]. This fact may be treated as evidence of the existence of an inertial contribution to the nonlinear optical response of silver colloids. In [20], the variation of the particle temperature due to radiation absorption and the associated variation of ε were treated as the inertial mechanism of nonlinearity of aggregated colloids of silver. The clarification of the relative importance of thermal and mechanical inertia calls for additional investigation.

We will now discuss the differences in the dependences of local and nonlocal responses on the degree of aggregation. Unlike the IFE, OKE, and linear absorption, the effect of nonlinear optical activity at low intensity increases approximately eighty times for colloid no. 1 compared with colloid no. 2 and, for $I > 2$ MW/cm^2, it decreases several times. In our opinion, this difference is associated with the fact that, in the case of nonlocal effects, it is the size of the region occupied by the resonant mode relative to the wavelength that is important and, consequently, the size of aggregates in the colloid. As was already mentioned, the presence of small aggregates resonant to the field is sufficient to amplify local effects. For illustration, we will turn to the general form of solution of a set of equations of interacting dipoles.

The optical activity effects are associated with different responses of the medium to the radiation with clockwise and anticlockwise polarizations,

$$\mathbf{E}_\pm = E_0(\hat{\mathbf{x}}\cos(k_z z - \omega t) \pm \hat{\mathbf{y}}\sin(k_z z - \omega t))/\sqrt{2}, \quad (49)$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are coordinate unit vectors. The solution of the set of equations of bound dipoles is expressed in terms of the components of eigenvectors $|n)$ of the interaction matrix V [24],

$$V|n) = \nu_n|n), \quad (50)$$

where ν_n denotes the eigenvalues. For a cluster consisting of N particles, the eigenvector has $3N$ components

[2] For gold (and some types of silver) colloids, $\chi^{(3)}$ first increases somewhat, as the threshold intensity is exceeded [19], and then decreases.

$n_{i\beta} \equiv (i\beta|n)$ corresponding to the amplitude of the plasmon mode on the ith particle with Cartesian component β (β assumes the values $\{x, y, z\}$). The dipole moment induced by the field of a wave with circular polarization on the ith particle is

$$(d_{i\beta})_{R,L} = \sum_{n=1}^{3N} n_{i\beta} C_n \times \sum_{j=1}^{N} [n_{jx}\cos(k_z z_j - \omega t) \pm n_{jy}\sin(k_z z_j - \omega t)] E_0/\sqrt{2}, \quad (51)$$

where

$$C_n = \left[\left(\sum_{j',\alpha} n_{j'\alpha}^2\right)(\alpha_0^{-1} + \nu_n)\right]^{-1}$$

is the resonance-normalization factor of the nth mode, $\alpha_0^{-1} = X + i\delta$ is the polarizability of an isolated monomer, and δ and X are parameters which have the meaning of the width of the resonant mode and its detuning from the resonance frequency of an isolated particle. The main contribution to the dipole moment of a particle is made by the resonant modes for which $\mathrm{Re}(\alpha_0^{-1} + \nu_n) < \delta$.

For the extinction cross section

$$\sigma_e = \frac{4\pi k}{|E_0|^2}\mathrm{Im}\sum_{i=1}^{N} \mathbf{d}_i \cdot \mathbf{E}^*(z_i) \quad (52)$$

in view of Eqs. (49) and (51), we have

$$(\sigma_e)_{R,L} = \pi k \mathrm{Im}\sum_{n=1}^{3N} C_n \sum_{i,j=1}^{N} [(n_{ix}n_{jx} + n_{iy}n_{jy}) \times \cos k_z(z_i - z_j) \pm (n_{iy}n_{jx} - n_{ix}n_{jy})\sin k_z(z_i - z_j)]. \quad (53)$$

The difference in extinction for clockwise- and anticlockwise-polarized radiation is expressed as

$$\Delta\sigma \equiv (\sigma_e)_R - (\sigma_e)_L = 2\pi k \mathrm{Im}\sum_{n=1}^{3N} C_n \times \sum_{i,j=1}^{N} (n_{iy}n_{jx} - n_{ix}n_{jy})\sin k_z(z_i - z_j). \quad (54)$$

One can see that zero contribution to $\Delta\sigma$ is produced by the terms with $i = j$, as well as at $n_{iy} = n_{ix}$ and $n_{jy} = n_{jx}$ (isotropic case). Therefore, resonant modes of a cluster with a high anisotropic factor

$$f = (n_{iy}n_{jx} - n_{ix}n_{jy})$$

which cover distances comparable with the wavelength (for increasing $\sin k_z(z_i - z_j)$ are of importance from the standpoint of optical activity effects. This means that the magnitude of nonlocal effects must increase with the emergence of large aggregates.

Not only does the constant $\mathrm{Re}\Gamma^{(3)}$ of nonlinear optical activity increase in the case of strong aggregation, but it increases in approximately the same manner relative to $\mathrm{Re}\chi^{(3)}$ ($\mathrm{Re}\Gamma^{(3)}/\mathrm{Re}\chi^{(3)}$ increased by a factor of 30). This means that the rise of nonlinear optical activity occurs mainly due to the effective geometric factor (including the anisotropic factor) rather than due to the factor of local field amplification (the latter must have a similar effect on $\Gamma^{(3)}$ and on $\chi^{(3)}$).

6. CONCLUSION

We have studied polarization effects in an aggregated colloidal silver solution, caused by the cubic nonlinearity of optical response in view of spatial dispersion. It was for the first time that the inverse Faraday effect, the optical Kerr effect, and the self-rotation of the polarization ellipse were observed for fractal clusters. The constant of nonlinear optical activity was measured under the same conditions.

The results of measurements of the tensor $\hat{\chi}^{(3)}$ of local nonlinear susceptibility have demonstrated that no decrease in nonlinear constants is observed for Ag(PVP) colloid at least at intensities of $\lesssim 30$ MW/cm^2. Apparently, one can say that no appreciable modification of local configurations of monomers occurs in this colloid in the above-identified range of intensities. In other words, those groups of several monomers for which the condition of collective response is valid largely remain unchanged; by virtue of this, a high magnitude of nonlinear response is retained. At the same time, the decrease in the nonlocal nonlinearity with the intensity above some value (2–3 MW/cm^2) is indicative of threshold variation of the structure of aggregates over sizes of the order of wavelength.

The results of measurements of nonlinear optical activity have demonstrated a considerable increase in the spatial scale of nonlocality of the correlation between polarization and exciting field upon transition from a weakly aggregated colloid to a more aggregated one. In this case, unlike the effects associated with $\hat{\chi}^{(3)}$, the value of nonlinear optical activity decreases considerably at $I \gtrsim 3$ MW/cm^2. Based on the results of analysis, one can assume a decrease in the effective geometric factor at high intensities, which is responsible for nonlocal effects.

The existing theories of nonlinear-optical properties of colloidal aggregates of metals are used to analyze collective modes associated with induced dipole moments, while assuming the particles proper to be stationary and rigidly secured. Our results lead to a general physical inference about the need to include in the treatment the mechanical motion of particles as a result of dynamic interaction of dipoles. Such motion may show up in the magnitude of nonlinear constants, in the time properties of responses, in polarization effects, and in the processes of photomodification of clusters.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (grant nos. 99-02-16670 and 01-02-06047), the Program for Support of Leading Scientific Schools of the Russian Federation (grant no. 00-15-96808), and the Civilian Research and Development Foundation (CRDF), grant RE1-2229.

APPENDIX

Let two spherical particles (dimer) of mass m and radius R be located at a distance $2R + D$ from each other (D is the minimal distance between the surfaces of the particles). Newton's equation for their relative motion is

$$\ddot{D} + \tau^{-1}\dot{D} = \mu^{-1}F(D, t), \quad \mu = m/2, \quad \tau = m/c_1, \tag{A.1}$$

where μ is the reduced mass, $c_1 = 6\pi\eta R$ is a coefficient in the Stokes formula, and η is the viscosity coefficient of the medium. According to Eq. (A.1), the relaxation time defined by the viscosity of the medium is $\tau = m/c_1 \approx 100$ ps ($R = 7$ nm, $\eta = 10^{-3}$ Pa s). The conditions of validity of the Stokes formula differ from our conditions; however, for estimation calculations, the friction force $\mu\tau^{-1}\dot{D}$ in (A.1) is sufficient. The force $F(D, t)$ is related to the particle interaction potential,

$$U = U_0 + U_{EM}, \quad F = F_0 + F_{EM}, \quad F_0 = -\frac{\partial U_0}{\partial D}, \quad F_{EM} = -\frac{\partial U_{EM}}{\partial D}. \tag{A.2}$$

Here, U_{EM} is defined by radiation, and U_0 is the potential in the absence of an external field. With coagulation colloidal aggregates in mind, we will assume that $U_0(D)$ is a potential curve with a well providing for the dimer stability and with a barrier at small distances D which prevents the particles from adhering to one another [25]. For typical conditions, the potential well depth is $\Delta U_0 \approx 0.2$ eV. Our measurements using electron microphotographs gave the mean distance $D \approx 1.5$ nm.

We will analyze Eq. (A.1) from the standpoint of the magnitude of possible particle shift (i.e., variation of D) under the effect of the light-induced force F_{EM}. In so doing, we will ignore the variation of the "undisturbed" potential U_0. In other words, we will assume that the potential well $U_0(D)$ has a fairly gentle slope compared with the scale of variation of $U_{EM}(D)$. We leave only $F_{EM}(D, t)$ in the right-hand part of Eq. (A.1) to derive the solution

$$\dot{D}(t) = \frac{1}{\mu}\int_0^t \exp\left(-\frac{t-t'}{\tau}\right)F_{EM}[D(t'), t']dt'. \tag{A.3}$$

If the light pulse duration τ_p is much shorter than τ, the following approximation is valid:

$$\dot{D}(t) \approx \frac{1}{\mu}e^{-t/\tau}\int_0^t F_{EM}[D(t'), t']dt'. \tag{A.4}$$

In the opposite case of $\tau_p \gg \tau$, the approximate solution is

$$\dot{D}(t) \approx \frac{\tau}{\mu}F_{EM}[D(t), t](1 - e^{-t/\tau}). \tag{A.5}$$

Note that the estimation of the shift during some time $t_0 > \max\{\tau_p, \tau\}$ in both limiting cases gives one and the same result,

$$\Delta D \approx \frac{2}{c_1}\int_0^{t_0} F_{EM}[D(t'), t']dt'. \tag{A.6}$$

The potential of interaction of two metal particles $U_{EM}(D, \omega, \theta)$ induced by an electromagnetic field was calculated in [21, 22]. In the dipole approximation,

$$U_{EM}(D, \omega, \theta) = -\frac{R^3}{2} \times \mathrm{Re}[\alpha_\parallel(D, \omega)\cos^2\theta + \alpha_\perp(D, \omega)\sin^2\theta - \alpha_0(\omega)]|E_0|^2,$$
$$\alpha_\parallel = \frac{\alpha_0}{1 - \alpha_0/4\xi^3}, \quad \alpha_\perp = \frac{\alpha_0}{1 + \alpha_0/8\xi^3}, \tag{A.7}$$
$$\alpha_0(\omega) = \frac{\varepsilon - \varepsilon_h}{\varepsilon + 2\varepsilon_h}, \quad \xi = 1 + \frac{D}{2R},$$

where θ is the angle between the vector **E** and the dimer axis. In view of the dependence of U_{EM} on θ, the moment of forces $-\partial U_{EM}/\partial\theta$ is active, which turns the dimer axis of maximal polarizability toward **E**. The relaxation time for angular motion coincides with τ by the order of magnitude. For polarization effects, the angular motion is significant (Kerr effect); however, we will not treat this motion in more detail and will restrict ourselves to the estimation of the variation of D.

In the case of Ag colloid [21],

$$U_{EM}(D, \omega, \theta)\ [\mathrm{eV}] \approx 2.2 \times 10^{-2}\gamma(D, \omega, \theta)R^3 I, \tag{A.8}$$

where $\gamma(D, \omega, \theta)$ is the dimensionless energy, R is in μm, and I is the power density in W/cm^2. In the dipole approximation (A.7) for a low-frequency (compared with the frequency of plasmon resonance for a sphere, ω_{ps}) region, as D increases, a potential barrier first arises, and then a potential well whose position is found for a greater D than the minimum of $U_0(D)$.

The results of calculations of $\gamma(D, \omega, \theta)$ with due regard for higher order multipoles performed by Claro and Rojas [21] demonstrate that, in the $D < 2$ nm region, a minimum of the $U_{EM}(D, \omega, \theta)$ curve may form, which is absent from the dipole approximation.

We substitute the quantity

$$F_{EM} = 2.2 \times 10^{-2} \frac{\partial \gamma[D(t)]}{\partial D} R^3 I(t) \qquad \text{(A.9)}$$

into Eq. (A.6) and approximately assume $\partial\gamma/\partial D$ to be constant in some small interval δD to derive the result

$$\Delta D = 2.2 \times 10^{-2} \frac{1}{c} \frac{\partial \gamma[D(t)]}{\partial D} R^3 \int_0^\infty I(t) dt, \qquad \text{(A.10)}$$

which depends on the energy density in the pulse (under the experimental conditions, this energy density was 20 mJ/cm^2). Expression (A.10) at $\partial\gamma/\partial D = 7$ nm^{-1} gives the value of the shift under the effect of a 10-ns pulse of $\Delta D \approx 1$ nm.

It must be emphasized that the dissipation of energy due to viscous friction will bring about an increase in the temperature of the medium and metal particles, which, generally speaking, must also be taken into account because of the temperature dependence of ε.

REFERENCES

1. V. M. Shalaev, Phys. Rep. **272**, 61 (1996).
2. B. M. Smirnov, *Physics of Fractal Clusters* (Nauka, Moscow, 1991).
3. M. I. Stockman, Phys. Rev. E **56**, 6494 (1997).
4. V. P. Safonov, V. M. Shalaev, V. A. Markel, *et al.*, Phys. Rev. Lett. **80**, 1102 (1998).
5. A. V. Butenko, V. M. Shalaev, and M. I. Stockman, Zh. Éksp. Teor. Fiz. **94**, 107 (1988) [Sov. Phys. JETP **67**, 60 (1988)].
6. S. G. Rautian, V. P. Safonov, P. A. Chubakov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **47**, 200 (1988) [JETP Lett. **47**, 243 (1988)].
7. Yu. E. Danilova, N. N. Lepeshkin, S. G. Rautian, and V. P. Safonov, Physica A (Amsterdam) **241**, 231 (1997).
8. V. P. Drachev, S. V. Perminov, S. G. Rautian, and V. P. Safonov, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 618 (1998) [JETP Lett. **68**, 651 (1998)].
9. P. Maker, R. Terhune, and C. Savage, Phys. Rev. Lett. **12**, 507 (1964).
10. V. N. Arutyunyan, T. A. Parazyan, G. G. Adonpts, *et al.*, Zh. Éksp. Teor. Fiz. **68**, 44 (1975) [Sov. Phys. JETP **41**, 22 (1975)].
11. A. M. Bonch-Bruevich, N. N. Kostin, and V. A. Khodovoĭ, Pis'ma Zh. Éksp. Teor. Fiz. **3**, 425 (1966) [JETP Lett. **3**, 279 (1966)].
12. S. A. Akhmanov, G. A. Lyakhov, V. A. Makarov, and V. I. Zharikov, Opt. Acta **29**, 1359 (1982).
13. S. A. Akhmanov, N. I. Zheludev, and R. S. Zadoyan, Zh. Éksp. Teor. Fiz. **91**, 984 (1986) [Sov. Phys. JETP **64**, 579 (1986)].
14. R. W. Boyd, *Nonlinear Optics* (Academic, San Diego, 1992).
15. H. Hirai, J. Macromol. Sci., Chem. A **A13**, 633 (1979).
16. S. V. Karpov, A. K. Popov, S. G. Rautian, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **48**, 528 (1988) [JETP Lett. **48**, 571 (1988)].
17. Yu. É. Danilova, V. P. Drachev, S. V. Perminov, and V. P. Safonov, Izv. Ross. Akad. Nauk, Ser. Fiz. **60**, 18 (1996).
18. A. V. Butenko, Yu. E. Danilova, S. V. Karpov, *et al.*, Z. Phys. D **17**, 283 (1990).
19. Yu. É. Danilova, S. G. Rautian, and V. P. Safonov, Izv. Ross. Akad. Nauk, Ser. Fiz. **60**, 56 (1996).
20. V. P. Safonov, Yu. E. Danilova, V. P. Drachev, and S. V. Perminov, in *Optics of Nanostructured Materials*, Ed. by T. F. George and V. A. Markel (Wiley, New York, 2001), p. 283.
21. F. Claro and R. Rojas, Appl. Phys. Lett. **65**, 2743 (1994).
22. F. Claro, Physica A (Amsterdam) **241**, 223 (1997).
23. A. N. Azarenkov, G. V. Al'tshuler, N. R. Belashenkov, and S. A. Kozlov, Kvantovaya Élektron. (Moscow) **20**, 733 (1993).
24. V. M. Shalaev, *Nonlinear Optics of Random Media: Fractal Composites and Metal-Dielectric Films* (Springer-Verlag, Berlin, 1999).
25. Yu. Frolov, *Colloid Chemistry* (Khimiya, Moscow, 1987).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 916–926.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1068–1079.
Original Russian Text Copyright © 2002 by Zelensky, Mironov.

ATOMS, SPECTRA, RADIATION

Electromagnetically Induced Transparency in Degenerate Two-Level Systems

I. V. Zelensky* and V. A. Mironov

Institute of Applied Physics, Russian Academy of Sciences, ul. Ul'yanova 46, Nizhni Novgorod, 603950 Russia
**e-mail: zelensky@appl.sci-nnov.ru*

Received December 23, 2001

Abstract—The evolution of an electromagnetic wave with slowly varying polarization, which interacts resonantly with the medium formed by degenerate two-level atoms, is studied using the wave function approach under the conditions of electromagnetically induced transparency. It is shown that the amplitude of the wave field propagates at the velocity of light in such a medium. The equation obtained for the polarization parameter has a solution in the form of a simple wave. The breaking length is determined. It is shown that the velocity of propagation of polarization waves may be much smaller than the velocity of light. The proposed approach is common for two-level systems with an arbitrary degeneracy. The case of a system with Zeeman degeneracy is analyzed in detail. The dependence of the velocity of propagation of the polarization structure on the amplitude and polarization is determined for an arbitrary level degeneracy. The evolution of the polarization structure in such a medium is discussed. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The discovery of coherent population trapping and electromagnetically induced transparency associated with it [1–3] stimulated the development of new tendencies in the optics of resonance media. The formation of a transparency window upon resonant interaction of wave fields in multilevel systems is obviously accompanied by a noticeable elongation of the path of propagation of laser radiation. Strong dispersion under the conditions of electromagnetically induced transparency leads to an anomalously strong deceleration of a probe pulse in the field of the driving wave. Under special conditions of radiation control, "light stoppage" (i.e., in fact, the realization of optical memory) is possible [4, 5]. The combination of the low velocity of the wave packet propagation and the elongation of the path over which fields interact under the resonance conditions makes the electromagnetically induced transparency mode promising for the investigation of nonlinear effects. The analysis of relevant processes for a small number of photons [6] is of special interest.

A theoretical analysis of this mode is usually carried out on the basis of scalar equations for the field. The influence of electromagnetic wave polarization on transition probabilities is obvious and is naturally manifested during experimental investigations of electromagnetically induced transparency in "three-level" systems. The role of polarization becomes decisive in systems with degeneracy. The behavior of the system in this case considerably depends on the mutual orientation of polarizations of optical fields [7–9]. This is associated with population trapping in the dark state, viz., the superposition of sublevels which does not interact with the field. The existence of the dark state and the trapping of population in it is the general property of degenerate systems, which has been studied extensively. In particular, coherent population trapping in a two-level system exhibiting Zeeman degeneracy and interacting with polarized radiation was studied in detail in the early publications devoted to the theory of this effect [10, 11].

Thus, at the first stage of investigation of the features of polarization effects, it is natural to consider the electromagnetically induced transparency mode by using the simple model of a two-level degenerate system. In this case, the excitation channels are separated due to different polarizations acting on the system of electromagnetic waves, and the emerging configurations are determined by the selection rules for transitions between sublevels. It is important to note that, in the case of accompanying wave propagation, the closeness (and even equality in the case of exact resonance) of the frequencies of the fields used in such systems makes it possible to eliminate almost completely the Doppler broadening of the two-photon transition corresponding to electromagnetically induced transparency. In addition, the possibility of using a common source of interacting waves makes it possible to considerably improve the coherence of exciting effects, which produces a positive effect on the observed characteristics. This renders degenerate systems convenient objects for observing electromagnetically induced transparency. In particular, the so-called storage of light [4] was obtained on the transition $5s_{1/2}$, $F = 2 \longleftrightarrow 5p_{1/2}$, $F = 1$ for ^{87}Rb using polarized radiation. It should be noted that the effect of electromagnetically induced transparency in systems with Zeeman degeneracy may be used

in spectroscopic applications [8] as well as in the measurements of weak magnetic fields [12]. The application of this model for an analysis of excitation of the polarization-squeezed state of light [13] is of special interest.

The study of electromagnetically induced transparency involves, as a rule, the analysis of the behavior of a weak probe wave in the presence of a high-intensity driving wave. In a degenerate system, such an approach is modified as follows. The field of any pre-set polarization can be regarded as driving, and its variation, as a probe wave. The present work aims at an analysis of evolution of the wave field with a slowly varying polarization interacting resonantly with the medium of degenerate two-level atoms. We will study the behavior of the intensity and parameter of polarization during the propagation of a wave through such a medium. The approach proposed by us was found to be convenient for describing two-level degenerate systems irrespective of the number and the specific structure of the sublevels. The paper has the following structure. In Section 2, basic equations are presented and the approximations of the model used in the present work are discussed. Section 3 is devoted to an analysis of processes in the Λ scheme with polarization-aided separation of lower levels. In Section 4, an arbitrary multiply degenerate two-level system is considered. In Section 5, the general approach developed in the present work is applied to the important particular case of a two-level system with Zeeman degeneracy. A brief review and discussion of the obtained results are carried out in Conclusions.

2. FORMULATION OF THE PROBLEM: BASIC EQUATIONS

Let us consider a two-level degenerate system (Fig. 1) resonantly interacting with the field

$$\mathbf{E}(z, t) = \boldsymbol{\varepsilon}(z, t)\exp(-i\omega t + ikz) + \text{c.c.}, \qquad (2.1)$$

where ω is the transition frequency and $\boldsymbol{\varepsilon}(z, t)$ is the slow amplitude. We will seek the wave function of the system in the form

$$\psi(\mathbf{r}, t) = \sum_{i=1}^{N} a_i\psi_i(\mathbf{r}) + \sum_{j=1}^{M} b_j\chi_j(\mathbf{r})\exp(-i\omega t + ikz), \qquad (2.2)$$

where the origin is taken at the energy of the lower level, $\{\psi_1(\mathbf{r}), \ldots, \psi_N(\mathbf{r})\}$, $\{\chi_1(\mathbf{r}), \ldots, \chi_M(\mathbf{r})\}$, are certain orthonormal bases at the lower and upper energy levels, respectively, and a_i and b_j are nonstationary probability amplitudes.

The system of equations for the probability amplitudes a_i and b_j, supplemented by taking into account the spontaneous decay of the upper level, can be written in the form

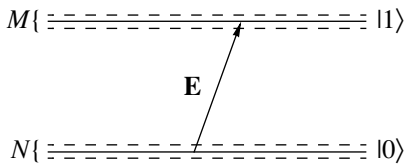

Fig. 1. Two-level degenerate system resonantly interacting with the field **E**; *M* and *N* are the degeneracies of the upper and lower energy levels, respectively.

$$\dot{a}_i = \frac{i}{\hbar}\sum_{j=1}^{M} b_j(\mathbf{d}_{ij}\cdot\boldsymbol{\varepsilon})^*, \qquad (2.3)$$

$$\dot{b}_j = \frac{i}{\hbar}\sum_{i=1}^{N} a_i(\mathbf{d}_{ij}\cdot\boldsymbol{\varepsilon}) - \gamma b_j, \qquad (2.4)$$

where $\mathbf{d}_{ij} = \int\psi_i^*\mathbf{d}\chi_j d^3r$ are the dipole moments of the corresponding transitions, γ is the decay constant, the term $-\gamma b_j$ was introduced for describing the relaxation of the population of the upper level, and the dot denotes the partial derivative with respect to time.

In order to describe the space–time evolution of the field, we use the truncated wave equation

$$\left[\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right]\boldsymbol{\varepsilon} = i\nu\sum_{i,j} a_i^* b_j\mathbf{d}_{ij}, \qquad (2.5)$$

where $\nu = 2\pi N\omega/c$, *N* being the concentration of atoms.

It is convenient for subsequent analysis to decompose the slow field amplitude $\boldsymbol{\varepsilon}$ in a certain orthonormal basis $\mathbf{e}_1$, $\mathbf{e}_2$ in a plane perpendicular to the direction of propagation:

$$\boldsymbol{\varepsilon} = \varepsilon_1\mathbf{e}_1 + \varepsilon_2\mathbf{e}_2. \qquad (2.6)$$

As a result, Eqs. (2.3)–(2.5) will be written in the form

$$\left[\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right]\varepsilon_1 = i\nu(\mathbf{a}\cdot\hat{D}_1\mathbf{b}), \qquad (2.7)$$

$$\left[\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right]\varepsilon_2 = i\nu(\mathbf{a}\cdot\hat{D}_2\mathbf{b}), \qquad (2.8)$$

$$\dot{\mathbf{a}} = \frac{i}{\hbar}(\varepsilon_1^*\hat{D}_1 + \varepsilon_2^*\hat{D}_2)\mathbf{b}, \qquad (2.9)$$

$$\dot{\mathbf{b}} = \frac{i}{\hbar}(\varepsilon_1\hat{D}_1^+ + \varepsilon_2\hat{D}_2^+)\mathbf{a} - \gamma\mathbf{b}, \qquad (2.10)$$

where $(\hat{D}_1)_{ij} = (\mathbf{e}_1\cdot\mathbf{d}_{ij})$, $(\hat{D}_2)_{ij} = (\mathbf{e}_2\cdot\mathbf{d}_{ij})$ (the scalar product is introduced here as $(\mathbf{p}\cdot\mathbf{q}) = p_i^* q_i$).

These equations lead to the following conclusion for a steady-state field ($\varepsilon_1 = \text{const}$, $\varepsilon_2 = \text{const}$). If the degeneracy of the lower level exceeds the degeneracy of the upper level, the lower level always contains a state which does not interact with the field (dark state) and satisfies the relation

$$(\varepsilon_1 \hat{D}_1^+ + \varepsilon_2 \hat{D}_2^+)\mathbf{a} = 0. \quad (2.11)$$

This relation corresponds to the vacant upper level

$$\mathbf{b} = 0. \quad (2.12)$$

In this case, the right-hand sides in Eqs. (2.7), (2.8) for the field are equated to zero; i.e., the medium does not interact with the field. Relaxation processes in the system ensure the stability of the dark state. Thus, in the case of steady-state fields, coherent population trapping in the dark state and electromagnetically induced transparency emerge in the system. In the subsequent analysis, we will describe the evolution of the wave field which is slow over the time of stabilization of coherent population trappings by using the adiabatic approximation for studying the system response.

It should be noted that the approach proposed for describing the evolution of the quantum system with the simplified inclusion of relaxation cannot be used for obtaining a correct description of the processes of profound rearrangement of population in the system and the establishment of coherent population trapping associated with it. Indeed, the term $-\gamma\mathbf{b}$ in Eq. (2.10) actually describes the depopulation of the system. In this case, the relaxation of population from the upper level to the lower level is disregarded as well as the relaxation processes within the levels. However, the analysis of the corresponding processes in systems of the Λ type on the basis of the density matrix formalism leads to the following conclusion [1]. If the system is in the state of coherent population trapping at the initial instant, a slow (on the scale of relaxation time) variation of the fields is accompanied by an adiabatic rearrangement of the dark state. To be more precise, in fields stronger than the threshold field for the coherent population trapping, the evolution of the system is independent of the relaxation constants.

3. POLARIZATION Λ SCHEME

In order to illustrate the features of the problem, we first consider the case when the lower level is doubly degenerate, while the upper level is not degenerate, i.e., the so-called Λ scheme of energy levels with the polarization-aided separation of excitation channels. In this case, we can easily obtain an explicit equation for the dark state. For the sake of simplicity, we assume that the dipole moments of the transitions between the lower sublevels are equal and orthogonal ($|\mathbf{d}_1| = |\mathbf{d}_2| = d$, $(\mathbf{d}_1 \cdot \mathbf{d}_2) = 0$). In the components along the orthogonal directions,

$$\mathbf{e}_1 = \frac{\mathbf{d}_1}{d}, \quad \mathbf{e}_2 = \frac{\mathbf{d}_2}{d}. \quad (3.1)$$

Equations (2.7)–(2.10) can be written in the form

$$\left[\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right]\varepsilon_1 = i\nu d a_1^* b, \quad (3.2)$$

$$\left[\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right]\varepsilon_2 = i\nu d a_2^* b, \quad (3.3)$$

$$\dot{a}_1 = \frac{id}{\hbar}\varepsilon_1^* b, \quad (3.4)$$

$$\dot{a}_2 = \frac{id}{\hbar}\varepsilon_2^* b, \quad (3.5)$$

$$\dot{b} = \frac{id}{\hbar}(\varepsilon_1 a_1 + \varepsilon_2 a_2)b - \gamma b. \quad (3.6)$$

In the case of steady-state fields, the system of constituent equations (3.4)–(3.6) has eigenvalues $\lambda_1 = 0$, $\lambda_{2,3} = -\gamma/2 \pm \sqrt{\gamma^2/4 - d^2|\varepsilon|^2/\hbar^2}$, where $|\varepsilon|^2 = |\varepsilon_1|^2 + |\varepsilon_2|^2$. It can be seen that $\mathrm{Re}\lambda_{2,3} < 0$. Thus, during the time $\tau_R = 1/\min(|\mathrm{Re}\lambda_{2,3}|)$, the dark state, which does not interact with the field and corresponds to zero eigenvalue, will set in, for which

$$a_1\varepsilon_1 + a_2\varepsilon_2 = 0, \quad b = 0. \quad (3.7)$$

For fields much weaker than the fields saturating the transition, $\Omega \ll \gamma$ ($\Omega = |\varepsilon|d/\hbar$ is the Rabi frequency), the time of stabilization of the dark state is given by

$$\tau_R = \gamma/\Omega^2. \quad (3.8)$$

It should be noted that, in more realistic models, we must take into account the processes of relaxation in the lower level, which ultimately lead to degradation of coherent population trapping (3.7). In this case, the lifetime of the dark state can be estimated as

$$\tau_s = \Gamma^{-1}, \quad (3.9)$$

where Γ is the constant of relaxation processes in the lower level.

A comparison of Eqs. (3.8) and (3.9) shows that the field-induced process of stabilization of the state (3.7) dominates over its "spontaneous" decay (3.9) in the fields

$$\Omega^2 \gg \Gamma\gamma. \quad (3.10)$$

It is this threshold condition (imposed on the field) of coherent population trapping that appears when relaxation is taken into account correctly on the basis of the density matrix formalism in Λ-type systems [1].

For fields much stronger than the threshold field (3.10) and smooth over the τ_R scale, we can assume that the quasi-stationary relations (3.7) are satisfied in the initial approximation (in slowness). Since Eqs. (3.2) and (3.3) for the field contain the products $a_1^* b$ and $a_2^* b$, we will be interested in the nonstationary correction to the population of the upper level only. It should be noted that relations (3.4), (3.5), and (3.7) lead to the conservation of population in the system, which should naturally be normalized to unity:

$$|a_1|^2 + |a_2|^2 = 1. \tag{3.11}$$

Taking into account this relation, we have the following relations in the zeroth order of perturbation theory:

$$a_1^{(0)} = \kappa\varepsilon_2/\varepsilon, \quad a_2^{(0)} = -\kappa\varepsilon_1/\varepsilon, \quad b^{(0)} = 0, \tag{3.12}$$

where κ is a certain phase factor, $|\kappa| = 1$.

Substituting $a_1^{(0)}$ and $a_2^{(0)}$ into Eqs. (3.4) and (3.5), we can easily find the nonstationary correction to b:

$$b = \frac{\hbar k}{id|\varepsilon|^3}\left(\varepsilon_1\frac{\partial}{\partial t}\varepsilon_2 - \varepsilon_2\frac{\partial}{\partial t}\varepsilon_1\right). \tag{3.13}$$

Finally, we arrive at the following self-consistent equations for the field:

$$\left[\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right]\varepsilon_1 = \frac{\hbar\nu}{|\varepsilon|^4}\left(\varepsilon_1\frac{\partial}{\partial t}\varepsilon_2 - \varepsilon_2\frac{\partial}{\partial t}\varepsilon_1\right)\varepsilon_2^*, \tag{3.14}$$

$$\left[\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right]\varepsilon_2 = -\frac{\hbar\nu}{|\varepsilon|^4}\left(\varepsilon_1\frac{\partial}{\partial t}\varepsilon_2 - \varepsilon_2\frac{\partial}{\partial t}\varepsilon_1\right)\varepsilon_1^*. \tag{3.15}$$

As applied to a three-level Λ scheme, the system of equations (3.2)–(3.6) describes the interaction of bichromatic radiation in a resonant medium [14–17]. The peculiarities of the dynamics of an accompanying Raman-type interaction were investigated analytically and numerically in [14–16]. Among other things, it was proved that the adiabatic approximation (3.14), (3.15) correctly describes the evolution of the central part of the pulses.

As applied to our case, it is convenient to write Eqs. (3.14) and (3.15) in terms of the polarization parameter

$$q = \varepsilon_1/\varepsilon_2 \tag{3.16}$$

and intensity $|\varepsilon|^2$. Finally, we have

$$\left[\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right]|\varepsilon|^2 = 0, \tag{3.17}$$

$$\left[\frac{\partial}{\partial z} + \left(\frac{1}{c} + \frac{\hbar\nu}{|\varepsilon|^2}\right)\frac{\partial}{\partial t}\right]q = 0. \tag{3.18}$$

It can be seen that the energy of an electromagnetic wave propagates in a system with the velocity of light (as in the conventional Λ scheme), while the polarization wave propagates at a certain effective velocity

$$V_{\text{eff}} = \left(\frac{1}{c} + \frac{\hbar\nu}{|\varepsilon|^2}\right)^{-1} = c\left(1 + \frac{W_{\max}}{4W_E}\right)^{-1}, \tag{3.19}$$

where $W_{\max} = N\hbar\omega$ is the maximum energy density which can be "pumped" to the medium and $W_E = |\varepsilon|^2/8\pi$ is the electromagnetic energy density. Thus, for $W_{\max} \gg W_E$, the velocity of the polarization wave turns out to be considerably smaller than the velocity of light.

It should be noted that expression (3.19) coincides with the expression for the velocity of a probe wave under the conditions of electromagnetically induced transparency in the presence of a powerful driving wave whose intensity is much higher than the threshold intensity (3.10) obtained in the framework of the density matrix formalism [18]. This is obvious for a small change in polarization which can be regarded as a low-intensity probe wave. In the case of a strong change in polarization, the occurring processes can be interpreted as resonance Raman scattering of an electromagnetic wave of a certain polarization into another wave at the low-frequency coherence induced in the medium. The adiabaticity condition imposes the following limitation on the length L_F of the rearrangement wave front:

$$L_F \gg L_{\min} = \tau_R V_{\text{eff}}. \tag{3.20}$$

For fields much weaker than the saturation field

$$L_{\min} = \frac{\gamma\hbar c}{2\pi N\omega d^2}, \tag{3.21}$$

or in the case in the case of radiation-induced relaxation $\gamma = 4\omega^3 d^2/3\hbar c^3$, we obtain

$$L_{\min} = \frac{8\pi}{3\lambda^2 N}, \tag{3.22}$$

where λ is the wavelength.

In particular, for experiments with rubidium ($D1$ and $D2$ lines), we have $L_{\min}(T = 300\ \text{K}) \approx 4 \times 10^{-1}$ cm and $L_{\min}(T = 350\ \text{K}) \approx 4 \times 10^{-3}$ cm.

4. TWO-LEVEL SYSTEM WITH MULTIPLE LEVEL DEGENERACY

Let us generalize the results obtained above in the adiabatic approximation to a multiply degenerate two-level system. We assume that, in the zeroth approximation, the atomic system is in the dark state; i.e., relations (2.11) and (2.12) are satisfied. As in the previous section, we will determine nonstationary corrections to the population of the upper level. It should be noted that Eqs. (2.9) and (2.11) lead to a condition of conservation

of population in the system, which is similar to relation (3.11):

$$|\mathbf{a}|^2 = 1. \tag{4.1}$$

We will describe the space–time evolution of the wave field in terms of the polarization parameter q and intensity $|\varepsilon|^2$. Using relations (2.7), (2.8), and (2.11), we can easily find that the behavior of the intensity is described, as in the case of a Λ scheme, by Eq. (3.17).

We will rewrite Eqs. (2.7)–(2.9) and (2.11), singling out the polarization parameter in explicit form:

$$\left[\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right]q = -\frac{i\nu}{\varepsilon_2}(q(\mathbf{a}\cdot\hat{D}_2\mathbf{b}) - (\mathbf{a}\cdot\hat{D}_1\mathbf{b})), \tag{4.2}$$

$$\dot{\mathbf{a}} = \frac{i}{\hbar}\varepsilon_2^*(q^*\hat{D}_1 + \hat{D}_2)\mathbf{b}, \tag{4.3}$$

$$(q\hat{D}_1^+ + \hat{D}_2^+)\mathbf{a} = 0. \tag{4.4}$$

In the case when Eq. (4.4) has a one-dimensional space of solutions, on account of the population conservation condition (4.1), we have

$$\mathbf{a} = \kappa\mathbf{a}_0, \tag{4.5}$$

where $\mathbf{a}_0(q)$ is a certain normalized solution of Eq. (4.4) and κ is an arbitrary phase factor, $|\boldsymbol{\kappa}| = 1$.

In order to find the corrections to the population of the upper level, which are associated with the rearrangement of the dark state (4.5), we differentiate relation (4.4) with respect to time:

$$\dot{q}\hat{D}_1^+\mathbf{a} + (q\hat{D}_1^+ + \hat{D}_2^+)\dot{\mathbf{a}} = 0. \tag{4.6}$$

Taking into account relation (4.3), we obtain

$$\mathbf{b} = \frac{i\hbar}{\varepsilon_2^*}\dot{q}T^{-1}\hat{D}_1^+\mathbf{a}, \tag{4.7}$$

where

$$T(q) = (q\hat{D}_1^+ + \hat{D}_2^+)(q^*\hat{D}_1 + \hat{D}_2). \tag{4.8}$$

Substituting the obtained expression (4.7) into Eq. (4.2) for the polarization parameter and using relation (4.4), we obtain

$$\left[\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right]q = -\frac{\hbar\nu}{|\varepsilon|^2}(1+|q|^2) \times\frac{\partial q}{\partial t}((\hat{D}_1^+\mathbf{a}\cdot T^{-1}\hat{D}_1^+\mathbf{a}) + (\hat{D}_2^+\mathbf{a}\cdot T^{-1}\hat{D}_2^+\mathbf{a})). \tag{4.9}$$

The substitution of $\mathbf{a}$ in the form (4.5) into Eq. (4.9) shows that the expression contains only the modulus of the phase factor κ, which is equal to unity. Finally, we can write

$$\left[\frac{\partial}{\partial z} + \left(\frac{1}{c} + \frac{\hbar\nu}{|\varepsilon|^2}u(q)\right)\frac{\partial}{\partial t}\right]q = 0, \tag{4.10}$$

where

$$u(q) = (1+|q|^2) \times((\hat{D}_1^+\mathbf{a}_0\cdot T^{-1}\hat{D}_1^+\mathbf{a}_0) + (\hat{D}_2^+\mathbf{a}_0\cdot T^{-1}\hat{D}_2^+\mathbf{a}_0)), \tag{4.11}$$

$\mathbf{a}_0(q)$ being a certain (any) normalized solution of Eq. (4.4).

It follows hence that the polarization structure of the field propagates at a certain effective velocity depending on the intensity and (in the general case) the direction of polarization:

$$V_{\text{eff}} = c\left(1 + \frac{W_{\max}}{4W_E}u(q)\right)^{-1}. \tag{4.12}$$

Thus, the analysis of the evolution of the wave field is reduced to determining the dependence of the deceleration factor $u(q)$ on the polarization of the wave.

5. BEHAVIOR OF POLARIZATION OF AN ELECTROMAGNETIC WAVE UNDER THE CONDITIONS OF ELECTROMAGNETICALLY INDUCED TRANSPARENCY IN A TWO-LEVEL QUANTUM SYSTEM WITH ZEEMAN DEGENERACY

By way of an application of the theory developed in the previous section, we consider the resonant interaction of polarized radiation with a quantum system with Zeeman degeneracy. Let the lower and upper energy levels of the two-level system under investigation correspond to the states with the angular momenta L_0 and L_1, respectively, while in zero magnetic field these levels are characterized by degeneracy $2L_0 + 1$ and $2L_1 + 1$ in the angular momentum component (the so-called Zeeman degeneracy, or degeneracy in the magnetic quantum number). This example is of practical importance, in particular, for transitions between the sublevels of the hyperfine structure, which are widely used for obtaining electromagnetically induced transparency. In accordance with the selection rules, transitions with

$$\Delta L = L_1 - L_0 = -1, 0, +1 \tag{5.1}$$

are possible. An analysis of the coherent population trapping in the given case is carried out in [10]. As applied to the problem under investigation, the results obtained in [10] can be presented as follows. The system of sublevels splits into two noninteracting subsystems (if we disregard the relaxation processes between the sublevels). In this case,

(a) for $\Delta L = +1$, no coherent population trapping can take place in either of the subsystems;

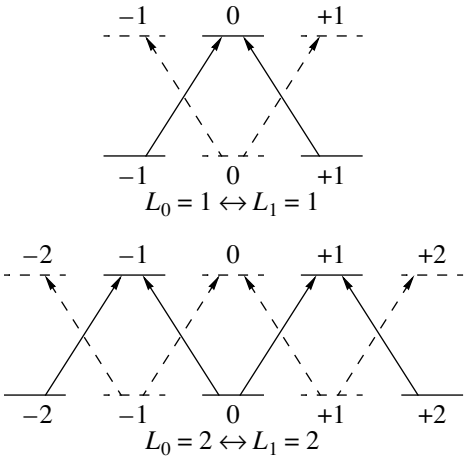

Fig. 2. Diagrams of transitions with $\Delta L = 0$ and an integer angular momentum. The subsystems in which coherent population trapping is possible (Λ-type) are presented by solid lines, while the subsystems in which it is impossible (V-type) are shown by dashed lines.

(b) for $\Delta L = 0$, no coherent population trapping is possible in either of the systems with a half-integer momentum for an arbitrary field polarization; in systems with integer spins, except for the special case of $L_0 = 0 \longleftrightarrow L_1 = 0$, when transition is forbidden, coherent population trapping is impossible in one of the systems (V-type system) and possible in the other (Λ-type) system;

(c) for $\Delta L = -1$, coherent population trapping is possible in both subsystems (Λ-type systems); in this case, the subsystems make additive contributions to the equation for the field and, accordingly, to the expression for the deceleration factor. (In the special case of $L_0 = 1 \longleftrightarrow L_1 = 0$, the Λ scheme + the "pocket" $|L_0 = 1, L_{0z} = 0\rangle$ does not interact with the field of any polarization.)

In the framework of the proposed approach, it is possible to consider the subsystems separately. Their responses appear additively in the expression for the field and, accordingly, for the retardation factor (4.11). As a result, the expression for this factor assumes the form

$$u(q) = n_a u_a(q) + n_b u_b(q). \tag{5.2}$$

Here, $u_a(q)$ and $u_b(q)$ are the deceleration factors and n_a and n_b are the populations of the subsystems a and b, respectively. It should be noted that, in accordance with Eq. (4.1), the quantities n_a and n_b are conserved.

Let us consider, by way of an illustration, the diagrams of some transitions in which coherent population trapping is possible (Figs. 2 and 3). In these figures, the numbers denote the angular momentum components corresponding to the sublevels. The quantization axis is chosen along the direction of propagation of the wave, and, hence, the field component along the quantization axis is equal to zero. Transitions with $\Delta L_z = -1$ and $\Delta L_z = +1$ correspond to the left (σ_-) and right (σ_+) circular polarizations. It is important to note that, in all the subsystems under investigation for which a dark state exists (Λ-type systems), the number of lower sublevels is larger by unity than the number of the upper sublevels, and the dimensionality of the dark state is equal to unity.

Let us now use the approach developed by us for systems with a one-dimensional dark state for analyzing two-level systems with Zeeman degeneracy.

For describing the transitions between magnetic sublevels, it is convenient to choose the right and left circular polarizations of the electric field as the basis polarizations:

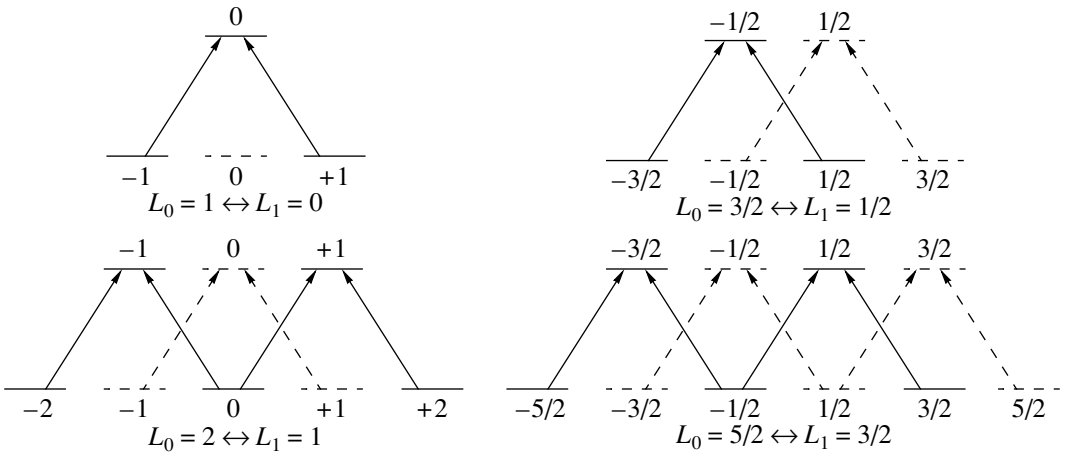

Fig. 3. Diagrams of some transitions with $\Delta L = -1$. In both subsystems (depicted by solid and dashed lines), coherent captures of population is possible (Λ-type). The special case is $L_0 = 1 \longleftrightarrow L_1 = 0$. This transition splits into the Λ diagram (solid line) and a "pocket," viz., isolated sublevel $|L_0 = 1, L_{0z} = 0\rangle$ (dashed line).

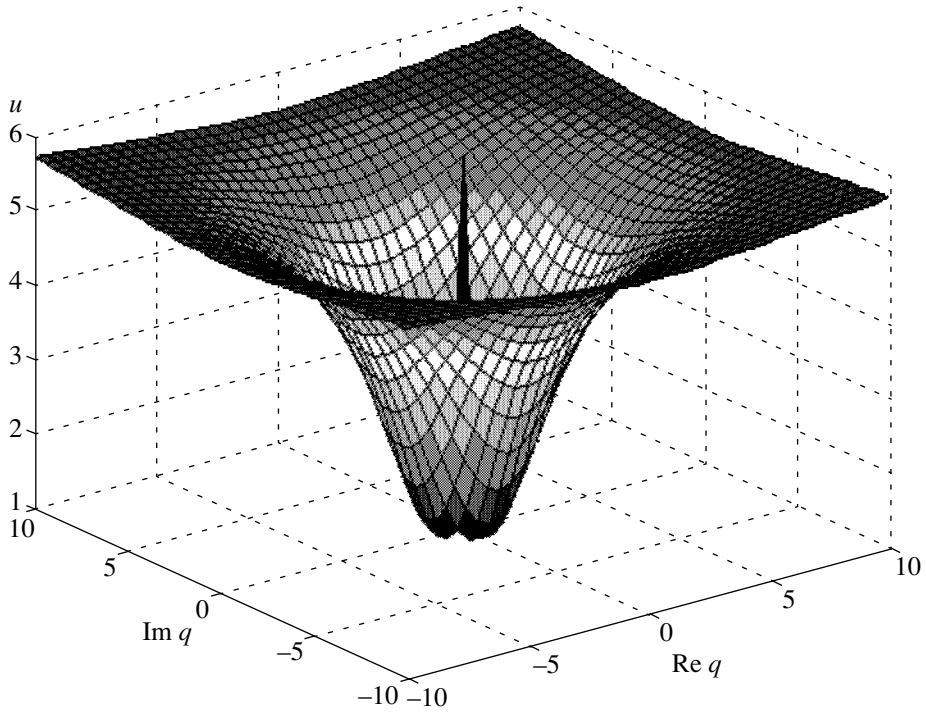

Fig. 4. Dependence of the deceleration factor on the polarization parameter for one of the subsystems of the transition $L_0 = 2 \longleftrightarrow L_1 = 1$ (which corresponds to the subsystem depicted by the solid line in Fig. 3).

$$\mathbf{e}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}, \tag{5.3}$$

$$\mathbf{e}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}. \tag{5.4}$$

In this case, for the dipole moments of transitions, we obtain

$$(\hat{D}_1)_{ij} = \mathbf{e}_1 \cdot \mathbf{d}_{ij} = \frac{1}{\sqrt{2}}\hat{d}_{-ij} = \frac{1}{\sqrt{2}}\int \psi_i^* \hat{d}_- \chi_j d^3 r, \tag{5.5}$$

$$(\hat{D}_2)_{ij} = \mathbf{e}_2 \cdot \mathbf{d}_{ij} = \frac{1}{\sqrt{2}}\hat{d}_{+ij} = \frac{1}{\sqrt{2}}\int \psi_i^* \hat{d}_+ \chi_j d^3 r, \tag{5.6}$$

where

$$\hat{d}_+ = \hat{d}_x + i\hat{d}_y, \tag{5.7}$$

$$\hat{d}_- = \hat{d}_x - i\hat{d}_y. \tag{5.8}$$

It turns out that with such a choice, the retardation factor depends only on the modulus of the polarization parameter,

$$u(q) = u(|q|). \tag{5.9}$$

This is apparently associated with the axial symmetry of the eigenfunctions of the magnetic sublevels relative to the quantization axis (which was chosen so that it coincides with the direction z of wave propagation). By way of an illustration, Fig. 4 shows the curve $u(q)$ calculated by formulas (4.11), (4.8), and (4.4) for one of two subsystems of the Λ type, which is formed by three sublevels of the lower level and by two sublevels of the upper transition $L_0 = 2 \longleftrightarrow L_1 = 1$ (the subsystem depicted by the solid line in Fig. 3). The radial symmetry of $u(q)$ confirms the above conclusion (5.9). In addition, it was found that the deceleration factor is independent of the absolute value of the dipole moment and is determined only by the relation between the moments of transitions between different sublevels. This can easily be explained as follows. If we multiply the matrices $\hat{D}_1$ and $\hat{D}_2$ by a certain factor, it will be canceled out after the substitution into formulas (4.8) and (4.11). It is also clear that, since the deceleration factor is a function of the modulus of q only, it does not change after the multiplication of one of the matrices $\hat{D}$ by a number whose magnitude is equal to unity. Indeed, such a change in the matrix of dipole moments is equivalent to the multiplication of the polarization parameter q in expressions (4.4), (4.8), and (4.11) defining the function $u(q)$ by the corresponding phase factor. The nonzero matrix elements of transitions between the Zeeman sublevels are given by [19]

$$\begin{aligned} &\langle n', L, M-1|\hat{d}_-|n, L, M\rangle \\ &= \sqrt{\frac{(L-M+1)(L+M)}{L(L+1)(2L+1)}}\langle n', L\|\hat{d}\|n, L\rangle, \end{aligned} \tag{5.10}$$

$$\langle n', L, M-1|\hat{d}_-|n, L-1, M\rangle = \sqrt{\frac{(L-M+1)(L-M)}{L(2L-1)(2L+1)}}\langle n', L\|\hat{d}\|n, L-1\rangle, \quad (5.11)$$

$$\langle n', L-1, M-1|\hat{d}_-|n, L, M\rangle = -\sqrt{\frac{(L+M-1)(L+M)}{L(2L-1)(2L+1)}}\langle n', L-1\|\hat{d}\|n, L\rangle, \quad (5.12)$$

$$\langle n', L', M'|\hat{d}_+|n, L, M\rangle = \langle n, L, M|\hat{d}_-|n', L', M'\rangle^*, \quad (5.13)$$

where $\langle n', L'\|\hat{d}\|n, L\rangle = \langle n, L\|\hat{d}\|n', L'\rangle^*$ are the reduced matrix elements.

It follows from the above arguments that the deceleration factor $u(q)$ is independent of the magnitude of reduced matrix elements. Thus, the behavior of the wave field in the adiabatic approximation is the same for different substances irrespective of the dipole moment of the transition.

We will now consider the results of calculations of the deceleration factor $u(q)$ on the basis of Eqs. (5.10)–(5.13) for various transitions under the conditions of coherent capture.

(a) $\Delta L = 0$; L is an integer. In this case, we assume that optical evacuation of population from the subsystem possessing no dark state takes place during the establishment of coherent population trapping (see Fig. 2). In this case, the response of the medium is determined by the subsystem with coherent population trapping. The form of the deceleration factor corresponding to such subsystems is shown in Fig. 5 for several small values of angular momentum. It can be seen that the deceleration factor possesses symmetry relative to the change of left circular polarization to right polarization, and vice versa (which corresponds to the substitution $q \longrightarrow 1/q$). Such a symmetry is in accord with the initial symmetry of the system (see Fig. 3). For the transition $L_0 = 1 \longleftrightarrow L_1 = 1$, the value of $u(|q|)$ is equal to unity. For the remaining transitions, the peak of the deceleration factor and, hence, the minimum of the velocity of the polarization wave correspond to the linear polarization ($|q| = 1$), while the minimum of the deceleration factor and the maximum of velocity correspond to circular polarizations ($|q| = 0$, $|q| = \infty$).

(b) $\Delta L = -1$. In this case, both subsystems (see Fig. 3) make additive contributions to deceleration (5.2). The dependences of the deceleration factor on the modulus of the polarization parameter for two different subsystems of some transitions with half-integer angular momenta are shown in Figs. 6a and 6b, respectively. It can be seen that, when the right circular polarization is replaced by the left polarization ($q \longrightarrow 1/q$), Fig. 6a is transformed into Fig. 6b, which corresponds to the initial symmetry of the subsystems. Figure 3 shows that, for a half-integer momentum, one subsystem is

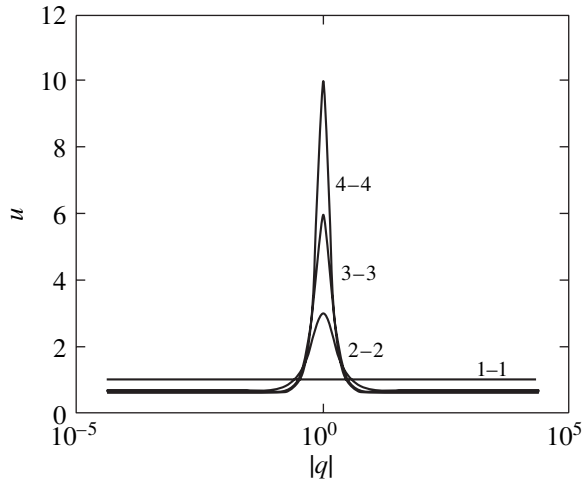

Fig. 5. Dependence of the deceleration factor on the modulus of the polarization parameter for some transitions with $\Delta L = 0$ and with an integer angular momentum.

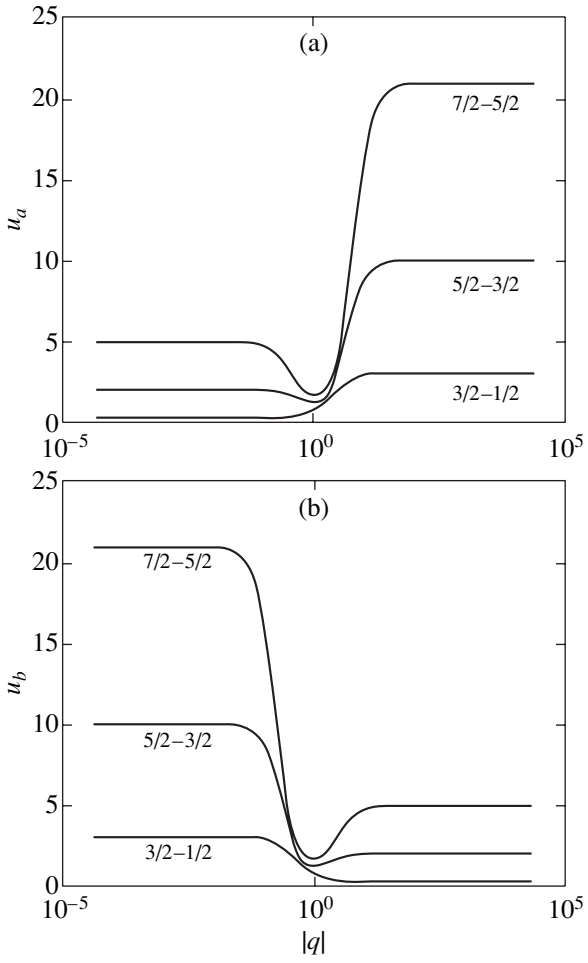

Fig. 6. Deceleration factors for one of the subsystems of certain transitions with $\Delta L = -1$ and a half-integer angular momentum, corresponding to solid (a) and dashed (b) lines in Fig. 3).

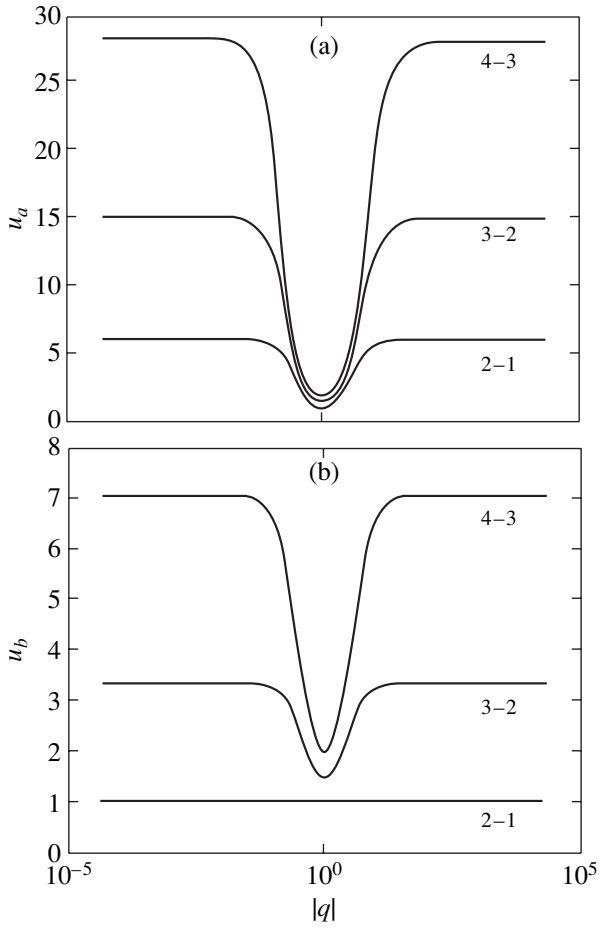

Fig. 7. Deceleration factors for one of the subsystems of certain transitions with $\Delta L = -1$ and an integer angular momentum, corresponding to solid (a) and dashed (b) lines in Fig. 3.

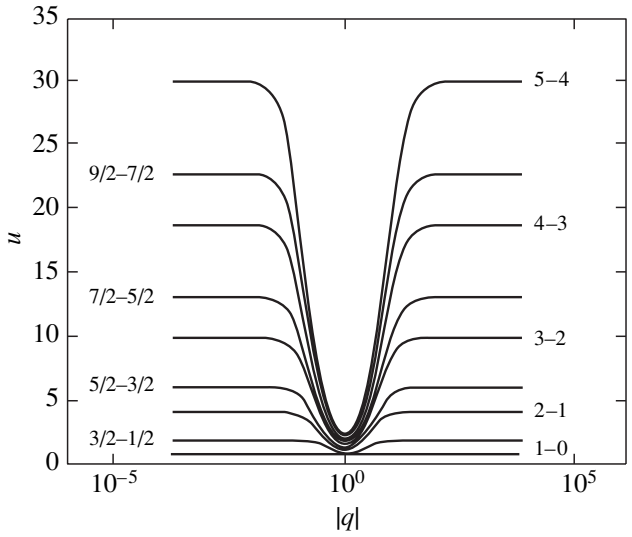

Fig. 8. Total deceleration factor as a function of the modulus of the polarization parameter for some transitions with $\Delta L = -1$.

transformed into the other, and vice versa, as a result of such a replacement.

Similar dependences for transitions with integer momenta are shown in Figs. 7a and 7b; in this case, the subsystems (see Fig. 3) and the corresponding deceleration factors are symmetric relative to the substitution of the left circular polarization for the right polarization. Special attention should be paid to the transition $L_0 = 1 \longleftrightarrow L_1 = 0$, which is not depicted in Figs. 7a and 7b and which decays (see Fig. 3) into the Λ diagram for which $u_\Lambda(|q|) = 1$ and a "pocket," viz., the sublevel $|L_0 = 1, L_{0,z} = 0\rangle$, which does not interact with a field of any polarization.

In accordance with Eq. (5.2), in order to obtain the total deceleration factor of the corresponding transition, we must add the deceleration factors of the subsystems constituting this transition with the weight factors equal to the populations of these subsystems. It is natural to assume that the population will be divided between the subsystems in proportion to the number of sublevels at their lower level. The dependence of the deceleration factor on the magnitude of the polarization parameter obtained under this assumption is presented in Fig. 8. As in the case of transitions with $\Delta L = 0$, the total retardation factor is symmetric relative to the replacement of the left circular polarization by the right polarization, and vice versa ($q \longrightarrow 1/q$), which corresponds to the initial symmetry of the quantum system. For the transition $L_0 = 1 \longleftrightarrow L_1 = 0$, we have $u(|q|) = \text{const}$. For the remaining transitions, the maximum of the deceleration factor and, hence, the minimum of the velocity of polarization wave correspond to circular polarizations ($|q| = 0$, $|q| = \infty$), while the minimum of the deceleration factor and the maximum of velocity correspond to the linear polarization ($|q| = 1$).

Thus, we find that, for all transitions except $L_0 = 1 \longleftrightarrow L_1 = 0$ and $L_0 = 1 \longleftrightarrow L_1 = 1$, the deceleration factor depends on the direction of polarization; in this case, the deceleration in the propagation of a polarization pulse is accompanied by a change in its shape. For our subsequent analysis, we consider the case when deceleration is significant; i.e., $W_E \ll W_{\max}$. It should be noted that this situation is most important for the practical realization of the predicted effects. In this case, we can disregard unity in expression (4.12) for the velocity of the polarization wave and, accordingly, $1/c$ in Eq. (4.10). Taking this into account, we can write the Eq. (4.10) for the polarization parameter as

$$\left[\frac{\partial}{\partial z} + \frac{W_{\max}}{4W_E}u(q)\frac{1}{c}\frac{\partial}{\partial t}\right]q = 0. \qquad (5.14)$$

In the case of a constant field intensity, Eq. (5.14) has a solution in the form of a simple wave:

$$q = q\left(z - \frac{4W_E}{W_{\max}u(q)}ct\right). \qquad (5.15)$$

For transitions with $\Delta L = 0$, the peak of $u(q)$ corresponds to the linear polarization ($|q| = 1$), while its minimum corresponds to the linear polarization ($|q| = 0$, $|q| = \infty$); for transitions with $\Delta L = -1$, the opposite situation is observed. Let us consider for definiteness the transition with $\Delta L = -1$, for which the front of rearrangement of the linear polarization into a circular polarization is extended during the propagation and is contracted upon a transition from a circular to the linear polarization.

The breaking length can be estimated as the distance over which the "rapid" polarization catches up with the "slow" polarization,

$$L = \frac{4W_E}{W_{\max}} \frac{1}{u_{\max} - u_{\min}} cT, \qquad (5.16)$$

where T *i*s the time of variation of polarization at the entrance in the medium.

Over paths shorter than the breaking length, dispersion and dissipation effects must be taken into consideration.

6. CONCLUSIONS

We have analyzed the evolution of an electromagnetic wave with a slowly varying polarization upon the resonant interaction with the medium formed by degenerate two-level atoms under the conditions of electromagnetically induced transparency. It was found that the wave field amplitude propagates at the velocity of light, while the polarization wave propagates at a certain effective velocity which generally depends on the intensity and polarization and may be considerably smaller than the velocity of light. We propose a method of determining this velocity for systems with a one-dimensional dark state. The knowledge of the dependence of deceleration on polarization makes it possible to describe the change in the shape of a polarization pulse as a result of its passage through the medium. It is interesting to note that the behavior of polarization is independent of the magnitude of the dipole moment of the transition and is determined only by the relation between the moments of transitions between different sublevels. The proposed approach is applied for analyzing a two-level system with Zeeman degeneracy, which is often used in experiments on electromagnetically induced transparency. It was found that for all systems of this type, in which electromagnetically induced transparency is possible (except $1 \longleftrightarrow 0$ and $1 \longleftrightarrow 1$ transitions), the velocity of propagation of a polarization pulse is a function of polarization, which leads to deformation of the pulse during its propagation. The form of the velocity of propagation was analyzed for transitions with a small value of the angular momentum.

In the present work, we disregarded the relaxation processes in the lower level, which inevitably emerge in real systems. It is known that the finite lifetime of low-frequency coherence leads to a threshold [1] for the emergence of coherent population trapping and electromagnetically induced transparency associated with it. However, the fact that the results obtained in this work are the consequence of coherent population trapping in the dark state allow us to assume that these results are also valid for fields of a strength exceeding the threshold value considerably.

We can propose the following scheme for observing the effect predicted in this work, which is similar in many respects to that used in [4]. Use should be made of a laser with a certain fixed polarization and a Pockels cell forming radiation with a polarization varying in time. By comparing the polarization structure of the wave passing through the resonant medium and outside this medium, one can observe a delay and variation in the form of a polarization pulse. Since the velocity of propagation of a polarization wave coincides with the velocity of a probe wave under the conditions of electromagnetically induced transparency, the conditions for observing the polarization effects considered by us here correspond to the conditions of other experiments with degenerate systems (see, for example, [4]). The estimates of the polarization rearrangement front length, which sets the limit on the sample size, are given (for rubidium) at the end of Section 3.

ACKNOWLEDGMENTS

The authors are grateful to R. A. Akhmedzhanov, R. L. Kolesov, A. G. Litvak, and V. E. Semenov for fruitful discussions.

This work was supported by the Russian Foundation for Basic Research (project nos. 01-02-17388 and 01-02-17779).

REFERENCES

1. B. D. Agap'ev, M. B. Gornyĭ, B. G. Matisov, and Yu. V. Rozhdestvenskiĭ, Usp. Fiz. Nauk **163** (9), 1 (1993) [Phys. Usp. **36**, 763 (1993)].
2. M. O. Scully and M. S. Zubairy, *Quantum Optics* (Cambridge Univ. Press, Cambridge, 1997).
3. M. D. Lukin, P. R. Hemmer, and M. O. Scully, Adv. At., Mol., Opt. Phys. **42**, 347 (2000).
4. D. F. Phillips, A. Fleischhauer, A. Mair, *et al.*, Phys. Rev. Lett. **86** (5), 783 (2001).
5. O. Kocharovskaya, Y. Rostovtsev, and M. O. Scully, Phys. Rev. Lett. **86**, 628 (2001).
6. M. D. Lukin and A. Imamoglu, Phys. Rev. Lett. **84**, 1419 (2000); S. E. Harris and A. Yamamoto, Phys. Rev. Lett. **81**, 3611 (1998); S. E. Harris, Phys. Rev. Lett. **82**, 4611 (1998).
7. A. V. Durrant, H. X. Chen, S. A. Hopkins, and J. A. Vaccaro, Opt. Commun. **151**, 136 (1998).
8. D. McGloin, M. H. Dunn, and D. J. Fulton, Phys. Rev. A **62**, 053802 (2000).
9. Y. C. Chen, Y. W. Chen, J. J. Su, *et al.*, Phys. Rev. A **63**, 043808 (2001).

10. V. S. Smirnov, A. M. Tumaĭkin, and V. I. Yudin, Zh. Éksp. Teor. Fiz. **96**, 1613 (1989) [Sov. Phys. JETP **69**, 913 (1989)].
11. A. M. Tumaĭkin and V. I. Yudin, Zh. Éksp. Teor. Fiz. **98**, 81 (1990) [Sov. Phys. JETP **71**, 43 (1990)].
12. I. Novikova, A. B. Matsko, and G. R. Welch, Opt. Lett. **26** (13), 1016 (2001).
13. A. S. Chirkin, A. A. Orlov, and D. Yu. Parashchuk, Kvantovaya Élektron. (Moscow) **20**, 999 (1993); D. N. Klyshko, Zh. Éksp. Teor. Fiz. **111**, 1955 (1997) [JETP **84**, 1065 (1997)].
14. S. E. Harris, Phys. Rev. Lett. **70**, 552 (1993); **72**, 52 (1994).
15. J. H. Eberly, M. L. Pons, and H. R. Haq, Phys. Rev. Lett. **72**, 56 (1994); R. Grobe, F. T. Hioe, and J. H. Eberly, Phys. Rev. Lett. **73**, 3183 (1994).
16. A. V. Andreev, Zh. Éksp. Teor. Fiz. **113**, 747 (1998) [JETP **86**, 412 (1998)].
17. V. G. Arkhipkin, D. V. Manushkin, and V. P. Timofeev, Kvantovaya Élektron. (Moscow) **25**, 1084 (1998); V. G. Arkhipkin and I. V. Timofeev, Kvantovaya Élektron. (Moscow) **30**, 180 (2000).
18. S. E. Harris, J. E. Field, and A. Kasapi, Phys. Rev. A **46**, R29 (1992).
19. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1989, 4th ed.; Pergamon, New York, 1977, 3rd ed.).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 927–942.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1080–1097.
Original Russian Text Copyright © 2002 by Tsytovich.

PLASMA, GASES

Collective Plasma Corrections to Thermonuclear Reaction Rates in Dense Plasmas

V. N. Tsytovich*

Institute of General Physics, Russian Academy of Sciences, ul. Vavilova 38, Moscow, 119991 Russia
**e-mail: tsytov@td.lpi.ac.ru*
Received October 11, 2001

Abstract—General kinetic equations are derived for nuclear reactions in dense plasmas by taking into account first-order collective plasma effects. We show that, apart from the corrections proportional to the product of the charges Z_i and Z_j of two reacting nuclei i and j, new corrections comparable in magnitude and proportional to the squares of the nuclear charges Z_i^2 and Z_j^2 arise. The Salpeter corrections [1] to the nuclear reaction probabilities due to the plasma screening of the interaction potential are shown to be at least a factor of r/d (r is the nuclear size and d is the Debye screening length) smaller than those assumed previously. These are zero in the approximation where the terms of order r/d are disregarded. The correlation corrections proportional to Z_iZ_j have a different physical meaning than those in [1], can have a different sign, and arise for reactions with zero Salpeter corrections. For the correlation corrections that substitute for the previously used Salpeter corrections, strong correlations are difficult to describe analytically. The interpolation formulas between weak and strong Salpeter screenings previously used in many astrophysical applications are inapplicable, because the interpolation formulas between weak and strong correlations cannot yet be obtained. We found a new type of corrections that are proportional to the squares of the charges of reacting nuclei. These are attributable to a change in the collective electrostatic self-energy of the plasma system during nuclear reactions. Plasma corrections for the hydrogen-cycle nuclear reactions are numerically calculated for the temperature, density, and abundances in the solar interior. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In his widely known paper, Salpeter [1] showed that the probability of nuclear reactions in dense plasmas for a Debye-screened Coulomb potential is appreciably higher than that for an unscreened Coulomb potential. This effect was called the plasma screening of thermonuclear reactions. It was widely used to describe stellar-evolution models [2] and nuclear reactions in the solar interior [3] (see reviews [4, 5]). For the hydrogen-cycle reactions in the solar interior, the effect contributes from 5 to 20% to the nuclear reaction rates. This contribution is large enough both for solar neutrinos and for the speed of solar sound oscillations, which are clearly detectable by currently available methods of solar seismology. Only 35 years later did Carraro *et al.* [9] notice that the static screening of nuclear reactions (as considered in [1]) is physically meaningless, because the reactions take place at energies much higher than the mean thermal energies (at the so-called Gamov energies), at which there is no static screening for plasma particles (see, e.g., [7]). For such energies, the screening becomes dynamic, vanishing in the limit of high velocities. Subsequently, many serious studies using a sophisticated diagram technique in quantum statistics appeared [8, 9]. These show that the corrections to the nuclear reaction rates must correspond precisely to static screening. As we show here, the authors of [8, 9] actually calculated an effect different from that considered in [1]. Although there are no errors in their calculations [8, 9], the physical interpretation of their results is inaccurate. The coincidence itself between the results of [8, 9] and [1] is apparently accidental and arises only in the zero approximation in small parameters, which are different in [1] and [8, 9]. In this paper, we obtain new, previously disregarded corrections to the nuclear reaction rates, which are proportional to the squares of the charges of reacting nuclei.

The debate on whether the screening of nuclear reactions is dynamic or static is still going on. To resolve this problem, it was necessary to abandon the original assumption [1, 6] that the interaction of nuclei is determined by an average potential. In [10–12], we derived the equations of nuclear kinetics in plasma from the first principles by averaging the microequations over plasma fluctuations. Only this approach is appropriate for systems of many particles. The approach of [1, 6] deals with only two reacting probe particles for which a screened potential is used without proof. The fluctuational approach, as applied to Coulomb collisions in plasma, was able to rigorously prove that collisions take place between dynamically screened particles [7]. Previously [10–12], we found a possible resolution to the dilemma of static or dynamic screening of nuclear reactions. More specifically, we obtained a cancellation of all static corrections. In this

paper, we develop and refine the method of deriving the kinetic equations for nuclear reactions in plasma used in [10–12] or, more specifically, we take into account the change in fluctuations through nuclear reactions. This dilemma is resolved here differently. We show that there is no Salpeter static screening at all and that, to a first approximation in a small parameter (the number of particles within a Debye sphere), the correlation effects in plasma lead to a result that matches the zero approximation of static screening. The existing paradox is resolved, because the physical interpretation of the effect changes radically: the correlation effects can be determined by static permittivity (which is well known for a number of processes in plasma physics, e.g., for wave scattering [7]), while the screening cannot be determined by the latter. These effects coincide only in the zero approximation and only for the corrections proportional to the product Z_iZ_j of the charges of reacting nuclei i and j. In the next approximations, the screening and correlation effects yield completely different results. Here, we consider only weak correlations, which would correspond to weak screening in the Salpeter approach.

A qualitatively new result of [10–12] is the detection of collective corrections proportional to the squares of the nuclear charges (Z_i^2 and Z_j^2), which were absent in all the previous approaches. These corrections are not related to the correlations of interacting nuclei; they are specific to systems in which direct thermonuclear reactions take place but inverse reactions are not possible. Such systems are open. Precisely these nuclear reactions involving neutrinos take place in stellar interiors and on the Sun if neutrinos are capable of freely leaving the region with nuclear reactions. Including this effect in the collision integral (which was ignored previously) gives additional contributions that depend on time derivatives and, thus, on nuclear reaction rates. This leads to a renormalization of the distribution functions for reacting nuclei and to corrections proportional to the squares of the nuclear charges. Like the correlation effects, these effects are related to changes in the distributions of nuclei but not to the nuclear reactions themselves; i.e., these are kinetic collective effects. The renormalization of the particle distributions is known to be a standard operation in any kinetic theory [13, 14], and the necessity of its use in the kinetics of nuclear reactions seems obvious.

Here, the time evolution of nuclear reactions in plasma is investigated in the statement that is most natural for any temporal problems. We assume that there were no nuclear reactions before the initial time $t = 0$ and analyze the asymptotic behavior of the system at large t. Landau used this statement of the problem to investigate the damping of plasma waves. As applied to nuclear reactions in plasma, it yields corrections quadratic in charges. Here, in contrast to [10–12], we assume the plasma fluctuations to be modified by nuclear reactions. The final corrections depend on the entire nuclear cycle and asymptotically differ from those obtained in [10–12].

Thus, in this study, we obtained the results in the approximation of small plasma corrections proportional both to the product of the charges of reacting nuclei and to their squares.

The general results are used for specific numerical calculations of the plasma corrections to the hydrogen-cycle nuclear reaction rates for the parameters of the current solar interior.

2. THE VANISHING OF SALPETER SCREENING

Recall the salient points of [1]. The interaction potential of the nuclei is assumed to be a Debye-screened Coulomb potential:

$$\phi(r) = \frac{Z_iZ_je^2}{r}\exp\left(-\frac{r}{d}\right) \approx \frac{Z_iZ_je^2}{r} - \frac{Z_iZ_je^2}{d}, \qquad (1)$$

where r is the distance between two nuclei, which is much smaller than the Debye screening length d. The correction to the Coulomb potential, the constant in the interaction energy, may be included in the energy E_r of relative motion of the nuclei. The nuclear reaction probability depends only on the latter:

$$\begin{aligned} w_{ij} &= w_{ij}\left(E_r + \frac{Z_iZ_je^2}{d}\right) \\ &\approx w_{ij}(E_r) + \frac{Z_iZ_je^2}{d}\frac{\partial}{\partial E_r}w_{ij}(E_r). \end{aligned} \qquad (2)$$

The thermonuclear reaction rate can be obtained by integrating the probability over the Maxwellian distribution. When integrating by parts, the derivative with respect to the relative energy reduces to the factor $1/T$ in the zero approximation in parameter

$$\frac{T}{E_r} \approx \frac{T}{E_G} \ll 1,$$

where E_G is the Gamov energy (the derivative of the phase factor has such a smallness compared to the derivative of the Maxwellian distribution). For the nuclear reaction rate R_{ij}, we have

$$R_{ij} = R_{ij}^{(0)}(1 + \Lambda_{ij}^{(S)}), \qquad (3)$$

$$\Lambda_{ij}^{(S)} = \frac{Z_iZ_je^2}{dT} = \frac{Z_iZ_je^2}{2\pi^2T}\int\left(1 - \frac{1}{\epsilon_{\mathbf{k},0}}\right)d\mathbf{k}, \qquad (4)$$

where $R_{ij}^{(0)}$ is the reaction rate without plasma corrections and $\epsilon_{\mathbf{k},0}$ is the static permittivity ($\epsilon_{\mathbf{k},0}|_{\omega=0}$). As was noted above, its presence in the final result for screening is physically unacceptable.

The error in this derivation is veiled. It stems from the fact that the screening polarization charge is not a

fixed spatial charge around the nuclei. This charge arises from the motion of the remaining particles that produce fluctuating potentials and fields for the entire system of particles. The nuclear reactions for a specific pair of nuclei take place in the external fluctuating potential ϕ produced by all the remaining nuclei. The average value of this potential does not determine the nuclear reaction rate, because the averaging time is much longer than the nuclear tunneling time. Inserting the fluctuating potential ϕ in the expression for the nuclear reaction probability, expanding it in terms of this potential, and averaging over fluctuations yield a value proportional to the fluctuating potential. However, this calculation, which formally gives (4), also seems to be incorrect. It is presented below. For the terms proportional to Z_iZ_j, the change in average probability is

$$\delta\langle w_{ij}\rangle = Z_iZ_je^2\langle\phi^2\rangle\frac{\partial^2}{\partial E_r^2}w_{ij}(E_r). \quad (5)$$

The fluctuating potential ϕ is determined by the dynamically screened particles, but the standard expression for the fluctuating potential together with the fluctuation–dissipation theorem and integration over the Maxwellian distribution again leads to expression (4) containing the static permittivity. This result shows that when using the standard theory of plasma fluctuations, the error is made somewhere before the averaging. The calculation yielding (4) and (5) uses the fact that there is a natural small parameter, namely, the ratio of the nuclear tunneling size to the fluctuation scale size. This small parameter is used below.

If we introduce the center-of-mass coordinate **R** for two reacting nuclei, their relative coordinate **r**, and the coordinates $\mathbf{r}_i$ and $\mathbf{r}_j$ of each of these nuclei and if take into account the fact that the coordinates of the two nuclei are almost equal in the nuclear reaction, $r \ll R$, then the following expansion can be used for the additional energy:

$$eZ_i\phi(\mathbf{r}_i) + eZ_j\phi(\mathbf{r}_j)$$
$$\approx e(Z_i + Z_j)\phi(\mathbf{R}) + e\mathbf{r}\cdot\mathbf{E}\frac{m_jZ_i - m_iZ_j}{m_i + m_j}. \quad (6)$$

The first term in the latter expression is a constant for the nuclear reactions, but it depends on the center-of-mass coordinate. Therefore, if the second term in (6) is disregarded when separating the variables, it contributes to the center-of-mass wave function, gives rise to a phase factor in the center-of-mass wave function for the system of reacting nuclei, and causes no change in the nuclear reaction probability. The second term in (6), which depends on the relative coordinate and electric field, is small in the ratio of the nuclear tunneling size to the fluctuation scale size. Thus, a constant energy shift arises not in the relative motion but in the translational motion, and there is actually no Salpeter screening of the nuclear reactions. The result (5) is obtained if the first term in (6) is erroneously considered as a perturbation in the relative motion of the nuclei. This result becomes obvious if the averaging is performed after the nuclear reaction and if the average potential is not used. This, in turn, requires that the tunneling time scale be much shorter than the fluctuation time scale (which determines the averaging time scale). The latter is satisfied to a good accuracy if, as usual, the reciprocal of the plasma frequency is taken as the fluctuation time scale. The central point here is the fact established in plasma physics that the screening is produced in fluctuations, which was disregarded in the approaches of [1, 6].

3. CORRELATION EFFECTS

The correlation effects describe the correlation between the states of two reacting nuclei. If the nuclei are close to each other more frequently, then the average reaction rate increases. This effect differs from that discussed in [1, 6], where the charge in the rates of the reactions themselves was considered. The kinetics of fluctuations in the system of reacting nuclei is important in describing the correlation effects. To investigate this effect, we will use microequations, as in [10–12], but in an improved form. This improvement is needed to construct a more detailed theory of correlations. Here, we do not present this theory but provide only its final result, which matches that previously obtained in the zero approximation [10–12]. Therefore, the above improvement is used only for a guaranteed justification of the correlation corrections found in [10–12]. The basic equation used in developing an improved theory of correlations is

$$\frac{\partial}{\partial t}f_i + \mathbf{v}\cdot\frac{\partial}{\partial\mathbf{r}}f_i + Z_ie\mathbf{E}\cdot\frac{\partial}{\partial\mathbf{p}}f_i = -\int w_{ij}f_{ij}\frac{d\mathbf{p}}{(2\pi)^3}, \quad (7)$$

where f_i and f_{ij} are the one- and two-particle distribution functions, respectively. The former and the latter can be obtained by integrating the total distribution over the variables of all particles except particle i and particles i and j, respectively. The approximation $f_{ij} \approx f_if_j$ was used in [10, 12]. Equation (7) is more accurate than that used in [10, 12]. This equation can be investigated by analyzing the equation for f_{ij} that can be derived by integration over all variables except i and j rather than over all variables except the variables of particle i, as in the derivation of Eq. (7). It allows the correlation effects to be studied in more detail. This approach is more detailed than that in [10, 12] but much more cumbersome. The correlation effects are contained in the approach used in [10, 12], because the average product of two one-particle distribution functions is not equal to the product of the average distributions. We present only the result of an extensive analysis of the correlation problem using Eq. (7): to a first nonzero approximation, the correlation corrections are equal to those obtained in [10, 12] by assuming that $f_{ij} \approx f_if_j$. This equality is not possible

in the next approximations. Recall that to obtain the correlation corrections, we introduce the fluctuation-averaged distributions

$$f_i = \Phi_i + \delta f_i, \quad \langle f_i \rangle = \Phi_i. \tag{8}$$

Averaging Eq. (7) and taking into account the fact that, to a first approximation, the nuclear reaction probability does not depend on the fluctuating potential (this dependence was taken into account in [10, 12]), we derive on the right-hand side

$$\langle f_i f_j \rangle = \Phi_i \Phi_j + \langle \delta f_i \delta f_j \rangle. \tag{9}$$

As the above studies showed, the second term in Eq. (9) correctly describes the correlations in the first nonzero approximation. The smallness of the reaction rate compared to the fluctuation frequency can be used to calculate these explicitly. Subsequently, standard expressions for fluctuations of the particle distributions in plasma in the absence of nuclear reactions can be used. Although general expressions can be written for any nonequilibrium distributions of nuclei, we present the result for equilibrium thermal distributions, where the fluctuation–dissipation theorem can be used to integrate the fluctuations over frequencies. In this limit, the corrections can be expressed in terms of static permittivity:

$$\langle \delta f_i \delta f_j \rangle = \frac{Z_i Z_j e^2}{2\pi^2 T} \Phi_i \Phi_j \int \frac{d\mathbf{k}}{k^2} \left(1 - \frac{1}{\epsilon_{\mathbf{k},0}}\right). \tag{10}$$

Here, we took into account the fact that the Coulomb field of two reacting nuclei should be disregarded in the fluctuating potential produced by all the remaining plasma particles. The correlation effect itself bears no relation to the possible change in the probability of nuclear reactions between two reacting nuclei: because of the correlations, the number of reacting nuclei proves to be large, on average. In [10, 12], apart from the correlations, we also took into account the change in probability, an effect that interfered with the correlation effect in [10, 12]. Since no change in probability is found in this study, we consider only the correlation corrections. The proportionality of the corrections to the product of the average distribution functions allows us to introduce some effective nuclear reaction probabilities, which lead to the same result in the equations for nuclear reactions in plasma as allowance for the correlations:

$$w_{ij}^{\text{eff}} = w_{ij}\left[1 + \frac{Z_i Z_j e^2}{2\pi^2 T} \int \frac{d\mathbf{k}}{k^2} \left(1 - \frac{1}{\epsilon_{bfk,0}}\right)\right] = w_{ij}(1 + \Lambda_{ij}^{(C)}). \tag{11}$$

The formal coincidence of the corrections $\Lambda_{ij}^{(C)}$ in (11) with the Salpeter corrections $\Lambda_{ij}^{(S)}$ (4) should not mislead us, because the corrections (11) have a different physical meaning. An example shows that these also differ quantitatively. There is no nuclear barrier in the reaction of ^{7}Be with electrons, and the Salpeter corrections are zero, while the correlation corrections are nonzero and are described by Eq. (11).

Since the calculations in [8, 9] were performed by statistically averaging the unperturbed probability over the electron and ion distributions, they allow for the correlations. Thus, the result obtained here also agrees with the result of [8, 9].

4. FLUCTUATIONS IN A TIME-EVOLVING SYSTEM

Since there are no inverse processes with neutrino absorption, the system is open and evolves with time. The plasma fluctuations are not stationary (as is usual in the absence of nuclear reactions), with the rate of their change with time being determined by the nuclear reaction rates. Although this rate is small compared to the characteristic fluctuation frequency, the effects related to the time evolution of plasma fluctuations must be taken into account when calculating all effects linear in nuclear reaction rate. Previously, these effects were ignored. The effects related to a collective electric field lead to plasma corrections. Therefore, let us first consider those effects for which the electric field is negligible. Since the total corrections are treated as small ones, we will take into account the effects related to collective fields by using perturbation theory; the absence of collective fields corresponds to the zero approximation of this theory. Denote this approximation by the superscript $^{(0)}$. The basic equation for linear fluctuations then takes the form

$$\frac{\partial \delta f_i^{(0)}(\mathbf{p})}{\partial t} + \mathbf{v} \cdot \frac{\partial \delta f_i^{(0)}(\mathbf{p})}{\partial \mathbf{r}} = -\int w_{ij}(\mathbf{p}, \mathbf{p}') \times (\delta f_i^{(0)}(\mathbf{p}) \Phi_j(\mathbf{p}') + \delta f_j^{(0)}(\mathbf{p}') \Phi_i(\mathbf{p})) \frac{d\mathbf{p}'}{(2\pi)^3}. \tag{12}$$

A similar equation can be written for $f_j^{(0)}(\mathbf{p}')$:

$$\frac{\partial \delta f_j^{(0)}(\mathbf{p}')}{\partial t} + \mathbf{v} \cdot \frac{\partial \delta f_j^{(0)}(\mathbf{p}')}{\partial \mathbf{r}} = -\int w_{ij}(\mathbf{p}, \mathbf{p}') \times (\delta f_i^{(0)}(\mathbf{p}) \Phi_j(\mathbf{p}') + \delta f_j^{(0)}(\mathbf{p}') \Phi_i(\mathbf{p})) \frac{d\mathbf{p}}{(2\pi)^3}. \tag{13}$$

Having written this system of equations for the Fourier components,

$$\delta f_{i,j} = \int \delta f_{i,j,\mathbf{k},\omega} \exp(i\mathbf{k} \cdot \mathbf{r} - \omega t) d\mathbf{k} d\omega,$$

we can derive an equation that contains only fluctuations in the distribution function of one of the nuclei:

$$(\omega - \mathbf{k}\cdot\mathbf{v} + i\nu_i(\mathbf{v})\delta f^{(0)}_{i,\mathbf{k},\omega}(\mathbf{v})) = -\Phi_i(\mathbf{v}) \times \int \frac{w_{ij}(\mathbf{p},\mathbf{p}')w_{ij}(\mathbf{p}'',\mathbf{p}')}{\omega - \mathbf{k}\cdot\mathbf{v}' + i\nu_j(\mathbf{v}')}\Phi_j(\mathbf{v}')\delta f^{(0)}_{i,\mathbf{k},\omega}(\mathbf{v}'')\frac{d\mathbf{p}'d\mathbf{p}''}{(2\pi)^6}, \quad (14)$$

where the quantities

$$\nu_i(\mathbf{p}) = \int w_{ij}(\mathbf{p},\mathbf{p}')\Phi_j(\mathbf{p}')\frac{d\mathbf{p}'}{(2\pi)^3},$$
$$\nu_j(\mathbf{p}') = \int w_{ij}(\mathbf{p},\mathbf{p}')\Phi_i(\mathbf{p})\frac{d\mathbf{p}}{(2\pi)^3} \quad (15)$$

describe the fluctuation damping due to the nuclear reactions. The right-hand side of Eq. (14) describes the correlation of fluctuations due to the nuclear reactions and their additional damping. Given that the characteristic fluctuation frequency is $kv_{Ti} \approx \omega_{pi}$ for k of the order of the inverse Debye length $1/d$, where ω_{pi} is the plasma frequency of ions i, this correlation can be easily estimated. The right-hand side of Eq. (14) is a factor of ν_i^2/ω_{pi} smaller than the damping on the left-hand side of Eq. (14) and may, therefore, be disregarded. For the spatial components of the fluctuations in the distribution function

$$\delta f^{(0)}_i(\mathbf{v},\mathbf{r},t) = \int \delta f^{(0)}_{i,\mathbf{k}}(\mathbf{v},t)\exp(i\mathbf{k}\cdot\mathbf{r})d\mathbf{k},$$

the corresponding equation can be written as

$$\left(\frac{\partial}{\partial t} + i\mathbf{k}\cdot\mathbf{v} + \nu_i(\mathbf{v})\right)\delta f^{(0)}_{i,\mathbf{k}}(\mathbf{v},t) = 0. \quad (16)$$

Here, we state a temporal problem with initial conditions for the system at time $t = 0$. This statement is necessary in an open system, which cannot come to a complete equilibrium because of the absence of inverse processes with neutrinos. We assume that the nuclear reactions are switched on at time $t = 0$ and consider their rate asymptotically at large times. Thus, we assume that $\nu_i = 0$ at $t < 0$. This statement of the problem is close to the actual situation in stars, where nuclear burning starts at a certain contraction stage of the protostellar cloud. For $t < 0$, the solution to (16) is then

$$\delta f^{(0)}_{i,\mathbf{k}}(\mathbf{p},t) = \delta f^{(0)}_{i,\mathbf{k}}(\mathbf{p})\exp(-i\mathbf{k}\cdot\mathbf{v}t), \quad (17)$$

while, for $t > 0$,

$$\delta f^{(0)}_{i,\mathbf{k}}(\mathbf{p},t) = \delta f^{(0)}_{i,\mathbf{k}}\exp(-i\mathbf{k}\cdot\mathbf{v}t - i\nu_i(\mathbf{v})t). \quad (18)$$

Below, we restrict ourselves to a spatially homogeneous problem. The average values of $\delta f_{i,\mathbf{k}}(\mathbf{p})$ must then be the same as those for a stationary (on average) system:

$$\langle \delta f^{(0)}_{i,\mathbf{k}}(\mathbf{p})\delta f^{(0)}_{j,\mathbf{k}'}(\mathbf{p}')\rangle = \Phi_i(\mathbf{p})\delta_{i,j}\delta(\mathbf{p}-\mathbf{p}')\delta(\mathbf{k}+\mathbf{k}'). \quad (19)$$

This leads to the following law for the averaging of temporal fluctuations in the presence of nuclear reactions used below:

$$\langle \delta f^{(0)}_{i,\mathbf{k},\omega}(\mathbf{p})\delta f^{(0)}_{j,\mathbf{k}',\omega'}(\mathbf{p}')\rangle = -\frac{1}{4\pi^2}\Phi_i(\mathbf{p})\delta_{i,j}\delta(\mathbf{p}-\mathbf{p}')\delta(\mathbf{k}+\mathbf{k}') \times \left[\frac{1}{\omega - \mathbf{k}\cdot\mathbf{v} - i0} - \frac{1}{\omega - \mathbf{k}\cdot\mathbf{v} + i\nu_i(\mathbf{v})}\right] \times \left[\frac{1}{\omega' - \mathbf{k}'\cdot\mathbf{v}' - i0} - \frac{1}{\omega' - \mathbf{k}'\cdot\mathbf{v}' + i\nu_i(\mathbf{v}')}\right]. \quad (20)$$

In the limit $\nu_{i,j} \longrightarrow 0$, the law (20) matches the standard averaging law for a stationary system (we denote the corresponding distributions by the superscript $^{(0,0)}$):

$$\langle \delta f^{(0,0)}_{i,\mathbf{k},\omega}(\mathbf{p})\delta f^{(0,0)}_{j,\mathbf{k}',\omega'}(\mathbf{p}')\rangle = \Phi_i(\mathbf{p})\delta_{i,j}\delta(\mathbf{p}-\mathbf{p}') \times \delta(\mathbf{k}+\mathbf{k}')\delta(\omega+\omega')\delta(\omega - \mathbf{k}\cdot\mathbf{v}). \quad (21)$$

The change in the fluctuation-averaged distribution of nuclei with time should be determined by using (20), which yields the change in nuclear reaction rates due to collective plasma effects. Apart from the change in fluctuations with time, the evolution of the average distribution function for nuclei is of considerable importance.

5. THE INFLUENCE OF TIME-EVOLVING FLUCTUATIONS ON THE NUCLEAR REACTION RATES

The basic equation for calculating the collective corrections to the nuclear reaction rates is the equation obtained by averaging (7) over plasma fluctuations:

$$\frac{\partial}{\partial t}\Phi_i = Z_i e\frac{\partial}{\partial \mathbf{p}}\langle \delta f_i \nabla\phi\rangle - \int w_{ij}(\Phi_i\Phi_j + \langle \delta f_i \delta f_j\rangle)\frac{d\mathbf{p}'}{(2\pi)^3}. \quad (22)$$

As in the calculation of fluctuations, we assume the average distribution to be homogeneous but time-dependent. This dependence is attributable to the nuclear reactions, the plasma corrections to which are investigated. In the absence of such a dependence, the first term on the right-hand side of Eq. (22) is known to lead to a collision integral that rapidly (on the collision time scale) approaches zero, making the particle distribution a thermal one. However, in the presence of time variations (which are proportional to the nuclear reaction rates in our case), this term gives an additional nonzero contribution proportional to the rate of change in fluctuations and to the rate of time evolution of the average distribution of nuclei. Our objective is to take into account effects of the first order in nuclear reaction rates, i.e., linear in the time derivatives of the average

distribution and linear in w_{ij}. The last term on the right-hand side of Eq. (22) contains the correlation corrections. Since the effect of time evolution and the correlation effects in our approximation linear in w_{ij} add up additively, we ignore the correlation effects in this section. Finally, if we disregard the corrections attributable to time evolution of the fluctuations, then Eq. (22) reduces to

$$\frac{\partial}{\partial t}\Phi_i^{(0)} = -\Phi_i^{(0)}\int w_{ij}\Phi_i'\frac{d\mathbf{p}'}{(2\pi)^3} = -\nu_i\Phi_i^{(0)}. \quad (23)$$

In this equation, the frequency ν_i should be assumed to be constant, because allowance for its time dependence would imply allowance for the corrections of higher orders in w_{ij}. For fluctuations, we have instead of Eq. (16)

$$-i(\omega - \mathbf{k}\cdot\mathbf{v} + i\nu_i)\delta f_{i,\mathbf{k},\omega} = Z_i e\left(\nabla\delta\phi\cdot\frac{\partial}{\partial\mathbf{p}}\Phi_i\right)_{\mathbf{k},\omega}, \quad (24)$$

where ϕ is the potential of the fluctuating electric field. The solution to the latter equation is

$$\delta f_{i,\mathbf{k},\omega} = \delta f^{(0)}_{i,\mathbf{k},\omega} - \frac{Z_i e}{\omega - \mathbf{k}\cdot\mathbf{v} + i\nu_i} \times\int\phi_{\mathbf{k},\omega-\omega'}\left(\mathbf{k}\cdot\frac{\partial}{\partial\mathbf{p}}\Phi_{i,\omega'}\right)d\omega', \quad (25)$$

where $\delta f^{(0)}_{i,\mathbf{k},\omega}$ is the solution to the homogeneous equation (24), which describes the fluctuations in a time-evolving system discussed in the previous section. Since the average distribution function varies with time much more slowly than the plasma fluctuations $\omega' \ll \omega$, the expansion in terms of ω' and ν_i can be written as

$$\phi_{\mathbf{k},\omega-\omega'} \approx \phi_{\mathbf{k},\omega} - \omega'\frac{\partial}{\partial\omega}\phi_{\mathbf{k},\omega}, \qquad \int\omega'\Phi_{i,\omega'}d\omega' \approx i\frac{\partial}{\partial t}\Phi_i. \quad (26)$$

Using (23), we obtain

$$\delta f_{i,\mathbf{k},\omega} = \delta f^{(0)}_{i,\mathbf{k},\omega} - \left(1 - i\frac{\partial}{\partial t}\frac{\partial}{\partial\omega}\right) \times\frac{Z_i e}{\omega - \mathbf{k}\cdot\mathbf{v} + i0}\phi_{\mathbf{k},\omega}\left(\mathbf{k}\cdot\frac{\partial}{\partial\mathbf{p}}\Phi_i\right). \quad (27)$$

We use the Poisson equation to derive an expression for the fluctuating potential in which the terms with a time derivative are considered based on perturbation theory:

$$\phi_{\mathbf{k},\omega} \approx \phi^{(0)}_{\mathbf{k},\omega} + \phi^{(1)}_{\mathbf{k},\omega} + \ldots,$$
$$\phi^{(0)}_{\mathbf{k},\omega} = \frac{4\pi}{k^2\epsilon^{(0)}_{\mathbf{k},\omega}}\sum_\alpha Z_\alpha e\int\delta f^{(0)}_{\mathbf{k},\omega}\frac{d\mathbf{p}}{(2\pi)^3}, \quad (28)$$

$$\phi^{(1)}_{\mathbf{k},\omega} = \frac{i}{\epsilon^{(0)}_{\mathbf{k},\omega}}\frac{\partial}{\partial\omega}\left(\phi^{(0)}_{\mathbf{k},\omega}\frac{\partial}{\partial t}\epsilon^{(0)}_{\mathbf{k},\omega}\right), \quad (29)$$

$$\epsilon^{(0)}_{\mathbf{k},\omega} = 1 + \frac{4\pi}{k^2}\sum_\alpha Z_\alpha^2 e^2\int\frac{1}{\omega - \mathbf{k}\cdot\mathbf{v} + i0} \times\left(\mathbf{k}\cdot\frac{\partial}{\partial\mathbf{p}}\Phi_\alpha\right)\frac{d\mathbf{p}}{(2\pi)^3}. \quad (30)$$

Here, the summation over α is extended to include all types of plasma particles: electrons and all ions, including the reacting nuclei. We have

$$\delta f_{i,\mathbf{k},\omega} = \delta f^{(0)}_{i,\mathbf{k},\omega} - \frac{Z_i e}{\omega - \mathbf{k}\cdot\mathbf{v} + i0}\phi^{(0)}_{\mathbf{k},\omega}\left(\mathbf{k}\cdot\frac{\partial}{\partial\mathbf{p}}\Phi_i\right) + \delta f^{(1)}_{i,\mathbf{k},\omega}, \quad (31)$$

$$\delta f^{(1)}_{i,\mathbf{k},\omega} = i\frac{\partial}{\partial t}\frac{\partial}{\partial\omega}\frac{Z_i e}{\omega - \mathbf{k}\cdot\mathbf{v} + i0}\phi^{(0)}_{\mathbf{k},\omega}\left(\mathbf{k}\cdot\frac{\partial}{\partial\mathbf{p}}\Phi_i\right) - \frac{Z_i e}{\omega - \mathbf{k}\cdot\mathbf{v} + i0}\phi^{(1)}_{\mathbf{k},\omega}\left(\mathbf{k}\cdot\frac{\partial}{\partial\mathbf{p}}\Phi_i\right). \quad (32)$$

The effects related to the time evolution of the system are described by the equation for the fluctuation-averaged distribution. Disregarding the correlations, we obtain from (22)

$$\frac{\partial}{\partial t}\Phi_i = Z_i e\frac{\partial}{\partial\mathbf{p}}\cdot\int i\mathbf{k}'\langle\delta f_{i,\mathbf{k},\omega}\phi_{\mathbf{k}',\omega'}\rangle \times\exp[i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{r} - i(\omega+\omega')t]d\mathbf{k}d\mathbf{k}'d\omega d\omega' - \int w_{ij}\Phi_i\Phi_j\frac{d\mathbf{p}'}{(2\pi)^3} = I_i^{(0)} + I_i^{(t)} - \int w_{ij}\Phi_i\Phi_j\frac{d\mathbf{p}'}{(2\pi)^3}, \quad (33)$$

where $I_i^{(0)}$ is determined by the time-evolving fluctuations $\delta f^{(0)}_{i,\mathbf{k},\omega}$,

$$I_i^{(0)} = -Z_i e\frac{\partial}{\partial\mathbf{p}}\cdot\int i\mathbf{k}\left\langle\left[\delta f^{(0)}_{i,\mathbf{k},\omega} - \frac{Z_i e}{\omega - \mathbf{k}\cdot\mathbf{v} + i0}\left(\mathbf{k}\cdot\frac{\partial\Phi_i}{\partial\mathbf{p}}\right)\phi^{(0)}_{\mathbf{k},\omega}\right]\phi^{(0)}_{\mathbf{k}',\omega'}\right\rangle \times\exp[-i(\omega+\omega')t]d\mathbf{k}d\mathbf{k}'d\omega d\omega', \quad (34)$$

and $I_i^{(t)}$ is determined by the time variation in the average distribution of reacting nuclei,

$$I_i^{(t)} = Z_i e\frac{\partial}{\partial\mathbf{p}}\cdot\int\mathbf{k}d\mathbf{k}d\mathbf{k}'\left\{\left\{\frac{\partial}{\partial t}\frac{\partial}{\partial\omega}\right.\right.$$

$$-\left[\frac{1}{\epsilon^{(0)}_{\mathbf{k}',\omega'}}\frac{\partial}{\partial\omega'}\frac{\partial}{\partial t}\epsilon^{(0)}_{\mathbf{k}',\omega'}+\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\frac{\partial}{\partial\omega}\frac{\partial}{\partial t}\epsilon^{(0)}_{\mathbf{k},\omega}\right]\Bigg\}$$

$$\times\frac{Z_ie}{\omega-\mathbf{k}\cdot\mathbf{v}+i0}\left(\mathbf{k}\cdot\frac{\partial\Phi_i}{\partial\mathbf{p}}\right)\langle\phi^{(0)}_{\mathbf{k},\omega}\phi^{(0)}_{\mathbf{k}',\omega'}\rangle \tag{35}$$

$$+\frac{1}{\epsilon^{(0)}_{\mathbf{k}',\omega'}}\frac{\partial}{\partial\omega'}\frac{\partial\epsilon^{(0)}_{\mathbf{k}',\omega'}}{\partial t}\langle\delta f^{(0)}_{i,\mathbf{k},\omega}\phi^{(0)}_{\mathbf{k}',\omega'}\rangle\Bigg\}$$

$$\times\exp[-(\omega+\omega')t]d\omega d\omega'.$$

Here, we took into account the fact that, according to (20) (homogeneity of the fluctuations), $\mathbf{k}' = -\mathbf{k}$. Whereas the time evolution of fluctuations must be included in Eq. (34), in Eqs. (35), which already contain a time derivative of the average distribution, it will suffice to use the approximate relation (21). In the latter case, we denote the corresponding correlation functions by the superscript (0,0), just like the distribution functions in (21). We obtain from (21)

$$\int\langle\phi^{(0,0)}_{\mathbf{k}',\omega'}\phi^{(0,0)}_{\mathbf{k},\omega}\rangle d\mathbf{k}'$$
$$=-\frac{T}{2\pi^3k^2\omega}\mathrm{Im}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\right)\delta(\omega+\omega'), \tag{36}$$

$$\int\langle f^{(0,0)}_{i,\mathbf{k}',\omega'}\phi^{(0,0)}_{\mathbf{k},\omega}\rangle d\mathbf{k}'$$
$$=\frac{Z_ie}{2\pi^2k^2\epsilon^{(0)}_{\mathbf{k},\omega}}\Phi_i\delta(\omega-\mathbf{k}\cdot\mathbf{v})\delta(\omega+\omega'). \tag{37}$$

Here, we assume that the average distribution function in the terms of (35) may be considered to be thermal (Maxwellian) with sufficient accuracy. The small deviations of the distributions from Maxwellian ones must be taken into account only in the terms that do not contain small time derivatives of the average distributions (see below). Substituting the approximate relations (36) and (37) for the correlation functions (35), we find that the first term in (35) containing the total derivative with respect to frequency becomes zero when integrating over frequencies and that

$$I_i^{(t)}=-\frac{Z_i^2e^2}{2\pi^3}\frac{\partial}{\partial\mathbf{p}}\cdot\int\frac{\mathbf{k}}{k^2}\Phi_i d\omega d\mathbf{k}$$

$$\times\Bigg\{-\pi\delta(\omega-\mathbf{k}\cdot\mathbf{v})\frac{\partial\epsilon^{(0)}_{\mathbf{k},\omega}}{\partial t}\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\frac{\partial}{\partial\omega}\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}$$

$$+\frac{\mathbf{k}\cdot\mathbf{v}}{\omega-\mathbf{k}\cdot\mathbf{v}+i0}\Bigg\{\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\frac{\partial}{\partial\omega}\left[\frac{\partial\epsilon^{(0)}_{\mathbf{k},\omega}}{\partial t}\frac{1}{\omega}\mathrm{Im}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\right)\right] \tag{38}$$

$$+\frac{1}{\omega}\mathrm{Im}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\right)\frac{\partial\epsilon^{(0)}_{-\mathbf{k},-\omega}}{\partial t}\frac{\partial}{\partial\omega}\frac{1}{\epsilon^{(0)}_{-\mathbf{k},-\omega}}\Bigg\}\Bigg\}.$$

To calculate the change in reaction rate due to the time evolution of the fluctuations described by relation (30), it should be borne in mind that these effects must be taken into account in the first nonvanishing order in reaction rates. In this limit, we derive the following relation from (20) [cf. (27) and (28)]:

$$\int\langle\delta f^{(0)}_{i,\mathbf{k},\omega}(\mathbf{p})\delta f^{(0)}_{j,\mathbf{k}',\omega'}(\mathbf{p}')\rangle e^{-(\omega+\omega')t}d\mathbf{k}'d\mathbf{p}'$$
$$=\delta_{i,j}\left[\Phi_i-\frac{i}{2}\frac{\partial\Phi_i}{\partial t}\left(\frac{\partial}{\partial\omega}+\frac{\partial}{\partial\omega'}\right)\right]\delta(\omega+\omega')\delta(\omega-\mathbf{k}\cdot\mathbf{v}). \tag{39}$$

The first term on the right-hand side of Eq. (39) leads to the standard Coulomb collision integral, which turns the particle distributions into thermal ones in very short time intervals. For a thermal distribution, the first term in (39) is exactly equal to zero. Below, we take into account the renormalization corrections, for which the first term in (39) is of the order of $1/N_d \ll 1$ and which may also be disregarded. In the corrections proportional to the time derivative of the average distribution, the latter may be assumed to be thermal. We obtain

$$I_i^{(0)}=-\frac{Z_i^2e^2}{4\pi^3}\frac{\partial}{\partial\mathbf{p}}\cdot\int\frac{\mathbf{k}}{k^2}d\mathbf{k}d\omega$$

$$\times\Bigg\{\pi\delta(\omega-\mathbf{k}\cdot\mathbf{v})\frac{\partial\Phi_i}{\partial t}\frac{\partial}{\partial\omega}\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}$$

$$+\frac{(\mathbf{k}\cdot\mathbf{v})\Phi_i}{(\omega-\mathbf{k}\cdot\mathbf{v}+i0)\epsilon^{(0)}_{\mathbf{k},\omega}} \tag{40}$$

$$\times\left[\frac{\partial}{\partial\omega}\frac{1}{\omega\epsilon^{(0)}_{-\mathbf{k},-\omega}}\frac{\partial}{\partial t}\mathrm{Im}\epsilon^{(0)}_{\mathbf{k},\omega}+\frac{1}{\omega}\frac{\partial}{\partial t}\mathrm{Im}\epsilon^{(0)}_{\mathbf{k},\omega}\frac{\partial}{\partial\omega}\frac{1}{\epsilon^{(0)}_{-\mathbf{k},-\omega}}\right]\Bigg\}.$$

This expression can be simplified as follows. The first term in square brackets can be integrated by parts over frequencies. The emerging combination with the second term in square brackets but without the derivative of the factor $1/(\omega-\mathbf{k}\cdot\mathbf{v}+i0)$ with respect to frequency,

$$-\frac{1}{\epsilon^{(0)}_{-\mathbf{k},-\omega}}\frac{\partial}{\partial\omega}\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}+\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\frac{\partial}{\partial\omega}\frac{1}{\epsilon^{(0)}_{-\mathbf{k},-\omega}}, \tag{41}$$

does not change when substituting $\omega\longrightarrow-\omega$ and $\mathbf{k}\longrightarrow-\mathbf{k}$. The remaining factors

$$\frac{\mathbf{k}(\mathbf{k}\cdot\mathbf{v})}{\omega(\partial/\partial t)\mathrm{Im}\epsilon_{\mathbf{k},\omega}}$$

do not change either during this substitution. Therefore, only the term $1/(\omega-\mathbf{k}\cdot\mathbf{v}+i0)$, which does not change during this substitution, i.e., $-i\pi\delta(\omega-\mathbf{k}\cdot\mathbf{v})$, remains in the expression. This implies that only the imaginary part remains in expression (41). In the remaining deriv-

ative of $1/(\omega - \mathbf{k}\cdot\mathbf{v} + i0)$ with respect to frequency, we can again perform integration by parts to give

$$I_i^{(0)} = -\frac{Z_i^2 e^2}{4\pi^3}\frac{\partial}{\partial \mathbf{p}}\cdot\int\frac{\mathbf{k}}{k^2}d\mathbf{k}d\omega$$
$$\times\left\{\pi\delta(\omega-\mathbf{k}\cdot\mathbf{v})\left[\frac{\partial\Phi_i}{\partial t}\frac{\partial}{\partial\omega}\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}+\frac{\partial}{\partial t}\mathrm{Im}\epsilon^{(0)}_{\mathbf{k},\omega}\right.\right.$$
$$\left.\times 2\left[\mathrm{Im}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\right)\frac{\partial}{\partial\omega}\mathrm{Re}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\right)-\mathrm{Re}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\right)\frac{\partial}{\partial\omega}\mathrm{Im}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\right)\right]\right] \quad (42)$$
$$\left.+\frac{\omega\Phi_i}{\omega-\mathbf{k}\cdot\mathbf{v}+i0}\frac{\partial}{\partial\omega}\left(\frac{1}{\omega\left|\epsilon^{(0)}_{\mathbf{k},\omega}\right|^2}\frac{\partial}{\partial t}\mathrm{Im}\epsilon^{(0)}_{\mathbf{k},\omega}\right)\right\}.$$

Similarly, we can simplify the first term in square brackets in (38). Integrating by parts over frequencies yields

$$I_i^{(t)} = -\frac{Z_i^2 e^2}{2\pi^3}\frac{\partial}{\partial\mathbf{p}}\cdot\int\frac{\mathbf{k}}{k^2}\Phi_i d\omega d\mathbf{k}$$
$$\times\left\{\pi\delta(\omega-\mathbf{k}\cdot\mathbf{v})\left[\frac{1}{\left|\epsilon^{(0)}_{\mathbf{k},\omega}\right|^2\epsilon^{(0)}_{\mathbf{k},\omega}}\frac{\partial\epsilon^{(0)}_{\mathbf{k},\omega}}{\partial t}\frac{\partial\epsilon^{(0)}_{\mathbf{k},\omega}}{\partial\omega}\right.\right.$$
$$-2\mathrm{Im}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\right)\left[\frac{\partial}{\partial\omega}\mathrm{Im}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k}\omega}}\right)\frac{\partial}{\partial t}\mathrm{Re}\epsilon^{(0)}_{\mathbf{k},\omega}\right. \quad (43)$$
$$\left.\left.+\frac{\partial}{\partial\omega}\mathrm{Re}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k}\omega}}\right)\frac{\partial}{\partial t}\mathrm{Im}\epsilon^{(0)}_{\mathbf{k},\omega}\right]\right]+\frac{\omega}{\omega-\mathbf{k}\cdot\mathbf{v}+i0}$$
$$\left.\times\frac{\partial}{\partial\omega}\left[\frac{1}{\omega}\mathrm{Im}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\right)\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\frac{\partial}{\partial t}\epsilon^{(0)}_{\mathbf{k},\omega}\right]\right\}.$$

The sum of (42) and (43) admits further simplifications. Separating out the terms with $\delta(\omega - \mathbf{k}\cdot\mathbf{v})$, we write $I_i = I_i^{(t)} + I_i^{(0)}$ as

$$I_i = \frac{Z_i^2 e^2}{4\pi^2}\frac{\partial}{\partial\mathbf{p}}\cdot\int\frac{\mathbf{k}}{k^2}d\omega d\mathbf{k}$$
$$\times\left\{\frac{\Phi_i}{\pi}\frac{\partial}{\partial t}\left[\mathrm{Re}\left(\frac{1}{\omega-\mathbf{k}\cdot\mathbf{v}+i0}\right)\omega\frac{\partial}{\partial\omega}\frac{1}{\omega}\mathrm{Im}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\right)\right.\right. \quad (44)$$
$$\left.+\mathrm{Im}\left(\frac{1}{\omega-\mathbf{k}\cdot\mathbf{v}+i0}\right)\omega\frac{\partial}{\partial\omega}\frac{1}{\omega}\mathrm{Re}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\right)\right]-\delta(\omega-\mathbf{k}\cdot\mathbf{v})$$
$$\left.\times\left[\frac{\partial\Phi_i}{\partial t}\frac{\partial}{\partial\omega}\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}+\frac{\Phi_i}{\left|\epsilon_{\mathbf{k},\omega}\right|^2}\omega\frac{\partial}{\partial\omega}\frac{1}{\omega}\frac{\partial}{\partial t}\mathrm{Re}\epsilon_{\mathbf{k},\omega}\right]\right\}.$$

The following equalities were used to derive these relations:

$$\mathrm{Re}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\frac{\partial\epsilon^{(0)}_{\mathbf{k},\omega}}{\partial t}\right) = -\frac{1}{2}\left|\epsilon^{(0)}_{\mathbf{k},\omega}\right|^2\frac{\partial}{\partial t}\frac{1}{\left|\epsilon^{(0)}_{\mathbf{k},\omega}\right|^2}, \quad (45)$$

$$\mathrm{Im}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\right)\mathrm{Im}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\frac{\partial\epsilon^{(0)}_{\mathbf{k},\omega}}{\partial t}\right)$$
$$=\frac{1}{2\left|\epsilon^{(0)}_{\mathbf{k},\omega}\right|^2}\frac{\partial\mathrm{Re}(\epsilon^{(0)}_{\mathbf{k},\omega})}{\partial t}+\frac{1}{2}\frac{\partial}{\partial t}\mathrm{Re}\left(\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\right). \quad (46)$$

The expression in the first square brackets of (44) can be transformed to

$$\mathrm{Im}\left(\frac{1}{\epsilon_{\mathbf{k},\omega}}\frac{\partial}{\partial\omega}\frac{1}{\omega-\mathbf{k}\cdot\mathbf{v}+i0}\right)$$
$$-\mathrm{Re}\left(\frac{1}{\omega}\right)\mathrm{Im}\left(\frac{1}{\epsilon_{\mathbf{k},\omega}}\frac{1}{\omega-\mathbf{k}\cdot\mathbf{v}+i0}\right). \quad (47)$$

After integration over frequencies, the first term in (47) becomes zero. To make sure that this is the case, we must factor the signs of the imaginary part and time integration outside the sign of the frequency integration and take into account the fact that the derivative of $1/(\omega - \mathbf{k}\cdot\mathbf{v} + i0)$ with respect to frequency has no poles in the upper half-plane of complex ω, just like $1/\epsilon^{(0)}_{\mathbf{k},\omega}$ (the latter due to the analytic properties of permittivity). The second term in (47) transforms to an expression without any poles in the upper half-plane of complex ω by adding and subtracting the corresponding expression with $\mathrm{Im}(1/\omega + i0) = -\pi\delta(\omega)$. Thus, the first square bracket in (44) transforms to

$$\int d\omega\pi\delta(\omega)\mathrm{Re}\frac{1}{(\mathbf{k}\cdot\mathbf{v})\epsilon_{\mathbf{k},0}} = \int d\omega\frac{\pi}{\omega\epsilon_{\mathbf{k},0}}\delta(\omega-\mathbf{k}\cdot\mathbf{v}). \quad (48)$$

Relation (48) is used only to write the final result in compact form containing $\delta(\omega - \mathbf{k}\cdot\mathbf{v})$ alone. Thus,

$$I_i = -\frac{Z_i^2}{4\pi^2}\frac{\partial}{\partial\mathbf{p}}\cdot\int\frac{\mathbf{k}}{k^2}d\omega d\mathbf{k}\delta(\omega-\mathbf{k}\cdot\mathbf{v}) \quad (49)$$
$$\times\left\{\frac{\partial\Phi_i}{\partial t}\frac{\partial}{\partial\omega}\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}-\frac{\Phi_i}{\omega}\frac{\partial}{\partial t}\frac{1}{\epsilon^{(0)}_{\mathbf{k},0}}+\frac{\Phi_i}{\left|\epsilon^{(0)}_{\mathbf{k},\omega}\right|^2}\omega\frac{\partial}{\partial\omega}\frac{1}{\omega}\frac{\partial}{\partial t}\mathrm{Re}(\epsilon^{(0)}_{\mathbf{k},\omega})\right\}.$$

6. RENORMALIZATION OF THE PARTICLE DISTRIBUTIONS

Just like the correlation effects, the effects due to the time evolution of fluctuations can be reduced to an effective change in the nuclear reaction probability. Note that the corresponding equation (33), which includes the effects of time evolution with the correla-

tions ignored, differs in form from the equation that is commonly used to describe nuclear reactions without plasma corrections. More specifically, the equation with plasma corrections is

$$\frac{\partial \Phi_i}{\partial t} = I_i^{(0)} + I_i^{(t)} - \int w_{ij}\Phi_i\Phi_j\frac{d\mathbf{p}'}{(2\pi)^3}, \quad (50)$$

while the equation without plasma corrections is

$$\frac{\partial \Phi_i}{\partial t} = -\int w_{ij}\Phi_i\Phi_j\frac{d\mathbf{p}'}{(2\pi)^3}. \quad (51)$$

Here, it is worth recalling that we restrict our analysis only to the first-order plasma corrections. With this accuracy, Eq. (50) can be reduced to the standard form by renormalizing the particle distribution function. Consider the problem of reducing Eq. (50) to form (51) by renormalizing the particle distribution and by introducing an effective probability. We introduce the renormalized distribution function $\Phi_i^{(R)}$ as a solution to the equation

$$\frac{\partial \Phi_i}{\partial t} - I_i^{(0)} - I_i^{(t)} = \frac{\partial \Phi_i^R}{\partial t}. \quad (52)$$

Assume that the solution to Eq. (52) is

$$\Phi_i = (1 + \Lambda_{i,\mathbf{p}}^R)\Phi_i^{(R)}. \quad (53)$$

Given that the corrections are small and that both distribution functions of reacting nuclei are renormalized, the final equation with correlation corrections in the standard form is

$$\frac{\partial \Phi_i^{(R)}}{\partial t} = -\int w_{ij}^{\text{eff}}\Phi_i^{(R)}\Phi_j^{(R)}\frac{d\mathbf{p}'}{(2\pi)^3}, \quad (54)$$

where

$$w_{ij}^{\text{eff}} = w_{ij}(1 + \Lambda_{ij}^{(S)} + \Lambda_{i,\mathbf{p}}^{(R)} + \Lambda_{j,\mathbf{p}'}^{(R)}). \quad (55)$$

When solving the equation for Φ_i^R, it should be borne in mind that the rate of particle redistribution in momenta is much higher than the nuclear reaction rate. Therefore, only the particle density depends on time, and the distribution function is a product of the time-dependent density and the momentum distribution. The time derivative of the permittivity in I_i is determined by all reacting nuclei and by the rates of change in their distribution functions with time, i.e., by the time derivatives of their densities. The dependence on momentum in the particle distribution is the same as that for ordinary permittivity.

Below, we consider the nuclear cycle (in applications to the solar interior, the hydrogen cycle) and deal with the asymptotic behavior of the system at large times when the rates of all reactions have already steadied and are determined by the slowest cycle reaction. By the time the reaction rates are equalized, the renormalized distribution has also steadied. As will be seen from the subsequent analysis, it describes the collective energy shift in the distribution of each type of nuclei. Since the rates of all reactions coincide in the asymptotic limit, the relative time derivatives in Eq. (50) are equal, which allows the renormalization coefficient to be easily determined. In this case, it should be kept in mind that we consider small corrections and there is no need to make a distinction between the renormalized and nonrenormalized distributions in the renormalization term. The expression for $\Lambda_i^{(R)}$ derived in this way is

$$\Lambda_{i,\mathbf{p}}^{(R)} = -\frac{Z_i^2e^2}{4\pi^2}\int\frac{\mathbf{k}d\mathbf{k}d\omega}{k^2}\left(\frac{\partial}{\partial\mathbf{p}} - \frac{\mathbf{v}}{T}\right)\delta(\omega - \mathbf{k}\cdot\mathbf{v}) \times \mathrm{Re}\left\{\frac{\partial}{\partial\omega}\frac{1}{\epsilon_{\mathbf{k},\omega}^{(0)}} + \frac{1}{\omega(\epsilon_{\mathbf{k},0}^{(0)})^2}(\tilde{\epsilon}_{\mathbf{k},0} - 1) + \frac{1}{|\epsilon_{\mathbf{k},\omega}^{(0)}|^2}\omega\frac{\partial}{\partial\omega}\frac{1}{\omega}(\tilde{\epsilon}_{\mathbf{k},\omega}^{(0)} - 1)\right\}, \quad (56)$$

where

$$\tilde{\epsilon}_{\mathbf{k},\omega} = 1 + \frac{4\pi}{k^2} \times \sum_{\tilde{j}}\int\frac{Z_{\tilde{j}}^2e^2}{\omega - \mathbf{k}\cdot\mathbf{v}' + i0}\left(\mathbf{k}\cdot\frac{\partial\Phi_{\tilde{j}}}{\partial\mathbf{p}'}\right)\frac{d\mathbf{p}'}{(2\pi)^3}. \quad (57)$$

Here, the summation is performed only over the nuclei $\tilde{j}$ involved in the nuclear cycle under consideration and over the electrons involved in the nuclear reactions (as is the case for the hydrogen cycle).

It should be emphasized that the probability w_{ij} depends on the relative nuclear energy E_r alone and has a sharp maximum near the Gamov energy E^G in the limit $E_r \gg T$. At the same time, the effective probability (55) also depends on the momentum of each of the reacting particles, i.e., on the particle velocities relative to the medium (plasma), which is natural enough for collective processes.

7. CHANGE IN THE COLLECTIVE ELECTROSTATIC ENERGY OF NUCLEI AND THE ENERGY SHIFT

Each nucleus i in plasma is surrounded by a polarization screening charge and has an additional self-energy $E_i^{(s)}$. The latter can be calculated only in a nondissipative medium if the imaginary part of the plasma permittivity may be disregarded. This does not correspond to our problem, because dissipative processes play a significant role. However, in the asymptotically steady state, it is possible to determine the rate of change in the total electrostatic energy of the system

per nucleus with time. This energy is proportional to the square of the nuclear charge and acts as an effective self-energy of the reacting nuclei. In a large system, the polarization charges around individual nuclei are produced by collective fluctuations of the remaining nuclei. In the presence of interacting and reacting nuclei in the system, when the associated dissipative processes do not allow the electrostatic self-energy of the nuclei to be unambiguously introduced, the change in the total electrostatic energy of the system cannot be equal to the sum of the self-energies of individual nuclei. However, after the rates of all reactions are equalized, the renormalization corrections reduce to an energy shift in their distribution and, in this sense, act as some effective self-energy. Before the rate equalization, this physical interpretation is not possible and the emerging corrections cannot be mathematically represented as some effective energy shift of individual nuclei.

The derived effective self-energy is not equal to the sum of the electrostatic self-energies of individual particles in nondissipative systems. The effective electrostatic self-energy of a single nucleus naturally depends on its velocity relative to the medium (plasma). The corrections introduced by the renormalization of the particle distribution per nucleus are similar to the electrostatic self-energy of individual nuclei, but their structure, magnitude, and sign are different. In contrast to the self-energy, the complete integral of the effective energy shift over momenta is zero. The sign and magnitude of the corrections are determined by the nuclear reactions, which show what nuclear energies and momenta give the largest contribution to the renormalizations.

In a dissipative system, we can analyze the rate of change in the total electrostatic energy of the system and can show that the derived renormalization corrections are actually determined by this rate. For the rate of change in the total electrostatic energy of the system with time, we have

$$\frac{dW^s}{dt} = \frac{1}{4\pi}\left\langle \mathbf{E}\cdot\frac{d\mathbf{D}}{dt}\right\rangle = \int\sum_i Z_i e\langle \delta f_{i,\mathbf{k},\omega}(-i\omega')\phi_{\mathbf{k}',\omega'}\rangle \times \exp[-i(\omega+\omega')t]d\mathbf{k}\mathbf{k}'d\omega d\omega'. \quad (58)$$

This relation should be compared with the first term on the right-hand side of Eq. (33). The change in particle energy can be calculated by multiplying the left-hand side of Eq. (33) by the energy of a single particle and by integration over the particle distribution. When integrating by parts, the derivative with respect to momenta leads to the factor $\mathbf{k}\cdot\mathbf{v}$ and, in view of (40) and (41), the product $\mathbf{k}\cdot\mathbf{v}$ can be substituted by ω. Thus, it is easy to see that (42) and (43) correspond to such changes in the particle distributions that describe the changes in the total electrostatic energy of the system due to its evolution with time through nuclear reactions. That is why the collective corrections to each nuclear reaction depend on the rates of other reactions of the entire cycle.

The renormalization terms described above can be easily interpreted as the appearance of an energy shift $\delta\epsilon_{i,\mathbf{p}}$ or as an addition to the nuclear energy,

$$\epsilon_{i,\mathbf{p}} = \frac{p^2}{2m_i}\Phi_i^R(\mathbf{p}) \propto \exp\left(-\frac{\epsilon_{i,\mathbf{p}}}{T}-\frac{\delta\epsilon_{i,\mathbf{p}}}{T}\right) \approx \exp\left(-\frac{\epsilon_{i,\mathbf{p}}}{T}\right)\left(1-\frac{\delta\epsilon_{i,\mathbf{p}}}{T}\right),$$

where $\epsilon_{i,\mathbf{p}}$ is the energy of particles i and $\delta\epsilon_{i,\mathbf{p}}$ is the collective energy shift of particles i. Given that $\Phi_i^R = \Phi_i(1-\Lambda_{i,\mathbf{p}}^{(R)})$, we obtain

$$\Lambda_{i,\mathbf{p}}^{(R)} = \frac{\delta\epsilon_{i,\mathbf{p}}-\delta\epsilon_i^{(0)}}{T}, \qquad \delta\epsilon_i^{(0)} = \frac{\int(\delta\epsilon_{i,\mathbf{p}}/T)\exp(-\epsilon_\mathbf{p}/T)d\mathbf{p}}{\int\exp(-\epsilon_\mathbf{p}/T)d\mathbf{p}}, \quad (59)$$

where $\delta\epsilon_i^{(0)}$ arises from the normalization. Relation (59) only assumes that the distribution of reacting nuclei remained thermal. Therefore, the corrections are attributable only to the appearance of an energy shift in the nuclei.

The integral of (56) over the thermal particle distribution is zero, implying that

$$\frac{\delta\epsilon_{i\mathbf{p}}}{T} = \Lambda_{i,\mathbf{p}}^{(R)}. \quad (60)$$

Expression (60) differs from that used in [10, 12], where we took into account the time evolution of the average distribution but ignored the time evolution of fluctuations and the collective change in distribution due to nuclear reactions, which lead to expressions (49) and (56). The latter are more general than the relations derived in [10, 12], because they include the change in fluctuations with time through nuclear reactions. For comparison with (56), we give the expression for the self-energy of a nondissipative system from [10, 12]:

$$\frac{E_{i,\mathbf{p}}^s}{T} = \frac{Z_i^2e^2}{4T\pi^2} \times\int d\mathbf{k}d\omega\delta(\omega-\mathbf{k}\cdot\mathbf{v})\omega^2\frac{\partial}{\partial\omega}\frac{1}{\omega\mathrm{Re}(\epsilon_{\mathbf{k},\omega}^{(0)})}. \quad (61)$$

In contrast to (57), the integral (61) over momenta is nonzero. The collective energy shift (60) clearly depends only on the magnitude of the particle velocity, i.e., on the particle energy. Therefore, it must change sign at a certain energy, because the integral over momenta becomes zero. In the integration with a weight determined by the nuclear reaction rate w_{ij}, the

relative energy is close to the Gamov energy and the translational energy is of the order of the thermal energy. However, angular averaging is also important in determining the effective sign of the energy shift.

Since the particle distribution is assumed to be thermal, the corrections can be interpreted only as an effective energy shift. The energy shift described above is collective and, to a certain measure, is an analog of the Lamb shift. The shift itself acts as an effective renormalized nuclear self-energy.

8. TRANSFORMATION OF THE CORRECTIONS

The ratio of the nuclear reaction rate R_{ij} with collective plasma corrections to their rate $R_{ij}^{(0)}$ in the absence of plasma is

$$\frac{R_{ij}}{R_{ij}^{(0)}} = \frac{\int w_{ij}^{\text{eff}}\Phi_i(\mathbf{p})\Phi_j(\mathbf{p}')d\mathbf{p}d\mathbf{p}'}{\int w_{ij}\Phi_i(\mathbf{p})\Phi_j(\mathbf{p}')d\mathbf{p}d\mathbf{p}'} = 1 + \Lambda^{(C)} + \Lambda^{(T)}, \quad (62)$$

where $\Lambda^{(C)}$ describes the correlation corrections and $\Lambda^{(T)}$ describes the corrections due to time evolution (the subscript T points to the time evolution). As a momentum-independent quantity, $\Lambda^{(C)}$ matches the previously derived expression (11), while $\Lambda^{(T)}$ is given by

$$\Lambda^{(T)} = \left[\int w_{ij}(\Lambda_{i,\mathbf{p}}^{(R)} + \Lambda_{j,\mathbf{p}'}^{(R)})\Phi_i(\mathbf{p})\Phi_j(\mathbf{p}')d\mathbf{p}d\mathbf{p}'\right] \times \left[\int w_{ij}\Phi_i(\mathbf{p})\Phi_j(\mathbf{p}')d\mathbf{p}d\mathbf{p}'\right]^{-1}. \quad (63)$$

By introducing the relative velocity $\mathbf{v}_r = \mathbf{v} - \mathbf{v}'$ and the center-of-mass velocity

$$\mathbf{V} = \frac{m_i\mathbf{v} + m_j\mathbf{v}'}{m_i + m_j}$$

of two reacting nuclei, expression (63) may be treated as an averaging over the relative motion and the center-of-mass motion:

$$\Lambda^{(T)} = \left[\int w_{ij}(\Lambda_{i,\mathbf{p}}^{(R)} + \Lambda_{j,\mathbf{p}'}^{(R)}) \times \exp\left[-\frac{E_r}{T} - \frac{V^2}{2(m_i + m_j)T}\right]d\mathbf{v}_\mathbf{r}d\mathbf{V}\right] \times \left[\int w_{ij}\exp\left[-\frac{E_r}{T} - \frac{V^2}{2(m_i + m_j)T}\right]d\mathbf{v}_\mathbf{r}d\mathbf{V}\right]^{-1}. \quad (64)$$

Depending on the relative energy $E_r = \mu_{ij}v_r^2/2$, the integrands have a sharp maximum near the Gamov energy $E_{ij}^G = \mu_{ij}(v_{ij}^G)^2/2$, where

$$\mu_{ij} = \frac{m_i m_j}{m_i + m_j}$$

is the reduced mass. Therefore, the integration over the relative energy reduces to its substitution by the Gamov energy. However, the sum $\Lambda_{i,\mathbf{p}}^{(R)} + \Lambda_{j,\mathbf{p}'}^{(R)}$ depends on the relative plasma particle momenta. Consequently, apart from the absolute values of the relative and translational energies, it depends on angles. Relation (64) contains the corresponding angular averaging. After separating out the angular dependence, we may perform integration over the absolute value of k by taking into account the fact that the $w/kv_{T\alpha}$ ratio ($v_{T\alpha}$ is the thermal velocity of particles α) does not depend on the absolute value of k in view of $\omega = \mathbf{k}\cdot\mathbf{v}$.

Simultaneously with the separation of the angular dependence, the corrections must be transformed to a form containing the relative mass fractions of various nuclei to explicitly calculate the corrections for actual applications to a plasma that is a mixture of various reacting nuclei. We assume the ions to be completely ionized, i.e., to be bare nuclei with their charges offset by free plasma electrons:

$$n_e = \sum_i Z_i n_i.$$

We use the permittivity in the form

$$\epsilon_{\mathbf{k},\omega}^{(0)} = 1 + \frac{1}{k^2d_e^2} + \sum_i \frac{1}{k^2d_i^2}W(s_i) = 1 + \frac{1}{k^2d^2(1 + Z_{\text{eff}})}\left(1 + \frac{\sum_i Z_i^2 n_i W(s_i)}{\sum_i Z_i n_i}\right), \quad (65)$$

where the summation is performed only over the ions and their velocity is assumed to be much lower than the thermal electron velocity. Therefore, the Debye screening approximation [the second term in the first equality in (65)] is used for the electron response. Here, d is the total Debye length and $W(s)$ is the standard plasma dispersion function,

$$\frac{1}{d^2} = \frac{1}{d_e^2} + \sum_i \frac{1}{d_i^2}, \qquad W(s) = 1 + s\exp(-s^2)\left(i\sqrt{\pi} - 2\int_0^s \exp t^2 dt\right) \quad (66)$$

and

$$s_i = \frac{\omega}{\sqrt{2}k v_{Ti}}, \quad 1 + Z_{\text{eff}} = \frac{d_e^2}{d^2},$$
$$Z_{\text{eff}} = \frac{\sum_i Z_i^2 n_i}{\sum_i Z_i n_i}. \tag{67}$$

Thus, both s_i and $W(s_i)$ depend on the angular variables alone. If

$$X_i = \frac{m_i n_i}{\sum_j m_j n_i}$$

describes the relative mass density of the nuclei, then

$$\epsilon^{(0)}_{\mathbf{k},\omega} = 1 + \frac{1}{k^2 d^2}\left(1 + \frac{\sum_i \frac{Z_i^2 X_i}{m_i}}{\sum_i \frac{Z_i X_i}{m_i}}\right)^{-1} \times \left(1 + \frac{\sum_i \frac{Z_i^2 X_i}{m_i} W(s_i)}{\sum_i \frac{Z_i X_i}{m_i}}\right). \tag{68}$$

Note that $\Lambda^{(R)}_{i,\mathbf{p}}$ contains sums over the types of ions j' of the functions that depend on k^2 and on the angular variables appearing in $s_{i,j,j'}$, while $\Lambda^{(R)}_{j,\mathbf{p}'}$ contains sums over the types of ions j' of the functions that depend on k^2 and on the angular variables appearing in $s_{i,j,j'}$, where

$$s_{i,j,j'} = \frac{\mathbf{k}\cdot\mathbf{v}}{k\sqrt{2}v_{Tj'}} = \sqrt{\frac{m_{j'}}{m_i + m_j}}\left(yx + \lambda_{ij} z \frac{m_j}{m_i + m_j}\right), \tag{69}$$

$$s_{j,i,j'} = \frac{\mathbf{k}\cdot\mathbf{v}'}{k\sqrt{2}v_{Tj'}} = \sqrt{\frac{m_{j'}}{m_i + m_j}}\left(yx - \lambda_{ij} z \frac{m_i}{m_i + m_j}\right). \tag{70}$$

Here, x is the cosine of the angle between vector $\mathbf{k}$ and the relative velocity $\mathbf{v}_r$ of the two reacting nuclei, z is the cosine of the angle between vector $\mathbf{k}$ and the translational velocity $\mathbf{V}$ of the two reacting nuclei, y is the normalized translational velocity, and λ_{ij} is the normalized velocity corresponding to the Gamov energy

$$y = \frac{V}{\sqrt{2T/(m_i + m_j)}}, \quad \lambda_{ij} = \frac{v^G_{ij}}{\sqrt{2T/(m_i + m_j)}}. \tag{71}$$

Using the relation

$$\omega\frac{\partial W(s)}{\partial\omega} = s\frac{dW(s)}{ds} = (1 - 2s^2)W(s) - 1 \tag{72}$$

and

$$\int d\omega \mathbf{k}\cdot\frac{\partial}{\partial\mathbf{p}}\delta(\omega - \mathbf{k}\cdot\mathbf{v})L_{\mathbf{k},\omega} = \int d\omega\delta(\omega - \mathbf{k}\cdot\mathbf{v})\frac{k^2 v_{Ti}^2}{T}\frac{\partial L_{\mathbf{k},\omega}}{\partial\omega}, \tag{73}$$

it is convenient to reduce Eq. (56) to a form containing only the operators $\omega\partial/\partial\omega$:

$$\Lambda^{(R)}_{i,\mathbf{p}} = \Lambda^{(T)}_{i,\mathbf{p}} + \tilde{\Lambda}^{(T)}_{i,\mathbf{p}}, \tag{74}$$

$$\Lambda^{(T)}_{i,\mathbf{p}} = \frac{Z_i^2 e^2}{4\pi^2 T} \times \int\frac{d\mathbf{k}}{k^2}\mathrm{Re}\left\{\left[1 + \frac{k^2 v_{Ti}^2}{\omega^2}\left(1 - \omega\frac{\partial}{\partial\omega}\right)\right]\omega\frac{\partial}{\partial\omega}\frac{1}{\epsilon^{(0)}_{\mathbf{k},\omega}}\right\}_{\omega = \mathbf{k}\cdot\mathbf{v}}, \tag{75}$$

$$\tilde{\Lambda}^{(T)}_{i,\mathbf{p}} = \frac{Z_i^2 e^2}{4\pi^2 T}\int\frac{d\mathbf{k}}{k^2}\mathrm{Re}\left\{\left(1 + \frac{k^2 v_{Ti}^2}{\omega^2}\right)\times\left[\frac{(\tilde{\epsilon}^{(0)}_{\mathbf{k},0} - 1)}{\left|\epsilon^{(0)}_{\mathbf{k},0}\right|^2} - \frac{\mathrm{Re}(\tilde{\epsilon}^{(0)}_{\mathbf{k},\omega} - 1)}{\left|\epsilon^{(0)}_{\mathbf{k},\omega}\right|^2}\right] + \frac{1}{\left|\epsilon^{(0)}_{\mathbf{k},\omega}\right|^2}\omega\frac{\partial}{\partial\omega}\mathrm{Re}(\tilde{\epsilon}^{(0)}_{\mathbf{k},\omega} - 1) + \frac{k^2 v_{Ti}^2}{\omega^2}\left(\omega\frac{\partial}{\partial\omega}\right)\times\left[2\frac{\mathrm{Re}(\tilde{\epsilon}^{(0)}_{\mathbf{k},\omega} - 1)}{\left|\epsilon^{(0)}_{\mathbf{k},\omega}\right|^2} - \frac{1}{\left|\epsilon^{(0)}_{\mathbf{k},\omega}\right|^2}\omega\frac{\partial}{\partial\omega}\mathrm{Re}(\tilde{\epsilon}^{(0)}_{\mathbf{k},\omega} - 1)\right]\right\}_{\omega = \mathbf{k}\cdot\mathbf{v}}. \tag{76}$$

The integration over k in (76) can be performed analytically, and finding the collective corrections reduces to averaging over y with the form factor

$$\frac{4}{\sqrt{\pi}}\int_0^\infty \ldots y^2\exp(-y^2)dy,$$

Table

No.	Reaction	$\lambda_{i,j}$	$\Lambda_{ij,N}^{(S)}$	$\Lambda_{ij,N}^{(C)}$	$\Lambda_{ij,N}^{(T)}$	$\Lambda_{ij,N}^{(C)} + \Lambda_{ij,N}^{(T)}$
1	$p + p$	4.280	1	1	0.357	1.357
2	$p + {}^2\mathrm{H}$	4.757	1	1	0.313	1.313
3	${}^3\mathrm{He} + {}^3\mathrm{He}$	8.150	4	4	0.98	4.98
4	${}^3\mathrm{He} + {}^4\mathrm{He}$	8.420	4	4	0.943	4.943
5	${}^7\mathrm{Li} + p$	10.234	3	3	0.676	3.676
6	${}^7\mathrm{Be} + p$	11.264	4	4	1.056	5.056
7	${}^7\mathrm{Be} + e$	0	0	–4	0.788	–3.212

over x with

$$(1/2)\int_{-1}^{1}\ldots dx,$$

and over z with

$$(1/2)\int_{-1}^{1}\ldots dz.$$

Using (66), we finally obtain

$$\Lambda^{(T)}(y,x,z) = -\frac{Z_i^2 e^2}{4\sqrt{\pi}Td\sqrt{1+Z_{\mathrm{eff}}}}\int_0^\infty y^2 e^{-y^2}dy \times \int_{-1}^{1}dx\int_{-1}^{1}dz\,\mathrm{Re}\{\Lambda_{ij}^{(T)} + \tilde{\Lambda}_{ij}^{(T)}\} + (i \longleftrightarrow j), \quad (77)$$

where $\Lambda_{ij}^{(T)}(y,x,z)$ and $\tilde{\Lambda}_{ij}^{(T)}(y,x,z)$ are determined by the averaging of (75) and (76), respectively; the expressions for $\Lambda_{ij}^{(T)}(y,x,z)$ and $\tilde{\Lambda}_{ij}^{(T)}(y,x,z)$ are given in the Appendix. Upon the change $i \longleftrightarrow j$, (A.1) and (A.2) will contain (70) instead of (69) and, accordingly, $s_{j,i,j}$ appears instead of $s_{i,j,i}$ and Z_j^2 appears instead of Z_i^2. Since the nuclear masses enter into the numerators and denominators of (A.1) and (A.2) with the same powers, the nuclear mass in (76) can be measured in proton masses; i.e., m in (69), (70) and (A.1), (A.2) corresponds to the atomic weight of the nuclei.

Relations (A.1) and (A.2) allow the rates of the nuclear reactions between nuclei i and j to be calculated for an arbitrary mixture of plasma nuclei. The summation over plasma ions j' includes both the reacting and nonreacting nuclei, while the summation over $\tilde{j}'$ includes only the reacting nuclei.

9. NUMERICAL RESULTS

Below, we present the results of our calculations for the central part of the Sun using the following parameters [3]: $X_{\mathrm{H}} = 0.3411$, $X_{\mathrm{He}} = 0.6387$, $X_{\mathrm{C}} = 0.00003$, $X_{\mathrm{N}} = 0.0063$, and $X_{\mathrm{O}} = 0.0085$. We have $Z_{\mathrm{eff}} = 2.551$ and $\sum_j Z_j X_j/m_j = 0.661$. The parameter e^2dT depends on temperature and density. For the temperature $T = 1.5$ keV and density $n = 5\times10^{25}$ cm^{-3} assumed in current solar models, we have $e^2/dT = 0.05$, which for the Salpeter corrections gives 5% for the p–p reactions (the beginning of the hydrogen cycle) and $4e^2/Td \approx 20\%$ for the reactions with Be (the end of the hydrogen cycle). In order not to be attached to specific temperature and density, it is appropriate to present the results of our numerical calculations for the corrections in units of e^2/Td for $\Lambda_{ij,N} \equiv \Lambda_{ij}Td/e^2$. The table gives the collective plasma corrections that were numerically calculated using expressions (77), (A.1), and (A.2). This table contains the assumed values for the relative Gamov energies λ_{ij}; the correlation corrections $\Lambda_{ij,N}^{(C)}$, $\Lambda_{ij,N}^{(T)}$, and $\Lambda_{ij,N}^{(S)}$; and the total corrections $\Lambda_{ij,N}^{(C)} + \Lambda_{ij,N}^{(T)}$.

10. DISCUSSION

Previously, it was assumed that all collective corrections were the Salpeter corrections and that they enhanced the rates of all reactions. As can be seen from the table, not all of the collective corrections lead to an increase in the reaction rates but only some of them. For example, the reaction rates decrease in the reaction with electron capture by ^{7}Be nuclei. The latter effect is important, because the reaction with ^{7}Be nuclei is a branching one (see the sixth column in the table). This branching has been and is still one of the outstanding problems for neutrino-producing reactions in the solar interior. The decrease in the rate of the reaction with electron capture by ^{7}Be nuclei causes a decrease in the number of B nuclei and in the number of neutrinos produced during their decay. For all the remaining hydrogen-cycle reactions, the reaction rates are larger than

those assumed previously. The net effect for these reactions is the sum of the suppression effect found in [10–12] and the effect due to time evolution through nuclear reactions with neutrinos freely leaving the system. The latter effect prevails, causing an enhancement of the reactions. On the other hand, the proven equality of the Salpeter corrections to zero is confirmed in that, if these corrections are nonzero, then their interference with the correlation corrections will lead to a result contradicting the observations. Indeed, in this case, the corrections are a factor of 4 larger than the Salpeter ones, i.e., about 80% for reactions 3 and 4 of the table, which is in conflict with solar seismology. The correlation effects lead to corrections that, to a certain extent, "restore" the result commonly used in standard solar models, but the total corrections given in the last column of the table differ from those used. Numerical calculations of solar models show that even relatively small corrections can appreciably affect the final parameters of an evolving nuclear system. The last row in the table indicates that the correlation corrections for the reactions of electron capture by ^{7}Be nuclei are significant and negative, with the zero Salpeter corrections giving a wrong answer. The correlation corrections for ^{7}Be nuclei must be taken into account when constructing solar models. Their inclusion is important in the existing problem of branching of nuclear reactions involving ^{7}Be nuclei.

Let us discuss the following question: With what accuracy can we disregard the terms that lead to zero Salpeter screening? This is determined by the possibility of disregarding the second term in Eq. (6). It describes the influence of the actually emerging, fluctuating electric fields on the thermonuclear reaction rates. The current theory of plasma fluctuations gives expressions for the latter effect that diverge at large k. If we use $k_{\max} \approx 1/r$ for our estimates, where r is the characteristic nuclear size, then we find that the effects related to fluctuating electric fields are a factor of r/R smaller than those taken into account here (R is the scale size of the fluctuating fields). If, alternatively, $k_{\max} \approx N_d/d$ (large angles in collisions), then the relative contribution of the second term in Eq. (6) is even smaller, being $\approx r^2 N_d/R^2$.

It is worth noting that the plasma density in laboratory laser experiments can be higher than that in the solar interior, while the temperatures are of the same order of magnitude. Then, e^2/dT is an order of magnitude larger than its value in the solar interior and the correlation corrections for the $D + T$ reactions are 50% instead of 5% in the solar interior. The total corrections calculated numerically by using the present theory are larger than the correlation corrections (which give 50%) by a factor of 1.194, which corresponds to an increase in the reaction rate to 51%.

According to our results, the physical meaning of the collective corrections is related to a change in the (correlation and time) distribution of nuclei but not to a change in the probabilities of nuclear reactions, as assumed previously. In this sense, the corrections are not specific to nuclear reactions and arise for any other reactions, in particular, chemical ones (however, for our numerical results to be applicable, their rates must be larger than the rate of fluctuations and the specific averaging results under opposite conditions will be different). This emphasizes a major physical difference between our calculated corrections and those that could be related to the screening of nuclear reactions taken into account previously. Thus, the paradox between dynamic and static screenings is resolved.

The zero approximation of weak Salpeter screening was assumed to be insufficient to describe the nuclear reactions in the solar interior and in stars at other evolutionary stages. The interpolation formulas that describe the nuclear reactions in the intermediate region between strong and weak screenings were commonly used for this purpose. It follows from our results that this approach is inapplicable. There is no screening, and the correlation effects must be described more accurately to derive the interpolation formulas. Here, we obtained the result for weak correlations by expansion in a small parameter, the number of particles within a Debye sphere. This parameter differs from that used previously for the Salpeter corrections—the ratio of thermal energy to Gamov energy. The correct interpolation formulas for the correlation corrections could be derived if the result for strong correlations were known. Although various methods were used to describe strong correlations in plasma, they are all based on several as yet unproven hypotheses. Even the problem of weak correlations, which we studied here, has not yet been completely solved. Using our results, we can propose a method for describing the correlation effects of the next order in parameter $1/N_d$, as well as formulate and analyze the corresponding equations. However, it is unlikely that the effects of strong correlations in general form, which could be used to derive the interpolation formulas, will be analyzed in detail in the foreseeable future. Thus, the accuracy of the currently available models for describing nuclear reactions during stellar evolution that use the Salpeter interpolation formulas is called into question.

ACKNOWLEDGMENTS

I express may gratitude to Prof. V.S. Berezinskiĭ and to all participants of the Gran Sasso workshop on solar neutrinos (Italy) for numerous discussions and for interest in this study. I am also grateful to Profs. K. Watanabe and T. Sato for the opportunity to perform more detailed numerical calculations of the plasma corrections to nuclear reactions during my visit to the NIFS (Tokyo, Japan).

APPENDIX

Our calculations yield the following expressions for $\Lambda_{ij}^{(T)}(y, x, z)$ and $\tilde{\Lambda}_{ij}^{(T)}(y, x, z)$, respectively:

$$\Lambda_{ij}^{(T)}(y, x, z) = \frac{\sum_{j'} \frac{Z_{j'}^2 X_{j'}}{m_{j'}} \left[\left(1 + \frac{3s_{i,j,j'}^2}{s_{i,j,i}^2} - 2s_{i,j,j'}^2 \left(1 + \frac{s_{i,j,j'}^2}{s_{i,j,i}^2} \right) \right) W(s_{i,j,j'}) - \left(1 + \frac{s_{i,j,j'}^2}{s_{i,j,i}^2} \right) \right]}{\sqrt{\sum_{j''} \frac{Z_{j''} X_{j''}}{m_{j''}} \left[\sum_{j'} \frac{Z_{j'} X_{j'}}{m_{j'}} + \sum_{j'} \frac{Z_{j'}^2 X_{j'}}{m_{j'}} W(s_{i,j,j'}) \right]}}$$
$$+ \frac{1}{4s_{i,j,i}^2} \frac{\left[\sum_{j'} \frac{Z_{j'}^2 X_{j'}}{m_j'} [(1 - 2s_{i,j,j'}^2) W(s_{i,j,j'}) - 1] \right]^2}{\sqrt{\sum_{j''} \frac{Z_{j''} X_{j''}}{m_{j''}} \left[\sum_{j'} \frac{Z_{j'} X_{j'}}{m_{j'}} + \sum_{j'} \frac{Z_{j'}^2 X_{j'}}{m_{j'}} W(s_{i,j,j'}) \right]^3}} \tag{A.1}$$

$$\tilde{\Lambda}_{ij}^{(T)}(y, x, z) = -\frac{1}{\sqrt{\sum_{j''} \frac{Z_{j''} X_{j''}}{m_{j''}}}} \left\{ \frac{\mathrm{Re}\left\{ \tilde{\sum_{\tilde{j}'}} \frac{Z_{\tilde{j}'}^2 X_{\tilde{j}'}}{m_{\tilde{j}'}} \left[\left(1 + \frac{3s_{i,j,\tilde{j}'}^2}{s_{i,j,i}^2} - 2s_{i,j,\tilde{j}'}^2 \left(1 + \frac{s_{i,j,\tilde{j}'}^2}{s_{i,j,i}^2} \right) \right) W(s_{i,j,\tilde{j}'}) - \left(1 + \frac{s_{i,j,\tilde{j}'}^2}{s_{i,j,i}^2} \right) \right] \right\}}{\mathrm{Re}\left\{ \sqrt{\sum_{j'} \frac{Z_{j'} X_{j'}}{m_{j'}} + \sum_{j'} \frac{Z_{j'}^2 X_{j'}}{m_{j'}} W(s_{i,j,j'})} \right\}} \right.$$
$$+ \frac{1}{2s_{i,j,i}^2} \mathrm{Re}\left\{ \frac{\mathrm{Re}\left\{ \tilde{\sum_{\tilde{j}'}} \frac{Z_{\tilde{j}'}^2 X_{\tilde{j}'}}{m_{\tilde{j}'}'} [(1 - 2s_{i,j,\tilde{j}'}^2) W(s_{i,j,\tilde{j}'}) - 1] \right\}}{\mathrm{Re}\left\{ \sqrt{\sum_{j'} \frac{Z_{j'} X_{j'}}{m_{j'}} + \sum_{j'} \frac{Z_{j'}^2 X_{j'}}{m_{j'}} W(s_{i,j,j'})} \right\}} + \frac{\sum_{j'} \frac{Z_{j'}^2 X_{j'}}{m_j'} [(1 - 2s_{i,j,j'}^2) W(s_{i,j,j'}) - 1]}{\sqrt{\sum_{j'} \frac{Z_{j'} X_{j'}}{m_{j'}} + \sum_{j'} \frac{Z_{j'}^2 X_{j'}}{m_{j'}} W(s_{i,j,j'})}} \right.$$
$$\left. \left. \times \frac{\mathrm{Re}\left\{ \tilde{\sum_{\tilde{j}'}} \frac{Z_{\tilde{j}'}^2 X_{\tilde{j}'}}{m_{\tilde{j}'}'} [(1 - 2s_{i,j,\tilde{j}'}^2) W(s_{i,j,\tilde{j}'}) - 1] \right\} - 2\mathrm{Re}\left\{ \tilde{\sum_{\tilde{j}'}} \frac{Z_{\tilde{j}'} X_{\tilde{j}'}}{m_{\tilde{j}'}} + \tilde{\sum_{\tilde{j}'}} \frac{Z_{\tilde{j}'}^2 X_{\tilde{j}'}}{m_{\tilde{j}'}} W(s_{i,j,\tilde{j}'}) \right\}}{2\left\{ \mathrm{Re}\left[\sqrt{\sum_{j'} \frac{Z_{j'} X_{j'}}{m_{j'}} + \sum_{j'} \frac{Z_{j'}^2 X_{j'}}{m_{j'}} W(s_{i,j,j'})} \right] \right\}} \right\} \right\} \tag{A.2}$$

$$- \mathrm{Re}\left\{ \left(1 + \frac{1}{2s_{i,j,i}^2} \right) \left[\frac{1 + \tilde{Z}_{\mathrm{eff}}}{\sqrt{1 + Z_{\mathrm{eff}}}} - \frac{\mathrm{Re}\left[\tilde{\sum_{\tilde{j}'}} \frac{Z_{\tilde{j}'} X_{j'}}{m_{\tilde{j}'}} + \tilde{\sum_{\tilde{j}'}} \frac{Z_{\tilde{j}'}^2 X_{\tilde{j}'}}{m_{\tilde{j}'}} W(s_{i,j,\tilde{j}'}) \right]}{\sqrt{\sum_{j''} \frac{Z_{j''} X_{j''}}{m_{j''}} \left[\sum_{j'} \frac{Z_{j'} X_{j'}}{m_{j'}} + \sum_{j'} \frac{Z_{j'}^2 X_{j'}}{m_{j'}} W(s_{i,j,j'}) \right]}} \right] \right\}.$$

Here, as above, the subscript $\tilde{j}$ allows only for the nuclei that are involved in the nuclear reactions and

$$\tilde{Z}_{\text{eff}} = \frac{\tilde{\sum_{\tilde{j}}} Z_{\tilde{j}}^2 X_{\tilde{j}}/m_{\tilde{j}}}{\sum_j Z_j X_j/m_j}.$$

REFERENCES

1. E. E. Salpeter, Aust. J. Phys. **7**, 373 (1954).
2. D. D. Clayton, *Principles of Stellar Evolution and Nucleosynthesis* (McGraw-Hill, New York, 1968).
3. J. N. Bahcall, *Neutrino Astrophysics* (Cambridge Univ. Press, Cambridge, 1989).
4. S. Ichimaru, Rev. Mod. Phys. **65**, 255 (1993).
5. G. Shaviv and N. J. Shaviv, Phys. Rep. **311**, 99 (1999).
6. C. Carraro, A. Schafer, and S. E. Koonin, Astrophys. J. **331**, 565 (1988).
7. V. N. Tsytovich, *Lectures on Nonlinear Plasma Kinetics* (Springer-Verlag, Berlin, 1995).
8. L. S. Brown and R. F. Sawyer, Rev. Mod. Phys. **69**, 411 (1997).
9. J. Weneser, Phys. Rev. D **52**, 640 (1995).
10. V. N. Tsytovich and M. Bornatici, Comments Mod. Phys., Part C **2**, 1 (2000).
11. V. N. Tsytovich, Astron. Astrophys. Lett. **356**, L57 (2000).
12. V. N. Tsytovich and M. Bornatichi, Fiz. Plazmy **26**, 804 (2000) [Plasma Phys. Rep. **26**, 840 (2000)].
13. L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1515 (1964) [Sov. Phys. JETP **20**, 1018 (1964)].
14. E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics* (Fizmatgiz, Moscow, 1979; Pergamon, Oxford, 1981).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 943–965.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1098–1123.
Original Russian Text Copyright © 2002 by Shevkunov.

PLASMA, GASES

A Dense Hydrogen Plasma Modeled by the Path Integral–Monte Carlo Method

S. V. Shevkunov

St. Petersburg State Technical University, St. Petersburg, 195251 Russia
e-mail: root@shevk.spb.ru; root@shevk.hop.stu.neva.ru

Received October 30, 2001

Abstract—An approach to the exact description of exchange in disordered quantum systems at finite temperatures is formulated in terms of Feynman path integrals, which eliminates rigid restrictions on the number of particles and allows numerical simulation of the equilibrium characteristics of the electron component of a dense plasma to be performed by the Monte Carlo method. The combinatorial weight factors for all diagrams involving the linked Feynman paths for systems including up to 50 electrons were determined, and the corresponding distributions over spin states were found. Based on the data obtained, a mechanism responsible for the formation of a spin state of a many-electron system at finite temperatures is studied. The combinatorial weights of the diagrams grow rapidly on the passage from short to long cycles of linked paths. In the system at a finite temperature, this growth is compensated by a decrease in statistical weights of the diagrams with long cycles. The proposed path-integral Monte Carlo formalism was used for the numerical simulation of a dense hydrogen plasma under the conditions corresponding to the boundary of degeneracy of the electron component. The periodic cell contained 100 electrons and 100 protons. The results of numerical calculations show that, despite a high temperature, a dense plasma is far from a classical regime. Energetics of the plasma is formed at small interparticle distances, where the quantum character of the electron motion plays a determining role. The degree of ionization and the correlation functions are calculated, and the correlation radius is estimated. The formation of proton–electron pairs in the course of plasma cooling is studied in detail. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

During the time that has passed since the original publication of Metropolis [1], extensive development of computer simulation techniques has given rise to an independent direction in statistical mechanics which is engaged in solving problems recognized as hopeless within the framework of traditional approaches. Description of the thermal and structural properties of a dense, strongly nonideal plasma occupies a special place among such problems because this system has a number of important applications, ranging from metal processing and magnetogasdynamics to controlled thermonuclear fusion [2].

As the particle number density decreases, the degree of collectivization in a system of charged particles increases due to a rapid expansion of the Debye sphere [3], rather than drops as in a system with short-range interactions. The Debye screening proves to be the only factor preventing the thermodynamic potentials from divergence. Using the idea of Mayer [4], according to which the group integrals are expanded over the clusters linked by simpler bonds, with their subsequent rearrangement, it is possible to write expressions for the virial coefficients in terms of converging cluster integrals. Summing the Fourier images of ring diagrams with the aid of the convolution theorem leads to the Debye mean force potential of the interaction between ions. The other contributions exhibit convolution into prototype diagrams with Debye screening in the links, which provides for their convergence. These components represent high-order corrections (in powers of the plasma density) to the limiting Debye law. The Mayer ideas were subsequently developed. Salpeter [5] suggested a method for passing to high-connectivity diagrams in the expansion over density powers, and Meeron [6] obtained cluster expansions for the mean force potential and a binary distribution function, while Stell and Lebowitz [7] obtained such expansions for the Ursel function. Friedman [8] suggested a generalization of the Mayer theory to a non-pairwise potential of the interaction between ions.

Despite later achievements in the development of analytical approaches, the Mayer theory still remains an asymptotic theory of the plasma of extremely low densities and high temperatures. The most significant particular results for ion systems were obtained using the methods of integral equations—a traditional approach in the theory of fluids—rather than using virial expansions. Martynov [9] analytically solved the Bogolyubov equation for the binary distribution function in a superposition approximation [10]. Rasaiah and Friedman [11] considered a solution to the integral equations for a system of ions in the hyperchain approximation; Carley [12] employed the Percus–

1063-7761/02/9405-0943$22.00 © 2002 MAIK "Nauka/Interperiodica"

Yewick integral equation with the hyperchain approximation in order to obtain a binary distribution function. The results obtained in [11, 12] refer to a domain where the Coulomb interaction can be considered as a small perturbation relative to the short-range repulsion between ions. The corresponding perturbation theory developed by Henderson and Blum [13] leads to a high-temperature expansion of the free energy, the leading term of which includes the Debye contribution and the spherical model corrections [14].

Even the first applications of the Monte Carlo method [15] led to progress in the study of dense ion systems, up to a transition to the condensed state [16–23]. A comparison of the results obtained by the Monte Carlo method to the numerical solution of the Percus–Yewick integral equation and the results obtained within the framework of a spherical model in the hyperchain approximation [23] showed that the latter approximation gives the results that are the closest to those obtained using the Monte Carlo simulations. In a system of charged particles, even the gas phase features extensive cluster formation as reflected by the equilibrium properties of the plasma [24–27]. Traversing the point of minimum correlation radius in the gas phase is accompanied by sharp changes in the equilibrium thermodynamic characteristics of the ion system, resembling the second-order phase transition. The region in the vicinity of a critical point of the ion plasma was exhaustively studied by Vorontsov-Velyaminov *et al.* [28–31]. Shiff [32–34] used the Monte Carlo method to study the nonsymmetric systems of classical charged particles. In the past decade, the interest of researchers in the ion plasma has not weakened. Wolf *et al.* [35] thoroughly studied the problem of taking into account the long-range interaction and screening; Caillol [36] developed the idea of replacing the periodic boundary conditions by modeling on a four-dimensional hypersphere surface. Investigations into the problem of a phase transition to the condensed phase were continued as well [37–44].

The formation of ion clusters was studied in detail by various researchers [45–56]. A broad region of stability of the ion chain clusters was found in the phase diagram of an electroneutral plasma. The ion systems with violated charge symmetry exhibit a reversed situation: the mean force potential of ion triplets favors their repulsion and hinders nucleation [53].

A considerable effort in the investigation of thermodynamic characteristics of classical charged particles led to the formation of a relatively complete pattern. On this background, the thermodynamic properties and structure of a dense ion–electron plasma offer a much less studied direction. Development of an effective computer simulation method for the study of quantum particles has begun only quite recently and has progressed rather slowly, encountering a number of basic difficulties. The approach to modeling quantum systems at finite temperatures is based on the Feynman representation of the quantum mechanics in terms of the path integrals [57]. Fosdick and Jordan [58, 59] applied the path integral Monte–Carlo (PIMC) method to modeling a simplest system of helium atoms. Attempts at simulating the electron plasma component were made in [60–63]. The main difficulty in the development of PIMC was related to the description of exchange in the electron system. The proposed approaches either referred to hypothetical "spinless fermions" [64–74] or completely ignored the exchange [75–78].

A principal requirement for the exact description of the permutation symmetry in systems at finite temperatures consists in constructing a complete set of antisymmetric wave functions. A simple symmetrization is insufficient for calculating the mean equilibrium values, while the trace of a statistical operator in the incomplete set representation is not a partition function [79]. Recently [80, 81], an exact procedure was suggested for constructing a complete set of wave functions and introducing spin into the PIMC formalism. Test calculations were performed for the electron shells of a hydrogen molecule [82, 83] and beryllium and lithium ions [84]; the spin state of an electron pair occurring in a microvoid (in the context of the problem of electrodes) was calculated in [85]. This study, which is a continuation of the previous publication [81], considers application of the ideas formulated there to the numerical calculation of the equilibrium properties of a dense hydrogen plasma.

2. THE PROBLEM OF DESCRIPTION OF THE PERMUTATION SYMMETRY IN A SYSTEM WITH A LARGE NUMBER OF QUANTUM PARTICLES

The complete wave function of a system of indistinguishable fermions must be antisymmetric with respect to simultaneous permutations of spin and coordinate variables. However, the coordinate part of the wave function in the general case is not antisymmetric: the symmetry type depends on the spin state. The coordinate part of the wave function becomes antisymmetric in the state with a maximum spin. This simplest particular case is usually treated in numerical statistical calculations [64–74], although the state of maximum spin usually corresponds to a relatively high energy and is not of much interest.

In contrast to the calculations of pure quantum states, the statistical description of a system implies summation over all accessible spin states and, hence, over all possible types of the permutation symmetry of the coordinate function. A basis set in the symmetrized space is constructed with the aid of the Young symmetry operators [86, 87]. Each particular Young's scheme generates its own basis of irreducible representation of the permutation group. Completeness of the combined basis set is confirmed by direct calculation of the total

dimensionality of all disjoint subspaces generated by all the possible Young's schemes [86].

It can be shown that the space of complete wave functions satisfying the requirement of permutation symmetry is exhausted by all direct products of the subspaces of spin and coordinate functions generated by the Young operators with dual (obtained by reflection from diagonal) schemes. For particles with the spin 1/2, each subspace of this kind corresponds to a particular eigenvalue of the square spin operator of the whole system [87].

Acting upon the wave function of an N-particle system, the Young operator converts this function into a linear combination including (depending on the spin state and the parity of N) from $2^{N/2}[(N/2)!]^2$ to $N!$ terms. The square of this yields the number of terms in the corresponding matrix elements of the statistical operator. For calculating the mean equilibrium values, the expression should be integrated with respect to all N variables. In the general case, a multidimensional integral cannot be calculated analytically, while a numerical realization of such calculations becomes impossible (because of the rapid growth of $N!$) even for a system of several fermions.

Previously [81], it was demonstrated that the set of diagrams of linked paths can be reduced to a much lower number without introducing approximations, after which the numerical integration becomes feasible. Upon such reduction, one can obtain the control tables of combinatorial weights and determine the distributions over spin states for each prototype diagram. However, the number of operations necessary for calculating the control tables grows faster than $N!$ and, hence, the calculation is realizable only for $N \le 10$. A method described below removes these restrictions.

A basic requirement posed upon any variant of the Monte Carlo method is that the modification must not introduce systematic errors. This requirement was taken into account in developing a variant of the method applicable to systems with large numbers of particles. Here, by saying "large number" we imply a situation excluding all the particle number limitations related to the description of exchange; only restrictions typical of the systems obeying classical statistics are retained. For modern computer facilities, the upper limit determined by these restrictions is on the order of several hundred to several thousand particles. The modeling of macroscopic systems requires using periodic boundary conditions so as to eliminate the problem of edge effects.

The action of the Young operator leads to the appearance of about $N!$ terms in the matrix element of the statistical operator, each of these terms representing a certain permutation of the coordinate variables. In terms of the path integral formalism, such a permutation corresponds to a certain structure of linked paths [81]. It would be senseless to attempt to calculate this linear combination numerically and, the more so, to integrate this combination in a system of, for example, 100 fermions: such a linear combination contains $100! \sim 10^{157}$ terms. The idea of an approach proposed in this paper consists in extending the concept of essential set to the procedure of symmetrization of the wave function. It is suggested to perform, parallel to wandering in the coordinate space, a Markov random walk process over elements of the set of all permutations generated by the Young symmetry operator. Although the combinatorial weights of all such permutations equal ± 1, the corresponding contributions to the partition function and canonical means strongly depend on a particular structure of linked paths generated by the given scheme.

The Markov random walk process in the combined space of the coordinate variables and various path linking modes (permutations) leads to an essential set selected from the vast number of terms contained in the Young symmetry operator. Permutations generating like structures of linked Feynman paths can be combined into classes, so that the random walk would proceed over such classes rather than over particular permutations. Various classes contain different numbers of elements and, hence, possess significantly different combinatorial weights. The essential set will involve primarily the most significant classes with respect to their combinatorial weights. It should be emphasized that less significant contributions are not rejected, but postponed. As the Markov random walk proceeds, the accumulated statistics involves the terms with smaller contributions; in the limit of an infinitely long process, the calculated quantum-mechanical mean values coincide with the corresponding exact values.

A change in the arrangement of variables over cells of the Young scheme leads onto to renumbering of the vertices of the diagram of linked Feynman paths of the quantum particles under consideration [81]. Since the results of integration of these diagrams is independent of the numbering of vertices, calculation of the canonical means reduces to the integration of the family of diagrams for any one arrangement of arguments in the Young scheme. Such an arrangement is conveniently selected in the form of the fundamental sequence, whereby the argument number coincides with the cell number [81].

The Young operator represents sequentially applied operators of pair symmetrization $(1 + \hat{n}_{ij})$ and antisymmetrization $(1 - \hat{n}_{ij})$ with respect to the argument number:

$$\hat{J}(S) = (1 - \hat{n}_a)(1 - \hat{n}_b)\ldots(1 + \hat{n}_v)(1 + \hat{n}_w). \quad (1)$$

Here, $\hat{n}_a$, $\hat{n}_b$, … are the operators of pair commutation over the numbers of arguments occurring in the same column of the Young scheme; $\hat{n}_v$, $\hat{n}_w$, … are the analogous operators for argument pairs occurring in the same row of the scheme [81]. After opening of the

parentheses, the product (1) acquires the following form:

$$\hat{J}(S) = \sum_n \alpha_S(n)\hat{N}_n, \tag{2}$$

where summation is performed over all $N!$ permutations of the argument numbers. The combinatorial weights $\alpha_S(n)$ of the permutations take one of three values: +1, –1, and 0.

If a permutation operator in the sum of Eq. (2) can be represented as the product

$$\hat{N}_n = \hat{N}_k\hat{N}_l\hat{N}_m, \tag{3}$$

where $\hat{N}_k$, $\hat{N}_l$ are the operators of argument permutation in the first and second columns of the Young scheme and $\hat{N}_m$ is the operator of argument permutation in the rows of this scheme, the corresponding weight can be written as

$$\alpha_S(n) = (-1)^{c(k)+c(l)},$$

where $c(k)$ and $c(l)$ are the parities of permutations $\hat{N}_k$ and $\hat{N}_l$, respectively. Any operator $\hat{N}_n$ which cannot be represented in the form (3) is in fact not contained in the symmetry operator $\hat{J}(S)$ corresponding to the eigenvalue of the square spin operator $S(S+1)$ and the fundamental arrangement of arguments over cells of the Young scheme. In this case, the corresponding combinatorial weight in Eq. (2) is $\alpha_S(n) = 0$. Equation (2) is essentially an expansion of the Young operator in terms of the unit vectors $\hat{N}_n$ in the operator space of all possible permutations, so that a vector of this space with the components

$$\{\alpha_S(n)\} = (\alpha_S(1), \alpha_S(2), \ldots, \alpha_S(N!))$$

uniquely determines the Young operator.

Previously [81], it was shown that the partition function of a system of N undistinguishable nonrelativistic fermions with the spin 1/2 can be written as

$$\begin{aligned} Z &= \frac{1}{N!}\sum_S (2S+1)\sum_n^S (-1)^{c(l)+c(k)} \\ &\times \int d^N r \left\langle \{\mathbf{r}_i\}\middle|\exp(-\beta\hat{H})\middle|\hat{N}_n\{\mathbf{r}_i\}\right\rangle \\ &= \frac{1}{N!}\sum_S (2S+1)\sum_n \alpha_S(n) \\ &\times \int d^N r \left\langle \{\mathbf{r}_i\}\middle|\exp(-\beta\hat{H})\middle|\hat{N}_n\{\mathbf{r}_i\}\right\rangle, \end{aligned} \tag{4}$$

where $\beta \equiv 1/k_B T$ is the reciprocal temperature and k_B is the Boltzmann constant; the symbol S over the summation sign indicates that summation is performed only over the permutation operators contained in the Young operator $\hat{J}(S)$. During the calculation of matrix elements in the integrand of (4), each permutation $\hat{N}_n$ generates a certain diagram of linked virtual numbered paths. The subsets of such diagrams differing only by numbering of the vertices form classes. These subsets can be combined with a common factor

$$\omega_S(\{\nu_i\}) = \sum_n^{\{\nu_i\}} \alpha_S(n), \tag{5}$$

and the summation over all $N!$ permutations can be replaced by summation over all classes with a prototype graph $\hat{\Pi}_{\{\nu_i\}}$ of the given class. Here, the multidimensional subscript $\{\nu_i\} = (\nu_1, \nu_2, \ldots, \nu_N)$ determines the diagram structure: ν_i is the number of cycles including i linked paths; the symbol $\{\nu_i\}$ over the summation sign in (5) indicates that summation is performed only over n values for which the operators $\hat{N}_n$ belong to the class with a prototype graph $\hat{\Pi}_{\{\nu_i\}}$ and enter into the symmetry operator $\hat{J}(S)$. In these terms,

$$\begin{aligned} Z &= \frac{1}{N!}\sum_{\{\nu_i\}} \omega(\{\nu_i\}) \\ &\times \int d^N \left\langle \{\mathbf{r}_i\}\middle|\exp(-\beta\hat{H})\middle|\hat{\Pi}_{\{\nu_i\}}\{\mathbf{r}_i\}\right\rangle, \end{aligned} \tag{6}$$

where

$$\omega(\{\nu_i\}) = \sum_S (2S+1)\omega_S(\{\nu_i\}).$$

The structure of expressions for the canonical mean values is analogous to that of the partition function (6).

It is possible to construct a procedure of wandering over particular permutations, but this pathway is not optimum since the Markov transitions between permutations belonging to the same class do not in fact improve the statistics. The random walk over permutations is expediently replaced by the random walk over permutation classes. In each class, a random (equiprobable) set of permutations is taken from the manifold of all permutations $\hat{N}_n$ upon checking for their simultaneous belonging to the given class $\{\nu_i\}$ and to the Young operator $\hat{J}(S)$. The permutation is given a combinatorial coefficient $\alpha_{S,\{\nu_i\}}(n) = \alpha_S(n)$ provided that both conditions are satisfied and $\alpha_{S,\{\nu_i\}}(n) = 0$ if at least one of the two conditions is violated. In each step of this random process, we obtain an unbiased estimate of the combinatorial coefficient $\omega_S(\{\nu_i\})$:

$$\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l) = \frac{N!}{l}\sum_{k=1}^{l} \alpha_{S,\{\nu_i\}}(n_k), \tag{7}$$

where l is the volume of set $\{n_k\}$. It is of basic importance that the approximate character of estimate (7) for $\omega_S(\{\nu_i\}, \{n_k\}, l)$ obtained using a finite set $\{n_k\}$ on a single step of the Markov process does not mean that the entire calculation procedure is approximate, provided that it is based on the combinatorial weights $\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l)$ rather than $\omega_S(\{\nu_i\})$. Indeed, expression (5) can be alternatively written as

$$\begin{aligned}\omega_S(\{\nu_i\}) &= \frac{1}{l(N!)^{l-1}}\sum_{\{n_k\}}\sum_{k=1}^{l^{\{\nu_i\}}}\alpha_S(n_k)\\ &= \frac{1}{l(N!)^{l-1}}\sum_{\{n_k\}}\sum_{k=1}^{l}\alpha_{S,\{\nu_i\}}(n_k)\\ &= \frac{1}{(N!)^{2l-1}}\sum_{\{n_k\}}\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l),\end{aligned} \tag{8}$$

where $\{n_k\} = n_1, n_2, \ldots, n_l$ is a set of l numbers from the complete series of integers from 1 to $N!$ (numbers in the set $\{n_k\}$ may be repeated). The symbol $\{n_k\}$ at the sum in (8) indicates that summation is performed over all possible sets; the coefficient $1/l(N!)^{l-1}$ takes into account the repeated terms. The last equality in (8) reduces calculation of the combinatorial coefficients $\omega_S(\{\nu_i\})$ to taking a sum over sets $\{n_k\}$ from the series of all possible permutations (instead of summing over permutations inside the class $\{\nu_i\}$ as in expression (5)). Upon substituting (8) into (6) and representing the matrix element in the integrand as a product of the matrix elements for which the analytical expressions are known [57], the partition function acquires the form

$$\begin{aligned}Z = \frac{1}{(N!)^{3l-1}}&\sum_{\{\nu_i\}}\sum_{\{n_k\}}\sum_{S}(2S+1)\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l)\\ &\times\int d^N r(1) d^N r(2)\ldots d^N r(M)\\ &\times\langle\{\mathbf{r}_i(M)\}|\exp\left(-\frac{\beta\hat{H}}{M}\right)|\{\mathbf{r}_i(1)\}\rangle\\ &\times\langle\{\mathbf{r}_i(1)\}|\exp\left(-\frac{\beta\hat{H}}{M}\right)|\{\mathbf{r}_i(2)\}\rangle\ldots\\ &\times\langle\{\mathbf{r}_i(M-1)\}|\exp\left(-\frac{\beta\hat{H}}{M}\right)|\hat{\Pi}_{\{\nu_i\}}\{\mathbf{r}_i(M)\}\rangle.\end{aligned} \tag{9}$$

Note that, despite being written in terms of coefficients $\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l)$ determined for finite sets, expression (9) is exact. Writing the partition function in the form of (9) allows the particle number limitations related to the exchange to be removed and the Markov random process for a system of several hundred or even several thousand particles to be constructed.

Each microstate of the Markov process based on the partition function (9) is determined as a combination of sets $X = (\{\mathbf{r}_i(j)\}, \{\nu_i\}, \{n_k\})$ representing coordinates of the vertices of broken (polygonal) paths $\{\{\mathbf{r}_i(j)\}$, the mode of linking nonnumbered paths into cycles $\{\nu_i\}$, and the sets of L permutations with numbers $\{n_k\}$ from all the possible $N!$ permutations. The values of matrix elements are determined by the set $(\{\mathbf{r}_i(j)\}, \{\nu_i\})$; the combinatorial coefficients are determined by the set $(\{\nu_i\}, \{n_k\})$ and are independent of $\{\mathbf{r}_i(j)\}$. According to the theory of Markov processes [15], the matrix of transition probabilities $p(X, Y)$ for a process with preset limiting distribution $\omega(X)$ is not uniquely determined. In order to obtain the preset limiting distribution, it is sufficient to satisfy the detailed balance condition

$$\omega(X)p(X, Y) = \omega(Y)p(Y, X) \tag{10}$$

and the ergodicity condition; the latter requires that a given microstate could be accessed from any other via a sequence of intermediate transitions with nonzero probabilities.

In the proposed method, the calculation algorithm consists in sequentially accomplishing transitions of five types:

1. Shifting one vertex of the polygonal path $\{\mathbf{r}_i(j)\}$, while retaining $\{\nu_i\}$ and $\{n_k\}$;

2. Shifting a path or a cycle of linked paths as a whole, while retaining $\{\nu_i\}$ and $\{n_k\}$;

3. Rotating a path or a cycle of linked paths as a whole, while retaining $\{\nu_i\}$ and $\{n_k\}$;

4. Forming a new cycle by linking two cycles or paths with simultaneously passing to a new set $\{n_k\}$ and changing positions of a part of the vertices in $\{\mathbf{r}_i(j)\}$;

5. Separating a cycle into two cycles or paths with simultaneously passing to a new set $\{n_k\}$ and changing positions of a part of the vertices in $\{\mathbf{r}_i(j)\}$.

The vertices, paths, and cycles to be shifted are selected in a random way. The probabilities of transitions are played using a random number generator according to the Metropolis algorithm [15]. Optimum values of the limiting spatial shift and rotation are determined by the program (with the aid of a special algorithmic feedback system) in the initial segment of the process. This segment is not involved in calculation of the equilibrium mean values. According to the algorithm, M steps of type 1 for one path are accompanied with one step of type 2 and three steps of type 3 (rotations about three Cartesian coordinate axes). Optimum frequencies of the steps of types 3 and 4 are also determined by the program in the initial segment of the random process. Besides quantum particles (electrons), the system contains heavy particles (protons) which are treated according to the standard Metropolis procedure of the classical particle simulation. All the Coulomb interactions of any particle with all others are explicitly taken into account: protons interact with each other and with all vertices of the paths of electrons; electrons

interact via vertices with identical numbers. In the regime of simulation with periodic boundary conditions, the interactions are taken into account by the closest image method [15].

The experience of numerical calculations showed that the frequency of transitions into new microstates via direct events of path linking and separation is usually too low for accumulating necessary statistics. For this reason, the probabilities of such transitions are selected using a special procedure. However, these probabilities still obey the detailed balance condition (10) that provides for the attainment of a limiting Gibbs distribution [15]. The path linking and separation events are played only for the pairs of vertices with identical numbers under the condition that a distance r from the separated vertex to the newly connected one is below certain fixed value $r_{\max}$. Linking via vertices with the numbers different from unity reduces to the exchange of path fragments. When r exceeds a certain fixed value $r_{\min}$, the linked vertices are brought closer to each other with simultaneous uniform stretching of the path over a fragment with the length $M_{\text{str}} \le M$. The optimum values of $r_{\max}$, $r_{\min}$, and M_{str} are established in the preliminary simulation stage using a special algorithmic feedback system.

The steps of types 4 and 5 lead to a change in the current values of combinatorial weights $\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l)$, which have to recalculated. The calculation of sum (7) is conducted as follows. First, the Young scheme of N empty cells is formed that corresponds to the given spin value S; the scheme comprises $j_1 = N/2 + S$ cells in the first column and $j_2 = N/2 - S$ cells in the second column. Then, numbers from 1 to N are equiprobably assigned to the cells with the aid of a random number generator. This assignment represents a particular permutation $\hat{N}_{n_k}$ selected randomly and equiprobably from $N!$ possible permutations. Finally, the combinatorial weight of the given permutation is determined by searching for and analyzing of the inverse permutation $\hat{N}_{n_k}^{-1}$ converting the given arrangement of numbers in cells of the Young scheme into the fundamental sequence. It was demonstrated previously [81] that symmetrization of any eigenfunction of the coordinate operator $f(\{\mathbf{r}_i\}; \{\mathbf{x}_i\})$ achieved by the action of a certain linear combination of the permutation operators $\hat{N}_n$ upon arguments $\{\mathbf{x}_i\}$ is equivalent to the action of the same linear combination of inverse operators $\hat{N}_n^{-1}$ on the eigenvalues $\{\mathbf{r}_i\}$. Therefore, the permutation operators $\hat{N}_n$ must act upon $\{\mathbf{r}_i\}$ in the reverse order: antisymmetrization with respect to numbers in the columns of the Young scheme, followed by symmetrization over the rows. This operator is referred to as the alternative Young operator.

Although the change of all permutations in the Young operator modifies the type of symmetry of the wave function, the values of matrix elements in the integrand of (6) remain unchanged. Indeed, the inverse operators correspond to the path diagrams with reoriented segments, but the orientations of segments do not influence the values of matrix elements of the statistical operator [81]. Owing to this invariance, there are two equivalent methods for calculating the weights $\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l)$. The difference is in the way of symmetrization with respect to $\{\mathbf{r}_i\}$ in the basis wave functions—with the Young operators or with the same operators in which all permutations are replaced by inverse ones. We will use the latter approach. In this case, the alternative Young operator contains a given permutation $\hat{N}_{n_k}$ provided that it has a structure of the type $\hat{N}_{n_k} = \hat{N}_m^{-1}\hat{N}_i^{-1}\hat{N}_k^{-1}$, where the first operator in the right-hand part produces permutation in the rows of the Young scheme and the latter two operators, in the first and second columns of this scheme. Therefore, the inverse permutation must possess a structure of the type

$$\hat{N}_{n_k}^{-1} = \hat{N}_k\hat{N}_i\hat{N}_m. \tag{11}$$

This is checked by sequentially commuting numbers in the rows of the Young scheme so as to transfer all the first j_1 numbers to the first column and the other j_2 numbers to the second column of the scheme. If this rearrangement is possible, the structure of the $\hat{N}_{n_k}^{-1}$ operator satisfies condition (11). Otherwise, permutation $\hat{N}_{n_k}$ is assigned the combinatorial weight

$$\alpha_{S,\{\nu_i\}}(n_k) = 0.$$

Once condition (11) is obeyed, the combinatorial weight of the $\hat{N}_{n_k}$ permutation is

$$\alpha_{S,\{\nu_i\}}(n_k) = (-1)^{c(k)+c(i)}.$$

Here, the parities $c(k)$ and $c(i)$ of permutations $\hat{N}_k$ and $\hat{N}_i$, respectively, are calculated as follows. Each permutation corresponds to a permutation graph, analogous to a diagram of linked Feynman paths. Any pair commutation leads either to the merge of two cycles into one or to the decay of one cycle into two; that is, the number of cycles in the graph always changes by unity. Therefore, the number of cycles $\gamma(k)$ in the graph is related to the number of pair commutations $c(k)$ contained in a given permutation $\hat{N}_k$ by the formula

$$(-1)^{N-c(k)} = (-1)^{\gamma(k)},$$

which allows the combinatorial coefficient of permutation $\hat{N}_{n_k}$ to be expressed via the numbers of cycles:

$$\alpha_{S,\{\nu_i\}}(n_k) = (-1)^{\gamma(k)+\gamma(i)}.$$

The numbers of cycles $\gamma(k)$ and $\gamma(i)$ in the permutations $\hat{N}_k$ and $\hat{N}_i$ are calculated numerically. For this purpose, one has to analyze the sequence of numbers in both columns of the Young scheme after accomplishing permutation $\hat{N}_m$. The presence of number i in the cell j corresponds to the path segment directed from the ith to jth vertex in the permutation graph. In this way, the graph structure is sequentially restored and the numbers of cycles for both permutations are determined. The entire procedure is repeated l times, and the sum in Eq. (7) is calculated. After accomplishing this calculation for all possible spin states $S = 0, \ldots, N/2$ (for even N) and $S = 1/2, \ldots, N/2$ (for odd N), the sum over spin states is taken with an allowance for degeneracy with respect to the eigenvalues of the spin projection operator:

$$\tilde{\omega}(\{\nu_i\}, \{n_k\}, l) = \sum_S (2S+1)\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l).$$

Unlike coefficients $\omega_S(\{\nu_i\})$, the values of $\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l)$ are only calculated in the course of the Markov random walk and not stored in the computer memory. As the set volume l increases, the latter values approach $\omega_S(\{\nu_i\})$ the faster, the greater N. In fact, essential information on $\omega_S(\{\nu_i\})$ is contained in a number of elements that is much smaller than $(N!)^l$—the number of elements in the right-hand part of Eq. (8). This circumstance makes the essential set method especially effective.

The probabilities of transitions between microstates $X = (\{\mathbf{r}_i(j)\}, \{\nu_i\}, \{n_k\})$ and $Y = (\{\mathbf{r}'_i(j)\}, \{\nu'_i\}, \{n'_k\})$ involved in the Markov process contain combinatorial weights in the form of ratios [15]

$$p(X, Y) = \min\left(\frac{|\tilde{\omega}(\{\nu'_i\}, \{n'_k\}, l)|}{|\tilde{\omega}(\{\nu_i\}, \{n_k\}, l)|}\frac{a(\{\mathbf{r}'_i(j)\}, \{\nu'_i\}, \beta)}{a(\{\mathbf{r}_i(j)\}, \{\nu_i\}, \beta)}, 1\right), \quad (12)$$

where $a(\{\mathbf{r}_i(j)\}, \{\nu_i\}, \beta)$ denotes the chain of high-temperature matrix elements of the density matrix operator in the integrand of Eq. (9).

As demonstrated above, constructing a fundamentally exact procedure does not require that $\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l)$ coincide with $\omega_S(\{\nu_i\})$. However, the relationship between the two quantities depending on the variation of l is of independent theoretical interest. A comparison of Eqs. (5) and (7) shows that the two expressions coincide for $l = N!$ and $\{n_k\} = 1, 2, 3, \ldots, N!$. It is naturally expected that $\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l) \longrightarrow \omega_S(\{\nu_i\})$ as $l \longrightarrow N!$ (uniformly or nonuniformly over $\{n_k\}$). The proof of this statement is omitted, but a reason for this behavior is evident, at least from the following fact. As the volume l of set $\{n_k\}$ increases, the relative fraction of the possible combinations containing repeated numbers in the sequence $\{n_k\}$ decreases; most of such combinations are close to the fundamental sequence $\{n_k\} = 1, 2, 3, \ldots$ for which Eqs. (5) and (7) exactly coincide.

From the standpoint of the efficacy of a calculation procedure, not only the fact of convergence $\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l) \longrightarrow \omega_S(\{\nu_i\})$ is important, but the rate of this process is significant as well. Only sufficiently rapid convergence would allow us to check for correctness of the numerical calculation of $\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l)$. Such a verification based on the identity (8) can only be performed for sufficiently small N and l because the right-hand part of this equation contains $(N!)^l l$ terms. The check implies a comparison of the numerical values of $\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l)$ and $\omega_S(\{\nu_i\})$ calculated using independent computer programs. This comparison is possible only for $N \le 10$, because the calculation of $\omega_S(\{\nu_i\})$ by direct summation over all possible permutations for greater N would require about 10^4–10^6 hours of medium-speed computer operation. Since the computation algorithm is independent of the particular N value, coincidence of the $\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l)$ and $\omega_S(\{\nu_i\})$ values for small N guarantees that the weights $\omega_S(\{\nu_i\})$ are correctly calculated using $\tilde{\omega}_S(\{\nu_i\}, \{n_k\}, l)$ for any N values (including those for which direct calculation is impossible).

In this study, such a verification was performed for various N values. The rate of convergence is illustrated by the data presented in Table 1. As can be seen, even a relatively small set of $l = 100$ provides for the weight factors differing from exact values by only 1–2%. In order to determine the value of one coefficient, it is necessary to take a set from all the possible Young's schemes: the set has a volume of $l(N/2 + 1)$ for even N and $l((N - 1)/2 + 1)$ for odd N. Since the exact calculation of equilibrium means does not require the weight factors to be calculated exactly in each step of the Markov process, there exists an optimum set volume l ensuring the obtainment of canonical means at a minimum computation time. An overly small l value hinders distribution of the computation effort between states in proportion to their contributions to the equilibrium means. On the contrary, overly large l values slow down accomplishing the procedure in each step. Apparently, the optimum l value decreases with growing N. The experience of calculations performed within the framework of this study shows that, for $N = 100$, the optimum set volume is about $l \approx 10$.

The data for $N = 10$ in Table 1 can illustrate the behavior of weight factors depending on the number of linked trajectories. As the table is traversed bottom to top, the number of linked paths increases and the cycle length grows. As can be seen, an increase in the cycle size is accompanied by the growth in absolute values of the combinatorial weights, with oscillations in their signs. The maximum weight is observed for a configuration where all electrons form a common cycle (the first row in Table 1).

Table 1. Generalized combinatorial weights $\omega'(\{\nu_i\}) = \omega(\{\nu_i\})/\omega(N00...0)$ of various modes of the distribution of nonnumbered virtual electron trajectories over cycles (ν_i is the number of cycles including i linked trajectories) in a system of $N = 10$ electrons: the results of calculations performed by method of random sampling in permutations with various set volumes l are compared to the exact values determined by direct summation for all spin states and permutations contained in the Young symmetry operators

	ν_i	$\omega'(\{\nu_i\})$			
		$l = 10$	$l = 100$	$l = 10^7$	Exact value
1	0000000001	-1.058×10^5	-1.101×10^5	-1.0086×10^5	-1.00800×10^5
2	1000000010	1.232×10^5	1.222×10^5	1.2097×10^5	1.20960×10^4
3	0100000100	6.867×10^4	6.212×10^4	6.3836×10^4	6.38400×10^4
4	2000000100	-8.001×10^4	-7.560×10^4	-7.3919×10^4	-7.39200×10^4
5	0010001000	6.048×10^4	4.752×10^4	4.8158×10^4	4.81600×10^4
6	1100001000	-7.344×10^4	-8.042×10^4	-7.8981×10^4	-7.89600×10^4
7	3000001000	3.624×10^4	3.125×10^4	3.0791×10^4	3.08000×10^4
8	0001010000	4.410×10^4	4.292×10^4	4.2057×10^4	4.20400×10^4
9	1010010000	-7.112×10^4	-6.128×10^4	-6.0707×10^4	-6.07200×10^4
10	0200010000	-2.268×10^4	-2.115×10^4	-2.1574×10^4	-2.15800×10^4
11	2100010000	4.368×10^4	4.726×10^4	5.0152×10^4	5.01600×10^4
12	4000010000	-1.190×10^4	-9.961×10^3	-9.8953×10^3	-9.90000×10^3
13	0000200000	1.956×10^4	2.109×10^4	2.0264×10^4	2.02560×10^4
14	1001100000	-5.897×10^4	-5.116×10^4	-5.4729×10^4	-5.47200×10^4
15	0110100000	-3.427×10^4	-3.331×10^4	-3.4329×10^4	-3.43200×10^4
16	2010100000	3.058×10^4	3.891×10^4	3.9758×10^4	3.97600×10^4
17	1200100000	2.848×10^4	2.838×10^4	2.8201×10^4	2.82000×10^4
18	3100100000	-2.050×10^4	-2.222×10^4	-2.2073×10^4	-2.20800×10^4
19	500010000	2.570×10^3	2.693×10^3	2.6480×10^3	2.64800×10^3
20	0102000000	-1.638×10^4	-1.405×10^4	-1.5994×10^4	-1.59900×10^4
21	2002000000	1.937×10^4	1.945×10^4	1.8667×10^4	1.86600×10^4
22	0021000000	-1.218×10^4	-1.422×10^4	-1.4105×10^4	-1.41067×10^4
23	1111000000	4.242×10^4	4.628×10^4	4.6498×10^4	4.64800×10^4
24	3011000000	-1.876×10^4	-1.708×10^4	-1.8265×10^4	-1.82667×10^4
25	0301000000	5.145×10^3	5.324×10^3	5.4820×10^3	5.48000×10^3
26	2201000000	-1.780×10^4	-1.983×10^4	-1.9325×10^4	-1.93200×10^4
27	4101000000	7.140×10^3	7.996×10^3	7.6964×10^3	7.70000×10^3
28	6001000000	-5.915×10^2	-6.475×10^2	-6.2656×10^2	-6.26667×10^2
29	1030000000	7.280×10^3	7.031×10^3	6.7988×10^3	6.80000×10^3
30	0220000000	7.420×10^3	7.322×10^3	7.3064×10^3	7.30667×10^3
31	2120000000	-1.582×10^4	-1.644×10^4	-1.7002×10^4	-1.70000×10^4
32	4020000000	3.827×10^3	3.309×10^3	3.4001×10^3	3.40000×10^3
33	1310000000	-8.400×10^3	-7.924×10^3	-8.0659×10^3	-8.06000×10^3
34	3210000000	1.057×10^4	9.388×10^3	9.5715×10^3	9.56667×10^3
35	5110000000	-2.450×10^3	-2.283×10^3	-2.3407×10^3	-2.34000×10^3
36	7010000000	1.773×10^2	1.4733×10^2	1.4003×10^2	1.40000×10^2
37	0500000000	-3.045×10^2	-2.648×10^2	-2.8274×10^2	-2.82500×10^2
38	2400000000	2.074×10^3	1.706×10^3	1.6962×10^3	1.69500×10^3
39	4300000000	-1.313×10^3	-1.383×10^3	-1.3752×10^3	-1.37500×10^3
40	6200000000	3.780×10^2	3.614×10^2	3.4666×10^2	3.46667×10^2
41	8100000000	-3.038×10^1	-3.209×10^1	-3.2515×10^1	-3.25000×10^1
42	10000000000	1.000	1.000	1.0000	1.00000

The combinatorial coefficients presented in Table 1 are independent of the temperature. On heating, the probability of forming large cycles decreases because the region of spatial localization of each path exhibits narrowing. The resulting distribution over cycles is a compromise between two opposite trends: the increase in absolute values of the combinatorial coefficients and the decrease in the temperature weight factors $a(\{\mathbf{r}_i(j)\}, \{\nu_i\}, \beta)$ for long cycles. As the temperature grows, the second tendency begins to prevail and shifts the equilibrium toward shorter cycles. The competition of the two trends takes place on the background of partial compensation of the positive and negative weight factors. The role of this compensation increases with a decrease in the temperature, which leads to a growth in the weight of configurations involving high-connectivity diagrams.

3. SPIN

Table 2 presents the distributions over spin states

$$\Gamma_{\{\nu_i\}}(S) = (2S+1)\frac{\omega_S(\{\nu_i\})}{\omega\{\nu_i\}}$$

determined for various modes of linking the Feynman trajectories $\{\nu_i\}$ in a system of ten electrons. Each linking mode bears contributions from all possible spin states of the system, which are proportional to the weights presented in the table taken from [81]. At first glance, the configurations with strongly linked paths (i.e., those formed at low temperatures) must possess a lower spin and, hence, the corresponding distributions with respect to the spin must exhibit higher probabilities for the states with lower S values. However, calculations reveal a more complicated pattern.

As can be seen in the first row of Table 2, which corresponds to the formation of a longest cycle, the probabilities of states with small S values are zero. Traversing the table from top to bottom, one can see how the paths separate and the low-spin states appear. In a configuration with fully separated paths (bottom row in Table 2), the distribution over spins exhibits a trivial pattern determined by degeneracy with respect to the eigenvalues of the spin projection operator ($\propto 2S+1$). This high-temperature state corresponds to a transition to the classical limit with complete absence of the exchange. When the temperature is varied, the relative contributions from various path linking modes to the average spin change as well. At high temperatures, only the state presented in the last row exists.

At low temperature, a decisive factor is the mutual compensation of contributions with opposite signs. The prevalence of one or another mode of path linking depends both on the Coulomb repulsion between electrons and on the external field geometry. The resultant spin (and, hence, the permutation symmetry) of the system is determined by a particular configuration of the applied field. In a spatially homogeneous system, the spin usually decreases with the temperature. In a strongly inhomogeneous system, the spin of the ground quantum state can be nonzero. A example of such behavior is offered by spins of the atomic electron shells, which are maximum for elements from the middle of the Periodic Table.

In order to study the temperature-induced evolution of the spin state of a many-electron system, it is necessary to order the diagrams of linked paths with respect to a certain factor which is common of the diagrams exhibiting a similar response of the statistical weights to the temperature variations. There is no universal solution to this problem, since the distribution of weights depends on the external field geometry, and a rather laborious statistical modeling has to be performed in each particular case. Nevertheless, we may study common features in the spin behavior in sufficiently homogeneous systems upon ordering the diagrams with respect to the parameter of connectivity. With neglect of oscillations in the sign of the weight factors, the distribution over configurations of the virtual electron trajectories is close to the statistical distribution of classical polymer chains.

The simplest quantitative measure of connectivity is the number of cycles in the diagram:

$$\gamma(\{\nu_i\}) = \sum_{i=1}^{N}\nu_i.$$

The connectivity parameter

$$\alpha_C(\{\nu_i\}) \equiv \frac{N-\gamma}{N-1}$$

acquires the value $\alpha_C = 0$ for the diagrams with $\gamma = N$ including no linked trajectories. On the contrary, a diagram of maximum connectivity represents a single cycle including all the virtual trajectories linked in series, in which case $\gamma = 1$ and $\alpha_C = 1$. Intermediate cases correspond to $0 < \alpha_C < 1$. The diagrams characterized by the same values of connectivity parameter represent connectivity classes, the number of which is equal to the number of particles N in the system.

The signs of the total combinatorial weights of various connectivity classes

$$\omega'(\alpha_C) = \sum_{\{\nu_i\}}{}'\omega'(\{\nu_i\}) \qquad (13)$$

depend on the parity of $N-\gamma$. The prime sign at the sum in (13) indicates that only diagrams possessing the same connectivity are added:

$$\alpha_C = \frac{N-\sum_{i=1}^{N}\nu_i}{N-1}.$$

Table 2. Generalized combinatorial weights of various spin states contained in each mode of the distribution $\{\nu_i\}$ of nonnumbered virtual electron trajectories over cycles for a system of $N = 10$ electrons (S is the eigenvalue of the square spin operator; normalization condition: $\sum_S \Gamma_{\{\nu_i\}}(S) = 1$

	$\{\nu_i\}$	$\Gamma_{\{\nu_i\}}(S)$					
		$S = 0$	$S = 1$	$S = 2$	$S = 3$	$S = 4$	$S = 5$
1	0000000001	0	0	0	0	−0.100000	1.100000
2	1000000010	0	0	0	−0.018519	0	1.018519
3	0100000100	0	0	−0.006579	0.019737	−0.098684	1.085526
4	2000000100	0	0	−0.005682	−0.017046	0.085227	0.937500
5	0010001000	0	−0.003322	0.006645	0	−0.099668	1.096346
6	1100001000	0	−0.003039	0	0	0	1.003039
7	3000001000	0	−0.002597	−0.010390	0	0.155844	0.857143
8	0001010000	−0.002379	0.003330	0	0	−0.099905	1.098953
9	1010010000	−0.002196	0	0.006148	−0.018445	0	1.014493
10	0200010000	−0.002317	0.003244	−0.012975	0.0389249	−0.097312	1.070436
11	2100010000	−0.001994	−0.002791	0	0	0.083732	0.921053
12	4000010000	−0.001684	−0.007071	−0.009428	0.028283	0.212121	0.777778
13	0000200000	0.004739	0	0	0	−0.099526	1.094787
14	1001100000	0.002193	0.003070	0	−0.018421	0	1.013158
15	0110100000	0.004662	−0.003263	0	0.019580	−0.097902	1.076923
16	2010100000	0	0.002817	0	−0.016901	0.084507	0.929578
17	1200100000	0.002128	−0.002979	0	0.017872	0	0.982979
18	3100100000	−0.001812	−0.002536	0	0.015217	0.152174	0.836957
19	500010000	−0.004532	−0.010574	0	0.063444	0.253776	0.697885
20	0102000000	−0.004690	0.006567	−0.006567	0.019700	−0.098499	1.083490
21	2002000000	0.004019	0.005627	−0.005627	−0.016881	0.084405	0.928457
22	0021000000	−0.002363	−0.003308	0.013233	0	−0.099244	1.091682
23	1111000000	0	0	0.006024	0	0	0.993976
24	3011000000	0.003650	0.005110	−0.005110	0	0.153285	0.843066
25	0301000000	−0.009124	0.012774	−0.019161	0.057482	−0.095803	1.053832
26	2201000000	0	0	0.005435	0.016304	0.081522	0.896739
27	4101000000	0	0	0.004545	0.040909	0.204546	0.750000
28	6001000000	−0.005319	−0.007447	0.01862	0.100532	0.279255	0.614362
29	1030000000	−0.006536	0	0.018301	−0.018301	0	1.006536
30	0220000000	0.006843	−0.003193	0	0.038321	−0.095803	1.053832
31	2120000000	−0.009608	0.002745	0.010980	0	0.082353	0.905882
32	4020000000	0.004902	0.006863	0	0.027451	0.205882	0.754902
33	1310000000	0.004136	0	0.005790	0.034740	0	0.955335
34	3210000000	0	0.004879	0.014634	0.029268	0.146342	0.804878
35	5110000000	0.002849	0.007977	0.019943	0.071795	0.239316	0.658120
36	7010000000	0	0.009524	0.047619	0.133333	0.285714	0.523810
37	0500000000	−0.022124	0.030974	−0.030973	0.092920	−0.092920	1.022124
38	2400000000	0.003687	0.005162	0.015487	0.046460	0.077434	0.851770
39	4300000000	0.003030	0.012727	0.029697	0.063636	0.190901	0.700000
40	6200000000	0.007212	0.023558	0.050481	0.111058	0.252404	0.555288
41	8100000000	0.028205	0.043590	0.089744	0.161539	0.269231	0.423077
42	10000000000	0.027777	0.083333	0.138889	0.194444	0.250000	0.305556

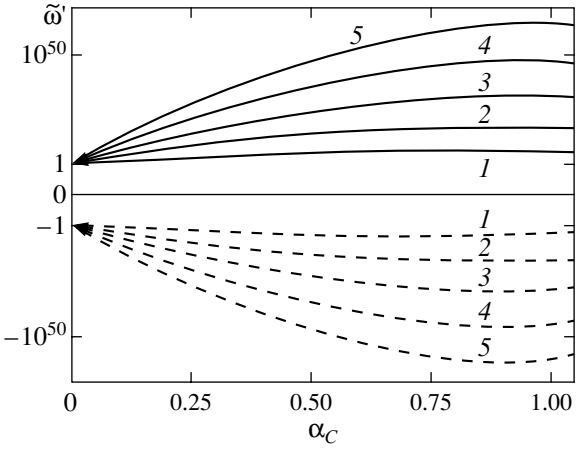

Fig. 1. Plots of the generalized combinatorial weights $\tilde{\omega}'(\alpha_C, \{n_k\}, l)$ (logarithmic scale) versus the degree of connectivity α_C for various numbers of electrons in the system: N = 10 (*1*), 20 (*2*), 30 (*3*), 40 (*4*), 50 (*5*). Solid and dashed curves refer to the diagrams with positive and negative weight coefficients.

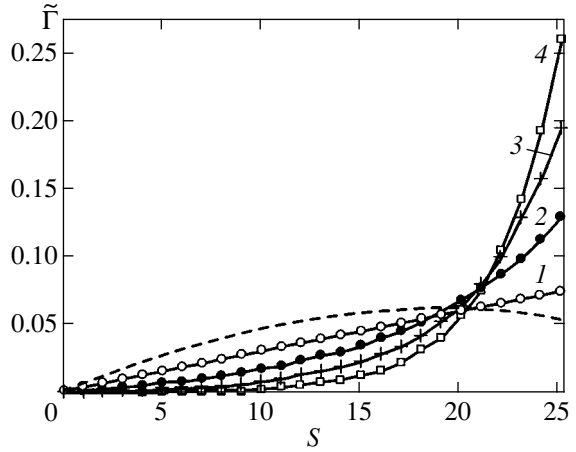

Fig. 2. The combinatorial distributions $\tilde{\Gamma}(\alpha_C, S, \{n_k\}, l)$ over spin states contained in all diagrams with positive combinatorial weights and a fixed connectivity parameter α_C = 0 (*1*), 0.0408 (*2*), 0.0816 (*3*), and 0.1224 (*4*) for a system of 50 electrons. The dashed curve shows (schematically) an expected form of the resulting distribution at a sufficiently high temperature.

The data in Table 3 illustrate the relations between various classes for the systems with different N. The numerical calculation was conducted by direct summation over all diagrams with nonnumbered vertices, with the weights $\omega'(\{\nu_i\})$ calculated by the random sampling method. For the final sample set $\{n_k\}$ with a volume of l, formula (13) can be written as

$$\tilde{\omega}'(\alpha_C, \{n_k\}, l) = \sum_{\{\nu_i\}}{}' \omega'(\{\nu_i\}, \{n_k\}, l). \qquad (14)$$

Figure 1 shows the plots of $\omega'(\{\nu_i\})$ calculated by formula (14) for 10, 20, 30, 40, and 50 electrons. The total set used for calculating each curve included $(1–5) \times 10^{10}$ diagrams. For example, in the system of 50 electrons, the values of $\tilde{\omega}'(\{\nu_i\}, \{n_k\}, l)$ with l = 10 000 were calculated by the random sampling method for each of the 204 226 diagrams with nonnumbered vertices. The volume of the random sampling set (taken over all spin states) used to calculate each value of the weight was $l(N/2) = 2.5 \times 10^5$. Then, the connectivity parameter was calculated for each of the 204 226 diagrams and the weights of diagrams possessing equal α_C values were added. These calculations required 10 hours of processor operation (IBM Pentium III processor with a clock frequency of 600 MHz).

As can be seen from Fig. 1, the generalized combinatorial weights of the diagrams rapidly grow in absolute value with increasing connectivity parameter. As the temperature increases, the statistical weights proportional to $a(\{\mathbf{r}_i(j)\}, \{\nu_i\}, \beta)$ of high-connectivity diagrams exhibit a rapid decrease, whereby only low-connectivity diagrams survive. As the temperature drops, the average length of the Feynman path increases and the low-connectivity diagrams lose their advantage. Under these conditions, the contribution from high-connectivity diagrams increases. Simultaneously, the statistical weight coefficients become more sensitive with respect to positions of the virtual trajectories in space $\{\mathbf{r}_i(j)\}$. For this reason, the electron exchange correlations tend to increase at low temperatures.

A mechanism of decrease in the spin of the system on cooling is elucidated by analysis of the distributions over spin states, which are presented in Fig. 2 for the diagrams with positive combinatorial weight factors. Analogous distributions are obtained for the diagrams with negative combinatorial weights as well. At a sufficiently high temperature, only the diagram with completely separated paths ($\alpha_C = 0$) and a linear distribution over spin states survives (Fig. 2, curve *1*). The formation of cycles of linked paths (the growth in α_C) with decreasing temperature is accompanied by a redistri-

Table 3. The numbers of diagrams of linked Feynman trajectories generated by application of the Young symmetry operators to systems with various numbers of electrons N

N	Number of diagram for numbered trajectories	Number of diagram for nonnumbered trajectories	Number of classes with different α_C values
5	120	7	5
10	362 800	42	10
20	2.433×10^{18}	627	20
30	2.653×10^{32}	5604	30
40	8.159×10^{47}	37 338	40
50	3.041×10^{64}	204 226	50

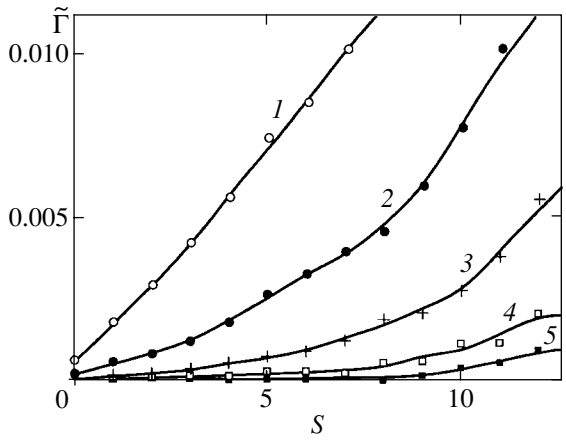

Fig. 3. The same as in Fig. 2, calculated for lower values of the spin and higher values of the connectivity parameter α_C = 0.0408 (*1*), 0.0816 (*2*), 0.1224 (*3*), 0.1633 (*4*), and 0.2041 (*5*).

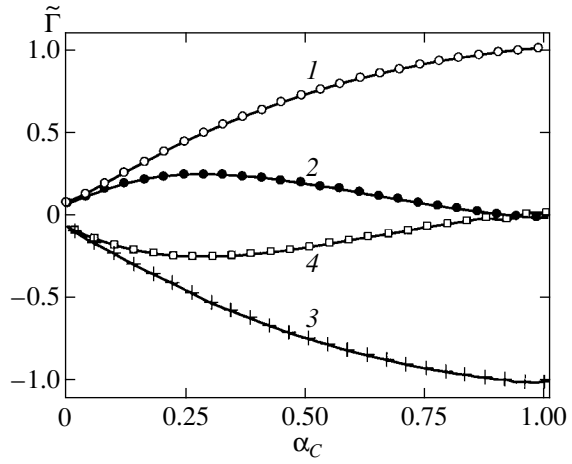

Fig. 4. The plots of $\tilde{\Gamma}$ (α_C, *S*, $\{n_k\}$, *l*) in a system of 50 electrons versus connectivity parameter α_C for two spin states with the maximum *S* values 25 (*1*, *3*) and 24 (*2*, *4*), calculated for the diagrams with (*1*, *2*) positive and (*3*, *4*) negative generalized weights.

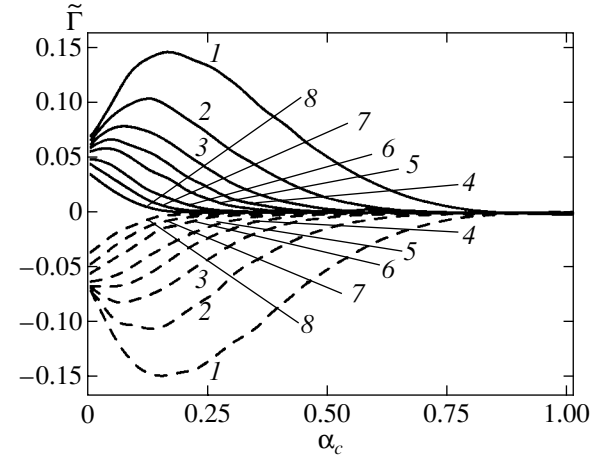

Fig. 5. The same as in Fig. 4, calculated for the states with lower values of the spin *S* = 23 (*1*), 22 (*2*), 21 (*3*), 20 (*4*), 19 (*5*), 17 (*6*), 15 (*7*), and 12 (*8*) using the diagrams with positive (solid curves) and negative (dashed curves) generalized weights.

bution in favor of greater *S* values (Figs. 2 and 3, curves *2*–*4*). The resulting distribution is obtained by summing the partial distributions for various α_C with the corresponding weights and signs. The most significant contribution (after that for $\alpha_C = 0$) to the mean values is due to the diagrams with minimum α_C (with a single pair of linked paths). The sign of the generalized weight of these diagrams is opposite to that of the diagrams with $\alpha_C = 0$; in Fig. 2, the corresponding distribution over the spin states is subtracted (with a certain relative weight below unity) from curve *1*. As a result, the probability of states (primarily those with large *S* values) decreases and the resulting curve becomes convex (see dashed curve in Fig. 2), which corresponds to the redistribution of weights toward low *S* values. Further decrease in the temperature leads to a redistribution of the statistical weights toward high-connectivity diagrams. For a sufficiently homogeneous system, we may expect that the tendency of the spin to decrease will be retained. In a strongly inhomogeneous external field, the diagrams complementary to the field geometry will be favored, the resulting distribution over the spin states can become more complicated, and the resulting spin may be nonzero.

With increasing connectivity, the weights of the states with maximum possible spins ($S = N/2$ and $S = N/2 - 1$) monotonically increase in absolute value, whereas the weights of the spin states with intermediate *S* values exhibit a extremum in the region of $\alpha_C < 0.25$. As the spin *S* decreases further, the monotonic increase changes to monotonic decrease (Figs. 4 and 5).

4. THE PROBLEM OF "NEGATIVE SIGNS"

Among the difficulties encountered in the development of computer modeling of the systems composed of quantum particles, most widely discussed is the so-called problem of "negative signs" [64–78]. The efficacy of the computational procedure can be significantly increased, provided that microstates giving close contributions with opposite (oscillating) signs would be excluded from the Markov random walk process. The exact result will be obtained by preliminarily taking a sum over all possible $\{\nu_i\}$ for each particular arrangement of the vertices of polygonal paths. In this case, the mutual compensation of coefficients with opposite signs reduces the weight of nonphysical path configurations and naturally displaces these paths from the random walk process. For large *N* values, the problem can be solved by passing from the summation over $\{\nu_i\}$ to taking random sets from the possible $\{\nu_i\}$.

In this study, we will use a method of displacing nonphysical microstates, which allows a simple numerical realization. The idea of this method consists in modifying the form of $a(\{\mathbf{r}_i(j)\}, \{\nu_i\}, \beta)$ so that the probabilities of microstates not obeying the Pauli principle would become low and these states would be not visited in the course of the Markov random walk pro-

cess. Obviously, such a modification unavoidably leads to a certain systematic error, because it is impossible to determine with absolute precision the boundaries of compensated microstates. As a result, the displacement of impermissible states will interfere with a part of physically allowed virtual trajectories. This uncertainty can be reduced to zero by narrowing the regions of displaced microstates so that their boundaries would occur beyond the displacement region. In this case, not all of such states will be excluded from the Markov process and a part of the computational facilities will be spent for visiting these states. The wider the region of displaced states, the higher the probability of interfering with physically realizable states, but the lower the statistical error. There exists an optimum regime, whereby the aforementioned systematic error is significantly smaller than the statistical error related to visiting nonphysical states. For each particular situation, the optimum regime has to be determined by trials.

The displacement of nonphysical microstates is provided by replacing $a(\{\mathbf{r}_i(j)\}, \{\nu_i\}, \beta)$ with a modified value

$$a^{\mathrm{mod}}(\{\mathbf{r}_i(j)\}, \{\nu_i\}, \beta) = a(\{\mathbf{r}_i(j)\}, \{\nu_i\}, \beta)\exp(-V_D(\{\mathbf{r}_i(j)\})), \tag{15}$$

with the directing potential

$$V_D(\{\mathbf{r}_i(j)\}) = \sum_{l<m}^{M}\sum_{j=1}\Bigg(\nu_D(|\mathbf{r}_l(j)-\mathbf{r}_m(j+1)|) + \sum_{l<m}^{M}\sum_{j=1}(\nu_D(|\mathbf{r}_l(j)-\mathbf{r}_m(j-1)|)\Bigg). \tag{16}$$

Here, $\mathbf{r}_n(M+1) = \mathbf{r}_k(1)$ for the linked trajectories with the numbers n and k,

$$\nu_D(r) = \exp\left(b_0\left(1-\frac{r}{r_0}\right)\right),$$

and b_0 and r_0 are the parameters controlling rigidity and the characteristic radius of interaction. Optimum values of the parameters b_0 and r_0 are set by the calculation program in the initial segment of the Markov process with the aid of a special feedback system stipulated in the algorithm; these values are adjusted so as to provide that the fraction of visited microstates with oscillating signs would not exceed 25% of the volume of accumulated statistics. The directing potential (16) leads to a decrease in the statistical weights of the states with closely spaced vertices by which the trajectories are linked. These configurations generate the contributions with oscillating signs and correspond to the states with strongly overlapping wave functions. Under the conditions of weakly degenerate plasma, there is no need for the repulsive potential and the program automatically sets $r \approx 0$, which corresponds to switching off the displacement mechanism.

When the temperature decreases, the spin of the system also exhibits a tendency to decrease. In terms of the one-electron approximation, this corresponds to "coupling" of the spins of a fraction of the electrons. The coincidence of the coordinate wave functions of two electrons occurring in different spin states does not contradict the Pauli principle. The states with overlapping wave functions may possess large statistical weights. For the low-temperature calculations, Eq. (16) should be replaced by a modified repulsive potential, in which the term $\nu_D(r)$ is operative only between trajectories of the same parity:

$$\begin{aligned}V_D(\{\mathbf{r}_i(j)\}) &= \sum_{l,m}^{M}\sum_{j=1}(\nu_D(|\mathbf{r}_{2l}(j)-\mathbf{r}_{2m}(j+1)|)) \\ &+ \sum_{l,m}^{M}\sum_{j=1}(\nu_D(|\mathbf{r}_{2l}(j)-\mathbf{r}_{2m}(j-1)|)) \\ &+ \sum_{l,m}^{M}\sum_{j=1}(\nu_D(|\mathbf{r}_{2l+1}(j)-\mathbf{r}_{2m+1}(j+1)|)) \\ &+ \sum_{l,m}^{M}\sum_{j=1}(\nu_D(|\mathbf{r}_{2l+1}(j)-\mathbf{r}_{2m+1}(j-1)|)).\end{aligned} \tag{17}$$

5. ESTIMATORS OF QUANTUM-MECHANICAL OBSERVABLES IN SYSTEMS WITH EXCHANGE

Calculation of the canonical means of various quantum-mechanical observables reduces to the integration of estimators representing the functions of position of the Feynman paths in space. The procedure of symmetrization with respect to permutations complicates both the distribution function and the estimators. This problem will be considered in the following section. The quantum-mechanical equilibrium means are expressed through the product of operators

$$\langle A\rangle = \frac{\mathrm{tr}(\hat{A}\exp(-\beta\hat{H}))}{\mathrm{tr}(\exp(-\beta\hat{H}))}, \tag{18}$$

where $\hat{H}$ is the Hamiltonian and $\beta = 1/k_BT$ is the reciprocal temperature. Note that the denominator of (18) contains only the diagonal matrix elements of the statistical operator, while the numerator contains both diagonal and nondiagonal matrix elements. The presence of nondiagonal matrix elements implies that, in the path integral representation, the functionals entering into the numerator of (18) must be calculated over open trajectories. The nondiagonal matrix elements in the numerator can be eliminated by finding an operator $\hat{B}$ which is diagonal in the coordinate representation and obeys the condition

$$\mathrm{tr}(\hat{A}\exp(-\beta\hat{H})) = \mathrm{tr}(\hat{B}\exp(-\beta\hat{H})).$$

The diagonal matrix elements of $\hat{B}$ represent an estimator of the operator $\hat{A}$. Let us consider a general approach to constructing estimators in a system with exchange in application to a spatial distribution function of the electron density. In the purely quantum state described by the wave function $\Psi(\{\mathbf{x}_i\})$, the electron density distribution is described by the function

$$\rho(\mathbf{r}) = \int \Psi(\{\mathbf{x}_i\}) \times \sum_{k=1}^{N} \delta(\mathbf{x}_k - \mathbf{r}) \Psi^*(\{\mathbf{x}_i\}) d\mathbf{x}_1 d\mathbf{x}_2 \ldots d\mathbf{x}_N. \quad (19)$$

Despite the fact that expression (19) is apparently similar to the matrix elements of an operator, the sum of Dirac delta functions in the integrand of (19) is not a operator acting upon the wave function, since multiplying by the delta function without subsequent integration makes no sense. The problem of finding an operator for the distribution (19) can be solved by introducing an auxiliary function $\rho_\omega(\mathbf{r})$ for which such a quantum-mechanical operator does exist:

$$\rho_\omega(\mathbf{r}) = \int \Psi(\{x_i\}) \frac{1}{\omega} \times \sum_{k=1}^{N} \theta_\omega(\mathbf{x}_k - \mathbf{r}) \Psi^*(\{\mathbf{x}_i\}) d\mathbf{x}_1 d\mathbf{x}_2 \ldots d\mathbf{x}_N. \quad (20)$$

Here, $\theta_\omega(\mathbf{x}) = 1$ in the ω vicinity of the point $x = 0$ and is zero everywhere outside this small vicinity. On the one hand, a comparison of (19) and (20) shows that

$$\rho(\mathbf{r}) = \lim_{\omega \to 0} \rho_\omega(\mathbf{r}).$$

On the other hand, function (20) is a matrix element of the operator

$$\hat{\Theta}_\omega(\mathbf{r}) = \frac{1}{\omega} \sum_{k=1}^{N} \theta_\omega(\mathbf{x}_k - \mathbf{r}), \quad (21)$$

for which the statistical averaging can be performed in a conventional manner in the representation symmetrized with respect to permutations. Using the notation from [81], we can write

$$\rho(\mathbf{r}) = \frac{\lim_{\omega \to 0} \mathrm{tr}(\hat{\Theta}_\omega(\mathbf{r}) \exp(-\beta \hat{H}))}{\mathrm{tr}(\exp(-\beta \hat{H}))} = \lim_{\omega \to 0} \sum_{S} \sum_{\{m_i\}} \sum_{\{m_i'\}}{}' \int d^N r \int d^N r' \quad (22)$$

$$\times \left\langle S, \{m_i\}, \{\mathbf{r}_i\} \middle| \hat{\Theta}_\omega(\mathbf{r}) \middle| S, \{m_i'\}, \{\mathbf{r}_i'\} \right\rangle \left\langle S, \{m_i'\}, \{\mathbf{r}_i'\} \middle| \exp(-\beta \hat{H}) \middle| S, \{m_i\}, \{\mathbf{r}_i\} \right\rangle d^N r d^N r' / \mathrm{tr}(\exp(-\beta \hat{H})).$$

The prime signs on the sums and integrals indicate that both summation and integration are performed over the sets of $\{\{m_i\}, \{m_i'\}, \{\mathbf{r}_i\}$, and $\{\mathbf{r}_i'\}$ not related by permutations. Note that the expression (22) contains no matrix elements nondiagonal with respect to S because (i) the corresponding spin functions are orthogonal and (ii) the operator $\hat{\Theta}_\omega(\mathbf{r})$ does not act upon spin variables. Using a diagonalized form of the complete wave function [81, Eq. (7)]

$$\Psi(S, \{m_i\}, \{\mathbf{r}_i\}; \{\sigma_i\}, \{\mathbf{x}_i\}) = \sum_{n=1}^{N!} (-1)^{g(n)} \tilde{\chi}(S, \{m_i\}; \hat{P}_n\{\sigma_i\}) f(\{\mathbf{r}_i\}, \hat{P}_n\{\mathbf{x}_i\}) = \sum_{n=1}^{N!} (-1)^{g(n)} \tilde{\chi}(S, \hat{P}_n\{m_i\}; \{\sigma_i\}) f(\hat{P}_n\{\mathbf{r}_i\}, \{\mathbf{x}_i\}), \quad (23)$$

we can write the matrix elements of the operator $\hat{\Theta}_\omega(\mathbf{r})$ in (22) as

$$\rho(\mathbf{r}) = \lim_{\omega \to 0} \sum_{S} \sum_{\{m_i\}} \sum_{\{m_i'\}}{}' \sum_{n=1}^{N!} \sum_{k=1}^{N!} \sum_{\{\sigma_i = \pm 1/2\}} \int d^N r \int d^N r' \times (-1)^{g(n)} \tilde{\chi}(S, \hat{P}_n\{m_i\}; \{\sigma_i\}) \times (-1)^{g(k)} \tilde{\chi}(S, \hat{P}_k\{m_i'\}; \{\sigma_i\}) \times \langle \hat{P}_n\{\mathbf{r}_i\} | \hat{\Theta}_\omega(\mathbf{r}) | \hat{P}_k\{\mathbf{r}_i'\} \rangle \times \langle S, \{m_i'\}, \{\mathbf{r}_i'\} | \exp(-\beta \hat{H}) | S, \{m_i\}, \{\mathbf{r}_i\} \rangle / \mathrm{tr}(\exp(-\beta \hat{H})). \quad (24)$$

Now let us change the variables of summation $\{\tilde{m}_i\} \equiv \hat{P}_n\{m_i\}$, $\{\tilde{m}_i'\} \equiv \hat{P}_k\{m_i'\}$ and integration $\{\tilde{\mathbf{r}}_i\} \equiv \hat{P}_n\{\mathbf{r}_i\}$, $\{\tilde{\mathbf{r}}_i'\} \equiv \hat{P}_k\{\mathbf{r}_i'\}$ in each term with subscripts n and k in (24) and use the property of antisymmetry of the complete wave function in the matrix elements $\exp(-\beta \hat{H})$ with respect to these permutations. Then, the terms are no longer dependent on n and k and can be combined with the term $(N!)^2$:

$$\rho(\mathbf{r}) = \lim_{\omega\to 0}(N!)^2\sum_S{\sum_{\{\tilde m_i\}}}'{\sum_{\{\tilde m_i'\}}}'\sum_{\{\sigma_i=\pm 1/2\}}\int' d^N\tilde r\int' d^N\tilde r'$$
$$\times\tilde\chi(S,\{\tilde m_i\};\{\sigma_i\})\tilde\chi(S,\{\tilde m_i'\};\{\sigma_i\})$$
$$\times\langle\{\tilde{\mathbf{r}}_i\}|\hat\Theta_\omega(\mathbf{r})|\{\tilde{\mathbf{r}}_i'\}\rangle$$
$$\times\langle S,\{\tilde m_i'\},\{\tilde r_i'\}|\exp(-\beta\hat H)|S,\{\tilde m_i\},\{\tilde{\mathbf{r}}_i\}\rangle \quad (25)$$
$$/\mathrm{tr}(\exp(-\beta\hat H))$$
$$= \lim_{\omega\to 0}\sum_S{\sum_{\{m_i\}}}'\int' d^N\tilde r\int d^N\tilde r'\langle\{\tilde{\mathbf{r}}_i\}|\hat\Theta_\omega(\mathbf{r})|\{\tilde{\mathbf{r}}_i'\}\rangle$$
$$\times\langle S,\{\tilde m_i\},\{\tilde{\mathbf{r}}_i'\}|\exp(-\beta\hat H)|S,\{\tilde m_i\},\{\tilde{\mathbf{r}}_i\}\rangle$$
$$/\mathrm{tr}(\exp(-\beta\hat H)).$$

This expression is written with an allowance for the orthogonality of the modified spin functions $\tilde\chi(S,\{m_i\};\{\sigma_i\})$ for the $\{m_i\}$ sets not related by permutations (as a result, the summation over $\{m_i'\}$ disappears) and with the use of relations [81, Eqs. (10)–(12)]. Equation (25) is divided by $N!$ and the prime sign on the integral with respect to $\tilde r'$ is omitted.

In the limit $\omega\longrightarrow 0$, the matrix elements $\hat\Theta_\omega(\mathbf{r})$ in a nonsymmetrized coordinate representation take the form

$$\lim_{\omega\to 0}\langle\{\mathbf{r}_i\}\tilde\Theta_\omega|\{\mathbf{r}_i'\}\rangle = \sum_{k=1}^N\delta(\mathbf{r}_k-\mathbf{r})\prod_i\delta(\mathbf{r}_i-\mathbf{r}_i'). \quad (26)$$

Therefore, the integration with respect to primed coordinates retains only the diagonal matrix elements of $\exp(-\beta\hat H)$ in the representation symmetrized with respect to permutations:

$$\rho(\mathbf{r}) = \sum_S{\sum_{\{m_i\}}}'\int' d^N r\sum_{k=1}^N\delta(\mathbf{r}_k-\mathbf{r})\langle S,\{m_i\},\{\mathbf{r}_i\}|\exp(-\beta\hat H)|S,\{m_i\},\{\mathbf{r}_i\}\rangle/\mathrm{tr}(\exp(-\beta\hat H))$$
$$= \sum_S(2S+1)\sum_n W_S(\hat P_n)\int d^N r\sum_{k=1}^N\delta(\mathbf{r}_k-\mathbf{r})\langle\{\mathbf{r}_i\}|\exp(-\beta\hat H)|\hat P_n(\mathbf{r}_i)\rangle/\mathrm{tr}(\exp(-\beta\hat H)). \quad (27)$$

This expression is obtained using the same transformations as those employed in deriving [81, Eq. (9)]. Writing Eq. (27) in the form of a path integral (prior to the limiting transition) and taking into account invariance of the integral with respect to cyclic renumeration of the path vertices, we obtain an expression for the estimator of the spatial density distribution:

$$\rho_e(\mathbf{r}) = \frac{1}{M}\sum_{k=1}^N\sum_{j=1}^M\delta(\mathbf{r}_k(j)-\mathbf{r}). \quad (28)$$

The numerical averaging of (28) is performed by wandering in the space of trajectories $\{\mathbf{r}_i(j)\}$ and path linking modes $\{\nu_i\}$ with the distribution function

$$\omega(x) = \tilde\omega(\{\nu_i\},\{n_k\},M)a(\{\mathbf{r}_i(j)\},\{\nu_i\},\beta)$$
$$\propto\tilde\omega(\{\nu_i\},\{n_k\},M)$$
$$\times\exp\left\{-\left[\frac{Mm_0}{2\beta\hbar^2}\sum_{i=1}^N\sum_{j=1}^M(\mathbf{r}_i(j+1)-\mathbf{r}_i(j))^2\right.\right. \quad (29)$$
$$\left.\left.+\frac{\beta}{M}\sum_{j=1}^M V(\{\mathbf{r}_i(j)\})\right]\right\}.$$

6. NUMERICAL SIMULATION OF A DENSE HYDROGEN PLASMA

The numerical simulation was performed for two values of the plasma density: $\rho_1 = 0.696\times 10^{21}$ cm^{-3} (T = 37 130–137 860 K) and $\rho_2 = 10^{23}$ cm^{-3} (T = 347 220–1 493 000 K). In the first case, the plasma density is three order of magnitude smaller as compared to that of a condensed phase, while the second value is comparable to the density in the condensed state. The edge effects were eliminated by using periodic boundary conditions. All electrostatic interactions between particles were explicitly taken into account using the closest image method. The periodic cubic cell contained 100 protons and 100 electrons described by method of classical and quantum statistics, respectively. Each electron was represented by a closed Feynman trajectory approximated by a broken line with M vertices (M = 80, 160, or 320, depending on the temperature).

The Markov random walk process consisted in a sequence of steps shifting the trajectory vertices, shifting or rotating of the whole trajectories, linking trajectories into cycles, unlinking cycles, or shifting protons. Each simulation run included 200 to 900 million steps, of which the first 100 to 300 million were used to provide for the system thermalization, while the others

Table 4. The proton–electron correlation function $\rho_{pe}(r)$ (in units of 10^{-24} cm^{-3}) calculated by the path integral method for a hydrogen plasma with the density $\rho = 6.96 \times 10^{20}$ cm^{-3} at various temperatures (the number of segments in the virtual electron trajectory $M = 160$)

r, $\times 10^{-8}$ cm	37130 K	48270 K	62750 K	81570 K	106040 K	137860 K
0.05	0.737	0.455	0.290	0.112	0.0640	0.0418
0.15	0.420	0.253	0.159	0.0624	0.0356	0.0221
0.25	0.296	0.177	0.108	0.0429	0.0245	0.0140
0.35	0.209	0.124	0.0729	0.0304	0.0174	0.00929
0.45	0.149	0.0866	0.0495	0.0217	0.0126	0.00652
0.55	0.107	0.0602	0.0339	0.0157	0.00926	0.00490
0.65	0.0764	0.0420	0.0233	0.0114	0.00687	0.00383
0.75	0.0542	0.0295	0.0162	0.00844	0.00512	0.00308
0.85	0.0383	0.0211	0.0116	0.00635	0.00396	0.00257
0.95	0.0270	0.0151	0.00849	0.00494	0.00318	0.00220
1.05	0.0191	0.0110	0.00639	0.00394	0.00264	0.00194
1.15	0.0137	0.00809	0.00493	0.00322	0.00225	0.00174
1.25	0.00986	0.00615	0.00392	0.00270	0.00198	0.00159
1.35	0.00720	0.00481	0.00322	0.00234	0.00178	0.00148
1.45	0.00534	0.00385	0.00272	0.00206	0.00163	0.00138
1.55	0.00405	0.00314	0.00236	0.00185	0.00152	0.00130
1.65	0.00315	0.00264	0.00209	0.00170	0.00143	0.00124
1.75	0.00251	0.00227	0.00189	0.00158	0.00136	0.00119
1.85	0.00206	0.00200	0.00173	0.00148	0.00129	0.00114
1.95	0.00174	0.00180	0.00160	0.00140	0.00124	0.00110
2.05	0.00151	0.00164	0.00150	0.00134	0.00119	0.00107
2.15	0.00135	0.00152	0.00142	0.00128	0.00115	0.00104
2.25	0.00123	0.00143	0.00135	0.00123	0.00111	0.00102
2.35	0.00114	0.00135	0.00129	0.00119	0.00108	0.00099
2.45	0.00107	0.00128	0.00124	0.00115	0.00105	0.00097
2.55	0.00102	0.00123	0.00120	0.00112	0.00102	0.00096
2.65	0.00099	0.00119	0.00116	0.00109	0.00100	0.00094
2.75	0.00096	0.00115	0.00112	0.00106	0.00098	0.00093
2.95	0.00092	0.00108	0.00106	0.00101	0.00095	0.00091
3.25	0.00088	0.00100	0.00099	0.00096	0.00092	0.00087
3.55	0.00086	0.00094	0.00094	0.00092	0.00088	0.00085
3.85	0.00084	0.00090	0.00090	0.00088	0.00086	0.00083
4.25	0.00082	0.00086	0.00086	0.00085	0.00083	0.00081
4.75	0.00079	0.00082	0.00083	0.00082	0.00080	0.00079

were used for the calculation of mean equilibrium values. The stage of calculating the mean values included 3000 to 10 000 events of trajectory linking and separation. The volume of random sampling for calculating the combinatorial coefficients in each step was $l = 10$ for each of 50 spin states. In the thermalization stage, the process was accelerated by one to two orders of magnitude using a procedure of vertex multiplication.

The statistical error of calculated mean values was estimated (as it was done in [81]) in a standard manner [15], using fluctuations of partial averages. For this purpose, the Markov random walk process in each run was separated into ten equal segments and the partial averages were calculated for each segment (the correlations between segments are negligibly small). The statistical error of the correlation functions did not exceed a level of about 1%. Correctness of the computational procedure was checked by simulating special trivial cases, whereby the given method is applied to the system at lower temperatures so as to obtain the ground quantum state for which either the analytical solution is known

(hydrogen atom) or reliable experimental data are available (hydrogen molecule). The results of such a comparison were reported previously [81]. In addition, the computation algorithm was tested by calculating a low-density system of 100 electrons and 100 fixed protons. On cooling, the system featured the formation of 100 weakly interacting hydrogen atoms and the energy acquired a value corresponding to the well-known ground quantum state of such system.

In the system with a density of 6.96×10^{20} cm^{-3}, the average distance between electrons is 11.3×10^{-8} cm; the thermal de Broglie electron wavelength at the temperature $T = 37\,130$ K is

$$\lambda = \frac{h}{\sqrt{2\pi m_e k_B T}} = 3.9 \times 10^{-8} \text{ cm}.$$

In the system with a density of 10^{23} cm^{-3}, the average distance between electrons is 2.2×10^{-8} cm and the thermal electron wavelength at $T = 347\,220$ K is 1.3×10^{-8} cm. In both cases, the exchange effects are significant despite relatively high temperatures.

The microstructural characteristics of the plasma are presented in the form of binary correlation functions. The proton–electron correlation function $\rho_{pe}(r)$ (Table 4) has the meaning of the density of probability to find an electron at the distance r from a proton. In the coordinate representation, with an allowance for degeneracy with respect to the eigenvalues of the spin projection operator,

$$\rho_{pe}(r) = \sum_S (2S+1) \int d\mathbf{r}_1 d\mathbf{r}_2 \ldots d\mathbf{r}_N d\mathbf{R}_1 \ldots d\mathbf{R}_N \times \langle \{\mathbf{r}_i\}, \{\mathbf{R}_i\}, S \Big| \Big| \sum_{k,n} \delta(r - |\mathbf{x}_k - \mathbf{X}_n|) \times \exp(-\beta \hat{H}) \Big| \Big| \{\mathbf{r}_i\}, \{\mathbf{R}_i\}, S \rangle / 4\pi r^2 \quad (30)$$

$$\times \sum_S (2S+1) \int d\mathbf{r}_1 d\mathbf{r}_2 \ldots d\mathbf{r}_N d\mathbf{R}_1 \ldots d\mathbf{R}_N \times \langle \{\mathbf{r}_i\}, \{\mathbf{R}_i\}, S || \exp(-\beta \hat{H}) || \{\mathbf{r}_i\}, \{\mathbf{R}_i\}, S \rangle,$$

where $\{\mathbf{x}_i\} \equiv \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ and $\{\mathbf{X}_i\}$ are the spatial variables of electrons and protons; $\{\mathbf{r}_i\}$ and $\{\mathbf{R}_i\}$ are the eigenvalues of the coordinate operator; and $\delta(r)$ is the one-dimensional Dirac delta function. The matrix elements in (30) are calculated in the representation of wave functions symmetrized with respect to permutations, in accordance with the eigenvalue of the square spin operator of the whole system. The electron–electron $\rho_{ee}(r)$ and proton–proton $\rho_{pp}(r)$ correlation functions are written by analogy to (30) with replacing $|\mathbf{x}_k - \mathbf{X}_n|$ by $|\mathbf{x}_k - \mathbf{x}_n|$ and $|\mathbf{X}_k - \mathbf{X}_n|$, respectively, and summing over $n \neq k$.

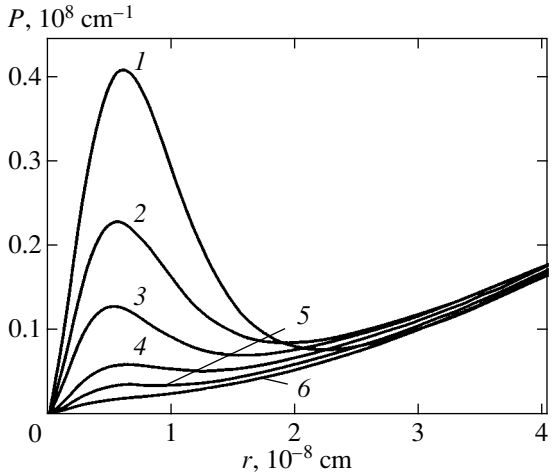

Fig. 6. Radial distributions of electrons near protons in a hydrogen plasma with the density $\rho = 0.696 \times 10^{21}$ cm^{-3} at various temperatures (K): (*1*) 37 130, (*2*) 48 270, (*3*) 62 750, (*4*) 81 570, (*5*) 106 040, (*6*) 137 860.

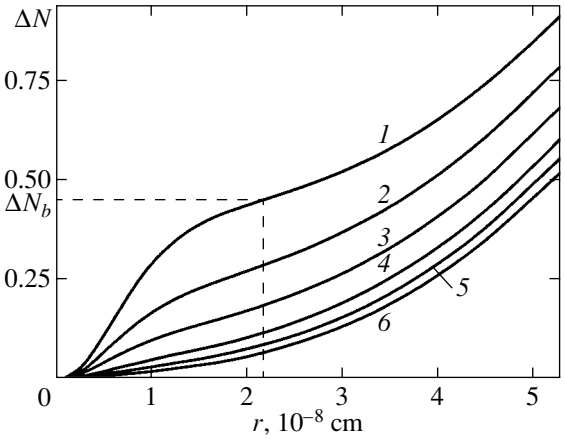

Fig. 7. Plots of the current coordination number of electrons near protons versus distance for a plasma with the density $\rho = 0.696 \times 10^{21}$ cm^{-3} at various temperatures (K): (*1*) 37 130, (*2*) 48 270, (*3*) 62 750, (*4*) 81 570, (*5*) 106 040, (*6*) 137 860.

In the plasma with the density ρ_1 at a temperature $T < 10^5$ K, the radial distribution of electrons in the field of protons exhibits a maximum (Fig. 6) at a distance of one Bohr radius, which corresponds to the bound proton–electron states. At a temperature of 37 130 K, the peak height amounts to 40% of the value in the ground state of the hydrogen atom. At a distance of approximately 2.2×10^{-8} cm, the radial distribution function shows a minimum, while the curve of the electron coordination numbers

$$\Delta N(r) = \int_0^r 4\pi x^2 \rho_{ie}(x) dx$$

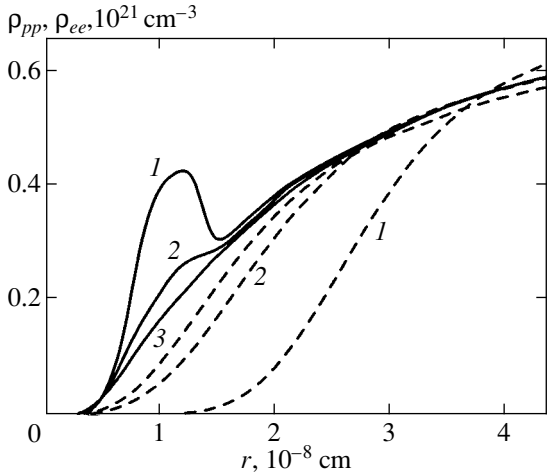

Fig. 8. The proton–proton (ρ_{pp}, solid curves) and electron–electron (ρ_{ee}, dashed curves) correlation functions for a plasma with the density $\rho = 0.696 \times 10^{21}$ cm^{-3} at various temperatures (K): (*1*) 37 130, (*2*) 48 270, (*3*) 81 570.

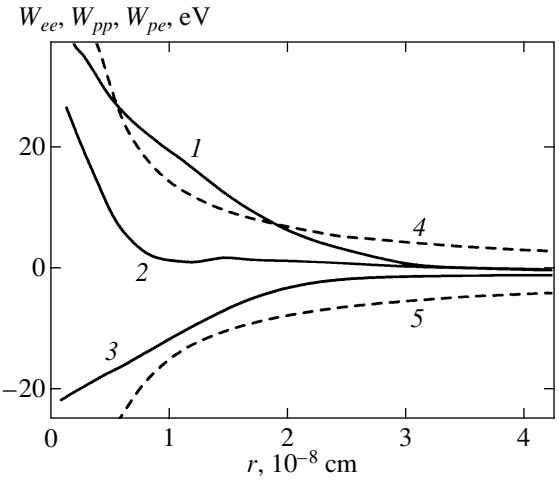

Fig. 9. The mean force potentials in a plasma with the density $\rho = 0.696 \times 10^{21}$ cm^{-3} and a temperature of 37 130 K (solid curves): (*1*) electron–electron (W_{ee}), (*2*) proton–proton (W_{pp}), (*3*) proton–electron (W_{pe}). Dashed curves show the energy of the Coulomb interaction between two classical point elementary charges of the (*4*) same and (*5*) opposite signs.

exhibits a bending point (Fig. 7). The degree of ionization can be estimated by calculating the area under the radial distribution peak in Fig. 6. At $T = 37\,130$ K, this value amounts to $1 - \Delta N_b = 0.55$ (Fig. 7).

At $T = 37\,130$ K, the system exhibits, in addition to the bound proton–electron states, the signs of more complex associates comprising two protons and two electrons—a prototype of the hydrogen molecule. In the proton–proton correlation function ρ_{pp} (Fig. 8), there appears a maximum at a distance of about 1.1×10^{-8} cm. This value is approximately 20% greater than the equilibrium distance between protons in the hydrogen molecule, which is evidence of strongly excited vibrational modes. These associates disappear almost completely at temperatures above 50 000 K, although the bound proton–electron states can be observed at temperatures up to 150 000 K (Fig. 6). Judging by the shape of the correlation functions (Fig. 8), a decrease in the temperature leads to an increase in the average distance between nearest electrons (the ρ_{ee} curves shift rightward), while the distance between nearest protons remains virtually unchanged. The number of protons at distances below 3×10^{-8} cm is greater than the number of electrons with the same spacing, and this effect increases on cooling.

At temperatures below 40 000 K, the requirement of spatial delocalization does not allow electrons to follow protons at a distance below 10^{-8} cm. As the protons approach one another, the electrons are driven to the periphery (rather than localizing between protons); nevertheless, the electrons still significantly compensate for the electrostatic repulsion of protons, even at distances less than the Bohr radius. This effect is due to the bound electron states which appear and disappear at the pair of protons in the course of thermal fluctuations. This is confirmed by the shape of the mean force potential (Fig. 9) described by the formula

$$W_{ee}(r) \equiv -k_B T \ln \frac{\rho_{ee}(r)}{\rho}, \qquad (31)$$

where ρ is the overall bulk number density of particles. This function has a meaning of interaction between two particles with an allowance for indirect contacts mediated by other particles of the system. Deviation of the $W_{ee}(r)$ value from the energy of the Coulomb interaction between point charges (e^2/r) reflects the quantum effects and the interaction with other particles. The quantum effects are manifested at distances shorter than the characteristic spatial delocalization length of a quantum particle, while the electrostatic screening by particles of the medium increases with the distance r.

At distances $r > 2 \times 10^{-8}$ cm, the interactions of all three types (W_{ee}, W_{pp}, and W_{pe}) are significantly decreased as a result of the screening effect. In the interval $0.5 \times 10^{-8} < r < 2 \times 10^{-8}$ cm, the effective repulsion between electrons is stronger than simple electrostatic interaction between classical point charges. This phenomenon is explained by the quantum character of electron motion. On the contrary, the interaction between protons is significantly screened by electrons. The signs of bound states, manifested by a weak minimum in the $W_{pp}(r)$ curve, can be traced to a distance of about 1.1×10^{-8} cm. Determining the correlation radius as a distance at which the mean force potential is on the order of $k_B T$ and the correlation function differs from the total particle number density by a factor of $\rho_{pe}/\rho \approx e$, we can infer from Table 4 that the proton–electron correlation radius at $T = 37\,130$ K amounts to 1.9×10^{-8} cm. As the temperature increases, the correlation radius decreases

Table 5. The proton–electron correlation function $\rho_{pe}(r)$ (in units of 10^{-24} cm^{-3}) calculated by the path integral method for a hydrogen plasma with the density $\rho = 1.0 \times 10^{23}$ cm^{-3} at various temperatures (the number of segments in the virtual electron trajectory $M = 80$)

r, $\times 10^{-8}$ cm	347220 K	416670 K	500000 K	600000 K	720000 K	1036800 K	1493000 K
0.045	0.639	0.551	0.476	0.409	0.359	0.255	0.199
0.075	0.575	0.495	0.434	0.372	0.331	0.244	0.194
0.105	0.521	0.453	0.399	0.342	0.305	0.232	0.189
0.135	0.474	0.414	0.363	0.315	0.280	0.217	0.179
0.165	0.432	0.377	0.330	0.288	0.257	0.202	0.170
0.195	0.392	0.345	0.300	0.263	0.235	0.188	0.160
0.225	0.357	0.315	0.275	0.243	0.218	0.176	0.152
0.285	0.299	0.266	0.233	0.209	0.190	0.158	0.139
0.345	0.253	0.228	0.203	0.183	0.169	0.145	0.130
0.405	0.220	0.199	0.180	0.165	0.154	0.136	0.124
0.465	0.195	0.177	0.163	0.152	0.143	0.129	0.120
0.525	0.175	0.162	0.151	0.142	0.135	0.124	0.117
0.585	0.160	0.150	0.142	0.134	0.129	0.120	0.114
0.645	0.149	0.141	0.134	0.129	0.125	0.117	0.112
0.705	0.140	0.134	0.129	0.124	0.121	0.115	0.110
0.765	0.134	0.128	0.125	0.121	0.118	0.129	0.110
0.825	0.128	0.125	0.121	0.118	0.116	0.111	0.109
0.885	0.124	0.121	0.118	0.116	0.114	0.110	0.108
0.945	0.121	0.118	0.116	0.114	0.112	0.109	0.107
1.005	0.118	0.115	0.114	0.112	0.111	0.108	0.106
1.065	0.116	0.114	0.112	0.111	0.109	0.107	0.106
1.125	0.114	0.112	0.111	0.110	0.108	0.106	0.105
1.185	0.112	0.110	0.110	0.108	0.107	0.106	0.105
1.245	0.110	0.109	0.108	0.108	0.107	0.105	0.104
1.305	0.109	0.108	0.107	0.107	0.106	0.104	0.103
1.365	0.108	0.107	0.106	0.106	0.105	0.104	0.103
1.425	0.107	0.106	0.106	0.105	0.105	0.104	0.103
1.485	0.106	0.106	0.105	0.105	0.104	0.103	0.103
1.545	0.106	0.105	0.105	0.104	0.104	0.103	0.103
1.605	0.105	0.105	0.104	0.104	0.104	0.103	0.103
1.665	0.104	0.104	0.104	0.104	0.103	0.102	0.102
1.725	0.104	0.104	0.103	0.103	0.103	0.102	0.102
1.965	0.102	0.102	0.102	0.102	0.102	0.102	0.101
2.475	0.101	0.101	0.101	0.101	0.101	0.101	0.101

to reach 1.3×10^{-8} cm at $T = 137\,860$ K. A comparison of the curves in Fig. 9 shows that the correlations between electrons are stronger than those between protons and electrons, although the correlation radii are close. Minimum correlations are observed between protons. The correlation radius of proton positions is close to 10^{-8} cm and remains virtually unchanged in the temperature range studied.

The above data confirm correctness of the periodic cell size selection (52.375×10^{-8} cm) and applicability of the closest image method. In this case, the electrostatic forces are strongly screened, representing essentially the short-range interactions. The range of these interactions falls within the interval of distances at which the quantum character of the motion of particles plays a determining role. Under these conditions, the caloric characteristics of the whole system are formed at interparticle distances for which the quantum nature of the electron interaction component is dominating. Even a decrease in the plasma density by two to three orders of magnitude does not render the system more

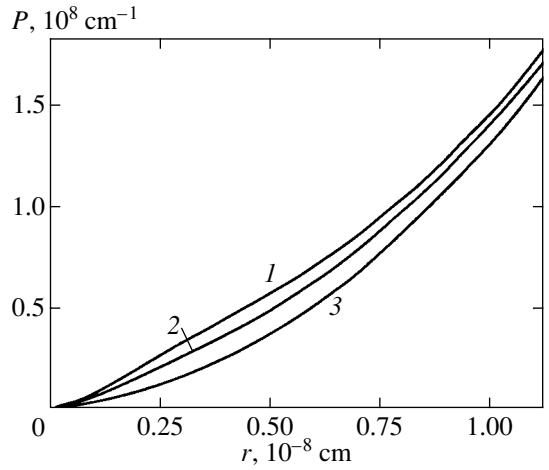

Fig. 10. Radial distributions of electrons near protons in a plasma with the density $\rho = 10^{23}$ cm^{-3} at various temperatures (K): (*1*) 347 220, (*2*) 500 000, (*3*) 1 493 000.

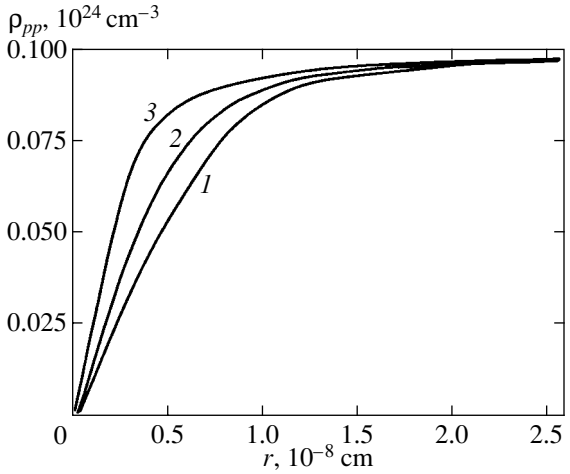

Fig. 11. The proton–proton correlation functions for a plasma with the density $\rho = 10^{23}$ cm^{-3} at various temperatures (K): (*1*) 347 220, (*2*) 600 000, (*3*) 1 493 000.

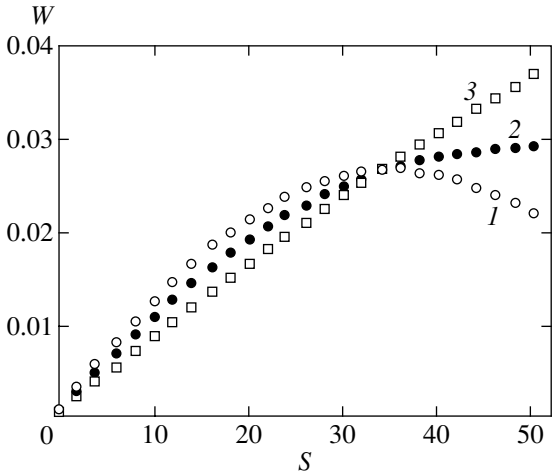

Fig. 12. Distributions of the square spin operator with respect to the spin *S* within a periodic cell for a plasma with the density $\rho = 10^{23}$ cm^{-3} at various temperatures (K): (*1*) 347 220, (*2*) 416 670, (*3*) 1 493 000.

"classical" since, as a result of screening, the energetics of the system is still formed at small distances. This circumstance is especially important if we take into account that, at these plasma densities, the thermal wavelength of electrons is much smaller than the average distance between nearest neighbors and, according to the formal criterion, the approximation of classical statistics should be applicable. Our data indicate that the classical description is already incorrect in this domain of conditions.

A detailed description of the exchange effects becomes decisive in the case of a plasma with the density approaching that in the condensed state. Such states arise in the initial stage of laser-induced heating of a deuterium–tritium target in the process of controlled thermonuclear fusion. The hydrogen plasma with a density of 10^{23} cm^{-3} and a temperature above 3×10^5 K reveals no signs of the bound proton–electron states. The correlation functions (Table 5) and the radial distributions of electron density near protons are monotonic, although a thorough analysis still reveals a very weak rise in the radial distribution curves in the region of *r* values about one Bohr radius (Fig. 10). The electron density near protons weakly changes with the temperature: a fivefold increase in the temperature leads only to a 10% decrease in the electron density within a sphere with a diameter of doubled Bohr radius around a proton (Table 5). Most sensitive to the temperature are the correlations between ions, for which the corresponding change amounts to 60% (Fig. 11). However, the temperature sensitivity rapidly drops with increasing distance.

In a dense plasma, the correlation radii are small and decrease with the temperature. At a temperature of 377 220 K, the proton–electron correlations possess a characteristic radius of 0.35×10^{-8} cm (Table 5), while the proton–proton correlation radius does not exceed 0.2×10^{-8} cm. At a distance of 10^{-8} cm, the proton–proton correlation function differs only by 10% from the overall particle number density in the system (Fig. 11). A comparison of the correlation radius and the average distance between neighboring particles shows that, in plasmas with the densities $\rho_1 = 6.96 \times 10^{20}$ cm^{-3} and $\rho_2 = 10^{23}$ cm^{-3}, the former value is smaller than the latter by a factor of five and six, respectively. This indicates that the heating and compression of the plasma leads to an almost proportional decrease in the correlation radius. At the same time, the correlation radius in the dense plasma decreases by half relative to the thermal wavelength of electrons. Despite a higher temperature, the latter regime corresponds to a more pronounced quantum character of the motion of particles in the plasma.

In this context, it is interesting to trace the effect of the temperature on the type of permutation symmetry of the electron wave function. The type of permutation symmetry is uniquely related to the total spin of the system. Therefore, the average square spin serves as a

good indicator of changes in symmetry of the wave function. Figure 12 shows distributions of the square spin operator with respect to the total spin S for electrons in the periodic cell. At high temperatures, the distribution is described by a linearly increasing function, the growth of which is related to degeneracy with respect to the eigenvalues of the spin projection operator ($\propto 2S + 1$). This behavior shows a statistical equivalence of various spin states and, hence, of various types of permutation symmetry. However, the linearity is already violated at $T = 416\,670$ K and the states with smaller S acquire a greater statistical weight, which is indicative of the partial "coupling" of the electron spins. An unexpected result is that the spin coupling in a high-density plasma begins only at such a high temperature. It should be noted that a decrease in the system spin on cooling is not an evident fact and depends on the configuration of the field applied to the system.

7. CONCLUSION

The exact statistical description of many-body systems of quantum particles with exchange requires constructing a complete set of basis functions symmetrized with respect to permutations. The symmetrization leads to a catastrophic multiplication of the diagrams of linked Feynman paths, making the exact calculation practically impossible in systems containing more than ten fermions.

In this study, a method allowing these difficulties to be surmounted without introducing essential approximations is formulated. The new approach consists in applying the method of essential set to calculating linear combinations of permutations generated by the Young symmetry operators. The results of numerical calculations showed that, despite the huge number of diagrams, the method provides for a sufficiently rapid convergence. In a system containing about 50 particles, even a relative small set covering on the order of 10^{10} diagrams from the total of 3×10^{64} allows the combinatorial weight coefficients for various modes of linking Feynman paths, as well as the distributions over spin states, to be obtained with high precision. The statistical calculation procedure remains fundamentally exact at finite temperatures, irrespective of the volume of sampling from the linear combinations in each step of the Monte Carlo procedure, because subsequent steps enrich the statistics with independent sets. This circumstance allows minimum sets to be taken in each step, thus practically eliminating restrictions on the number of particles in the system.

Using the proposed approach, all combinatorial weights for many-body exchange integrals and the corresponding distributions over spin states were calculated for a system containing up to 50 electrons and a hydrogen plasma with 100 electrons in a periodic cell was simulated. A significant feature of the developed method is that the Monte Carlo simulation is possible even in cases when a very large number of particles hinders the determination of all weights: a fundamentally exact simulation procedure can be realized using finite sets of permutations.

An analysis of the results showed evidence of the mutual compensation of the positive and negative contributions to the partition function being the main mechanism of the spin state formation in a system of fermions at finite temperatures. A decrease in the system spin on cooling results from the partial compensation of states with large S values by negative contributions from the diagrams with long cycles of linked paths. The statistical weight of such paths increases with decreasing temperature.

The calculations on microscopic level reported in this paper were performed under conditions presenting considerable difficulties for traditional approaches. Strong interparticle correlations and the quantum character of electron motion play a dominating role. Despite the long-range character of the Coulomb forces, the interaction under these conditions is strongly screened and becomes effectively short-range. This circumstance allows the calculations to be conducted using relatively small periodic cells. Although the temperatures are relatively high, the energy characteristics of the dense plasma are formed over small interparticle distances, where the quantum character of electron motion plays a determining role. At these temperatures, a decrease in the plasma density does not render it more "classical," since the plasma energetics is formed over small interparticle distances even when the formal criterion based on the thermal de Broglie electron wavelength predicts the classical regime.

The motion of particles in a dense plasma is strongly correlated, but the correlation radii are small and comparable with atomic dimensions. The correlation radii in a high-density hydrogen plasma decrease with increasing temperature, thus changing in the opposite direction with respect to the Debye radius in a rare ion plasma. The quantum character of electron motion renders the electron–electron correlations less sensitive to the temperature. The proton–proton correlations are the most temperature-sensitive. In a dense plasma, the coupling of electron spins begins on cooling to a temperature on the order of several hundred thousand kelvin.

REFERENCES

1. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, and H. A. Teller, J. Chem. Phys. **21**, 1087 (1953).
2. In *International Thermonuclear Experimental Reactor (ITER)* (Int. Atomic Energy Agency, Vienna, 1990–1993).
3. P. Debye and E. Huckel, Phys. Z. **24**, 185 (1923).
4. J. E. Mayer, J. Chem. Phys. **18**, 1426 (1950).
5. E. Salpeter, Ann. Phys. **5**, 183 (1958).
6. E. J. Meeron, J. Chem. Phys. **28**, 630 (1958).
7. G. Stell and J. L. Lebowitz, J. Chem. Phys. **49**, 3706 (1968).

8. H. L. Friedman, Mol. Phys. **2**, 23 (1959).
9. G. A. Martynov, Usp. Fiz. Nauk **91**, 455 (1967) [Sov. Phys. Usp. **10**, 171 (1967)].
10. J. G. Kirkwood and J. C. Poirier, J. Phys. Chem. **58**, 591 (1954).
11. J. C. Rasaiah and H. L. Friedman, J. Chem. Phys. **48**, 2742 (1968); **50**, 3965 (1969).
12. D. D. Carley, J. Chem. Phys. **46**, 3783 (1967).
13. D. Henderson and L. Blum, Mol. Phys. **40**, 1509 (1980).
14. E. Waisman and J. L. Lebowitz, J. Chem. Phys. **52**, 4307 (1970).
15. V. M. Zamalin, G. E. Norman, and V. S. Filinov, *The Monte Carlo Method in Statistical Thermodynamics* (Nauka, Moscow, 1977).
16. P. N. Vorontsov-Vel'yaminov, A. M. El'yashevich, and A. K. Kron, Élektrokhimiya **2**, 708 (1966).
17. P. N. Vorontsov-Vel'yaminov and A. M. El'yashevich, Élektrokhimiya **4**, 1430 (1968).
18. D. N. Card and J. P. Valleau, J. Chem. Phys. **52**, 6232 (1970).
19. P. N. Vorontsov-Veliaminov, A. M. Eliashevich, J. C. Rasaiah, and H. L. Friedman, J. Chem. Phys. **52**, 1013 (1970).
20. D. J. Adams and I. R. McDonald, J. Phys. C **7**, 2761 (1974).
21. Ch. Margheritis and C. Sinistri, Z. Naturforsch. A **30**, 83 (1975).
22. G. Ciccotti and G. Jacucci, Phys. Rev. A **13**, 426 (1976).
23. J. C. Rasaiah, D. N. Card, and J. P. Valleau, J. Chem. Phys. **56**, 248 (1972).
24. V. P. Chasovskikh and P. N. Vorontsov-Vel'yaminov, Teplofiz. Vys. Temp. **14**, 379 (1976).
25. P. N. Vorontsov-Vel'yaminov and V. P. Chasovskikh, Vestn. Leningr. Univ. **10** (2), 30 (1975).
26. E. V. Zelener, G. É. Norman, and V. S. Filinov, Teplofiz. Vys. Temp. **11**, 922 (1973).
27. P. N. Vorontsov-Vel'yaminov and V. K. Shiff, Teplofiz. Vys. Temp. **15**, 1137 (1977).
28. P. N. Vorontsov-Vel'yaminov, A. M. El'yashevich, L. A. Morgenshtern, and V. P. Chasovskikh, Teplofiz. Vys. Temp. **8**, 277 (1970).
29. P. N. Vorontsov-Vel'yaminov and V. P. Chasovskikh, Teplofiz. Vys. Temp. **13**, 1153 (1975).
30. P. N. Vorontsov-Vel'yaminov and V. K. Shiff, Available from VINITI No. 3783-78 (1978).
31. P. N. Vorontsov-Vel'yaminov and V. K. Shiff, Available from VINITI No. 2086-79 (1979).
32. V. K. Shiff, Available from No. 3074-80 (1980).
33. V. K. Shiff, Teplofiz. Vys. Temp. **24**, 1020 (1986).
34. V. K. Shiff, Teplofiz. Vys. Temp. **26**, 1072 (1988).
35. D. Wolf, P. Keblinski, S. R. Phillpot, and J. Eggebrecht, J. Chem. Phys. **110**, 8254 (1999).
36. J. M. Caillol, J. Chem. Phys. **99**, 8953 (1993); **111**, 6528 (1999).
37. C. Caccamo and G. Malescio, J. Chem. Phys. **90**, 1091 (1989).
38. J. P. Valleau, J. Chem. Phys. **95**, 584 (1991).
39. J. P. Valleau, J. Comp. Phys. **96**, 193 (1991).
40. J. M. Caillol, J. Chem. Phys. **100**, 2161 (1994).
41. Y. Guissani and B. Guillot, J. Chem. Phys. **101**, 490 (1994).
42. G. Orkoulas and A. Z. Panagiotopoulos, J. Chem. Phys. **101**, 1452 (1994).
43. G. Manificat and J.-M. Caillol, J. Chem. Phys. **103**, 4266 (1995).
44. Ph. J. Camp and G. N. Patey, J. Chem. Phys. **111**, 9000 (1999).
45. P. N. Vorontsov-Vel'yaminov and S. V. Shevkunov, Fiz. Plazmy **4**, 1354 (1978) [Sov. J. Plasma Phys. **4**, 756 (1978)].
46. S. V. Shevkunov and P. N. Vorontsov-Vel'yaminov, Teplofiz. Vys. Temp. **21**, 625 (1983).
47. S. V. Shevkunov and P. N. Vorontsov-Vel'yaminov, Teplofiz. Vys. Temp. **20**, 1025 (1982).
48. S. V. Shevkunov and P. N. Vorontsov-Vel'yaminov, Available from VINITI No. 2652-79 (1979).
49. S. V. Shevkunov and P. N. Vorontsov-Vel'yaminov, Khim. Fiz. **1**, 83 (1983).
50. S. V. Shevkunov, P. N. Vorontsov-Vel'yaminov, and N. B. Gromova, Teplofiz. Vys. Temp. **24**, 998 (1986).
51. A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, Mol. Simul. **6**, 143 (1991).
52. S. V. Shevkunov, Teplofiz. Vys. Temp. **31**, 704 (1993).
53. S. V. Shevkunov, Zh. Éksp. Teor. Fiz. **104**, 3032 (1993) [JETP **77**, 413 (1993)].
54. J.-M. Caillol and J.-J. Weis, J. Chem. Phys. **102**, 7610 (1995).
55. A. Matro and D. L. Freeman, J. Chem. Phys. **104**, 8690 (1996).
56. K. S. Pitzer, J. Chem. Phys. **104**, 6724 (1996).
57. R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965; Mir, Moscow, 1968).
58. L. D. Fosdick and H. F. Jordan, Phys. Rev. **143**, 58 (1966).
59. H. F. Jordan and L. D. Fosdick, Phys. Rev. **171**, 128 (1968).
60. V. M. Zamalin and G. É. Norman, Zh. Vychisl. Mat. Mat. Fiz. **13**, 408 (1973).
61. V. S. Filinov, Teplofiz. Vys. Temp. **11**, 871 (1973); **13**, 251 (1975); **14**, 245 (1976).
62. V. S. Filinov, Phys. Lett. A **54**, 259 (1975).
63. V. S. Filinov and G. E. Norman, Phys. Lett. A **55**, 219 (1975).
64. K. S. Schweizer, R. M. Stratt, D. Chandler, and P. G. Wolynes, J. Chem. Phys. **75**, 1347 (1981).
65. M. Imada, J. Phys. Soc. Jpn. **53**, 2861 (1984).
66. D. Chandler and P. G. Wolynes, J. Chem. Phys. **74**, 4078 (1981).
67. E. L. Pollock and D. M. Ceperley, Phys. Rev. B **30**, 2555 (1984).
68. S. R. White and J. W. Wilkins, Phys. Rev. B **37**, 5024 (1988).
69. H. Q. Lin and J. E. Hirsch, Phys. Rev. B **34**, 1964 (1986).
70. Ch. Chakravarty, J. Chem. Phys. **99**, 8038 (1993).
71. P. Beran, Phys. Rev. B **59**, 9725 (1999).
72. M. Skorobogatiy and J. D. Joannopoulos, Phys. Rev. B **60**, 1433 (1999).

73. P.-N. Roy and G. A. Voth, J. Chem. Phys. **110**, 3647 (1999).
74. P.-N. Roy, S. Jang, and G. A. Voth, J. Chem. Phys. **111**, 5303 (1999).
75. E. V. Shuryak and O. V. Zhirov, Nucl. Phys. B **242**, 393 (1984).
76. L. M. Sese and R. Ledesma, J. Chem. Phys. **102**, 3776 (1995).
77. C. Chakravarty, J. Chem. Phys. **102**, 956 (1995).
78. Y. J. Wong and G. V. Chester, Phys. Rev. B **37**, 9590 (1988).
79. A. P. Lyubartsev and P. N. Vorontsov-Velyaminov, Phys. Rev. A **48**, 4075 (1993).
80. S. V. Shevkunov, Available from VINITI No. 6370-B87 (1987).
81. S. V. Shevkunov, Zh. Éksp. Teor. Fiz. **118**, 36 (2000) [JETP **91**, 31 (2000)].
82. S. V. Shevkunov, in *Modern Problems of Statistical Physics*, Ed. by I. R. Yukhnovskiĭ (Naukova Dumka, Kiev, 1989), Vol. 1, p. 379.
83. S. V. Shevkunov, Dokl. Akad. Nauk **369**, 43 (1999) [Dokl. Phys. **44**, 730 (1999)].
84. S. V. Shevkunov, Teplofiz. Vys. Temp. **29**, 45 (1991).
85. S. V. Shevkunov and P. N. Vorontsov-Velyaminov, Mol. Simul. **7**, 249 (1991).
86. H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950; Nauka, Moscow, 1983).
87. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1974; Pergamon, New York, 1977).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 966–976.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1124–1138.
Original Russian Text Copyright © 2002 by Skobelev, Faenov, Magunov, Pikuz, Boldarev, Gasilov, Abdallach, Jr., Junkel-Vives, Auguste, Dobosz, d'Oliveira, Hulin, Monot, Blasco, Dorchies, Caillaud, Bonte, Stenz, Salin, Loboda, Litvinenko, Popova, Baidin, Sharkov.

PLASMA, GASES

X-ray Spectroscopy Diagnostic of a Plasma Produced by Femtosecond Laser Pulses Irradiating a Cluster Target

I. Yu. Skobelev[a, *], A. Ya. Faenov[a], A. I. Magunov[a], T. A. Pikuz[a], A. S. Boldarev[b], V. A. Gasilov[b], J. Abdallach, Jr.[c], G. C. Junkel-Vives[†c], T. Auguste[d], S. Dobosz[d], P. d'Oliveira[d], S. Hulin[d], P. Monot[d], F. Blasco[e], F. Dorchies[e], T. Caillaud[e], C. Bonte[e], C. Stenz[e], F. Salin[e], P. A. Loboda[f], I. A. Litvinenko[f], V. V. Popova[f], G. V. Baidin[f], and B. Yu. Sharkov[g]

[a]*Center of Data on Spectra of Multiply Charged Ions, All-Russia Research Institute of Physicotechnical and Radio Engineering Measurements, Mendeleevo, Moscow oblast, 141570 Russia*
[b]*Institute for Mathematical Modeling, Russian Academy of Sciences, Moscow, 125047 Russia*
[c]*Los Alamos National Laboratory, P. O. Box 1663, Los Alamos, New Mexico 87545, USA*
[d]*CEA, Centre D'Etudes de Saclay, DSM/DRECAM, Service des Photons, Atomes et Molécules, 91191 Gif-sur-Yvette, France*
[e]*CELIA, Université Bordeaux, 33405 Talence, France*
[f]*Russian Federal Nuclear Center, All-Russia Research Institute of Technical Physics, Snezhinsk, Chelyabinsk oblast, 456770 Russia*
[g]*Institute of Theoretical and Experimental Physics, Russian Academy of Sciences, Moscow, 117259 Russia*
**e-mail: skobelev@orc.ru*

Received December 26, 2001

Abstract—The parameters of a plasma produced upon the interaction of ultrashort laser pulses with cluster targets are measured by the methods of X-ray spectroscopy. The dependence of the plasma parameters on the initial properties of a cluster target (the design of a supersonic nozzle, the average size of clusters, the spatial inhomogeneity) and the laser pulse properties (its duration and contrast) is studied. The plasma diagnostics is performed using the model of formation of emission spectra, which was proposed earlier and includes a number of fitting parameters, which provide good agreement with experimental spectra. The systematic experimental studies performed by us showed that our model of cluster heating by ultrashort pulses is indeed a physical model, and the fitting parameters represent the average values of plasma parameters in the corresponding space–time regions. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The interaction of high-power ultrashort laser pulses with clusters attracts considerable recent attention [1–9]. The production of a plasma in cluster targets combines the main mechanisms of plasma formation in gaseous and solid targets irradiated by lasers (ionization by an optical field and resonance absorption). It is known that ultrashort laser pulses are strongly absorbed by cluster targets. This property is especially important for efficient generation of X-rays, which can be used in microlithography, medicine, and biology. Unfortunately, the experimental results that have been obtained so far are not systematic because they depend on many parameters of a laser pulse and irradiated targets. This complicates the analysis of various physical processes in the plasma being produced and a detailed verification of new theoretical models [10, 11] that take into account the specific features of plasma formation by ultrashort pulses.

We have considered the numerical model of cluster formation in a supersonic gas jet in papers [12–14]. It has been shown that a detailed study of two-phase gasdynamic processes in a nozzle forming the jet allows us to obtain spatial distributions for all parameters of the clusters, which are required for the correct calculation of the cluster heating by ultrashort laser pulses. The calculations performed for nozzles of different designs at different gas pressures showed that a virtually homogeneous cluster target can be formed in some cases, whereas in other cases the distributions are not only inhomogeneous but also even nonmonotonic.

We proposed in these papers a simple physical model of the cluster target heating by femtosecond laser pulses with a picosecond prepulse, which allows us to calculate rather simply the X-ray emission spectrum of the plasma. The model is valid if the condition

$$\tau_{\text{prepulse}} \leq \tau_{\text{expansion}} \quad (1)$$

is fulfilled, where τ_{prepulse} is the duration of the laser prepulse and $\tau_{\text{expansion}}$ is the cluster lifetime. The condi-

[†]Deceased.

tion (1) means that clusters are destroyed only partially during the prepulse, and when the main femtosecond pulse arrives, the dense regions that continue to absorb laser radiation remain in the produced preplasma. The electron temperature in these regions drastically increases under the action of the main pulse (up to several kiloelectronvolts for the laser power density of 10^{18} W/cm^2), and after the end of the main pulse, the hot electrons partially penetrate into a much colder intercluster plasma with the electron temperature of 100–200 eV. Therefore, the X-ray emission spectrum of the plasma produced in the regime $\tau_{\text{prepulse}} \sim \tau_{\text{expansion}}$ should be calculated taking into account the presence of less dense (with the electron density N_{e1} lower than the critical density $N_{e,\,\text{cr}}$) and more dense (with the electron density $N_{e2} \geq N_{e,\,\text{cr}}$) plasma regions heated to different temperatures ($T_{e1} < T_{e2}$). In each of the regions, a fraction f_i of hot electrons will be present with the mean energy E_0, which substantially exceeds temperature T_{e1} and T_{e2}, the relative amount f_2 of hot electrons in the dense region being greater than their amount f_1 in the rarefied plasma region. The simple estimates of the collision ionization probability show that, for multiply charged ions with $Z \sim 10$–20 and typical parameters of femtosecond lasers ($\tau_{\text{las}} \sim 30$–$60$ fs, $q_{\text{las}} \sim 10^{17}$–$10^{18}$ W/cm^2, contrast $\sim 10^5$), the ionization state of the plasma in regions 1 and 2 will approximately correspond to its electron temperature, whereas the influence of hot electrons on the ionization state can be neglected in the first approximation. This means that the time-integrated emission spectrum of the plasma calculated within the framework of this model should depend on eight free parameters N_{e1}, N_{e2}, T_{e1}, T_{e2}, f_1, f_2, E_0, and α (the ratio of contributions from regions 1 and 2 to the total spectrum) whose values can be found from the best fit of the experimental spectrum. According to physical concepts about the plasma formation, we can expect that the values of these parameters should satisfy the conditions

$$N_{e1} < N_{e,\,\text{cr}} \leq N_{e2}, \quad T_{e1} < T_{e2} \ll E_0, \quad f_1 < f_2 \ll 1.$$

Moreover, because the temperature T_{e1} is attained due to the action of the prepulse with a typical flux density $\sim 10^{12}$–10^{13} W/cm^2, its value should be 100–200 eV, while the energy E_0 of hot electrons produced during the action of the main pulse should be of the order of kiloelectronvolts. Note that the correspondence of the ionization state of the plasma to temperatures T_{e1} and T_{e2} is a rather crude approximation, whose applicability depends on the values of electron densities N_{e1} and N_{e2}. When the plasma density is low, the ionization state of the plasma has no time to achieve a stationary value during the action of the prepulse, and the values of T_{e1} and T_{e2} used in the calculation of the emission spectrum will correspond to the ionization temperature of the plasma rather than to the electron temperature. Because $N_{e1} < N_{e2}$, this remark refers in a greater extent to plasma region 1 and, hence, to the physical meaning of the temperature T_{e1}.

In this paper, we used the model described above (see details in [13, 14]) to interpret experiments performed with supersonic nozzles of different types using heating laser pulses with different parameters. In this way, we studied the dependence of the parameters of the plasma being produced on the initial properties of a cluster target (mean cluster size, spatial inhomogeneity) and on the properties of the laser pulse (its duration and contrast). In most experiments described below, we used argon as a working target substance; in some cases, other gases were also used (CO_2, Kr, and Xe).

2. EXPERIMENTAL SETUPS AND METHODS OF CALCULATION

We used in our experiments two femtosecond laser setups at Centre D'Etudes de Saclay (France) and Université Bordeaux (France).

In the first setup, a 0.8-μm, 10 TW UHI10 Ti:sapphire laser was used [12]. The duration and power of the main pulse were 60 fs and 0.6 J, respectively. The duration of the prepulse was about 1 ps, and its contrast was of the order of 10^5. The laser beam was focused on a cluster target with an off-axis parabolic mirror. The beam diameter in the focal plane was about 25 μm, which provided the laser power density on the target up to 10^{18} W/cm^2.

The second setup also used a Ti:sapphire laser, which had a lower power (the main pulse energy and duration were 0.015 J and 20 fs, respectively) [15]. The duration of the main pulse in this laser could be varied in a broad range from 20 fs to several picoseconds. The pulse contrast also could be varied from 10 to 10^6. An off-axis parabolic mirror focused laser radiation into a spot of diameter 12 μm, providing the power density up to 4×10^{17} W/cm^2.

In both cases, two types of a supersonic nozzle were used: the Laval nozzle and a cone nozzle. The gas pressure in a valve could be varied up to 100 atm. By using different pressures and different nozzles, we could vary in a broad range both the average size of clusters and the spatial distribution of their concentration in the interaction region.

The X-ray emission of the plasma was detected simultaneously with several spectrographs with spherical mica crystals (radii of curvature were 100 and 150 mm). The spectrographs were arranged in the FSPR-2D scheme (see [16–18]) and were tuned to the spectral ranges containing resonance transitions in the He- and H-like Ar XVII and Ar XVIII ions. The spectral resolving power was $\lambda/\delta\lambda \sim 4000$ for a spatial resolution better than 80 μm. The spectra were recorded either on a photographic film or using a CCD camera.

The parameters of cluster targets were calculated using the codes developed at the Institute for Mathematical Modeling, Russian Academy of Sciences, which were described in detail in papers [13, 19–21].

Kinetic calculations used for simulating the emission spectra of plasma were performed employing the codes developed at the Los Alamos National Laboratory (USA). In these calculations, multiply charged argon ions with the number of bound electrons up to $m = 4$ were taken into account and all the atomic configurations with principal quantum numbers $n < 6$ were considered, including autoionization states. The kinetic matrix included 1500 levels and took into account all possible radiative, collision, and autoionization processes. The rates of collision processes were calculated using the model electron distribution function [22, 23], which included the hot component with the average energy $E_0 = 5$ keV. The emission spectra were calculated for experimental spectral regions using kinetic calculations performed within the framework of the simplified model of the plasma dynamics described above. The shape of spectral lines was either assumed to be instrumental or calculated taking into account the Stark and Doppler broadening.

The shapes of the spectral lines of argon ions were determined taking into account the Stark shift in the ion microfield, the collision broadening caused by elastic collisions of electrons with ions, and the Doppler broadening [24]. The latter was taken into account together with the spectral resolution by introducing the effective ion temperature. The distribution function of the ion microfield was used taking into account ion correlations and the Debye screening [25–28].

3. EFFECT OF THE PROPERTIES OF A CLUSTER TARGET ON THE PLASMA PARAMETERS

The calculations of formation of clusters in a supersonic jet performed in this paper show that the main factors affecting the properties of the cluster target being produced, along with the type of gas used, are the pressure in the valve and the type of nozzle. Note that, while the initial gas pressure affects the properties of clusters quite obviously (the average size of clusters increases with pressure), a change in the nozzle design can lead to less obvious results. We performed calculations for nozzles of two types: a Laval nozzle, which is commonly used for formation of cluster-containing jets, and a cone nozzle. We found that, all other factors being the same, the cone nozzle allows one to obtain considerably larger clusters, which are more uniformly distributed over across the gas jet.

The initial properties of the cluster target should affect the radiative characteristics of the laser-induced plasma. In this section, we consider the influence of the target inhomogeneity (Section 3.1) and of the average size of clusters (Section 3.2) on the properties of cluster laser plasma.

3.1 Effect of the Inhomogeneous Distribution of the Cluster Concentration in a Gas Jet on the Radiative Properties of Laser Plasma

Figure 1 shows the results of calculations of cluster parameters in a gas jet emerging from a cone nozzle and a Laval nozzle. The coordinate r in this figure is directed across the gas jet, i.e., across the propagation direction of the laser pulse in the laser–cluster experiment. One can see from the figure that, when the Laval nozzle is used, we have a target that is strongly inhomogeneous along the laser beam: the distributions of cluster concentration and of the mean density of cluster atoms have local minima at the gas-jet axis ($r = 0$), and their maximum values are achieved at the distance $r \sim 1$ mm from the axis. This means that the laser pulse propagating perpendicular to the gas-jet axis passes through two regions with the maximum cluster density between which the region of low cluster density is located. According to our calculations, the distance between the dense regions should be about 2 mm. The heating of such an inhomogeneous target should result in the formation of a plasma with an inhomogeneous distributions of density and temperature and, hence, with an inhomogeneous emission.

The inhomogeneity of the plasma emission should correspond qualitatively to the initial distribution of the concentration of clusters, which determines the efficiency of their interaction with the laser pulse. However, because the plasma emission depends not only on its density but also on its temperature, while the temperature depends on the laser power density, upon focusing the laser beam into the gas jet, the laser power density will decrease with distance from the point $r = 0$. For this reason, the plasma emission distribution will be more strongly "pressed down" to the jet axis than the cluster concentration distribution. In other words, the distance between two emission maxima should be somewhat smaller than the distance between the maxima of the cluster concentration.

These arguments are illustrated by the experimental results that were obtained with Laval and cone nozzles used in setups at Bordeaux and Saclay [15, 29]. In these experiments, Kr, Xe, and CO_2 were used as working gases along with Ar. The plasma emission was detected with a spatial resolution in different spectral lines using X-ray pinhole cameras, and its continuous bremsstrahlung and photorecombination emission was also studied. The experimental results are presented in Fig. 2. One can see that the spatial distribution of the plasma emission is indeed nonmonotonic and the distance between the maxima is approximately 0.5–1 mm, which is somewhat smaller than the distance between the maxima in the cluster concentration distribution. Note that the coincidence of the results obtained for the different spectral lines of different ions means that nonmonotonic emission is related to the general parameters of the plasma such as its temperature and density and

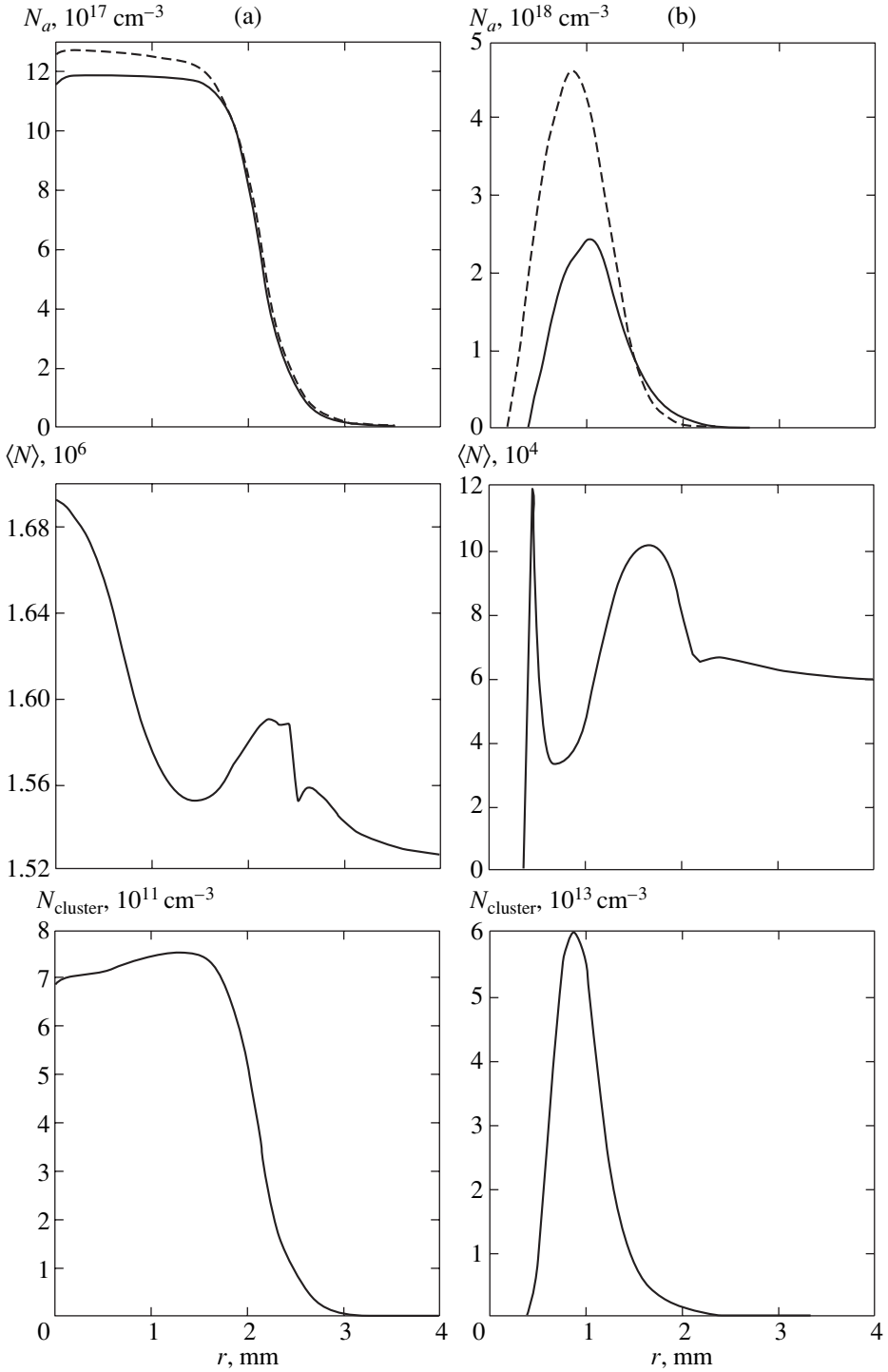

Fig. 1. Calculations of the cluster formation for (a) a cone nozzle and (b) a Laval nozzle at a distance of 1.5 mm from the nozzle outlet. Presented are the density N_a of cluster atoms in a target, the average number $\langle N \rangle$ of atoms in a cluster, and the density N_{cluster} for argon (solid curves) and krypton (dashed curves). The initial gas pressure in a valve is 40 atm.

does not reflect the specific features of the sublevel kinetics of a specific ion.

The gas-dynamic calculations of the cluster formation in a cone nozzle (Fig. 1) showed that the cluster distribution at the nozzle output is virtually homogeneous. In this case, the plasma emission should be maximal at the gas-jet axis (where the laser power density is maximal) and will continuously decrease with the distance from the point $r = 0$. This was also confirmed experimentally. The corresponding spectrograms and densitograms are presented in Fig. 2.

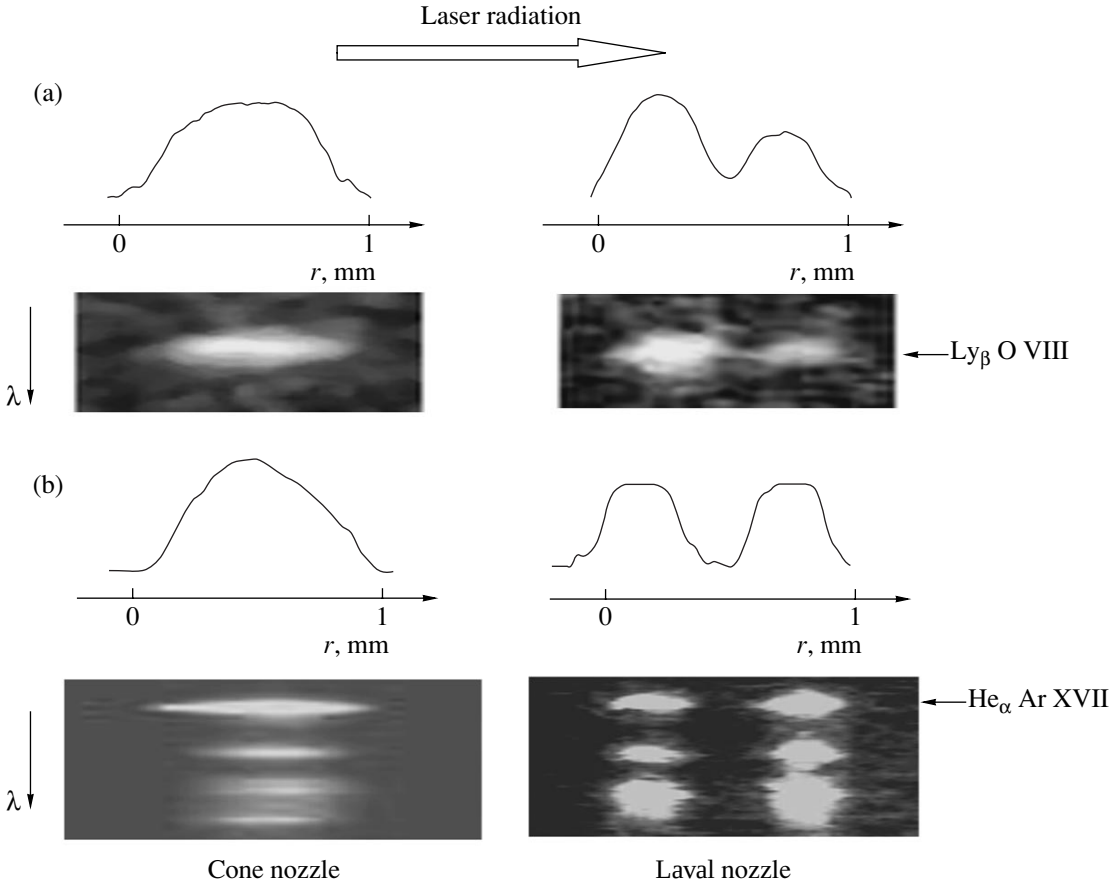

Fig. 2. Spatial distributions of plasma emission measured using a spectrograph with a one-dimensional spatial resolution for the Laval and cone nozzles in experiments with (a) CO_2 and (b) argon clusters.

It follows from the theoretical and experimental studies considered above that the use of a cone nozzle for the cluster target formation is preferable in most cases because, first, a more homogeneous target can be produced and, second, clusters of considerably larger sizes can be obtained. In this case, by using long-focus objectives, we can easily obtain a very long (~4–6 mm) plasma object. Figure 3 shows, for example, the results obtained for CO_2 and Xe clusters using interferometry and an X-ray pinhole camera. One can see that plasma stripes had the length of about 4 mm and the width less than 0.7 mm.

3.2. Effect of the Average Size of Clusters on Plasma Parameters

The simplified model of the laser–cluster interaction considered above allows us to make some qualitative conclusions about the dependence of plasma parameters on the average size of clusters. According to this model, an increase in the cluster size should first of all result in the increase in the role of dense plasma regions because now the laser prepulse will destroy a smaller part of the cluster. This should lead, first, to the increase in the fraction of hot electrons, which are formed upon the interaction of the main femtosecond laser pulse with the dense plasma and, second, to the rise of the average density N_{e2} and, to a lesser extent, N_{e1}. In turn, the increase in the average density will reduce the time of establishment of the ionization equilibrium, and the ionization temperature will approach the electron temperature with increasing N_{e2} and N_{e1}; i.e., it will also increase.

The qualitative conclusions made above are confirmed by the results of numerical simulation of the emission spectra of the argon plasma produced upon the interaction of a femtosecond laser pulse with clusters of different sizes [30]. To obtain the most homogeneous cluster target, a cone nozzle was used, and the average cluster size was changed by changing the initial gas pressure in a valve in the range from 15 to 100 atm. The calculations of the cluster formation showed that the average number of atoms in the cluster was varied in this case from 2×10^5 to 2×10^7. The plasma parameters, which were determined by fitting the experimental spectra to the model spectra, are presented in Table 1, while the quality of fitting is well illustrated by Fig. 4, where the experimental spectra obtained at pressures of 30 and 100 atm are shown together with the model spectra.

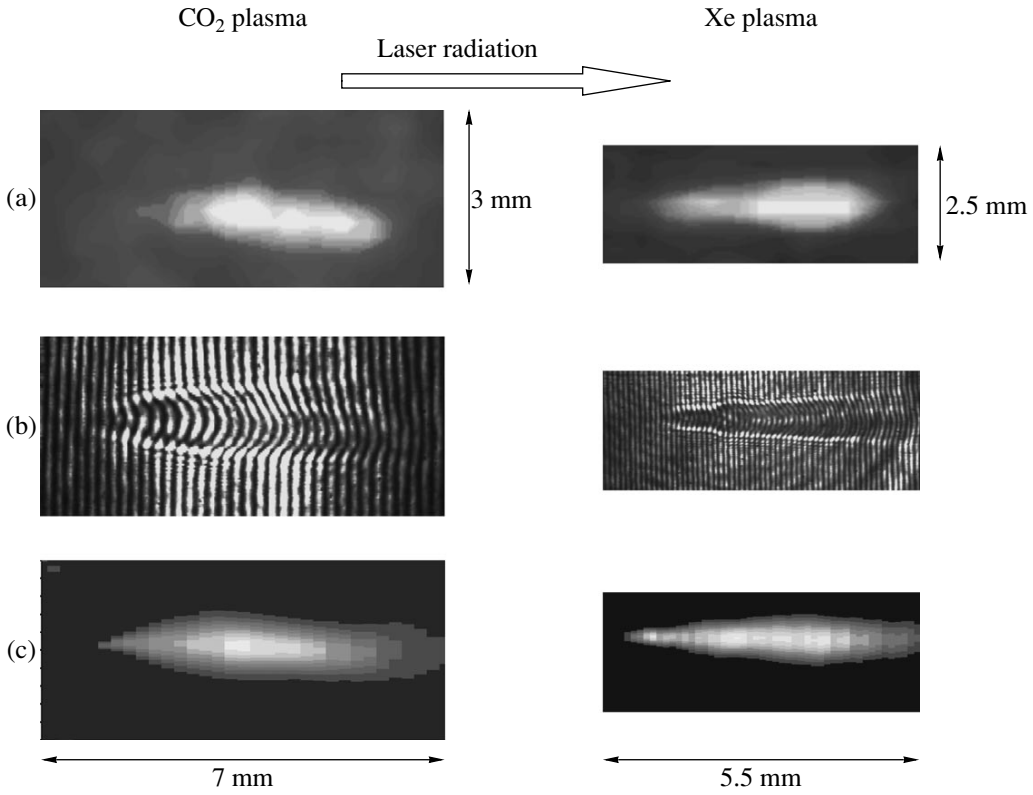

Fig. 3. Cluster plasma produced using a long-focus objective and a cone nozzle: (a) X-ray pinhole images; (b) interference patterns; (c) electron density distribution.

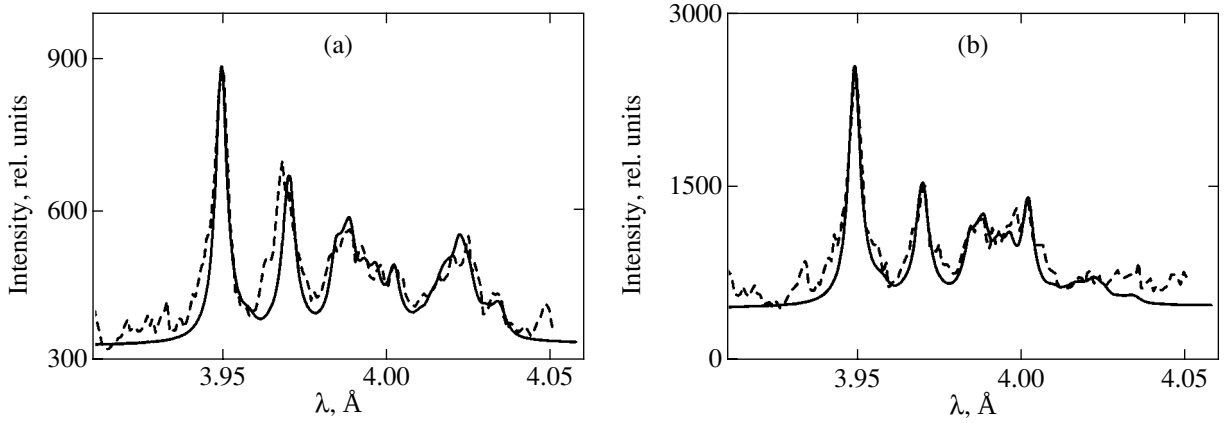

Fig. 4. Simulations of spectra in the region of the He_α line of Ar XVII ion at gas pressures of (a) 30 and (b) 100 atm (the values of plasma parameters are presented in Table 1). The solid curves are model spectra, and the dashed curves are experimental spectra.

One can see from Table 1 that at the initial gas pressure equal to 100 atm, a large amount of hot electrons are present in dense plasma regions (at a concentration of 2×10^{19} cm^{-3}). Because the ionization rate of the He-like Ar XVII ion by such electrons is (for $E_0 = 5$ keV)

$$\langle v\sigma_{ion}(\text{Ar XVII} - \text{Ar XVIII})\rangle \sim 10^{-11}\ \text{cm}^3\ \text{s}^{-1},$$

a noticeable amount of the H-like Ar XVIII ions should be formed in the dense plasma regions during their lifetime of the order of 1 ps:

$$N(\text{Ar XVIII})/N(\text{Ar XVII}) \approx 2 \times 10^{-4}.$$

These ions can be detected by observing the Ly_α emission line. Because the rates of excitation of the

Table 1. Effect of the average size of argon clusters on the plasma parameters

Gas pressure, atm	Average number of atoms in a cluster	Rarefied plasma (region 1)			Dense plasma (region 2)		
		T_{e1}, eV	N_{e1}, cm^{-3}	f_1N_{e1}, cm^{-3}	T_{e2}, eV	N_{e2}, cm^{-3}	f_2N_{e2}, cm^{-3}
15	2×10^5	100	5×10^{19}	1.5×10^{13}	215	3×10^{20}	0.9×10^{17}
30	10^6	160	5×10^{19}	5×10^{13}	230	5×10^{20}	3.5×10^{17}
70	7×10^6	140	2×10^{20}	2×10^{14}	230	2×10^{21}	6×10^{17}
100	1.6×10^7	195	1.5×10^{21}	1.2×10^{17}	400	2×10^{21}	2×10^{19}

$1s^2\,{}^1S_0$–$1s2p^1P_1$ and $1s^2S$–$2p^2P$ transitions by hot electrons are of the same order of magnitude, the intensity of the Ly_α line will be very low (approximately 2×10^{-4} of the He_α line intensity), but nevertheless it is sufficient for detecting this line with a CCD camera having a broad dynamic range. In this connection, we preformed special measurements [31] of the emission spectrum in the region $\lambda = 3.7$–3.8 Å (the wavelengths of the $Ly_{\alpha1}$ and $Ly_{\alpha2}$ components of the Ar XVIII ion are 3.73110 and 3.73652 Å, respectively). In complete agreement with the estimates made above for the initial gas pressure equal to 100 atm, we detected a rather weak resonance emission line of the H-like argon (Fig. 5). The intensities of the Ly_α and He_α lines were 0.2 and 2000 relative units; i.e., their ratio was approximately equal to 10^{-4}, which is very close to the expected ratio of the concentrations of the H- and He-like ions. We failed to detect the Ly_α line at lower gas pressures, i.e., for lower cluster sizes.

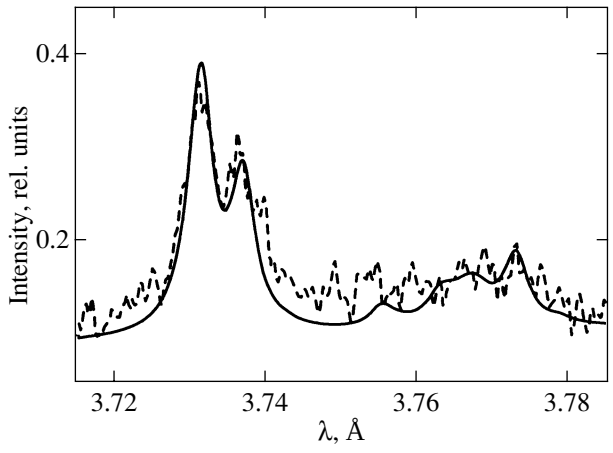

Fig. 5. Spectrum of the resonance Ly_α emission line of the H-like Ar XVIII ion observed at a gas pressure of 100 atm (dashed curve) and the model spectrum (solid curve) for the plasma parameters presented in Table 1.

4. EFFECT OF THE LASER PULSE PROPERTIES ON PLASMA CHARACTERISTICS

In this section, we consider the influence of the duration and contrast of a laser pulse on the parameters of the laser-induced plasma.

4.1. Dependence of the Plasma Parameters on the Contrast of a Femtosecond Laser Pulse

The simplified model of the laser–cluster interaction predicts that a change in the laser pulse contrast will affect first of all the temperature of the plasma and the density of its densest regions (i.e., the value of N_{e2}). The increase in the contrast should obviously be accompanied by a decrease in the temperature because the laser prepulse power density decreases, whereas the value of N_{e2} should increase because the prepulse will destroy a smaller part of the cluster.

To verify these qualitative predictions, we performed a series of experiments in which the energy and duration of the main femtosecond pulse and the average size of clusters were fixed, while the pulse contrast was varied in a broad range

$$q_{\text{pluse}}/q_{\text{prepluse}} = 5 \times 10^1\text{–}5 \times 10^3$$

[32]. The duration of the main pulse was 45 fs. The clusters were formed using a cone nozzle with an initial gas pressure of 63 atm; the interaction region was located at a distance of 1.5 mm from the nozzle outlet. The parameters of the plasma, as in the previous case, were determined by fitting the emission spectrum of plasma in the region of the He_α Ar XVII line by the model spectrum. One can see from the results presented in Table 2 that the plasma temperature monotonically decreases with increasing pulse contrast, while the density N_{e2} increases and exceeds the critical density $N_{e,\,\text{cr}} = 1.7 \times 10^{21}$ cm^{-3} when the pulse contrast $\geq 10^3$. Note that the latter result agrees with the case of the interaction of ultrashort pulses with solid targets [33–35] when the supercritical-density plasma was also observed only at high contrasts of laser pulses.

Table 2. Effect of the laser pulse contrast on the plasma parameters

Contrast	500	100	350	1000	5000	⩾5000
T_{e2}, eV	250	222	215	215	205	215
N_{e2}, cm^{-3}	3×10^{20}	5×10^{20}	2×10^{21}	5×10^{21}	7×10^{21}	2×10^{22}
f_2N_{e2}, cm^{-3}	5.1×10^{17}	3.5×10^{17}	1.4×10^{18}	8.5×10^{17}	1.2×10^{18}	3.4×10^{18}

Table 3. Effect of the laser pulse duration on the plasma parameters

τ_{las}, fs	Rarefied plasma (region 1)			Dense plasma (region 2)		
	T_{e1}, eV	N_{e1}, cm^{-3}	f_1N_{e1}, cm^{-3}	T_{e2}, eV	N_{e2}, cm^{-3}	f_2N_{e2}, cm^{-3}
1100	130	3×10^{19}	3×10^{11}	200	3×10^{21}	9×10^{17}
700	130	3×10^{19}	9×10^{12}	215	4×10^{21}	1.2×10^{18}
45	130	3×10^{19}	3×10^{12}	200	10^{22}	7×10^{17}

4.2. Effect of the Laser Pulse Duration on the X-ray Emission Spectrum and Parameters of a Cluster Plasma

We studied the effect of the pulse duration on the plasma parameters by fixing the pulse energy at 15 mJ, so that the change in the pulse duration from 45 fs to 1.1 ps was accompanied by the change in the power density from 10^{17} to 3×10^{15} W/cm^2. As in the previous case, clusters were produced using a cone nozzle with an initial argon pressure of 60 atm.

The plasma parameters determined by fitting the spectra of the He_α line of Ar XVII and of its dielectronic satellites are presented in Table 3. Analysis of the rates of excitation of levels of multiply charged argon ions by thermal and hot electrons showed that even a small amount of hot 5-keV electrons makes a substantial contributions not only to excitation of the satellites but also to the resonance line itself. Because the energy $E_0 = 5$ keV is of the order of the ionization potential of the He-like Ar XVII ion, such electrons should also excite efficiently the $1snp^1P_1$ Rydberg states of this ion with $n > 2$, whose radiative decay results in emission of the highest terms in the resonance series of the He-like argon ion. We detected such emission lines for $n = 3$–10 in the fifth reflection order of the spectrograph crystal in the 3.0- to 3.4-Å spectral range containing the $1snp^1P_1$–$1s^{21}S_0$ lines with $n > 2$. Figure 6a shows an example of the spectrogram.

By observing the Rydberg transitions, we can perform additional independent diagnostics of plasma because the shape of Rydberg lines strongly depends on the plasma density and to a lesser extent on its temperature. In this paper (see also [36]), we used the Rydberg lines to verify the correctness of determining plasma parameters by the method described above, i.e., from the spectra of the resonance line and its satellites. For this purpose, the profiles of Rydberg transitions were calculated for the plasma parameters presented in Table 3, with an allowance for two spatial regions. The shapes of spectral lines of Ar XVII were calculated taking into account the Stark shift in the ion microfield, the collision broadening caused by elastic collisions of electrons with ions, and the Doppler broadening [24]. The latter was accounted for together with the spectral resolution, which corresponded to the effective ion temperature $T_i = 2$ keV. We used the distribution function of the ion microfield taking into account the ion correlations and Debye screening [25–28].

One can see from Fig. 6b that the results of independent calculations with the fitting parameters of the plasma are in good agreement with the experimental data. These results show that the electron density of the plasma that makes the dominant contribution to the observed spectrum increases with shortening of the laser pulse. In addition, the relative contribution of the subcritical electron density (see also [36]) decreases when shorter pulses are used. As the pulse duration is increased up to 1 ps, laser radiation is absorbed simultaneously with the cluster decay, resulting in a decrease in the dominating electron density of the emitting plasma. The electron temperature remains very low for any pulse duration, and it is insufficient for ionization and efficient excitation of the levels of the He-like argon ion. Therefore, the emission lines can be related only to the Rydberg levels populated due to excitation by a small amount of hot electrons. Because the electron temperature is established in the plasma rather slowly (especially at gas densities), the effect of hot electrons on the plasma relaxation kinetics can be manifested for a longer time than the laser pulse duration.

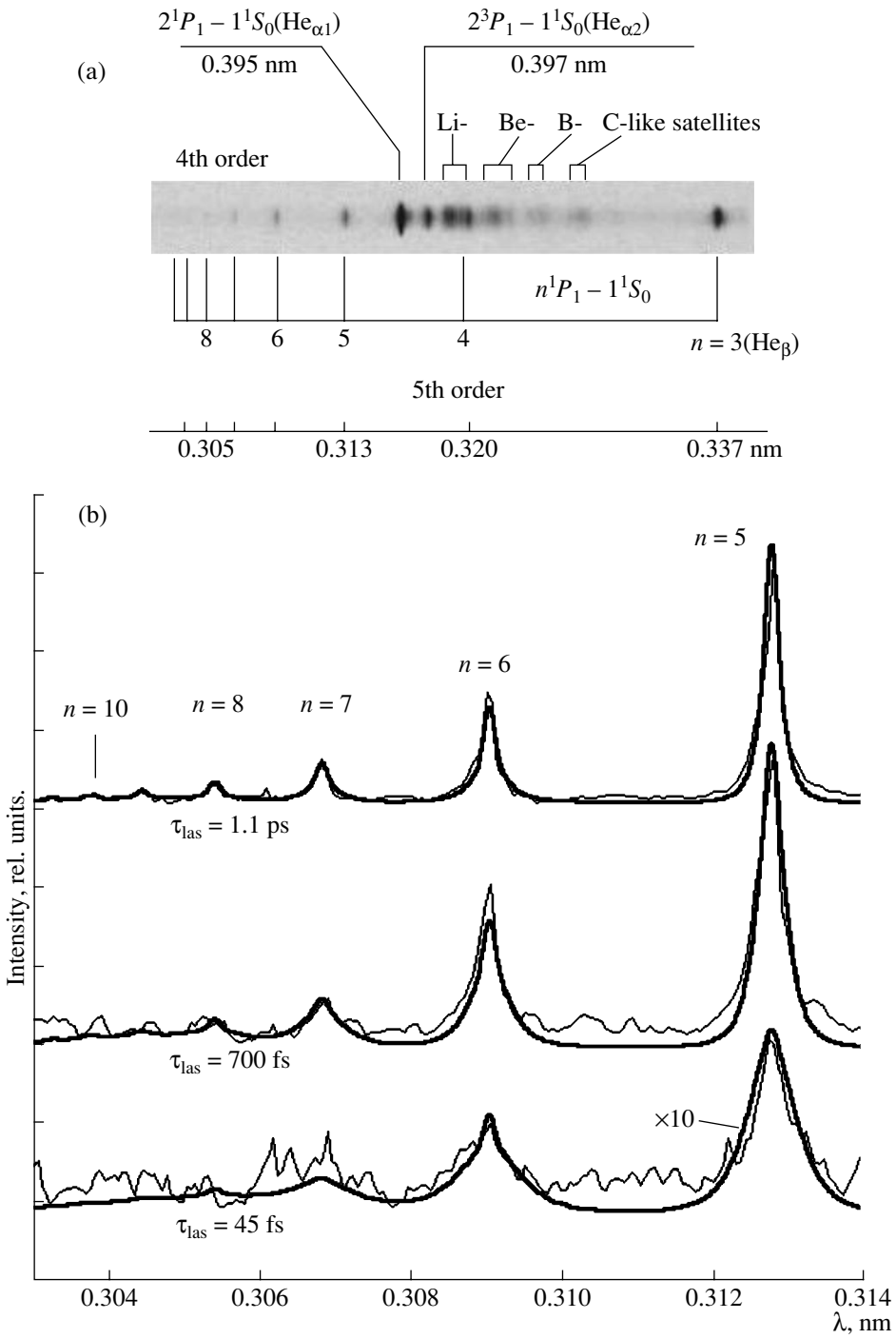

Fig. 6. (a) Spectrogram of Rydberg transitions in the He-like argon ion; (b) comparison of the measured spectrum of the argon plasma (thin curves) with the model spectrum (thick curves) for the n^1P_1–1^1S_0 transitions ($n \geq 5$) in Ar XVII for laser pulse durations of 1.1 ps, 700 fs, and 45 fs.

Therefore, our model of the cluster heating by ultrashort laser pulses correctly describes not only the intensities of resonance lines and their satellites but also the shapes of spectral transitions from the Rydberg levels of the He-like argon ion. The results of the measurements and calculations show the electron density of a plasma increases with shortening of the laser pulse. The X-ray emission is mainly determined by a plasma with a supercritical electron density.

5. CONCLUSIONS

The model [12–14] describing the formation of the emission spectra of a plasma produced upon the inter-

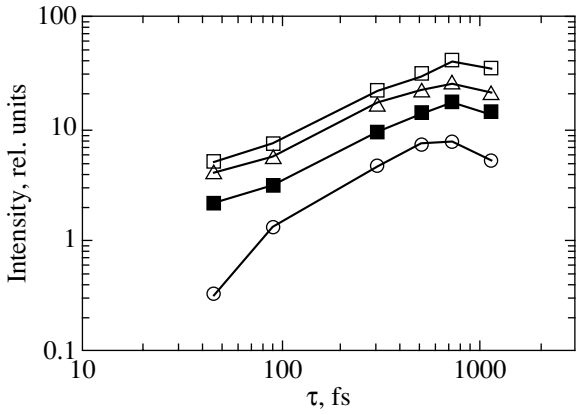

Fig. 7. Dependence of the intensity of the $He_{\alpha 1}$ (□), $He_{\alpha 2}$ (△), He_{β} (■), and He_{δ} (○) emission lines of the He-like Ar XVII ion on the duration of laser pulse with a fixed energy.

action of femtosecond laser pulses with clusters includes a number of fitting parameters, which allow one to reproduce experimental spectra quite accurately. However, for these parameters to have not only mathematical but also physical meaning, their values should vary reasonably depending on the laser–cluster interaction. The systematic experimental study performed in this paper for clusters of different sizes at different durations and contrasts of laser pulses have shown that this is the case. As noted above, the change in the interaction parameters resulted in all cases in the physically reasonable change in the values of fitting parameters. Therefore, our model of the cluster heating by ultrashort pulses is indeed a physical model, and the fitting parameters represent the average values of the plasma parameters in the corresponding space–time regions.

Our study allows us also to make some conclusions concerning the use of a laser–cluster plasma in applied problems.

For example, the optimization of heating for enhancing the brightness of an X-ray lithographic source is especially important in the development of this source. According to the results of our measurements presented in Fig. 7, the brightness is a nonmonotonic function of the laser pulse duration. The position of the maximum brightness depends on the cluster size and the laser pulse contrast. For the experimental conditions of Fig. 7, the optimal pulse duration is 700 fs.

Because the laser–cluster interaction in the $\tau_{prepulse} \sim \tau_{expansion}$ regime allows one to obtain a strongly ionized plasma, such a plasma source can be used to obtain lasing at X-ray transitions in multiply charged Ne- and Ni-like ions where the optimal conditions for obtaining large absolute values of the population inversion are naturally satisfied (see, for example, [37–40]). In this case, it is important to produce a sufficiently extended and homogeneous plasma. As shown in Section 3.1, a homogeneous extended plasma can be obtained by

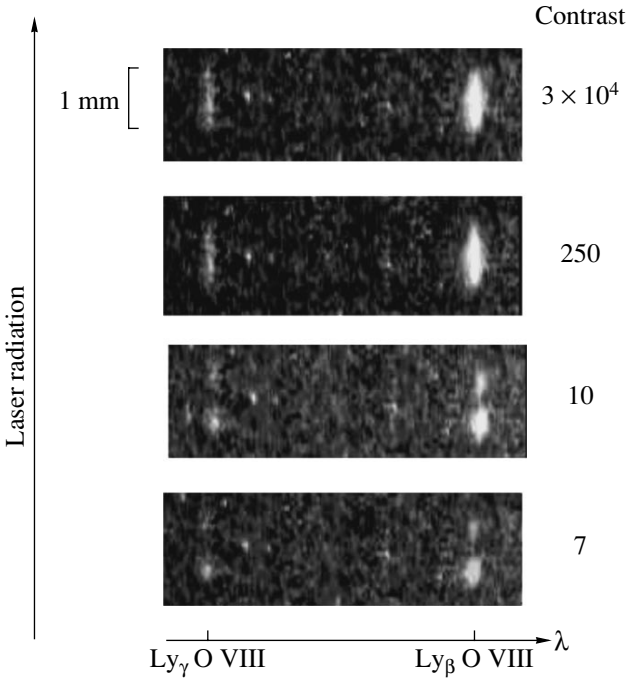

Fig. 8. Spatial distributions of the CO_2 cluster plasma emission observed upon excitation by high- and low-contrast laser pulses.

using a cone nozzle for the formation of a cluster target. Note, however, that upon heating of clusters by a low-contrast laser pulse, the homogeneous plasma cannot be produced even using a homogeneous target. This is explained by the fact that, in the case of the low contrast, the prepulse power density in the focal plane proves to be so large that the prepulse destroys completely the clusters located near the gas-jet axis. In this case, as for an inhomogeneous cluster target, the produced plasma can be strongly inhomogeneous (Fig. 8). Therefore, the plasma that can be employed for the creation of X-ray lasers can be produced only using high-contrast laser pulses and homogeneous cluster targets.

ACKNOWLEDGMENTS

This work was partially supported by the US Department of Energy, the Fond Européen de Développement Economique Régional et Conseil Régional d'Aquitaine (France), the NATO grant PST.CLG.977637, the CRDF grant no. RP1-2328-ME-02, and the INTAS Foundation grant no. 2001-IB-233.

REFERENCES

1. T. Ditmire, T. Donnelly, A. M. Rubenchik, *et al.*, Phys. Rev. A **53**, 3379 (1996).
2. T. Ditmire, J. Zwelback, V. P. Yanovsky, *et al.*, Nature **398**, 490 (1999).
3. A. McPherson, B. D. Tompson, A. B. Borisov, *et al.*, Nature **370**, 631 (1994).

4. T. Ditmire, R. A. Smith, J. W. G. Smith, and M. H. R. Hutchinson, Phys. Rev. Lett. **78**, 3121 (1997).
5. T. Ditmire, J. W. G. Tisch, E. Springate, *et al.*, Phys. Rev. Lett. **78**, 2732 (1997).
6. J. Zweiback, R. A. Smith, T. E. Cowan, *et al.*, Phys. Rev. Lett. **84**, 2634 (2000).
7. E. Parra, T. Alexeev, J. Fan, *et al.*, Phys. Rev. E **62**, R5931 (2000).
8. E. Lamour, S. Dreuil, J.-C. Gauthier, *et al.*, Proc. SPIE **4504**, 97 (2001).
9. J.-P. Rozet, M. Cornille, S. Dobosz, *et al.*, Phys. Scr. T **92**, 113 (2001).
10. I. Last and J. Jortner, Phys. Rev. A **62**, 013201 (2000).
11. V. P. Krainov and A. S. Roshchupkin, J. Phys. B **34**, L297 (2001).
12. T. Auguste, P. D'Oliveira, S. Hulin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **72**, 54 (2000) [JETP Lett. **72**, 38 (2000)].
13. I. Yu. Skobelev, A. Ya. Faenov, A. I. Magunov, *et al.*, Zh. Éksp. Teor. Fiz. **121**, 88 (2002) [JETP **94**, 73 (2002)].
14. J. Abdallah, Jr., A. Ya. Faenov, I. Yu. Skobelev, *et al.*, Phys. Rev. A **63**, 032706 (2001).
15. C. Stenz, V. Bagnoud, F. Blasco, *et al.*, Kvantovaya Élektron. (Moscow) **30**, 721 (2000).
16. I. Yu. Skobelev, A. Ya. Faenov, B. A. Bryunetkin, *et al.*, Zh. Éksp. Teor. Fiz. **108**, 1263 (1995) [JETP **81**, 692 (1995)].
17. T. A. Pikuz, A. Ya. Faenov, S. A. Pikuz, *et al.*, J. X-ray Sci. Technol. **5**, 323 (1995).
18. B. K. F. Young, A. L. Osterheld, D. F. Price, *et al.*, Rev. Sci. Instrum. **69**, 4049 (1998).
19. A. S. Boldarev, V. A. Gasilov, F. Blasco, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 583 (2001) [JETP Lett. **73**, 514 (2001)].
20. G. C. Junkel-Vives, J. Abdallah, Jr., T. Auguste, *et al.*, Phys. Rev. E **65**, 036410 (2002).
21. A. Y. Faenov, I. Y. Skobelev, A. I. Magunov, *et al.*, Proc. SPIE **4504**, 121 (2001).
22. J. Abdallah, Jr., R. E. H. Clark, A. Ya. Faenov, *et al.*, J. Quant. Spectrosc. Radiat. Transf. **62**, 85 (1999).
23. J. Abdallah, Jr., A. Ya. Faenov, T. A. Pikuz, *et al.*, J. Quant. Spectrosc. Radiat. Transf. **62**, 1 (1999).
24. H. R. Griem, *Spectral Line Broadening by Plasmas* (Academic, New York, 1974; Mir, Moscow, 1978).
25. J. T. O'Brien and C. F. Hooper, Jr., Phys. Rev. A **5**, 867 (1972).
26. R. J. Tigle and C. F. Hooper, Jr., Phys. Rev. A **14**, 1514 (1976).
27. P. A. Loboda, V. V. Popova, and M. K. Shinkarev, J. Quant. Spectrosc. Radiat. Transf. **58**, 757 (1997).
28. P. A. Loboda, I. A. Litvinenko, G. V. Baydin, *et al.*, Laser Part. Beams **18**, 275 (2000).
29. S. Dobosz, M. Schmitd, M. Perdrix, *et al.*, Zh. Éksp. Teor. Fiz. **115**, 2051 (1999) [JETP **88**, 1122 (1999)].
30. G. C. Junkel-Vives, J. Abdallah, Jr., F. Blasco, *et al.*, J. Quant. Spectrosc. Radiat. Transf. **71**, 417 (2001).
31. G. C. Junkel-Vives, J. Abdallah, Jr., F. Blasco, *et al.*, Phys. Rev. A **64**, 021201 (2001).
32. F. Blasco, C. Stenz, F. Salin, A. Ya. Faenov, A. I. Magunov, T. A. Pikuz, I. Yu. Skobelev, G. C. Junkel-Vives, and J. Abdallah, Jr., private communication.
33. A. Ya. Faenov, A. I. Magunov, T. A. Pikuz, *et al.*, Phys. Scr. T **80**, 536 (1999).
34. A. Maksimchuk, M. Nantel, G. Ma, *et al.*, J. Quant. Spectrosc. Radiat. Transf. **65**, 367 (2000).
35. F. B. Rosmej, U. N. Funk, M. Geissel, *et al.*, J. Quant. Spectrosc. Radiat. Transf. **65**, 477 (2000).
36. A. I. Magunov, T. A. Pikuz, I. Yu. Skobelev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 412 (2001) [JETP Lett. **74**, 375 (2001)].
37. P. V. Nickles, V. N. Shlyaptsev, M. Kalachnikov, *et al.*, Phys. Rev. Lett. **78**, 2748 (1997).
38. J. Dunn, A. L. Osterheld, V. N. Shlyaptsev, *et al.*, in *Proceedings of 11th APS Topical Conference "Atomic Processes in Plasmas"*, Ed. by E. Oks and M. S. Pindzola; AIP Conf. Proc. **443**, 106 (1998).
39. J. Dunn, A. L. Osterheld, V. N. Shlyaptsev, *et al.*, in *Proceedings of the 6th International Conference on X-ray Lasers, Kyoto, Japan, 1998*, Ed. by Y. Kato, H. Takuma, and H. Daido; Inst. Phys. Conf. Ser. **159**, 51 (1998).
40. J. Dunn, A. L. Osterheld, R. Shepherd, *et al.*, Phys. Rev. Lett. **80**, 2825 (1998).

Translated by M. Sapozhnikov

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 977–984.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1139–1148.
Original Russian Text Copyright © 2002 by Aleksandrov, Voronov, Vtyurin, Goryainov, Zamkova, Zinenko, Krylov.

SOLIDS
Structure

Lattice Dynamics and Hydrostatic-Pressure-Induced Phase Transitions in ScF_3

K. S. Aleksandrov[a], V. N. Voronov[a], A. N. Vtyurin[b], S. V. Goryainov[b], N. G. Zamkova[a], V. I. Zinenko[a, *], and A. S. Krylov[a]

[a] *Kirenskiĭ Institute of Physics, Siberian Division, Russian Academy of Sciences, Akademgorodok, Krasnoyarsk, 660036 Russia*

[b] *Joint Institute of Geology, Geophysics, and Mineralogy, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia*

**e-mail: zvi@iph.krasn.ru*

Received November 13, 2001

Abstract—New phase transitions induced by hydrostatic pressure in a cubic (under standard conditions) ScF_3 crystal are discovered by the methods of polarization microscopy and Raman scattering. The space groups $R\bar{3}c$ for $Z = 2$ and *Pnma* for $Z = 4$ are proposed for the high-pressure phases. A nonempirical computation of the lattice dynamics of the crystal is carried out. It is shown that, under normal pressure, the cubic phase is stable down to $T = 0$ K, while the application of a hydrostatic pressure gives rise to a phonon branch in the vibrational spectrum (between points *R* and *M* of the Brillouin zone) with negative values of squares of frequencies. The condensation of soft mode R_5 at the boundary point of the Brillouin zone leads to rhombohedral distortion of the cubic structure with the unit cell volume doubling. The calculated frequencies at $q = 0$ of the ScF_3 lattice in the distorted rhombohedral phase are real-valued; the number and position of frequencies active in Raman scattering are in accord with the experimental values. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Fluorides of trivalent metals MeF_3 with an ideal or distorted structure of α-ReO_3 belong to the family of perovskite-like compounds with the general formula ABX_3, in which one of the cation sites is vacant (Fig. 1). Like all perovskites, these substances experience consecutive phase transformations under external effects. The presence of bulk cavities in the structure makes it possible to modify the physical properties of these crystals smoothly by creating structural disorder or by introducing impurities, which makes these crystals interesting objects for studying the mechanisms of phase transitions (see, for example, [1]) and also makes it possible to find their practical applications [2, 3].

Among other compounds with the chemical formula MeF_3, scandium fluoride is apparently the least studied. The most comprehensive reviews devoted to the description of structural phase transitions in perovskites [4, 5] contain no information on this material. In the structural database [6], information is given on three different structures (cubic, rhombohedral, and orthorhombic) of ScF_3 under normal conditions; however, special stability tests of these phases [7] revealed that the orthorhombic phase under normal conditions is metastable, while the cubic phase was not detected at all.

It was noted in [5, 8, 9] that the phase diagram of these crystals is very sensitive to structural defects and impurities, which creates additional difficulties in their investigations. In addition, at least some of the phase transitions in crystals belonging to this family are ferroelastic [1, 4], and the presence of growth stresses in the samples synthesized at high temperatures may also considerably affect their behavior upon a change in external conditions.

In our earlier publication [10], we analyzed the vibrational spectrum of the lattice for the cubic modification of ScF_3 at low temperatures down to 4 K, but no phase transitions were detected. The present work aims at studying phase transitions in a ScF_3 crystal under pressure by using Raman spectroscopy combined with

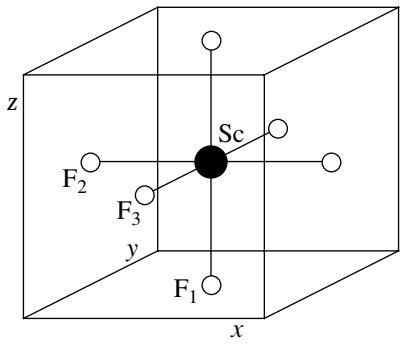

Fig. 1. Structure of the cubic phase of ScF_3.

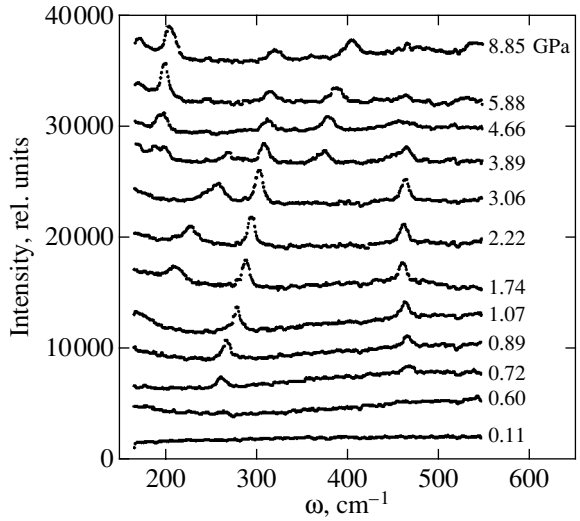

Fig. 2. Variation of the Raman spectrum of ScF_3 upon an increase in pressure.

polarization microscopy on samples synthesized under the conditions ensuring minimal stresses during their growth and at establishing the origin of the lattice instability appearing in this case by using nonempirical calculations of the frequency spectrum of lattice vibrations in the framework of a microscopic model of an ionic crystal.

2. SYNTHESIS AND STRUCTURE OF SINGLE CRYSTALS

We could not find in the literature any technique for growing ScF_3 single crystals; however, the synthesis of similar single crystals of iron and aluminum fluorides from flux in melt is described in [11–13]. It should be noted that this method makes it possible to lower the synthesis temperature and, hence, to reduce the probability of emergence of stresses during the crystal growth. We used lithium fluoride as the solvent. Attempts were made to use other compounds, but the single crystals grown in this case were too small (less than 1 mm^3).

The flux–melt containing 40 mol % ScF_3 was hermetically sealed in an oxygen-free atmosphere in a platinum ampule with a wall thickness of 0.2 mm. Over a period of 14 days, the ampule was lowered at a rate of 20 mm/day in a vertical tube furnace with an axial temperature gradient of 10–20 K/cm from the temperature region of 1400 K.

After cooling and opening the ampule, we discovered a cylindrical sample in it. The lower transparent part of the sample having a diameter of 10 mm and a height of 7 mm did not contain any defects or inclusions that could be seen in the microscope. The X-ray structural analysis proved the correspondence of the obtained crystal to the structure of the cubic phase of ScF_3 with the unit cell parameter $a_0 = 4.01$ Å (a comparison was made against the data presented in [6]). The observation in a polarization microscope revealed the optical isotropy of the crystal, which is also in accord with the cubic symmetry.

3. VARIATION OF RAMAN SPECTRA

The vibrational representation of the space group $Pm\bar{3}m$ of the cubic phase for the center of the Brillouin zone has the form

$$\Gamma_0 = F_{2u} + 3F_{1u}, \quad (1)$$

all vibration being inactive in the Raman spectrum.

A similar expansion for the rhombohedral structure has the form

$$\Gamma_1 = \mathbf{A}_{1g} + 2A_{2g} + 3\mathbf{E}_g + 2A_{1u} + 3A_{2u} + 5E_u, \quad (2)$$

while, for the orthorhombic structure, we have

$$\Gamma_1 = 7\mathbf{A}_g + 5\mathbf{B}_{1g} + 7\mathbf{B}_{2g} + 5\mathbf{B}_{3g} + 5A_u + 7B_{1u} + 5B_{2u} + 7B_{3u}. \quad (3)$$

In expansions (2) and (3), the active modes in the Raman spectrum are singled out.

A comparison of expressions (1)–(3) shows that the selection rules for these structures differ considerably and, hence, these structures should be easily diagnosed from the form of their Raman spectra.

We studied ScF_3 samples under a high (up to 9 GPa) hydrostatic pressure at room temperature on a setup with diamond anvils, which is similar to that used in [14, 15]; the diameter of the cell containing the sample was 0.25 mm, and its height was 0.1 mm. The pressure was determined to within 0.05 GPa from the shift in the luminescence band of a ruby microcrystal [15, 16] placed next to the sample. Mixtures of ethyl and methyl alcohols or glycerene were used as the pressure-transmitting medium. The Raman spectra were generated by radiation emitted by an Ar^+ laser (514.5 nm, 0.5 W) and were recorded by a multichannel spectrometer OMARS 89 (Dilor). In view of the small size of the sample and strong diffuse scattering, the high-frequency (150–600 cm^{-1}) region of the spectrum was recorded. The domain structure and birefringence effects in the sample were observed simultaneously with the help of a polarization microscope.

Under the normal pressure, the crystal has no Raman spectrum; it is optically isotropic and is darkened in crossed polarizers (slight field blooming emerges due to anisotropic mechanical stresses appearing in diamond anvils). Under a pressure of 0.7 GPa, the spectrum acquires two spectral lines (at 260 and 465 cm^{-1}; see Fig. 2), simultaneously, the sample placed between two crossed polarizers is bleached (Fig. 3), indicating the emergence of optical anisotropy. Some samples exhibit splitting into coarse domains (of the

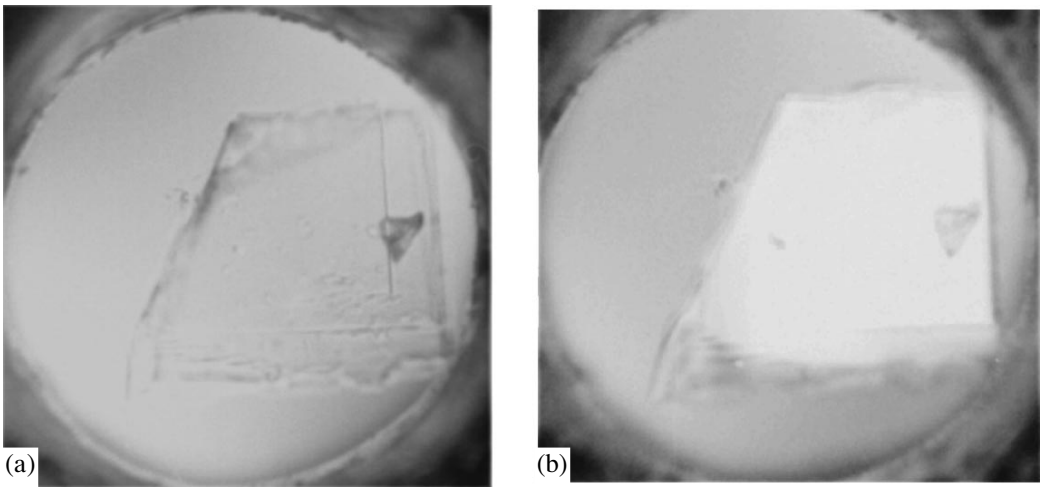

Fig. 3. Microphotographs of a sample before (left, 0.07 GPa) and after (right, 0.72 GPa) the first phase transition. The polarizers of the microscope are crossed. The cell diameter is 0.25 mm.

order of 0.02–0.05 mm) of irregular shape. Upon a further increase in pressure, the intensity of these lines increases monotonically. The frequency of the 260-cm^{-1} line also increases, attaining a value of 300 cm^{-1} under a pressure of 3.8 GPa. In the low-frequency region, one more line appears, shifting monotonically from 180 cm^{-1} at 1.4 GPa to 260 cm^{-1} at 3.8 GPa. This is accompanied by an enhancement of the birefringence effect and a change in the interference coloring of the sample associated with a change in the shape of the optical indicatrix.

The changes occurring up to values of 3.8 GPa are reversible and can be reproduced in different samples taken from the same product of crystallization and with different pressure-transmitting liquids (Fig. 4). Within the experimental error indicated above, no hysteresis effects are observed. The form of the domain structure determined to a considerable extent by defects at the sample boundaries changes from sample to sample and as afunction of the pressure variation rate; the monodomain state can be obtained in well-faceted microcrystals subjected to a slowly increasing pressure.

A further increase in pressure leads to one more transition (at 3.8 GPa). The crystal acquires a complex system of a large number of small (less than 0.01 mm) domains which can be seen through a microscope. The boundaries of these domains strongly scatter light (Fig. 5). Simultaneously, the form of the Raman spectrum changes sharply: some lines disappear, and a considerable number of new lines and bands consisting apparently of a several closely spaced profiles are formed (see Figs. 3 and 4). It should be noted that the pressure of 3.8 GPa corresponding to the transition point is in accord with the pressure of transition from the rhombohedral to the orthorhombic phase of the crystal under investigation, which was observed earlier in [8, 9]. Under a further increase in pressure, the high-frequency (> 200 cm^{-1}) part of the spectrum changes insignificantly, while, below 200 cm^{-1}, a shift of one of the lines towards higher frequencies, its intersection with another line, and, probably, the emergence of one more line in the vicinity of 160 cm^{-1} under a pressure of 7 GPa are observed. The system of domain walls emerging during the transition is slightly modified upon an increase in pressure; the total number of domains slightly decreases, but the system does not disappear completely. The crystal remains optically anisotropic, although strong scattering at domain walls complicates the observation of the effects associated with it.

A decrease in the pressure on a sample in this phase does not lead to a reverse transition (Fig. 6). The system of domain walls and the general nature of the spectrum are preserved, although the spectrum displays the softening of at least one vibration in the low-frequency

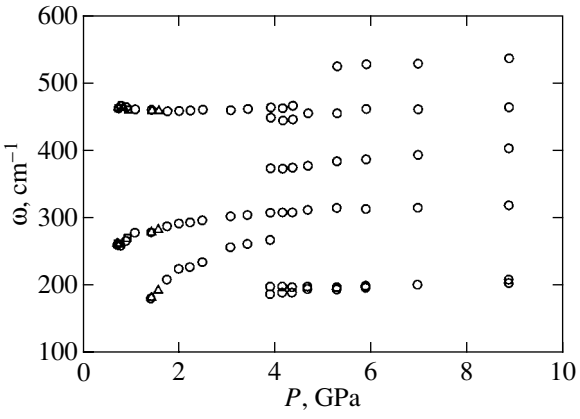

Fig. 4. Pressure dependences of the frequencies of experimentally observed lines. Circles correspond to values obtained under increasing pressure, while squares and triangles correspond to values obtained on different samples in the second phase under decreasing pressure.

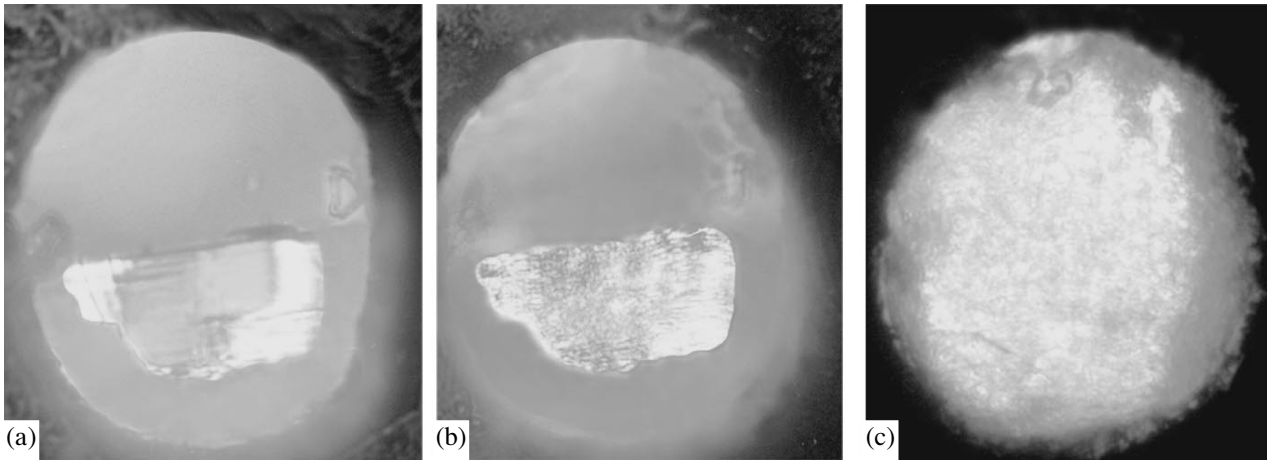

Fig. 5. Microphotographs of a sample before (a, 3.06 GPa) and after (b, 3.89 GPa) second phase transition and after its fracture upon a rapid decrease in pressure (c, 0.1 GPa).

part, which may be due to the emergence of lattice instability. Under a slow (over several hours) decrease in pressure from 1 GPa down to normal pressure, the sample can be preserved (the lower spectrum in Fig. 6 was measured in air on the sample extracted from the cell). A more rapid decrease in pressure below 1 GPa leads to crystal breakdown (see Fig. 5c), indicating the presence of strong mechanical stresses in the sample.

4. LATTICE DYNAMICS

In order to calculate the vibrational spectrum of the ScF_3 crystal lattice, we used a nonempirical model of an ionic crystal, generalizing the Gordon–Kim approximation by taking into account the effect of crystal lattice on the deformability and polarizability of ions [17]. The expression for the dynamic matrix was given in [18]. In the framework of this model, the lattice dynamics of AlF_3, GaF_3, and InF_3 crystals in the cubic phase, which are isomorphic to ScF_3, was calculated earlier and it was shown that the vibration spectrum of these crystals does not contain imaginary frequencies, which indicates the stability of the cubic phase [8].

The equilibrium value of the lattice parameter of the ScF_3 crystal was determined from the minimum of the total energy of the crystal as a function of volume. Table 1 contains this value together with the experimental value as well as the calculated values of polarizabilities of scandium and fluorine ions, high-frequency permittivity ε_∞, and Born dynamic charges. It can be seen that the calculated unit cell parameter is 5% smaller than the experimental value. Unfortunately, the experimental value of ε_∞ for this crystal is unknown, but the obtained value is typical of perovskites contain-

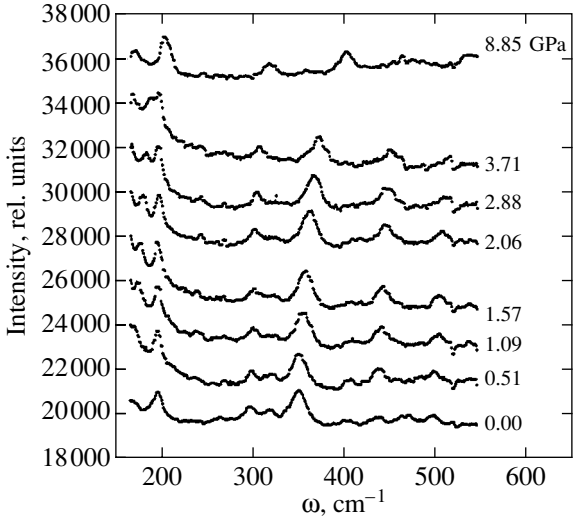

Fig. 6. Variation of the Raman spectrum of ScF_3 upon a decrease in pressure.

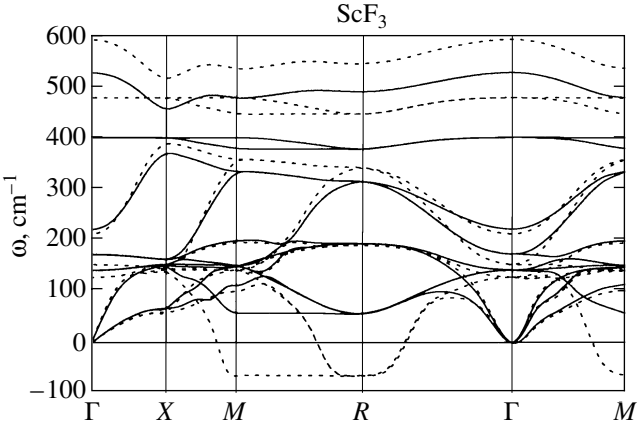

Fig. 7. Results of calculation of the phonon spectrum of the cubic phase of ScF_3. Solid curves correspond to the unit cell parameter a = 7.22 at. units (normal pressure) and dashed curves correspond to a = 7.06 at. units.

ing fluorine. The effective charge tensor for the Sc ion is isotropic in accordance with the symmetry of the position of this ion in the cubic phase, and its value is close to the nominal value of the ion charge (+3). For a fluorine ion, there are two tensor components corresponding to the displacement of the F ion parallel ($Z^*_{\parallel,F}$) and perpendicular ($Z^*_{\perp,F}$) to the Sc–F bond.

The calculated spectrum of ScF_3 lattice vibrations in the cubic phase is shown in Fig. 7 for a unit cell volume of 52.1 Å^3 (corresponding to the applied pressure $P \approx$ 6 GPa). The vibrational spectrum contains no imaginary frequencies (which explains the structural stability down to temperatures of 4 K). However, it contains a weak dispersion branch (between points R and M in the Brillouin zone) with an anomalously low frequency. In this vibrational branch, the triply degenerate mode R_5 at point R and nondegenerate modes in the $R \longrightarrow M$ direction (including point M) correspond to vibrations in which fluorine ions are displaced [19]. The structural phase transitions in most halogenides with the perovskite structure [4], including MeF_3 crystals (Me = Al, Ga, In, …), are associated precisely with the mode condensation of this vibrational branch [1, 2].

We also calculated the vibrational spectrum of the cubic phase of the ScF_3 lattice for nonequilibrium decreasing values of the unit cell parameter, which corresponds to the application of a hydrostatic pressure to the crystal. The value of pressure was estimated from the numerical differentiation of the total energy of the crystal with respect to the volume and from the calculated bulk compression modulus $B = (C_{11} + 2C_{12})/3$ in terms of elastic constants whose values were obtained from the dispersion dependence of acoustic vibrational branches for $\mathbf{q} \longrightarrow 0$ (which are given in Table 1). The curve corresponding to the equation of state of ScF_3 is shown in Fig. 8.

Figure 7 shows that the most significant changes in the lattice vibrational spectrum upon the application of a hydrostatic pressure to the crystal occur in the regions of high and low frequencies of optical vibrational modes. The high-frequency vibrational modes become "harder" upon a decrease in volume, while the branch of lattice vibrations (between points R and M of the Brillouin zone) with anomalously low values of frequency becomes "softer" upon the application of pressure, and the cubic structure of the crystal becomes unstable.

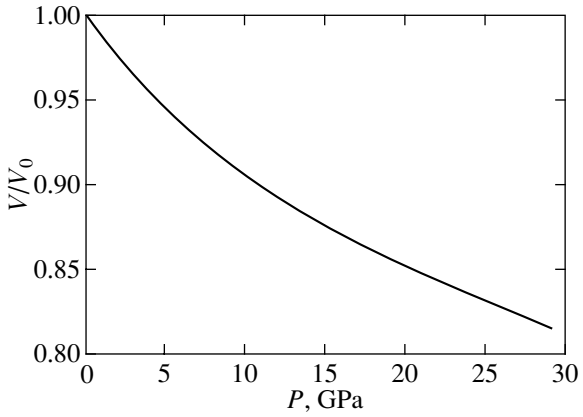

Fig. 8. Calculated equation of state of ScF_3.

5. DISCUSSION

The results of our calculations show that, in accordance with the available experimental data [10], the cubic phase of the ScF_3 crystal under normal pressure remains stable down to $T = 0$ K. At the same time, according to the results of our experiments and an analysis of lattice dynamics, the application of a hydrostatic pressure leads to cubic phase instability. The calculations of the frequencies of lattice vibrations under the action of a hydrostatic pressure show that the softest mode is the triply degenerate mode R_5 belonging to the boundary point $R = \pi/a(1, 1, 1)$ of the Brullouin zone of the cubic phase. Consequently, it is natural to assume that the phase transition observed under a pressure equal to 0.7 GPa is associated with the condensation of precisely this mode.

The solid lines in Fig. 9 show the volume dependence of the squared frequency $\omega^2(R_5)$ of this mode. A decrease in the unit cell volume (increase of pressure) leads to a linear decrease in the value of $\omega^2(R_5)$, so that $\omega^2(R_5) = 0$ for $P \approx 2.5$ GPa (see Fig. 9). The lower experimental value of pressure corresponding to the transition may be due to nonideality of the samples used in experiments; according to [8], structural defects lead to instability of the cubic phase.

A displacement of fluorine ions in the triply degenerate R_5 mode corresponds to the "rotation" of the octahedron ScF_6 [19] about the spatial diagonal of the cubic unit cell. The irreducible representation R_5 appears in the vibrational representation of the crystal only once; consequently, the expression for $\omega^2(R_5)$ in terms of the

Table 1. Results of calculations of the main physical parameters of the cubic phase of ScF_3

a_0, Å, experiment	a_0, Å, theory	Z^*_{Sc}	$Z^*_{\parallel F}$	$Z^*_{\perp F}$	α_{Sc}, Å^3	α_F, Å^3	ε_∞	C_{11}, GPa	C_{12}, GPa	C_{44}, GPa	B, GPa
4.01	3.82	3.36	–0.71	–1.95	0.27	0.72	1.75	172.7	18.9	18.6	70.2

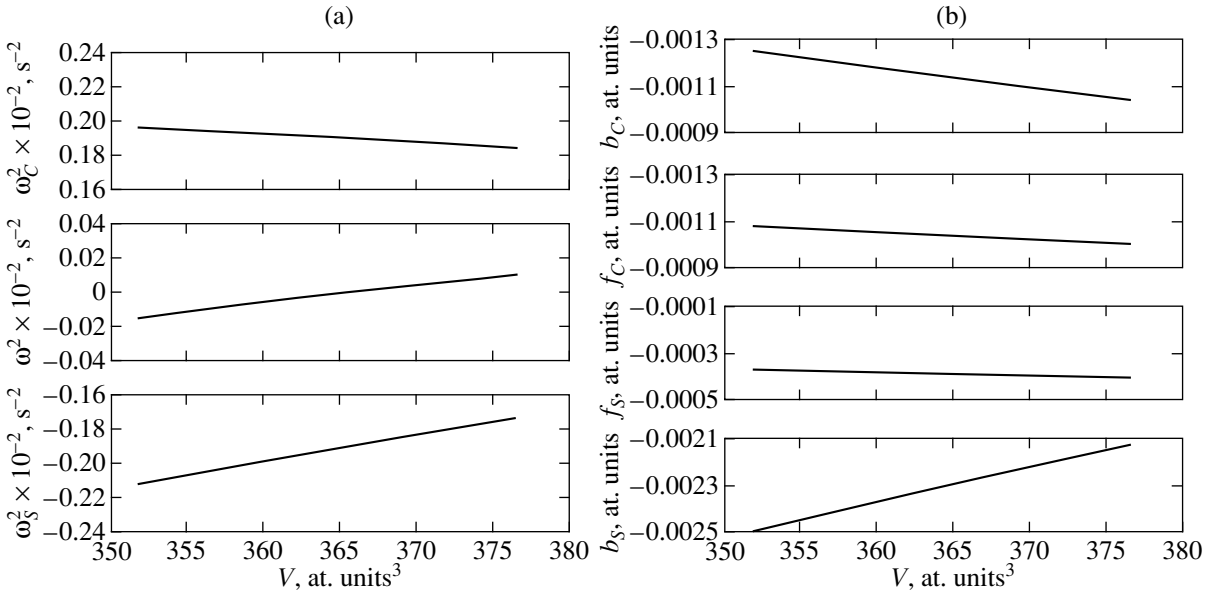

Fig. 9. Calculated dependence of (a) the squared frequency of the soft mode R_5 and (b) dynamic matrix elements determining it on the unit cell volume.

elements of the dynamic matrix can be obtained in analytic form. We can separate the contribution ω_C^2 from the Coulomb interaction of ions to frequency so that the remaining term ω_S^2 includes the short-range and long-range dipole–dipole contributions:

$$\omega^2(R_5) = \omega_C^2 + \omega_S^2. \quad (4)$$

Each of these terms can, in turn, be decomposed into two parts corresponding to different elements of the dynamic matrix of the crystal:

$$\omega_C^2 = b_C - f_C, \quad \omega_S^2 = b_S - f_S, \quad (5)$$

$$b = D^{xx}_{F_1-F_1}\left(\mathbf{q} = \frac{\pi}{a}(1, 1, 1)\right), \quad (6)$$

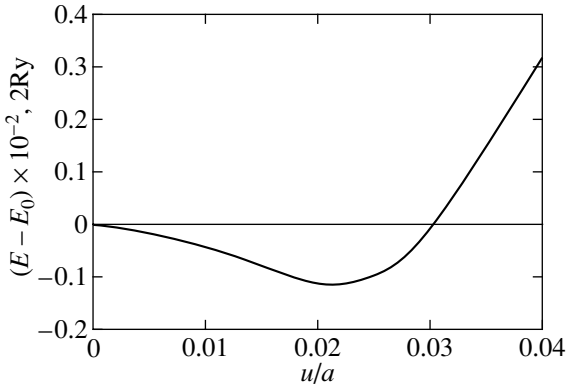

Fig. 10. Dependence of the total energy of a crystal with the doubled unit cell on the displacement of fluorine ions from the equilibrium position of the cubic phase: $E_0 = -2148.9854$; 2Ry is the total energy of the undistorted phase.

$$f = D^{yz}_{F_1-F_2}\left(\mathbf{q} = \frac{\pi}{a}(1, 1, 1)\right), \quad (7)$$

where b and f correspond to longitudinal and transverse force constants of interaction of fluorine ions, respectively. The volume dependence of the quantities appearing in expressions (4) and (5) is also shown in Fig. 9. It can be seen that the Coulomb and short-range contributions to $\omega^2(R_5)$ have opposite signs, and the positive Coulomb contribution exceeds in magnitude the negative short-range contribution under normal pressure, thus ensuring the stability of the cubic phase. As the unit cell volume decreases, both contributions increase in magnitude, but the absolute value of the short-range contribution increases more rapidly than that of the Coulomb contribution, leading to instability of the cubic phase. Figure 9b, presenting the volume dependences of the Coulomb and short-range contributions to the dynamic matrix elements, shows that an increase in the negative short-range contribution to $\omega^2(R_5)$ is mainly associated with the stronger volume dependence of the short-range and dipole–dipole contributions to the diagonal element $D^{xx}_{F_1-F_1}$ ($\mathbf{q} = \pi/a(1, 1, 1)$) of the dynamic matrix. It should be noted that the value of b_S and its dependence on pressure are mainly determined by the contribution from the long-range dipole–dipole interactions.

The structural distortions associated with the condensation of the triply degenerate mode R_5 lead to a rhombohedral distortion of the crystal structure and stabilize the lattice. Figure 10 shows the dependence of the total energy of a crystal with the doubled unit cell on the displacement of fluorine ions from the equilibrium

position of the cubic phase under pressure $P = 6$ GPa:

$$\tilde{u}^x_{F_1} = -\tilde{u}^y_{F_1} = \tilde{u}^y_{F_2} = -\tilde{u}^z_{F_2} = -\tilde{u}^x_{F_3} = \tilde{u}^z_{F_3} = u, \quad (8)$$

$$\tilde{u}^\alpha_{F_i} = \frac{u^\alpha_{F_i}}{2a}\exp(i\mathbf{q}_R\mathbf{r}), \quad (9)$$

where a is the lattice parameter of the cubic phase, $\mathbf{r} = m_1\mathbf{a}_1 + m_2\mathbf{a}_2 + m_3\mathbf{a}_3$ is the translation vector, and $\mathbf{q}_R = \pi/a(1, 1, 1)$. It can be seen that the total energy minimum corresponds to the displacement $u \approx 0.025a$ (0.7 Å) of fluorine ions.

Table 2 contains the values of calculated limiting frequencies of vibrations in the rhombohedral phase in which the coordinates of atoms correspond to the obtained values of the displacement of fluorine ions. The table also gives for comparison the values of frequency at points $\Gamma(\mathbf{q} = (0, 0, 0)$ and $R(\mathbf{q} = \pi/a(1, 1, 1))$ of the undistorted cubic phase as well as the compatibility relations.

After a transition, four lines must be activated in the spectrum (see relation (2)). Two of these lines (A_{1g} and E_g) correspond to the restored soft mode which was split after the transition and, hence, correspond to low frequencies, which must depend considerably on pressure. The three lines observed experimentally above $P = 0.7$ GPa (the frequencies corresponding to these lines are given in Table 2 in parentheses) are in good agreement with this description; the line corresponding to the lowest frequency is noticeably shifted upwards upon an increase in pressure, and the emerging low-frequency wing apparently corresponds to the second mode being restored, whose frequency lies below 150 cm^{-1}. The calculated and experimentally obtained frequencies are in satisfactory agreement. It should be noted that the position of lines correlates well with the frequencies in the Raman spectrum for rhombohedral phases of fluorides of some other trivalent metals [20]. On the whole, we may conclude that the first high-pressure phase is rhombohedral with the space group $R\bar{3}c$, $Z = 2$.

The second point of transition corresponding to a pressure of 3.8 GPa is in good agreement with the transition from the rhombohedral to the orthorhombic (D^{16}_{2h}, $Z = 4$) phase, which was observed earlier in [9]. The strong increase in the number of lines in the Raman spectrum (3) is also in accord with these observations. In [9], a strong diffusion-controlled X-ray scattering was detected above this transition point, which gradually decreased upon an increase in pressure. This correlates well with the emergence of a developed system of domain walls observed through the microscope and with the increase in the domain size under an increase in pressure. The type of this transition (which is predominantly a first-order transition) and the existence of a considerable hysteresis upon a decrease in pressure are also in accord with the results obtained in [9].

Table 2. Compatibility relations and values of vibrational frequencies in the cubic and rhombohedral phases (experimentally measured values of frequencies are given in parentheses)

Cubic phase		Rhombohedral phase	
Frequency, cm^{-1}	Symmetry of vibrations	Symmetry of vibrations	Frequency, cm^{-1}
590	F_{1u}–LO	A_{2u}	557
477	F_{1u}–TO	E_u	442
154	F_{1u}–LO	A_{2u}	210
211	F_{1u}–TO	E_u	164
126	F_{2u}	A_{1u}	148
		E_u	132
542	R_1	A_{2g}	513
341	R_{10}	A_{1u}	327
		E_u	330
445	R_3	E_g	412 (465)
188	R_4	A_{2g}	190
		E_g	198 (260)
65i	R_5	A_{1g}	79 (180)
		E_g	34

6. CONCLUSIONS

Thus, we have carried out experimental studies and numerical calculations of lattice dynamics in a ScF_3 crystal induced by a hydrostatic pressure.

The structural phase transition from the cubic to the rhombohedral phase observed experimentally for the first time under a pressure of 0.7 GPa is in accord with the results of nonempirical calculations of the lattice vibrational spectrum: the calculated spectrum of the crystal under normal pressure contains no imaginary frequencies, which indicates the stability of the structure. However, the vibrational spectrum contains a branch (between points R and M of the Brillouin zone) with anomalously low frequencies. As the unit cell volume decreases (which corresponds to an increase in pressure), the frequencies corresponding to this branch decrease and the lattice becomes unstable at a fairly high pressure. The reason for this decrease in frequency and, hence, in the emergence of lattice instability is the violation of the balance of the Coulomb interactions, on the one hand, and the sum of short-range and dipole–dipole interactions, on the other hand.

The calculated spectrum of the distorted rhombohedral formed as a result of a transition is in good agreement with the experimental spectrum. The experimental Raman spectrum displays the restoration of a soft mode; the frequency of the second expected soft mode apparently lies below the investigated frequency range, and only a wing of this mode is observed.

The pressure and the general type of the second transition from the rhombohedral to the orthorhombic phase investigated by us coincide with those observed earlier [9]; the Raman spectrum of this high-pressure phase is obtained for the first time. The transition is predominantly a first-order transition and is accompanied by strong hysteresis effect. The formation of a complex system of domain walls is apparently responsible for the strong diffuse X-ray scattering (reported in [9]) in this phase. The mechanism of transition to the second high-pressure phase, its structure, and lattice dynamics require further investigations.

ACKNOWLEDGMENTS

The authors are grateful to A.P. Shebanin for his help in experimental measurements.

The work was carried out under financial support from the Russian Foundation for Basic Research (project nos. 00-02-17792 and 00-15-96790) and partly from the INTAS (grant no. 97-10177).

REFERENCES

1. Ph. Daniel, A. Bulou, M. Rousseau, *et al.*, J. Phys.: Condens. Matter **2**, 5663 (1990).
2. K. Rotereau, Ph. Daniel, and J. Y. Gesland, J. Phys. Chem. Solids **59**, 969 (1998).
3. K. Rotereau, Ph. Daniel, A. Desert, and J. Y. Gesland, J. Phys.: Condens. Matter **10**, 1431 (1998).
4. K. S. Aleksandrov, A. T. Anistratov, B. V. Beznosikov, and N. V. Fedoseeva, *Phase Transitions in ABX_3 Crystals* (Nauka, Novosibirsk, 1981).
5. D. Babel and A. Tressaud, in *Inorganic Solid Fluorides* (Academic, London, 1985), p. 77.
6. *Powder Diffraction Data* (International Center on Diffraction Data, 1999), nos. 75-0877, 46-1243, 44-1096, 43-1145, 32-0989, 17-0836.
7. M. M. Aleksandrova, N. A. Bendeliani, V. D. Blank, and T. I. Dyuzheva, Izv. Akad. Nauk SSSR, Neorg. Mater. **26**, 1028 (1990).
8. V. I. Zinenko and N. G. Zamkova, Fiz. Tverd. Tela (St. Petersburg) **42**, 1310 (2000) [Phys. Solid State **42**, 1348 (2000)].
9. N. A. Bendeliani, É. Ya. Atabaeva, and V. M. Agotkov, Izv. Akad. Nauk SSSR, Neorg. Mater. **19**, 816 (1983).
10. K. S. Aleksandov, V. N. Voronov, A. Bulou, *et al.*, in *Abstracts of 6th Japan-CIS Symposium on Ferroelectricity, Noda, Japan, 1998*, p. 152.
11. B. M. Wanklyn, J. Cryst. Growth **5**, 279 (1969).
12. B. M. Wanklyn, J. Mater. Sci. **14**, 1447 (1979).
13. V. A. Timofeeva, *Growth of Crystals from Solutions-Melts* (Nauka, Moscow, 1978).
14. Q. Wang, G. Ripault, and A. Bulou, Phase Transit. **53**, 1 (1995).
15. S. V. Goryainov and I. A. Belitsky, Phys. Chem. Miner. **22**, 443 (1995).
16. R. G. Munro, G. J. Piermarini, S. Block, and W. B. Holzapfel, J. Appl. Phys. **57**, 165 (1985).
17. O. V. Ivanov and E. G. Maksimov, Zh. Éksp. Teor. Fiz. **108**, 1841 (1995) [JETP **81**, 1008 (1995)].
18. V. I. Zinenko, N. G. Zamkova, and S. N. Sofronova, Zh. Éksp. Teor. Fiz. **111**, 1742 (1998) [JETP **87**, 944 (1998)].
19. R. Cowley, Phys. Rev. **134**, 981 (1964).
20. P. Daniel, A. Bulou, M. Rousseau, *et al.*, Phys. Rev. B **42**, 10545 (1990).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 985–991.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1149–1156.
Original Russian Text Copyright © 2002 by Levchenko, Morozov, Sigov.

SOLIDS
Structure

Phase Diagram of Multilayer Magnetic Structures

V. D. Levchenko[a], A. I. Morozov[b, *], and A. S. Sigov[b]

[a]*Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Miusskaya pl. 4, Moscow, 125047 Russia*
[b]*Moscow State Institute of Radioengineering, Electronics, and Automation (Technical University), pr. Vernadskogo 78, Moscow, 117454 Russia*
**e-mail: mor-alexandr@yandex.ru*

Received November 15, 2001

Abstract—Multilayer ferromagnet–layered antiferromagnet (Fe/Cr) structures frustrated because of roughness of interlayer boundaries were studied by mathematical modeling methods. The phase diagram of a three-layer system (plotted as film thickness versus the degree of roughness of the interfaces) was obtained, and the order parameter distributions in each phase were determined. The character of phase transitions in this system was studied. The applicability range of the Slonczewski magnetic proximity model was determined. © *2002 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

Multilayer magnetic structures have been attracting close attention of researchers since the discovery of the giant magnetoresistance phenomenon in them [1]. In the past decade, a huge number of works concerned with these structures have been published; for instance, see reviews [2, 3]. Recently, interest of researchers has shifted to multilayer ferromagnet–layered antiferromagnet structures. According to the recent neutron diffraction experiments [4, 5], examples of such structures are Fe/Cr multilayers in which a chromium layer of thickness $d < 4.5$ nm is a set of ferromagnetic planes with antiparallel spin orientations in neighboring planes. Chromium atom spins lie in these planes, which, in turn, are on average parallel to interlayer boundaries. A similar structure was observed in manganese layers in Fe/Mn structures [6, 7].

Exchange between ferromagnetic layers in multilayer ferromagnet–layered antiferromagnet structures is caused by the interaction through the antiferromagnetic order parameter, and the Riderman–Kittel–Kasuya–Yoshida interaction is only a small addition to this interaction far from the Néel temperature [8]. Slonczewski suggested a phenomenological magnetic proximity model to describe the interaction through the antiferromagnetic order parameter [9]. Within this model, ferromagnetic layers are considered magnetized virtually uniformly, and substantial order parameter distortions are assumed to occur only in antiferromagnetic layers [9].

The presence of atomic steps that change the thickness of the antiferromagnet by one monoatomic layer at interlayer boundaries causes frustrations in the ferromagnet–antiferromagnet system (Fig. 1a). A uniform order parameter distribution in layers ceases to correspond to an energy minimum.

If the distance between atomic steps on the surface of a layer (step width R) exceeds some critical value, separation of ferromagnetic layers into domains becomes energetically favorable [10, 11] (Fig. 1b). Domain boundaries in the plane of layers coincide with atomic step edges. Note that the R value substantially depends on technological conditions [12].

Recent studies of the state of a ferromagnetic iron film on a rough Cr(001) surface revealed the presence of several magnetic phases depending on film thickness and the degree of roughness (on the R value) [13].

The aforesaid shows that the problem of considering "thickness–roughness" phase diagrams of ferromagnet–layered antiferromagnet magnetic structures is topical. This problem is the subject matter of the present communication.

The paper is organized as follows. Section 2 presents a simple model that allows the system under study to be qualitatively described. The method for calculations is considered in Section 3. The phase diagram of a ferromagnet–antiferromagnet–ferromagnet three-layer structure, which can be generalized to multilayers, is obtained in Section 4. The conclusion summarizes the most important results of this work.

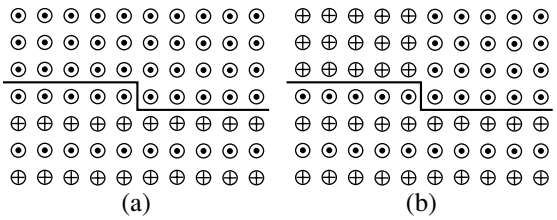

Fig. 1. Frustrations in the ferromagnet–layered antiferromagnet system caused by the presence of steps at the interlayer boundary.

2. MODEL DESCRIPTION

In describing multilayers, we will restrict ourselves to the mean field approximation. Let us introduce an order parameter for each layer. This will be the magnetization vector for magnetic layers and the antiferromagnetic vector equal to the difference of the magnetizations of sublattices for antiferromagnetic layers.

In magnetic layers dozens of angstrom units thick, atomic spins are known to lie in the plane of layers. Therefore, at $T < T_C, T_N$, where T_C is the Curie temperature of ferromagnetic layers and T_N is the Néel temperature of antiferromagnetic layers, we can characterize the local value of an order parameter lying in the plane of layers by angle θ between the order parameter and a selected axis also lying in the plane of layers. The order parameter modulus is assumed to be virtually constant in each layer.

Within the framework of these approximations, exchange energy W_i caused by inhomogeneity within the ith layer is written in the form

$$W_i = \frac{J_i S_i^2}{2b_i}\int (\nabla\theta_i)^2 dV, \qquad (1)$$

where the integral is over the volume of the layer, J_i is the exchange stiffness, S_i is the mean atomic spin value, and b_i is the interatomic distance.

Varying (1) with respect to parameter θ_i yields the equation that describes the order parameter distribution in the layer,

$$\Delta\theta_i = 0. \qquad (2)$$

A more thorough procedure should be applied to obtain boundary conditions. The $\theta_i - \theta_{i+1}$ difference can take on fairly large values because of frustrations at layer boundaries, whereas, within layers, frustrations are absent, θ_i varies smoothly, and the difference of θ_i in neighboring cells is small. For this reason, in exchange energy calculations by the X–Y model, the cosine of the difference of the θ_i angles in neighboring cells can be expanded into a power series if the cells occur in the same layer; this, however, cannot be done if the cells belong to different layers. We must therefore write the interaction energy between spins situated close to the interlayer boundary in the discrete representation, differentiate it with respect to angle θ_i of rotation of a certain spin, and then pass to the continuous representation. As a result, we obtain the boundary condition

$$\tilde{\Delta}\theta_i - \frac{\partial\theta_i}{\partial n} = \pm\frac{J_{f,af}S_{i+1}}{J_i S_i}\sin(\theta_i - \theta_{i+1}), \qquad (3)$$

where $\tilde{\Delta}$ is the two-dimensional Laplacian in the layer plane, $\partial/\partial n$ is the derivative in the direction of the outer normal to the layer, the $J_{f,af}$ exchange constant describes the interaction of spins in different layers, and all distances are reduced to the dimensionless form through dividing by the $b_i = b$ value considered equal for all layers. The signs on the right-hand side of (3) are opposite on opposite atomic step sides at the boundary. For a free surface, the right-hand side of (3) vanishes.

The exchange interaction energy between neighboring layers has the form

$$W_{i,i+1} = \pm\frac{J_{f,af}S_i S_{i+1}}{b^2}\int \cos(\theta_i - \theta_{i+1})dS, \qquad (4)$$

where the integration is over the boundary surface between layers i and $i + 1$, and the sign on the right-hand side of (4) is the same as in (3).

Varying the interlayer interaction energy written in the continuous representation with respect to θ_i yields an equation that differs from (3) by the absence of the first term on the left-hand side. This prevents passage from (3) to (2) when the bordering layers are identical.

Atomic steps divide the whole interlayer boundary surface into regions of two types. In regions of the first and second types, the boundary energy is minimum at $\theta_i = \theta_{i+1}$ and $\theta_i = \pi - \theta_{i+1}$, respectively.

Determining the distribution of order parameters in a multilayer requires solving the system of differential equations (2) with boundary conditions (3).

Consider the applicability of this simple model to real multilayers. The continuous representation is valid when the characteristic distances of the problem are much larger than interatomic distances. Layer thicknesses in multilayers and distances between atomic steps amount to dozens of angstrom units. It can be taken that these values are much larger than interatomic distances, and the continuous representation can therefore be used for qualitative consideration and obtaining order-of-magnitude estimates.

The model under consideration assumes exchange interaction to be isotropic, that is, the same in the plane of layers and in the perpendicular direction. A model with anisotropic interactions reduces to that considered above by renormalizing the length scale in one of two nonequivalent directions.

Interdiffusion of neighboring layer atoms only results in renormalizing the $J_{f,af}$ constant if the region of mixing includes one or two monolayers, that is, has the atomic thickness scale. This constant is found in microscopic calculations [14].

Equations (2) and (3) are written in the exchange approximation but can easily be generalized to systems with weak anisotropy in the plane of layers.

To summarize, the suggested model can be used to qualitatively describe the magnetic characteristics of multilayer ferromagnet–antiferromagnet structures.

3. METHOD FOR CALCULATIONS

Let the edges of atomic steps be rectilinear and parallel to each other. The x axis of the frame of reference lies in the plane of layers and is normal to step edges, and the z axis is perpendicular to the layers (a two-dimensional case).

The initial system of equations includes Laplace equations (2) for each plane layer, $-\infty < x < \infty$, $0 < z < a_i$, where $i = 1, \ldots, n$ are layer numbers, with nonlinear discontinuity conditions (3) at interlayer boundaries. For numerically solving this system, we will reduce it to a system of one-dimensional integral equations.

Let us impose the requirement that the $\theta_i(x, z)$ function be continuous in the region $0 < x < L$ and the $0 < z < a_i$, $\partial\theta_i/\partial x|_{x \to 0, L} \longrightarrow 0$ additional conditions be met. Extend the definition of $\theta_i(x, z)$ to period $2L$, namely, $\theta_i(-x, z) = \theta_i(x, z)$, $-L < x < L$.

Introduce a uniform mesh $\{x_j\}$ with step Δ^x in $-L < x < L$. Let us represent the $\theta_i(x, z)$ function and the right-hand side of boundary conditions (3) in the form of the Fourier series

$$\theta_i(x, z) = \Phi_{0, i}(z) + \sum_{k=1}^{N} \Phi_{k, i}(z) \exp\left(i\frac{\pi}{L}kx\right), \quad (5)$$

$$\sigma_i^{\pm}(x) \sin(\delta\theta_i^{\pm}(x)) = \Psi_{0, i}^{\pm} + \sum_{k=1}^{N} \Psi_{k, i}^{\pm} \exp\left(i\frac{\pi}{L}kx\right). \quad (6)$$

Here, $\sigma_i^{\pm}(x) = \pm J_{f, af} S_{i \pm 1}/J_i S_i$ is the step function which changes its value at atomic step boundaries and

$$\delta\theta_i \equiv \delta\theta_i^{+}(x) = -\delta\theta_{i+1}^{-}(x) = \theta_{i+1}|_{z=0} - \theta_i|_{z=a_i}.$$

Substituting (5) and (6) into the initial equations yields the ordinary differential equation

$$\frac{d^2\Phi_k}{dz^2} - A_k^2 \Phi_k = 0 \quad (7)$$

with the boundary conditions

$$\left(\frac{d\Phi_k}{dz} \pm A_k^2 \Phi_k\right)_{z=a, 0} = -\Psi_k^{\pm},$$
$$A_k = \frac{1}{\Delta^x}\sqrt{1 - \cos\left(\frac{\pi}{L}k\Delta^x\right)} \quad (8)$$

within layer i for each harmonic $k = 1, \ldots, N$ (here and throughout, index i is omitted if this cannot cause confusion). A solution to (7) is sought in the form

$$\Phi_k(z) = C_1^k \exp(A_k z) + C_2^k \exp(-A_k z).$$

The C_1^k and C_2^k constants can be found from boundary conditions (8). As a result, we obtain

$$\Phi_k(z) = K_k^{+}(z)\Psi_k^{+} + K_k^{-}(z)\Psi_k^{-}, \quad (9)$$

where

$$K_k^{+}(z) = dK_k \exp(-(a - z)A_k) \times [(1 + A_k) + (1 - A_k)\exp(-2zA_k)],$$

$$K_k^{-}(z) = dK_k \exp(-zA_k) \times [(1 + A_k) + (1 - A_k)\exp(-2(a - z)A_k)],$$

$$dK_k = -A_k^{-1}[(1 + A_k)^2 - (1 - A_k)^2 \exp(-2aA_k)]^{-1}.$$

Using the equation for the zeroth Fourier coefficients

$$\frac{d^2\Phi_0}{dz^2} = 0$$

with the boundary conditions

$$\left.\frac{d\Phi_0}{dz}\right|_{z=a, 0} = -\Psi_0^{\pm}$$

and taking into account that, for free boundaries of the top and bottom layers, $d\Phi_0/dz = 0$ and, for neighboring layers, the ratio between the functionals is given by

$$\Psi_{0, i-1}^{+}(\delta\theta_{i-1}^{+})/\Psi_{0, i}^{-}(\delta\theta_i^{-}) = \text{const},$$

we obtain $\Psi_0^{\pm} = 0$, or, selecting the mean angle

$$\overline{\delta\theta} = \frac{1}{2L}\int_{-L}^{L} \delta\theta(x)dx$$

and the variation

$$\widetilde{\delta\theta}(x) = \delta\theta(x) - \overline{\delta\theta},$$

we find

$$0 = \int_{-L}^{L} \sigma(x) \sin(\widetilde{\delta\theta}(x) + \overline{\delta\theta})dx = \cos\overline{\delta\theta} \times \int_{-L}^{L} \sigma(x) \sin\widetilde{\delta\theta}(x)dx + \sin\overline{\delta\theta}\int_{-L}^{L} \sigma(x)\cos\widetilde{\delta\theta}(x)dx.$$

Hence, the mean angle is

$$\overline{\delta\theta} = n\pi - \arctan\left(\frac{\int_{-L}^{L} \sigma(x)\sin\widetilde{\delta\theta}(x)dx}{\int_{-L}^{L} \sigma(x)\cos\widetilde{\delta\theta}(x)dx}\right). \quad (10)$$

Equations (5), (6) and (9), (10) can be used to obtain the sought integral equation for $\delta\theta_i$,

$$\delta\theta_i(x) = \overline{\delta\theta_i} + \sum_{k=1}^{N} \exp\left(i\frac{\pi}{L}kx\right)\frac{1}{2L}\int_{-L}^{L} \exp\left(-i\frac{\pi}{L}k\xi\right)d\xi$$

$$\times [K^-_{i+1,k}(0)\sigma^-_{i+1}\sin\delta\theta_i + K^+_{i+1,k}(0)\sigma^+_{i+1}\sin\delta\theta_{i+1}$$

$$- K^-_{i,k}(a_i)\sigma^-_i\sin\delta\theta_{i-1} - K^+_{i,k}(a_i)\sigma^+_i\sin\delta\theta_i]$$

$$\equiv \hat{I}(K, \delta\theta_i, \delta\theta_{i\pm1}).$$

This equation will be solved using the simple iterative procedure

$$\delta\theta_i^{n+1} = (1 - F_i(x))\delta\theta_i^n + F_i(x)\hat{I}(F_{k,i}K, \delta\theta_i^n, \delta\theta_{i\pm1}^n).$$

Here, $0 < F(x), F_k \le 1$ are the adaptive filters ensuring stability of the iterative procedure and increasing the rate of convergence. Iterations are performed until

$$\varepsilon = \max\left|\delta\theta_i^n(x) - \hat{I}(\delta\theta_i^n, \delta\theta_{i\pm1}^n)\right|$$

becomes smaller than a preset value (usually, $\varepsilon < 10^{-6}$).

The solution in the whole region can then be reconstructed by the formula

$$\theta_i(x, z) = \overline{\delta\theta_i} + \sum_{k=1}^{N}\exp\left(i\frac{\pi}{L}kx\right)$$

$$\times\frac{1}{2L}\int_{-L}^{L}\exp\left(-i\frac{\pi}{L}k\xi\right)d\xi$$

$$\times [K^-_{i,k}(z)\sigma^-_i\sin\delta\theta_{i-1} + K^+_{i,k}(z)\sigma^+_i\sin\delta\theta_i].$$

The starting approximation can be selected in the form

$$\delta\theta_i^0(x) = \pi\sum_j \pm\eta(x - x_j),$$

where $\eta(x - x_j)$ is the unit step function with a jump at defect (step edge x_j) positions, and the $\pm$ sign means that this function can enter into the sum with either plus or minus. As a result, we obtain about 2^{N_j} possible initial approximations, where N_j is the number of defects in the region to be calculated. Note that one of the solutions to the initial nonlinear equation (a local potential energy minimum) can correspond to each of the initial conditions. Finding the global minimum requires comparing the energies corresponding to all of the obtained solutions.

The solution depends on the a_i values; characteristic distance R between step edges (reduced to the dimensionless form through dividing by interatomic distance b); the parameter

$$\alpha_f = \frac{J_{f,af}S_{af}}{J_f S_f} \tag{11}$$

characterizing the ratio between the exchange interaction energies of the nearest spins belonging to different layers and to the ferromagnetic layer, respectively; and the α_{af} parameter determined by (11) with replacement of the *f* and *af* indices.

4. A FERROMAGNET–ANTIFERROMAGNET MULTILAYER

In [15], we obtained the phase diagram of a ferromagnetic film on an antiferromagnetic substrate.

Consider a three-layer system comprising two ferromagnetic layers separated by an antiferromagnetic interlayer. Because of the presence of a large number of parameters, we will restrict our consideration to equal thicknesses of all layers. In addition, in Fe/Cr multilayers that we are interested in, the magnetic stiffness of ferromagnetic layers exceeds that of antiferromagnetic layers; that is, the γ parameter equal to the ratio between the exchange energies in the film and in the substrate,

$$\gamma = \frac{J_f S_f^2}{J_{af}S_{af}^2} = \frac{\alpha_{af}}{\alpha_f}, \tag{12}$$

is much larger than one ($\gamma \gg 1$). Otherwise, if the interlayer is magnetically hard, interaction of ferromagnetic layers is weak and the problem reduces to a two-layer system.

A three-layer system can have three different states.

Phase A. At large parameter R values, all layers are separated into domains with parallel and antiparallel orientations of ferromagnetic layer magnetizations. Domain walls penetrate all three layers, and their coordinates in the plane of layers coincide with atomic step edges on both boundary surfaces. Ferromagnetic layer magnetizations experience rotations in opposite directions in the domain wall. The antiferromagnetic order parameter rotates together with the magnetization vector of the ferromagnetic layer whose boundary with the antiferromagnetic layer does not contain a step at a given place.

The structure and energy of a domain wall depend on the $\alpha_{af}a$ parameter. If $\alpha_{af}a \ll 1$, the $\theta_{f(af)}(z)$ dependence, that is, domain wall broadening, can be ignored, and the problem becomes one-dimensional.

The $|\nabla\theta_f|$ value in the domain wall is of the order of δ_f^{-1}. Energy w_1 per unit domain wall length can be estimated by (1). This gives

$$w_i \approx \frac{J_f S_f^2 a}{b\delta_f}. \tag{13}$$

At the same time, spins at the interface are frustrated in the region $|x| \lesssim \delta_f$, which increases the interaction energy between layers by

$$w_2 \approx \frac{J_{f,af}S_f S_{af}\delta_f}{b} \tag{14}$$

per unit domain wall length.

Minimizing the $w = w_1 + w_2$ sum, we find

$$\delta_f \approx \sqrt{a/\alpha_f}. \tag{15}$$

The domain wall width in the antiferromagnetic interlayer is $\delta_{af} \approx \sqrt{a/\alpha_{af}} = \delta_f/\sqrt{\gamma} \ll \delta_f$. The distribution of order parameters in the wall is shown in Fig. 2a; the energy of the wall per unit length is

$$w \approx \frac{J_f S_f^2}{b}\sqrt{a\alpha_f} \sim \frac{S_f}{b}\sqrt{aJ_f J_{f,af} S_f S_{af}}. \qquad (16)$$

Exact numerical calculations of δ_f and w in a wide range of α_f and a parameter values substantiate the estimates given above (this also refers to what follows).

However, if $\alpha_{af} a \gg 1$, the domain wall width in the antiferromagnetic interlayer considerably increases as the distance from the interlayer boundary containing an atomic step grows larger. The character of the distribution of order parameters in the domain wall is shown in Fig. 2b. Its parameters are found by estimations similar to those made for $\alpha_{af} a \ll 1$. The major contribution to wall energy is made by order parameter distortions in the antiferromagnet. In the region $|x| \lesssim a$, the $|\nabla\theta_{af}|$ value is inversely proportional to the distance from the step, whereas at $a \ll |x| \ll \delta'_f$, where δ'_f is the domain wall width in ferromagnetic layers, lines of constant θ_{af} values are virtually parallel to the interlayer boundaries (Fig. 2c). In this region, $|\nabla\theta_{af}| \approx a^{-1}$.

The smallest domain wall thickness in the antiferromagnet is $\delta_0^{af} = (1 + \alpha_{af})/\alpha_{af}$, the derivative near the step is given by $\partial\delta_{af}/\partial z \approx 1$, and the δ'_f value equals

$$\delta'_f \approx a\sqrt{\gamma} \gg a. \qquad (17)$$

The energy of the domain wall per unit length is

$$w \approx \frac{J_{af} S_{af}^2}{b}\left(\sqrt{\gamma} + \ln\frac{a}{\delta_0^{af}}\right). \qquad (18)$$

Clearly, δ_0^{af} is of the order of the interatomic distance, and the mean domain wall width amounts to dozens of angstrom units; that is, domain walls caused by frustrations are much narrower than usual domain walls in a ferromagnet whose width is determined by competition between exchange and anisotropy energies.

Phase B. Because the magnetic stiffness of ferromagnetic layers exceeds that of the antiferromagnetic interlayer ($\gamma \gg 1$), the transition to the state in which ferromagnetic layers are virtually uniform occurs at $R_c = \delta_f(\delta'_f)$ as R continuously decreases because of overlap of domain walls. Additional energy is related either to distortions in the antiferromagnetic interlayer or to the boundary energy. Close to the interlayer Néel temperature T_N (T_N is lower than the Curie temperature of the ferromagnet), we have $\gamma \propto T_N/(T_N - T)$. The

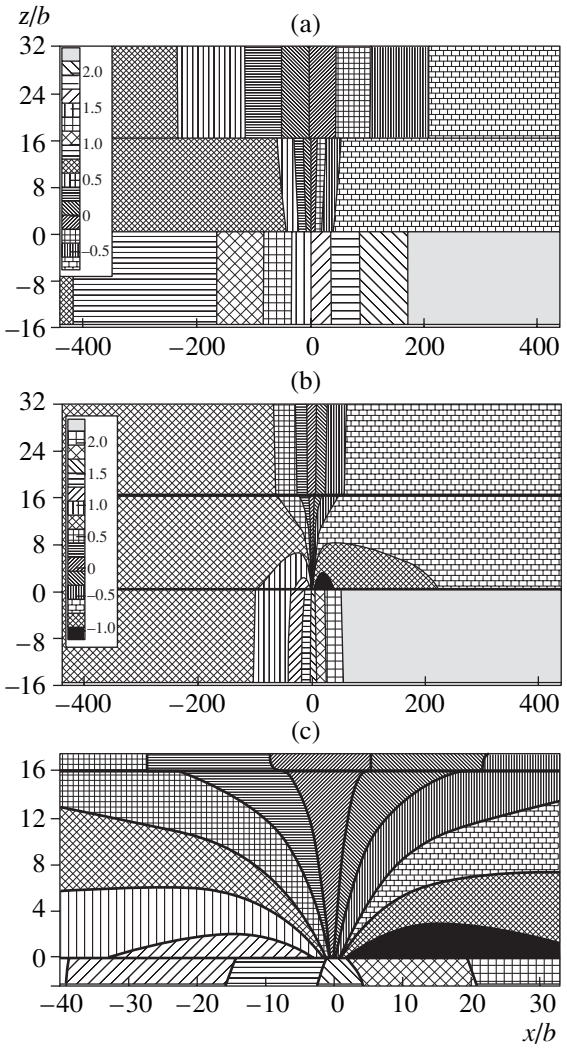

Fig. 2. Domain wall in three-layer system: (a) $\alpha_{af} a \ll 1$ and (b) $\alpha_{af} a \gg 1$. Different levels of θ_i angle of order parameter rotations in radians are shown by different hatching patterns (see scale). The $z = 0$ and 16 coordinates correspond to interlayer boundaries. The step is situated at $x = 0$, $z = 0$. The distributions were obtained for (a) $\gamma = 10$, $\alpha_{af} = 0.01$, and $a = 16$ and (b) $\gamma = 10$, $\alpha_{af} = 1$, and $a = 16$. The distribution in Fig. 2c is the central part of the distribution shown in Fig. 2b.

$A \longrightarrow B$ transition can therefore be effected by heating the system from the initial temperature $T_0 < T_N$.

Precisely in the region of R values where phase B exists, the Slonczewski magnetic proximity model is applicable [9].

In the region of values $\max(a, \delta_0^{af}) \ll R \ll R_c$, the dependence of system energy on angle ψ between the magnetizations of ferromagnetic layers is described by the formula

$$W = C_1\psi^2 + C_2(\pi - \psi)^2. \qquad (19)$$

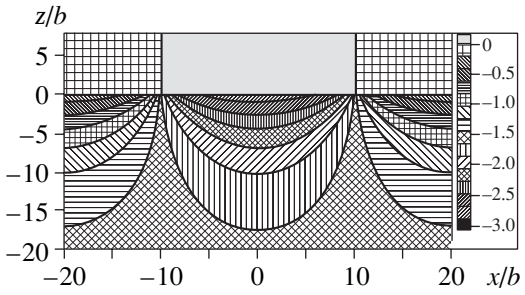

Fig. 3. Static spin vortex in the film–substrate system at $\gamma \gg 1$. Different levels of the angle of order parameter rotations in radians are shown by different hatching patterns (see scale). The distribution was obtained for $\gamma = 8$, $\alpha_{af} = 1$, and $a = 8$. The $z = 0$ value corresponds to the film–substrate boundary. Step edges are situated at $x = \pm 10$.

Estimation gives the following C_1 and C_2 constant values [8]:

$$C_{1,2} = \begin{cases} \dfrac{J_{af}S_{af}^2}{2a}\dfrac{\sigma_{1,2}}{b^2}, & \alpha_{af}a \gg 1, \\ \dfrac{4(\sqrt{2}-1)}{\pi^2}J_{af}S_fS_{af}\dfrac{\sigma_{1,2}}{b^2}, & \alpha_{af}a \ll 1, \end{cases} \quad (20)$$

where $\sigma_{1,2}$ is the area of regions of the first (second) type on the layer surface.

If $\sigma_1 = \sigma_2$, the energy minimum is attained at $\psi = \pi/2$; that is, mutually perpendicular orientation of the magnetizations of ferromagnetic layers exists in the absence of an external magnetic field.

In state A, system energy does not depend on the direction of order parameter rotations in domain walls. The situation is different in state B. Overlapping of domain walls removes degeneracy with respect to rotation directions, and a large number of metastable states arise. These states differ from each other in the direction and magnitude of the rotation angle of the antiferromagnetic order parameter in separate regions bounded by atomic steps.

What does happen when R decreases further? If $\alpha_{af}a \ll 1$, the system occurs in the region of weak distortions at $a \ll R \ll \delta_{af}$. In this region, order parameters are virtually uniform, and the magnetizations of the ferromagnetic layers remain perpendicular to each other, whereas the $C_{1,2}$ constants decrease compared with (20) by a factor of $(R/\delta_{af})^2$.

Phase C. Consider the $R \ll a$ region. All distortions are then concentrated close to interlayer boundaries, the interaction between ferromagnetic layers becomes weak, and the key role is played by the interaction energy between neighboring layers, which we considered in [15] for the example of a two-layer system.

As a result, the antiferromagnetic order parameter becomes oriented normally to the magnetizations of ferromagnetic layers at $\sigma_1 = \sigma_2$, and these magnetizations become collinear. This is state C.

If $\alpha_{af}a \gg 1$, static vortices are formed in the antiferromagnetic interlayer at $\delta_0^{af} \ll R \ll a$ close to the boundaries (Fig. 3). At smaller R values, the system goes into the region of weak distortions.

If $\alpha_{af}a \ll 1$, the transition from state B to C already occurs in the region of weak distortions. Like phase B, phase C is characterized by the presence of a large number of metastable states. As follows from the results of our modeling, the transition from phase B to C is a first-order phase transition. Both states coexist in a considerable range of R values, and their energies become equal at some $R^* \sim a$ (Fig. 4). The phase diagram of the three-layer system is shown in Fig. 5. In [16], the orientation of spins in a three-layer structure at $R \sim a$ was cal-

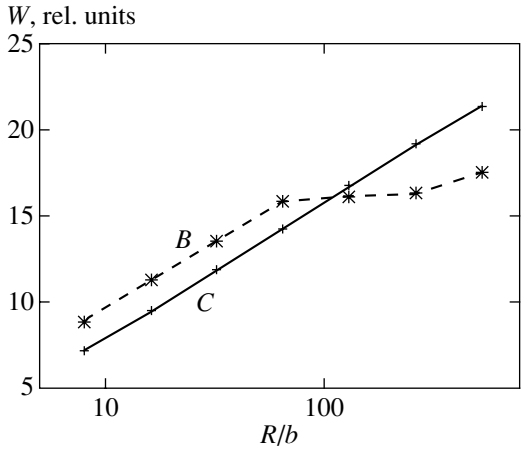

Fig. 4. Phase B and C energies as functions of distance R between steps ($a = 64$, $\alpha_f = 1/8$, and $\alpha_{af} = 1$).

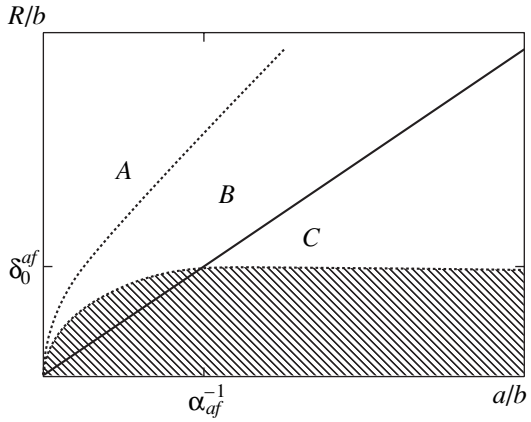

Fig. 5. Phase diagram of the three-layer system. For convenience, $R = a$ (solid) and $R = \delta_f$ and $R = \delta_0^{af}$ (dashed) lines are shown. The region of weak distortions is hatched.

culated using the discrete model. The magnetizations of neighboring ferromagnetic layers were assumed to be mutually orthogonal. The behavior of the system in the whole range of R values was not analyzed.

The R^* value is independent of temperature, and the B $\longrightarrow$ C phase transition cannot therefore be observed by varying the temperature of the system. The transition from the state with strong biquadratic exchange to the low-temperature state with a weak interaction between layers described in [17] has no bearing on the phase transition considered above. This transition occurs because the interaction of ferromagnetic layers through the antiferromagnetic order parameter decreases as the system approaches the Néel temperature and becomes equal to the interaction through spin polarization induced in the antiferromagnet (the Riderman–Kittel–Kasuya–Yoshida interaction) [8]. At a higher temperature, the latter interaction predominates.

5. CONCLUSION

The most important results and conclusions of this work are as follows.

(1) A simple model for determining the distribution of spins in frustrated layered ferromagnet–antiferromagnet structures was suggested.

(2) The thickness–roughness phase diagram for a three-layer ferromagnet–antiferromagnet–ferromagnet system was obtained.

(3) The transition from the polydomain (phase A) to the monodomain (phase B) state of ferromagnetic layers as the distance between atomic steps at the interlayer boundary decreased was shown to occur continuously. Strictly, this transition was shown to be not a phase transition.

(4) At the same time, the transition from the phase with a mutually orthogonal orientation of the magnetizations of neighboring ferromagnetic layers (phase B) to the phase with their collinear orientation (phase C) when the distance between atomic steps became smaller than the thickness of the antiferromagnetic interlayer was shown to be a first-order phase transition.

(5) The Slonczewski phenomenological magnetic proximity model was only valid for phase B. The parameters of this model in the whole range of its applicability were determined.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 00-02-17162 and 01-02-16716) and by a joint project of the US Civilian Research and Development Foundation (CRDF) and the Ministry of Education of the Russian Federation (grant VZ-010-0).

REFERENCES

1. M. N. Baibich, J. M. Broto, A. Fert, *et al.*, Phys. Rev. Lett. **61**, 2472 (1988).
2. D. T. Pierce, J. Unguris, R. J. Celotta, and M. D. Stiles, J. Magn. Magn. Mater. **200**, 290 (1999).
3. R. S. Fishman, J. Phys.: Condens. Matter **13**, R235 (2001).
4. A. Schreyer, C. F. Majkrzak, Th. Zeidler, *et al.*, Phys. Rev. Lett. **79**, 4914 (1997).
5. P. Bodeker, A. Schreyer, and H. Zabel, Phys. Rev. B **59**, 9408 (1999).
6. M. Chirita, G. Robins, R. L. Stamp, *et al.*, Phys. Rev. B **58**, 869 (1998).
7. S. Yan, R. Schreiber, F. Voges, *et al.*, Phys. Rev. B **59**, R11641 (1999).
8. A. I. Morozov and A. S. Sigov, Fiz. Tverd. Tela (St. Petersburg) **41**, 1240 (1999) [Phys. Solid State **41**, 1130 (1999)].
9. J. C. Slonczewski, J. Magn. Magn. Mater. **150**, 13 (1995).
10. A. Berger and H. Hopster, Phys. Rev. Lett. **73**, 193 (1994).
11. E. J. Escorcia-Aparicio, H. J. Choi, W. L. Ling, *et al.*, Phys. Rev. Lett. **81**, 2144 (1998).
12. C. M. Schmidt, D. E. Bürgler, D. M. Schaller, *et al.*, Phys. Rev. B **60**, 4158 (1999).
13. E. J. Escorcia-Aparicio, J. H. Wolfe, H. J. Choi, *et al.*, Phys. Rev. B **59**, 11892 (1999).
14. M. Freyss, D. Stoeffler, and H. Dreysse, Phys. Rev. B **56**, 6047 (1997).
15. V. D. Levchenko, A. I. Morozov, and A. S. Sigov, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 544 (2000) [JETP Lett. **71**, 373 (2000)].
16. C. Cornea and D. Stoeffler, Europhys. Lett. **49**, 217 (2000).
17. E. E. Fullerton, C. H. Sowers, and S. D. Bader, Phys. Rev. B **56**, 5468 (1997).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 94, No. 5, 2002, pp. 992–1007.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 121, No. 5, 2002, pp. 1157–1173.
Original Russian Text Copyright © 2002 by Kim, Roznova, Satanin, Stenberg.

SOLIDS
Electronic Properties

Interference of Quantum States in Electronic Waveguides with Impurities

C. S. Kim[a], O. N. Roznova[b], A. M. Satanin[b, *], and V. B. Stenberg[b]

[a]*Applied Physics, Yale University, New Haven, Connecticut 06520-8284, USA*
[b]*Nizhni Novgorod State University, pr. Gagarina 33, Nizhni Novgorod, 603600 Russia*
**e-mail: satanin@phys.unn.runnet.ru*

Received November 23, 2001

Abstract—Effects of interference between propagating and localized states in quasi-one-dimensional electronic waveguides containing finite-size attracting impurities (quantum dots) are investigated. The electron scattering matrix is calculated in the framework of the Feshbach theory [H. Feshbach, Ann. Phys. **5**, 357 (1958); Ann. Phys. **19**, 287 (1962)], when resonant states in closed channels are taken into account exactly, while nonresonant states are taken into account in perturbation theory. It is shown that finite-size attracting impurities may generate a series of asymmetric Fano resonances in the waveguide transmission. As a result of interference of electron states, the characteristics of resonances may oscillate upon a change in the impurity parameters. The conditions are determined under which the interference of an electron wave leads to a "collapse" and "swing" of Fano resonances. © *2002 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Quasi-one-dimensional electron waveguides with impurities (quantum dots) may display in principle new coherent effects, where a propagating electron wave interferes with a localized impurity state. In this case, asymmetric (Fano) resonances may exist against the background of the waveguide conductance (transmission) steps [3–5]. The interference effects closely associated with Fano resonances [6] have been actively studied in recent years both theoretically and experimentally. Fano resonances are universal by nature and were observed earlier in various systems: in photoionization of atoms [7], in electron and neutron scattering [8, 9], in Raman scattering [10], in photoabsorption in quantum wells and superlattices [11–13], and in electron transport through an interferometer [14]. Fano resonances are due to interference of waves emerging as a result of scattering at a heterostructure; they carry important information on its geometrical relief and intrinsic potential fields. An analysis of the mechanisms governing the transmission of electronic waveguides is important for applications since this effect can be used for creating resonance nanoelectronic instruments of a new generation [15]. The solution of this type of problems necessitates the development of a theory of scattering of electron states in waveguides with impurities. Earlier, isolated resonances were studied in optics, in the case of scattering of electrons from atoms and molecules, and in the theory of nuclear reactions. In [1, 2], a dispersion representation was obtained for the scattering matrix of multichannel systems. The resonance overlapping effects were considered for the first time in [16]. It should be noted that the pole structure of the scattering matrix, which determines the transmission peaks, is quite clear, while the effects associated with total reflection (existence of transmission dips) are studied insufficiently. In recent publications [17–19], the effect of short-range impurities on the transmission was mainly considered and it was shown that such impurities lead to the emergence of additional peaks and valleys against the background of transmission steps.

In the present work, we study the scattering of electron waves at finite-size impurities in a quasi-one-dimensional waveguide. A basically new aspect here is the possible interference (configuration interaction) of a propagating wave with a wave trapped in the region of impurity, which may change qualitatively the waveguide transmission. First, we formulate the resonance theory of scattering which can be used for solving a wide range of problems in the theory of electron transport in nanochannels. Some elements of the procedure applied here are based on the works of Feshbach [1, 2] and have already been used in [20, 21] for single resonances. The theory proposed here is based on the concept of the resonant group of states emerging in the wells split from the size-quantization subbands. The electron scattering matrix is calculated under the assumption that resonant states in closed channels can be included exactly, while nonresonant states are taken into account in perturbation theory. The theory developed in the present work makes it possible to describe interference between the propagating and localized states, resulting in the emergence of asymmetric resonances. In the case of a finite-size impurity, a series of Fano resonances may exist [22], and the interference of quantum states in open and closed scattering channels

may lead to oscillations of the characteristics of a Fano resonance upon a change in the parameters of the system [23]. For some specific models of impurities, we investigate in greater detail the conditions under which the effects of coherent interaction of states may lead to the collapse of Fano resonances, when the width of the resonances vanishes. It is shown that the effect under investigation is similar to annihilation of a particle (resonance) with a hole (resonance zero). We propose a simple numerical procedure enabling us to calculate the characteristics of a waveguide with impurities in a quasi-one-dimensional approximation and describe the results of simulation upon the variation of the impurity parameters.

2. MODEL OF AN ELECTRONIC WAVEGUIDE AND EQUATIONS

We will study the scattering of electron waves in a 2D electronic waveguide of width W arranged along the x axis. Let the confining potential in the transverse direction be described by the function $V_c(y)$. We will describe the potential of impurities by the function $V(x, y)$. The waveguide geometry is presented schematically in Fig. 1 showing the potential field lines in the waveguide.

The electron wave function can be determined from the Schrödinger equation

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\Psi(x, y) + V_c(y)\Psi(x, y) + V(x, y)\Psi(x, y) = E\Psi(x, y), \quad (1)$$

where m is the electron effective mass. For a channel free of impurities, $V(x, y) = 0$, and the solution to Eq. (1) in this case can be written in the form

$$\psi^{(0)}_{n,k}(x, y) = e^{ik_n x}\varphi_n(y), \qquad E = \frac{\hbar^2 k_n^2}{2m} + E_n, \quad (2)$$

where $\varphi_n(y)$ and E_n are defined by the solutions to the equation

$$\left\{-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial y^2} + V_c(y)\right\}\varphi_n(y) = E_n\varphi_n(y). \quad (3)$$

It is convenient to expand the wave function $\Psi(x, y)$ in the complete basis of functions describing the transverse motion:

$$\Psi(x, y) = \sum_{n=1}^{\infty}\psi_n(x)\varphi_n(y). \quad (4)$$

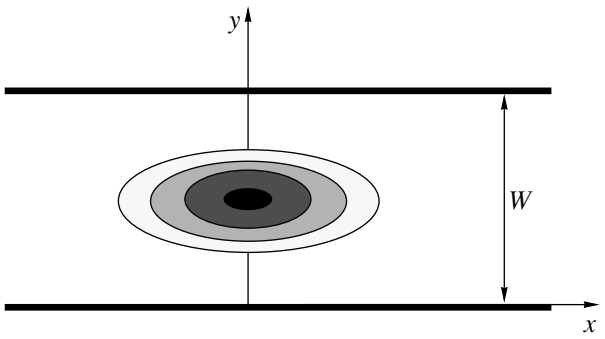

Fig. 1. Schematic diagram of a 2D electronic waveguide containing an attracting impurity. Constant-potential lines are depicted; different shades characterize the variation of the potential.

Substituting expansion (4) into Eq. (1), we obtain the equation for $\psi_n(x)$ in the form

$$-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi_n(x) + \sum_{n'=1}^{\infty}V_{nn'}(x)\psi_{n'}(x) = (E - E_n)\psi_n(x), \quad (5)$$

where

$$V_{nn'}(x) = \int\varphi_n(y)V(x, y)\varphi_{n'}(y)dy \quad (6)$$

are the matrix elements of the impurity potential. The system of equations (5) is completely equivalent to the 2D Schrödinger equation. It should be noted that an infinite system of coupled equations of type (5) is often encountered in physics (linear theory of interaction of waves) and the development of methods for solving such equations is of independent interest.

It was shown in [24, 25] that the calculation of the conductance of a waveguide with impurities in the ballistic mode is reduced to the solution of the scattering problem. We will be interested in the transmission amplitude $t_{nn'}$ describing the scattering of electrons from the channel with number n' into a channel with number n. The transmission amplitude can be determined from the solution to Eq. (5). The conductance measured by the two-probe method is determined by the Buttiker–Landauer formula [24, 25]

$$G = \frac{2e^2}{h}T, \quad T = \sum_{n, n'}\frac{k_n}{k_{n'}}|t_{nn'}|^2, \quad (7)$$

where T is the transmission of the waveguide, n and n' denote the channel numbers for incident and scattered waves, and summation is carried out over all states with energy E propagating in the quantum waveguide. It should be noted that the poles of the scattering amplitude $t_{nn'}(E)$ in the complex plane E correspond to levels or resonances, while branching points correspond to threshold singularities [26].

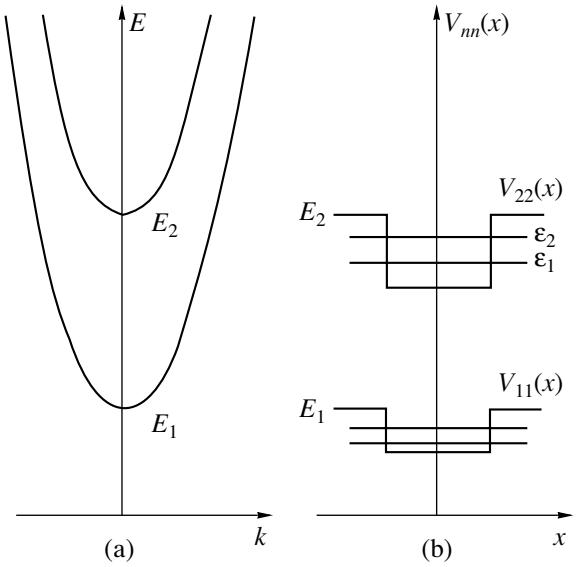

Fig. 2. (a) Dispersion relation for charge carriers in an ideal waveguide and (b) the diagram of discrete energy levels in effective wells.

3. SCATTERING MATRIX AND RESONANCE APPROXIMATION

The theory of resonances based on the relation between a local state and states in open channels was formulated long ago by Feshbach [1, 2]. In this section, we will derive the expression for the electron scattering matrix in a waveguide containing a finite-size impurity in the framework of the Feshbach theory. We will take into account an arbitrary number of possible bound states in the field of the impurity.

A. Scattering Matrix

We begin our analysis with the case of a single open channel; in other words, the energy of an electron being scattered lies in the interval $E_1 < E < E_2$ (Fig. 2a). Proceeding in accordance with [1], we retain in the system of equations (5) only the terms which correspond to the closest values of energy (resonance terms); the discarded terms can be taken into account later in perturbation theory. In other words, we retain in Eqs. (5) the terms containing ψ_1 and ψ_2:

$$\left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V_{11}(x)\right)\psi_1(x) + V_{12}(x)\psi_2(x) = (E - E_1)\psi_1(x), \tag{8}$$

$$\left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V_{22}(x)\right)\psi_2(x) + V_{21}(x)\psi_1(x) = (E - E_2)\psi_2(x), \tag{9}$$

under the assumption that the matrix element V_{12} describing the coupling between the channels is smaller than the interval between the quantization subbands.

It is more convenient to begin the analysis with Eq. (9). We disregard in Eq. (9) the term containing V_{21}, it assumes the form of the one-dimensional Schrödinger equation:

$$\left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V_{22}(x)\right)\psi_2(x) = (E - E_2)\psi_2(x). \tag{10}$$

If potential $V_{22}(x)$ is of the attracting type and has a large magnitude, bound states exist for certain values of energy E. Depending on the parameters of the well of $V_{22}(x)$ in the energy interval $E_2 - \max|V_{22}(x)| < E < E_2$, Eq. (10) may have a series of bound states u_j with energies $\mathscr{E}_j$ ($j = 1, \ldots, N$) (Fig. 2b), where the functions u_j are normalized by the conditions

$$\langle u_j | u_{j'} \rangle = \delta_{jj'}. \tag{11}$$

In addition to discrete energy levels, the field of potential $V_{22}(x)$ usually contains the states belonging to the continuous spectrum, which will be denoted by $u_\nu(x)$. Taking into account localized states and the states belonging to the continuum, we will seek the solution to Eq. (9) in the form

$$\psi_2(x) = \sum_{j=1}^{N} A_j u_j(x) + \int A_\nu u_\nu(x) d\nu \equiv \sum_\lambda A_\lambda u_\lambda(x), \tag{12}$$

where $A_\lambda = (A_j, A_\nu)$ are the amplitudes which are arbitrary so far and $\lambda = (j, \nu)$ is the complete set of quantum numbers, $u_\lambda = (u_j, u_\nu)$.

Substituting expression (12) into Eq. (9) and taking into account the orthogonality of states $u_\lambda(x)$, we obtain the formal equations for amplitudes A_λ:

$$(E - \mathscr{E}_\lambda)A_\lambda = \langle u_\lambda | V_{21} | \psi_1(x) \rangle. \tag{13}$$

Let us now consider Eq. (8). We write it in the form

$$\left[E - E_1 - \left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V_{11}(x)\right)\right]\psi_1(x) = V_{12}(x)\psi_2(x). \tag{14}$$

In the present case, we are dealing with the problem of electron scattering in the potential field $V_{11}(x)$. Formally, the solution Eq. (14) can be written using a Green's function:

$$|\psi_1\rangle = |\chi_E^+\rangle + G_1^+ V_{12} |\psi_2\rangle, \tag{15}$$

where $\langle x | \chi_E^+ \rangle$ is the solution to Eq. (14) with zero right-hand side. We have chosen the particular solution

corresponding to sources at $x \longrightarrow -\infty$. Substituting solution (15) into Eq. (13), we obtain the following explicit equation for amplitudes A_λ:

$$(E - \mathscr{E}_\lambda)A_\lambda - \sum_{\lambda'} U_{\lambda,\lambda'}A_{\lambda'} = F_\lambda, \tag{16}$$

where

$$U_{\lambda,\lambda'} = \langle u_\lambda | V_{21} G_1 V_{12} | u_{\lambda'} \rangle, \quad F_\lambda = \langle u_\lambda | V_{21} | \chi_E^+ \rangle. \tag{17}$$

Function ψ_1, in turn, can be obtained from Eq. (15) if we substitute into its right-hand side the expression for ψ_2 defined by formula (12):

$$|\psi_1\rangle = |\chi_E^+\rangle + \sum_\lambda G_1^+ V_{12} |u_\lambda\rangle A_\lambda. \tag{18}$$

In order to find the scattering matrix t_{11} determined by the asymptotic behavior of $\psi_1(x)$ for $x \longrightarrow +\infty$: $\psi_1(x) = t_{11} e^{ik_1 x}$, we require the following expression for the Green's function:

$$G_1^+(x, x') = \frac{m}{ik_1\hbar^2 t} \begin{cases} \chi_E^+(x)\chi_E^-(x'), & x > x', \\ \chi_E^+(x')\chi_E^-(x), & x < x', \end{cases} \tag{19}$$

where $\chi_E^+(x)$ and $\chi_E^-(x)$ are, respectively, the solutions to the equation

$$\left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V_{11}(x)\right)\chi_E^\pm(x) = (E - E_1)\chi_E^\pm(x) \tag{20}$$

with sources at $x \longrightarrow \pm\infty$. Function $\chi_E^+(x)$ has the following asymptotic forms:

$$\chi_E^+(x) = \begin{cases} te^{ik_1 x}, & x \longrightarrow +\infty, \\ e^{ik_1 x} + re^{-ik_1 x}, & x \longrightarrow -\infty, \end{cases} \tag{21}$$

while the asymptotic forms of $\chi_E^-(x)$ are given by

$$\chi_E^-(x) = \begin{cases} e^{-ik_1 x} + r'e^{ik_1 x}, & x \longrightarrow +\infty, \\ te^{-ik_1 x}, & x \longrightarrow -\infty, \end{cases} \tag{22}$$

where t is the transmission amplitude and r and r' are the amplitudes of reflection during scattering in the field $V_{11}(x)$.

After analyzing the asymptotic form of the wave function $\psi_1(x)$ for $x \longrightarrow +\infty$, proceeding from Eq. (18) and taking into account expressions (19) and (20), we obtain

$$\psi_1(x) = te^{ik_1 x} + \frac{me^{ik_1 x}}{ik_1\hbar^2}\sum_\lambda \langle (\chi_E^-)^* | V_{12} | u_\lambda \rangle A_\lambda. \tag{23}$$

Consequently, the scattering matrix in the channel $1 \longrightarrow 1$ is defined as

$$t_{11} = t\left\{1 + \frac{m}{ik_1\hbar^2 t}\sum_{\lambda,\lambda'} \langle (\chi_E^-)^* | V_{12} | u_\lambda \rangle \times (E - \mathscr{E} - U)^{-1}_{\lambda\lambda'} \langle u_{\lambda'} | V_{21} | \chi_E^+ \rangle \right\}. \tag{24}$$

This formula makes it possible to describe the interaction of resonances during the scattering of an electron from impurities in an electronic waveguide in the energy interval (E_1, E_2). It should be noted that, in the energy interval (E_2, E_3), we must take into account additional resonant states associated with channel 3; i.e., we must consider a system of three coupled equations, etc.

B. Resonance Approximation

If the electron energy is close to the group of energy levels in the potential well of $V_{22}(x)$, precisely these bound states will interfere with the incident wave most strongly. In this group of bound states, the most important are those for which the magnitudes of the matrix elements connecting these states with the states of the continuous spectrum are comparable with the interval between the nearest levels.

We retain in Eq. (16) only the resonant terms, omitting nonresonant states belonging to the continuous spectrum. Let us suppose that the number of amplitudes being retained is determined by the number N of bound states in the well. In this case, the explicit equation for determining amplitudes A_j assumes the form

$$\sum_{j'=1}^{N} M_{jj'}A_{j'} = F_j, \quad M_{jj'} = (E - \mathscr{E}_j)\delta_{jj'} - U_{jj'}, \tag{25}$$

where

$$U_{jj'} = \langle u_j | V_{12} G_1 V_{12} | u_{j'} \rangle, \quad F_j = \langle u_j | V_{21} | \chi_E^+ \rangle. \tag{26}$$

Consequently, in the resonance approximation, the scattering matrix in channel $1 \longrightarrow 1$ is defined as

$$t_{11} = t\left\{1 + \frac{m}{ik_1\hbar^2 t} \times \sum_{j,j'} \langle (\chi_E^-)^* | V_{12} | u_j \rangle (M^{-1})_{jj'} \langle u_{j'} | V_{21} | \chi_E^+ \rangle \right\}. \tag{27}$$

The representation for the scattering amplitude in form (27) is remarkable since it makes it possible to determine the resonant structure of the channel transmission

in the general case. The formal solution to system (25) can be written in the form

$$A_j = \frac{1}{\det M}\sum_{j'=1}^{N}(M^c)_{jj'}F_{j'}, \tag{28}$$

where M^c is the cofactor of matrix M. The scattering amplitude can be written in the form

$$t_{11}(E) = t(E)\frac{\mathcal{N}(E)}{\mathcal{D}(E)}, \tag{29}$$

where

$$\mathcal{N}(E) = \mathcal{D}(E) + \frac{m}{ik_1\hbar^2 t} \times \sum_{j,j'}\langle(\chi_E^-)^*|V_{12}|u_j\rangle(M^c)_{jj'}\langle u_{j'}|V_{21}|\chi_E^+\rangle, \tag{30}$$

$$\mathcal{D}(E) = \det M.$$

It can be seen from this expression that the poles of the scattering amplitude are determined by zeros of the function $\mathcal{D}(E)$, while the zeros of the amplitude are determined by zeros of the numerator of $\mathcal{N}(E)$.

The resonant structure of the scattering matrix depends to a considerable extent on the properties of matrix U, which, in turn, is determined by the properties of the Green's function and the impurity potential $V(x, y)$. We will specially consider the case when the matrix element possesses parity relative to reflections $x \longrightarrow -x$ ($V_{n,n'}(x) = V_{n,n'}(-x)$). We will also consider the case of a general-position potential, when $V_{n,n'}(x) \neq V_{n,n'}(-x)$. In accordance with the reciprocity theorem, the Green's function possesses the following property:

$$G_1(x, x') = G_1(x', x). \tag{31}$$

If, however, the impurity potential is such that $V_{11}(x) = V_{1,1}(-x)$, the equation for the Green's function leads to the additional relation

$$G_1(x, x') = G_1(-x, -x'). \tag{32}$$

Using the reciprocity theorem, we can easily verify that matrix U is symmetric ($U_{jj'} = U_{j'j}$). For the case when the impurity potential possesses parity, we can easily prove that the matrix elements of U between states with different symmetries (symmetric s and antisymmetric a) are equal to zero: $U_{sa} = 0$.

It can be seen from the above expressions that the behavior of the scattering amplitude is determined by the matrix elements $U_{jj'}$. Using formula (19) for the Green's function, we can easily transform the matrix element $U_{jj'}$ to the form

$$U_{jj'} = \frac{m}{ik_1\hbar^2 t}F_j R_{j'} + Q_{jj'}, \tag{33}$$

where

$$Q_{jj'} = -\frac{2m}{k_1\hbar^2|t|^2} \times \mathrm{Im}\left(\int_{-\infty}^{\infty}u_j(x)V_{12}(x)\chi^+(x)\int_x^{\infty}u_{j'}'(x')V_{12}(x')\chi^{+*}(x')\right), \tag{34}$$

$$R_j = \langle(\chi_E^-)^*|V_{12}|u_j\rangle. \tag{35}$$

Using the well-known relations between the wave functions of the scattering problem,

$$\chi_E^+ = r\chi_E^{+*} + t\chi_E^{-*}, \qquad \chi_E^- = r'\chi_E^{-*} + t\chi_E^{+*}, \tag{36}$$

the matrix elements $U_{jj'}$ can be presented in the form

$$U_{jj'} = \frac{m}{2ik_1\hbar^2}(F_jF_{j'}^* + R_jR_{j'}^*) + \frac{m}{2ik_1\hbar^2}\left(\frac{r}{t}R_jF_{j'}^* - \frac{r^*}{t^*}F_jR_{j'}^*\right) + \frac{1}{2}(Q_{jj'} + Q_{j'j}). \tag{37}$$

In particular, the diagonal elements can be written as

$$U_{jj} = -i\frac{m}{2k_1\hbar^2}(|F_j|^2 + |R_j|^2) + \frac{m}{k_1\hbar^2}\mathrm{Im}\left(\frac{r}{t}R_jF_j^*\right) + Q_{jj}, \tag{38}$$

where we have singled out the real and imaginary components of U_{jj}.

The form of the denominator of Eq. (29) can be determined in the general case. For this purpose, we reduce the complex symmetric matrix M to the diagonal form. The elements of the matrix performing a unitary transformation can be found by solving the system of homogeneous equations:

$$\sum_{j'=1}^{N}(\mathscr{E}_j\delta_{jj'} + U_{jj'})\Omega_{j'} = \tilde{\mathscr{E}}\Omega_j. \tag{39}$$

Let us suppose that a set of complex solutions $\tilde{\mathscr{E}}_\alpha$, which can be presented in the form

$$\tilde{\mathscr{E}}_\alpha = \mathscr{E}_\alpha^R - i\Gamma_\alpha, \tag{40}$$

has been determined. In this case, we can write the denominator of Eq. (30) in the form

$$\mathcal{D}(E) = \prod_\alpha(E - \mathscr{E}_\alpha^R + i\Gamma_\alpha). \tag{41}$$

It follows from the structure of the expression for matrix M that $\mathscr{E}_\alpha^R$ and Γ_α are in fact functions of

energy; however, the subsequent analysis will prove that these functions depend on energy only slightly in the case of a weak coupling between the channels.

The expression for $\mathcal{N}(E)$ appearing in the numerator of Eq. (29) may have zeros which indicate the possibility of total reflection of waves in a waveguide with impurities. Below, we will demonstrate by specific examples how these zeros are determined and will consider possible consequences for the channel transmission.

In the case when the well parameters are such that it contains a small number of energy levels (resonances), we must obtain the general expression for the scattering amplitude. Let us consider the case of a symmetric impurity. We assume that the well contains only one bound state. Using expressions (29) and (30) and taking into account Eq. (37), we obtain

$$t_{11}(E) = t(E)\frac{E-\mathcal{E}_1^0}{E-\mathcal{E}_1^R+i\Gamma_1}, \tag{42}$$

where the real resonance parameters have been introduced:

$$\mathcal{E}_1^0 = \mathcal{E}_1 + Q_{11}, \quad \mathcal{E}_1^R = \mathcal{E}_1 + Q_{11} + \Delta_{11},$$
$$\Delta_{11} = \frac{m}{\hbar^2 k_1}|F_1|^2\mathrm{Im}\left(\frac{r}{t}\right), \tag{43}$$

$$\Gamma_1 = \frac{m}{\hbar^2 k_1}|F_1|^2. \tag{44}$$

It follows from Eq. (42) that the transmission amplitude and, hence, the transmission have the structure of a Fano resonance: a peak of width Γ_1 at energy $\mathcal{E}_1^R$ and a zero at energy $\mathcal{E}_1^0$ occur against the potential background determined by the amplitude $t(E)$. It should be noted that the peak width is determined by the matrix element connecting a localized state with the continuum of states of band 1.

In the case when a symmetric well contains two energy levels, they can interact only through virtual transitions to the continuum. The transmission amplitude acquires the form

$$t_{11}(E) = t(E)\frac{(E-\mathcal{E}_1^0)(E-\mathcal{E}_2^0)+\gamma_1\gamma_2}{(E-\mathcal{E}_1^R+i\Gamma_1)(E-\mathcal{E}_2^R+i\Gamma_2)}, \tag{45}$$

where

$$\mathcal{E}_1^0 = \mathcal{E}_1 + Q_{11}, \quad \mathcal{E}_2^0 = \mathcal{E}_2 + Q_{22},$$
$$\gamma_1 = \Gamma_1 + \Delta_{11}, \quad \gamma_2 = \Gamma_2 + \Delta_{22},$$
$$\Delta_{11} = \frac{m}{\hbar^2 k_1}|F_1|^2\mathrm{Im}\left(\frac{r}{t}\right), \quad \Delta_{22} = \frac{m}{\hbar^2 k_1}|F_2|^2\mathrm{Im}\left(\frac{r}{t}\right),$$
$$\Gamma_1 = \frac{m}{\hbar^2 k_1}|F_1|^2, \quad \Gamma_2 = \frac{m}{\hbar^2 k_1}|F_2|^2,$$
$$\mathcal{E}_1^R = \mathcal{E}_1 + Q_{11} + \Delta_{11}, \quad \mathcal{E}_2^R = \mathcal{E}_2 + Q_{22} + \Delta_{22}.$$

It can be seen that the "initial" levels do not interact directly ($U_{12} = 0$), but are connected through the matrix element describing the transition from the local level to band 1. As a consequence, the positions of two zeros on the real energy axis depend on the parameter of coupling between localized states and the continuum. The scattering amplitude also has two poles. Thus, Fano resonances may interact effectively. It will be proved below that this may lead to a number of interesting consequences.

In the case of three energy levels, both the interaction of levels with the continuum and their direct interaction are possible since the matrix element between the states with the same parity differs from zero in this case ($U_{13} \neq 0$). The transmission amplitude can be written in the form

$$t_{11}(E) = t(E)\left(\frac{\mathcal{N}_{13}(E)}{\mathcal{D}_{13}(E)} - \frac{mF_2^2}{ik_1\hbar^2 t(E)(E-\mathcal{E}_2^R+i\Gamma_2)}\right), \tag{46}$$

where

$$\mathcal{N}_{13}(E) = (E-\mathcal{E}_1^0)(E-\mathcal{E}_3^0) - Q_{13}^2,$$
$$\mathcal{D}_{13}(E) = \mathcal{N}_{13}(E) + \mathcal{S}_{13}(E),$$
$$\mathcal{S}_{13}(E) = -\frac{m}{ik_1\hbar^2 t}$$
$$\times (2Q_{13}F_1F_3 + (E-\mathcal{E}_3^0)F_1^3 + (E-\mathcal{E}_1^0)F_3^2),$$
$$\mathcal{E}_2^R = \mathcal{E}_1 + Q_{22} + \Delta_{22}, \quad \Delta_{22} = \frac{m}{\hbar^2 k_1}|F_2|^2\mathrm{Im}\left(\frac{r}{t}\right),$$
$$\Gamma_2 = \frac{m}{\hbar^2 k_1}|F_2|^2.$$

It can easily be verified that the expression in the numerator of Eq. (46), which can be presented in the form

$$\mathcal{N}(E) = (E-\mathcal{E}_2^0)\mathcal{N}_{13}(E) - \frac{m}{\hbar^2 k_1 t}F_2^2\mathcal{S}_{13}(E),$$

is a real function having three zeros in the energy range under investigation, i.e., in the vicinity of the corresponding three energy levels in the well.

Finally, we can write the expression for the amplitude for a well containing four levels. In this case, we have

$$t_{11}(E) = t(E)\left(\frac{\mathcal{N}_{13}(E)\mathcal{N}_{24}(E) - \mathcal{S}_{13}(E)\mathcal{S}_{24}(E)}{\mathcal{D}_{13}(E)\mathcal{D}_{24}(E)}\right), \tag{47}$$

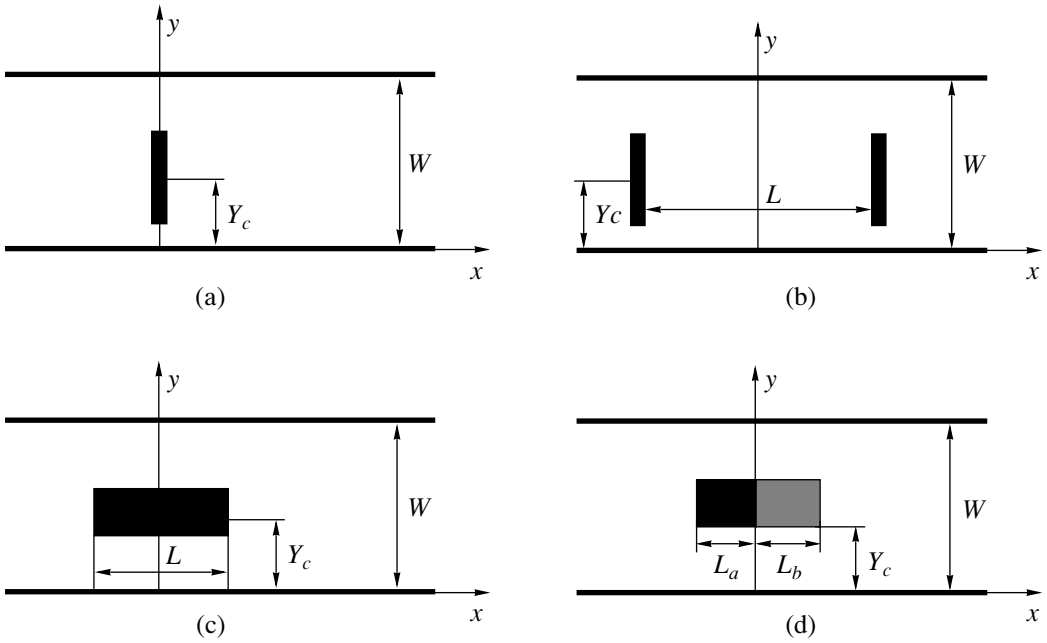

Fig. 3. Models of artificial impurities (quantum dots).

where the expression appearing in the formula has a structure similar to that for the above cases of interaction of resonance pairs. It can be seen that resonances directly interact in pairs. We can easily verify that the function

$$\mathcal{N}(E) = \mathcal{N}_{13}(E)\mathcal{N}_{24}(E) - \mathcal{S}_{13}(E)\mathcal{S}_{24}(E)$$

is real and find its four real zeros. In this case, the pole function $\mathcal{D}(E) = \mathcal{D}_{13}(E)\mathcal{D}_{24}(E)$ defines two pairs of poles.

In fact, the scattering matrix can be easily calculated in the resonance approximation by using numerical methods for any impurity. For this purpose, we must construct localized states in the one-dimensional well $V_{22}(x)$ and the scattering states in the well $V_{11}(x)$. Then, we must determine the matrix elements and solve the linear system of equations (25). The obtained expression makes it possible to study the dependence of transmission on the parameters of the scattering potential. The results of such calculations will be given below.

4. INTERFERENCE OF RESONANT STATES

We will now use the general expressions derived above for the scattering amplitude in the resonance approximation to study interesting interference effects in electronic waveguides with various types of scatterers.

A. Solitary Resonance

To begin with, we will analyze a resonance in the case of an impurity with a finite transverse dimension; the size of the impurity along the channel is assumed to be smaller than the electron wavelength (Fig. 3a). The matrix elements of the potential are defined as

$$V_{nn'}(x) = -\frac{\hbar^2}{m}\nu_{nn'}\delta(x), \tag{48}$$

where $\nu_{nn'} > 0$. In this case, a single level splits from the subband $n = 2$; this level interacts with the states of the subband $n = 1$. The solution to Eq. (10) may give only one bound state in the energy interval $E_1 < E < E_2$:

$$\mathcal{E}_1 = E_2 - \frac{\hbar^2\nu_{22}^2}{2m} \tag{49}$$

with the wave function

$$u_1(x) = \sqrt{\nu_{22}}\exp(-\nu_{22}|x|). \tag{50}$$

The scattering amplitude $t \equiv t(E)$ is determined by solving Eq. (20); it can be written in the form

$$t(E) = \frac{ik_1}{ik_1 + \nu_{11}}. \tag{51}$$

In this case, $t(E)$ is a monotonic function of energy since its pole lies outside the interval $E_1 < E < E_2$. The matrix elements determining the resonance structure of t_{11} are given by

$$F_1^2 = \frac{\hbar^2\nu_{12}^2\nu_{22}t}{m}, \quad U_{11} = \frac{\hbar^2\nu_{12}^2\nu_{22}t}{imk_1}. \tag{52}$$

As a result of substitution of the derived expressions into Eq. (29), we obtain

$$\mathcal{N}(E) = E - \mathcal{E}_1, \quad \mathcal{D}(E) = E - (\mathcal{E}_1 + \Delta_1) + i\Gamma_1, \tag{53}$$

where

$$\Delta_1 = \frac{\hbar^2 \nu_{12}^2 \nu_{22} \nu_{11}}{m(k_1^2 + \nu_{11}^2)}, \quad \Gamma_1 = \frac{\hbar^2 \nu_{12}^2 \nu_{22} k_1}{m(k_1^2 + \nu_{11}^2)}. \tag{54}$$

It should be emphasized that it is configuration interaction of a level in the well $V_{22}(x)$ with a state from the continuous spectrum of the subband $n = 1$ that leads to the emergence of an asymmetric resonance in transmission.

It is useful to compare the expression derived for the transmission amplitude in the resonance approximation with the expression obtained as a result of the exact solution to Eqs. (8) and (9) for potential (48). We can easily find that

$$t_{11}(E) = \frac{ik_1(ik_2 + \nu_{22})}{(ik_1 + \nu_{11})(ik_2 + \nu_{22}) - \nu_{12}^2}. \tag{55}$$

This expression differs formally from formula (29) with $\mathcal{N}(E)$ and $\mathcal{D}(E)$ defined by formulas (53), but it can be reduced to the same form by expanding the numerator and denominator of the expression for $t_{11}(E)$ in the vicinity of its zero and pole and omitting small nonresonant terms. The obtained result demonstrates the obvious fact that the resonance approximation correctly describes the transmission structure in the energy range $E_1 < E < E_2$.

B. Interference of a Resonant Pair

More interesting interference effects are observed when resonances can interact with one another. It is well known that Breit–Wigner resonances repel one another during their interaction, while Fano resonances may exhibit a new effect which was called earlier the collapse of resonances [23]. We will analyze the collapse of resonances in the framework of the theory developed above applied to two wells (see Fig. 3b) each of which can be described by a matrix of the type (48). For this purpose, we require explicit expressions for the corresponding matrix elements. The matrix elements $V_{n,n'}(x)$ for a two-well system are defined as

$$V_{n,n'}(x) = -\frac{\hbar^2}{m}\nu_{n,n'}\left(\delta\left(x - \frac{L}{2}\right) + \delta\left(x + \frac{L}{2}\right)\right). \tag{56}$$

In the case under investigation, parity is conserved and we can consider even and odd states of scattering separately. Solving Eq. (10) for even states, we obtain

$$u_s(x) = \begin{cases} a_s \cosh(|k_2|x), & |x| < L/2, \\ b_s e^{-|k_2||x|}, & |x| > L/2, \end{cases} \tag{57}$$

where a_s and b_s are constants determined by the boundary conditions and normalization. The condition for the existence of a symmetric energy level with energy $\mathcal{E}_s$ lying below E_2 has the form

$$\tanh(|\theta_2|) = \frac{2\nu_{22}L}{|\theta_2|} - 1. \tag{58}$$

Similarly, we can find the energy $\mathcal{E}_s$ for an odd state (this level exists only for certain parameters of the system). It is equally easy to solve the scattering problem for the field of two wells, i.e., to obtain wave amplitudes in the channel $n = 1$. For our subsequent analysis, we will need only the transmission and reflection amplitudes, which have the form

$$t = \frac{k_1^2}{e^{-2i\theta_1}(k_1 + i\nu_{11})^2 + e^{2i\theta_1}\nu_{11}^2},$$
$$r = -1 + \frac{t}{k_1}((k_1 + i\nu_{11})e^{-2i\theta_1} + i\nu_{11}e^{2i\theta_1}). \tag{59}$$

It can easily be seen that all the matrix elements determining the resonance scattering amplitude for the potential under investigation can be expressed in terms of *r*, *t*, and functions of bound states:

$$F_s = -\frac{\hbar^2 \nu_{12} u_s(L/2)}{m}(1 + r + t), \quad R_s = F_s, \tag{60}$$

$$U_{ss} = 2\frac{\hbar^2 \nu_{12}^2 u_s^2(L/2)}{mik_1}(1 + r + t) \tag{61}$$

(these are the explicit expressions for symmetric states only). Calculating the resonance transmission amplitude, we obtain

$$t_{11}(E) = t\left\{1 + \frac{m}{ik_1\hbar^2 t}\left(\frac{F_s^2}{E - \mathcal{E}_s - U_{ss}} - \frac{F_a^2}{E - \mathcal{E}_a - U_{aa}}\right)\right\}. \tag{62}$$

Using the explicit expressions for *r* and *t*, we separate the real and imaginary components of the matrix elements U_{ss} of a symmetric state:

$$U_{ss} = \Delta_s - i\Gamma_s,$$

$$\Delta_s = 4\frac{\hbar^2 \nu_{12}^2 u_s^2(L/2)}{m}$$

$$\times \frac{(k_1 \sin\theta_1 + 2\nu_{11}\cos\theta_1)\cos\theta_1}{(k_1 \sin\theta_1 + 2\nu_{11}\cos\theta_1)^2 + k_1^2\cos^2\theta_1}, \tag{63}$$

$$\Gamma_s = 4\frac{\hbar^2 \nu_{12}^2 u_s^2(L/2)}{m}$$

$$\times \frac{k_1 \cos^2\theta_1}{(k_1^2 \sin\theta_1 + 2\nu_{11}\cos\theta_1)^2 + k_1^2\cos^2\theta_1}.$$

Similarly, for an odd state, we have

$$U_{aa} = \Delta_a - i\Gamma_a,$$

$$\Delta_a = -4\frac{\hbar^2 \nu_{12}^2 u_a^2(L/2)}{m}$$

$$\times \frac{(k_1\cos\theta_1 - 2\nu_{11}\sin\theta_1)\sin\theta_1}{(k_1^2\cos\theta_1 - 2\nu_{11}\sin\theta_1)^2 + k_1^2\sin^2\theta_1}, \quad (64)$$

$$\Gamma_a = 4\frac{\hbar^2 \nu_{12}^2 u_a^2(L/2)}{m}$$

$$\times \frac{k_1 \sin^2\theta_1}{(k_1^2\cos\theta_1 - 2\nu_{11}\sin\theta_1)^2 + k_1^2\sin^2\theta_1}.$$

Consequently, the pole part of the scattering amplitude can be written in the form

$$\mathcal{D}(E) = (E - (\mathcal{E}_s + \Delta_s) + i\Gamma_s) \times (E - (\mathcal{E}_a + \Delta_a) + i\Gamma_a). \quad (65)$$

The expression appearing in the numerator of the transmission amplitude and determining its zeros has the form

$$\mathcal{N}(E) = (E - \mathcal{E}_s)(E - \mathcal{E}_a) - 4\frac{\hbar^2 \nu_{12}^2}{mk_1} \times \left((E - \mathcal{E}_a)u_s^2\left(\frac{L}{2}\right) - (E - \mathcal{E}_s)u_a^2\left(\frac{L}{2}\right)\right)\sin\theta_1\cos\theta_1. \quad (66)$$

It is important to note that the characteristics of resonances are essentially the oscillating functions of the parameters of the system. Let us change the distance between the wells. In this case, one of the widths, say, the width Γ_s of a symmetric resonance, may vanish. This is accompanied by the collapse of the resonance. It follows from relations (63) that this takes place under the following condition:

$$\cos\theta_1 = 0. \quad (67)$$

If this condition is satisfied, the pole shifts to the real axis and its energy exactly coincides with the energy of a zero.

Thus, the necessary condition for the collapse of resonances in the system is vanishing of matrix elements F_s or F_a. Let us consider it in greater detail for F_s. We present the wave function $\chi^+(x)$ in the form of a superposition of the symmetric and antisymmetric components:

$$\chi^+(x) = \chi_s^+(x) + \chi_a^+(x). \quad (68)$$

In accordance with relations (17), the matrix element can be written in the form

$$F_s = \int_{-\infty}^{\infty} dx u_s(x) V_{21}(x) \chi_s^+(x). \quad (69)$$

In the case of pointlike potentials, the contribution to the integral appears from neighborhoods of the points $x = \pm L/2$. In an open channel, we take the wave function in the form $\chi_s^+(x) = a\cos k_1 x$. The matrix element may vanish if the condition $\chi_s^+(\pm L/2) = a\cos k_1 L/2 = 0$ is satisfied. This means that the nodes of a wave in the scattering channel exactly coincide with the positions of impurities. Obviously, conditions (58) and (67) must be satisfied simultaneously for this purpose. This may occur if the levels in the well $V_{22}(x)$ emerging below E_2 intersect with virtual energy levels in a rectangular well of width L. This follows from expression (67).

C. Interference of Resonant States

Let us now consider the case when a potential well may generate a large number of resonances and analyze their interaction. We can expect that both nontrivial interaction between energy levels and the nontrivial behavior of asymmetric resonances are possible in such a system [23]. Let us consider a 2D impurity in the channel (see Fig. 3c). The impurity potential can be written as

$$V(x, y - Y_s) = -V_{\text{att}}\vartheta\left(\frac{L}{2} - |x|\right)\vartheta\left(\frac{W_{\text{att}}}{2} - |y - Y_s|\right), \quad (70)$$

where $\vartheta(x) = 0$ for $x < 0$ and $\vartheta(x) = 1$ for $x > 1$; $X_s = 0$ and Y_s are the coordinates of the center of the well; V_{att} is the well depth; and W_{att} is its transverse size. For such a well, the matrix elements $V_{nn'}$ can be easily found in explicit form. We will be interested in the transmission of an electronic waveguide containing the given scatterer. It should be noted that, in spite of the apparent simplicity, the given problem has no exact solution. Proceeding in accordance with the algorithm proposed above, we first find solutions for energy levels in the well $V_{22}(x)$. The energy levels for symmetric and antisymmetric states can be determined by solving the transcendental equations

$$\tan\theta_2 = \frac{\kappa}{q_2}, \quad \cot\theta_2 = -\frac{\kappa}{q_2}, \quad (71)$$

where

$$q_2 = \sqrt{\frac{2m}{\hbar^2}(E - E_2 + |V_{22}|)}, \quad \theta_2 = \frac{L}{2}q_2,$$

$$\kappa = \sqrt{\frac{2m}{\hbar^2}(E_2 - E)}.$$

The number of levels in a well is determined by the parameter $v = L\sqrt{m|V_{22}|/2\hbar^2}$. We can easily derive the wave functions $u_j^s(x)$ and $u_j^a(x)$ corresponding to energy levels. For example, for even states, we have

$$u_j^s(x) = \begin{cases} A_j^s \cos(q_2 x), & |x| < L/2, \\ B_j^s e^{-\kappa|x|}, & |x| > L/2, \end{cases} \tag{72}$$

where A_j^s and B_j^s are the constants determined by the boundary conditions and normalization (the corresponding quantities for odd states are denoted by A_j^a and B_j^a). The solution of the scattering problem gives

$$t = \frac{ik_1}{(q_1 \sin\theta_1 + ik_1 \cos\theta_1)(-q_1 \cos\theta_1 + ik_1 \sin\theta_1)}, \tag{73}$$

where q_1 is the wave vector of a particle in the well,

$$q_1 = \sqrt{\frac{2m}{\hbar^2}(E - E_1 + |V_{11}|)}, \quad \theta_1 = \frac{L}{2}q_1,$$

and k_1 is the wave vector in the scattering channel 1,

$$k_1 = \sqrt{\frac{2m}{\hbar^2}(E - E_1)}.$$

It follows from expression (73) that the potential scattering amplitude has the pole structure determined by reflections at the well edges. Such poles correspond to Breit–Wigner resonances.

We can write the expressions for the matrix elements determining the resonant structure of the scattering amplitude:

$$\begin{gathered} F_j^s(E) = 2V_{12}A_j^s C f_j^s(E), \quad R_j^s(E) = F_j^s(E), \\ F_j^a(E) = 2V_{12}A_j^a iS f_j^a(E), \\ R_j^a(E) = -F_j^{(a)}(E), \end{gathered} \tag{74}$$

where

$$\begin{aligned} C &= \left(\cos\theta_1 - i\frac{k_1}{q_1}\sin\theta_1\right)t, \\ S &= \left(-i\sin\theta_1 + \frac{k_1}{q_1}\cos\theta_1\right)t, \end{aligned} \tag{75}$$

and C and S are connected through the relation

$$CS = \frac{k_1}{q_1}t. \tag{76}$$

We have also introduced the functions

$$\begin{aligned} f_j^s(E) &= \frac{1}{q_2^2 - q_1^2} \\ &\times \left(q_2 \sin\frac{q_2 L}{2}\cos\frac{q_1 L}{2} - q_1 \cos\frac{q_2 L}{2}\sin\frac{q_1 L}{2}\right), \\ f_j^a(E) &= \frac{1}{q_2^2 - q_1^2} \\ &\times \left(q_2 \sin\frac{q_2 L}{2}\cos\frac{q_1 L}{2} + q_1 \cos\frac{q_2 L}{2}\sin\frac{q_1 L}{2}\right). \end{aligned} \tag{77}$$

Let us present the matrix elements $U_{jj'}$ in the form

$$U_{jj'} = \frac{m}{ik_1\hbar^2 t}F_j R_{j'} + Q_{jj'}, \tag{78}$$

where

$$\begin{aligned} Q_{jj'} &= \frac{2mV_{12}^2}{q_1\hbar^2} \\ &\times \int_{-L/2}^{L/2} u_j(x) \int_x^{L/2} u_j'(x') \sin q(x - x'), \end{aligned} \tag{79}$$

and $\mathrm{Im}\, Q_{jj'} = 0$. In the case of a symmetric well, the matrix elements $U_{jj'}$ between the states with different parities are equal to zero. It is convenient to introduce the matrix elements $U_{ss'}$ and $U_{aa'}$ which connect the pairs of symmetric (ss') and antisymmetric (aa') states:

$$\begin{aligned} U_{ss'} &= \frac{mV_{12}^2 A_s A_{s'}}{ik_1\hbar^2 t}C^2 f_s f_{s'} + Q_{ss'}, \\ U_{aa'} &= \frac{mV_{12}^2 A_a A_{a'}}{ik_1\hbar^2 t}S^2 f_a f_{a'} + Q_{aa'}. \end{aligned} \tag{80}$$

The expression for the transmission amplitude was derived in the general form in Section 2 for a small number of resonances. If the value of parameter v is such that $v < \pi/2$, the well contains only one energy level and, hence, one Fano resonance in transmission (42). In the given interval of v, the parameters of the resonance are monotonic functions of the well width. For $\pi/2 < v < \pi$, the transmission contains two Fano resonances (45), but the parameters of the resonances

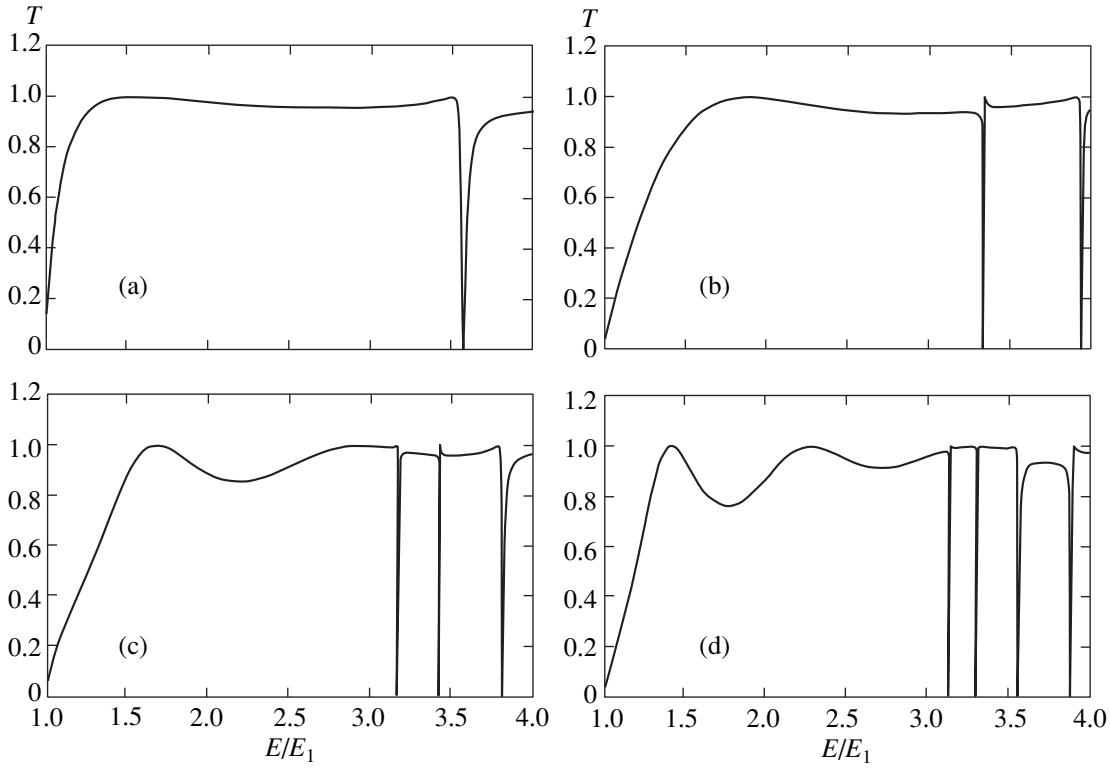

Fig. 4. Resonant structure of the transmission of a waveguide with an extended impurity for different values of L: $0.7W$, $1.3W$, $2.7W$, and $3.6W$ (the energy is measured in units of $E_1 = \pi^2\hbar^2/2mW^2$) and for the following impurity parameters: $V_{11} = 1.515E_1$, $V_{12} = 0.521E_1$, and $V_{22} = 1.0E_1$.

in this interval may oscillate since, in the given case, we have

$$\mathscr{E}_1^0 = \mathscr{E}_1 + Q_{11}, \quad \mathscr{E}_2^0 = \mathscr{E}_2 + Q_{22},$$

$$\mathscr{E}_1^R = \mathscr{E}_1 + Q_{11} + \frac{mV_{12}^2A_1^2}{k_1\hbar^2} f_1^2(E)\mathrm{Im}\left(\frac{C^2}{t}\right),$$

$$\mathscr{E}_2^R = \mathscr{E}_2 + Q_{22} + \frac{mV_{12}^2A_2^2}{k_1\hbar^2} f_2^2(E)\mathrm{Im}\left(\frac{C^2}{t}\right), \qquad (81)$$

$$\gamma_{12}^2 = \left(\frac{mV_{12}^2}{\hbar^2}\right)^2 q_1^2A_1^2A_2^2f_1^2f_2^2,$$

and the widths of the resonances are defined as

$$\Gamma_j = \frac{mV_{12}^2A_j^2}{k_1\hbar^2} f_j^2(E)\mathrm{Re}\left(\frac{C^2}{t}\right), \quad j = 1, 2. \qquad (82)$$

The number of Fano resonances is equal to three for $\pi < \nu < 3\pi/2$, four for $3\pi/2 < \nu < 2\pi$, etc.

Let us consider the results of numerical calculations in the case of a 2D impurity of rectangular shape (Fig. 3c). We fix the well depth and vary its length L. Figure 4 shows the probability of transmission through a waveguide with an extended symmetric attracting impurity as a function of energy E (E is measured in units of $E_1 = \pi^2\hbar^2/2mW^2$, W being the waveguide width) for various values of parameter L. It can be seen that for $L = 0.7W$, when potential $V_{22}(x)$ contains only one energy level, the transmission contains one Fano resonance. For $L = 1.3W$, $L = 2.7W$, and $L = 3.6W$, one can see two, three, and four Fano resonances, respectively. The above analysis shows that, by varying the impurity parameters (the parameter L in the present case), we change the number of energy levels and the number of resonances in the transmission. The zeros of the transmission amplitude always lie on the real axis, while its poles lie in the complex plane. We can also see the evolution of Breit–Wigner resonances emerging as a result of wave interference in the channel $n = 1$. For the chosen parameters, they lie far in the complex plane. When the Breit–Wigner resonances intersect Fano resonances, they interact effectively.

Let us now consider the possibility of the collapse of resonances in the given system. We will study the behavior of characteristics of resonances as functions of well parameters, e.g., its width L. An analysis of expression (82) shows that the function $\mathrm{Re}(C^2/t)$ cannot vanish. It can easily be verified that only the functions $f_j(E)$ may have zeros, which occurs when the matrix

elements F_s or F_a vanish. Let us consider in greater detail the matrix element

$$F_s = \int_{-\infty}^{\infty} dx u_s(x) V_{21}(x) \chi_s^+(x) = V_{12} A_s C \int_{-L/2}^{L/2} dx \cos q_2 x \cos q_1 x, \quad (83)$$

where we have used the expression $\chi_s^+(x) = C\cos q_1 x$ for the wave function in the well region. It follows from Eq. (83) that the integral may vanish if $\chi_s^+(x)$ changes its sign, and the positive contribution to the integral is exactly compensated by the negative contribution. This is possible if one of the equalities (71) holds for even and odd energy levels, respectively. Thus, the collapse for even states takes place if the following conditions are satisfied simultaneously:

$$\tan\theta_2 = \frac{\kappa}{q_2}, \quad \tan\theta_1 = \frac{\kappa}{q_1}, \quad (84)$$

the corresponding conditions for odd states being

$$\cot\theta_2 = -\frac{\kappa}{q_2}, \quad \cot\theta_1 = -\frac{\kappa}{q_1}. \quad (85)$$

Let us consider in greater detail the meaning of the derived expressions for even states. We will measure energy from E_2. Then, the first equality in (84) can be interpreted as the condition for the existence of energy levels in a well of depth $|V_{22}|$ and the second equality as the condition for the existence of energy levels in a well of depth $E_2 - E_1 + |V_{11}|$ (shown by dashed lines in Fig. 5). Let us now increase the size L of the well. Upon a change in parameter L, the energy levels of the deeper well descend more sharply and intersect the energy levels in the shallower well. The intersection of levels leads to the collapse of resonances. Figure 6 shows the graphical solution of equations for the two-parametric spectral problem (84) and (85). The intersection of continuous curves makes it possible to determine the critical parameters (E_i^c, L_i^c) for symmetric states (i = 1, 2, …). Similarly, the intersections of dashed curves give the critical parameters of asymmetric states of the system. For example, for the impurity parameters selected earlier, we can indicate several pairs of critical parameters for symmetric states: $(1.1562W, 3.2987E_1)$, $(2.2512W, 3.1188E_1)$, $(3.3310W, 3.0633E_1)$, etc. Figure 7 shows transmission as a function of energy for the following three values of the longitudinal impurity size L: $1.95W$, $L_2^c = 2.25116W$, and $2.45\,W$. It can be seen from the figure that, as L passes through the critical value, the matrix element of the interaction of states changes its sign, owing to which the collapse and swing (change from the "zero–pole" to the "pole–zero" positions) of Fano resonances can be observed.

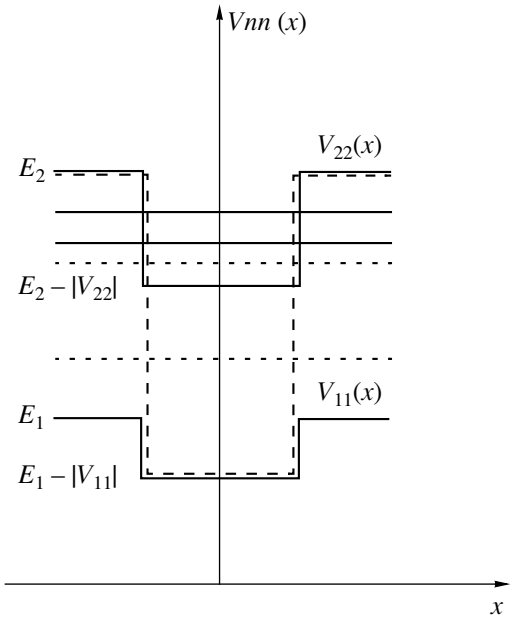

Fig. 5. Diagram of intersection of energy levels in effective potential wells $|V_{22}|$ and $E_2 - E_1 + |V_{11}|$.

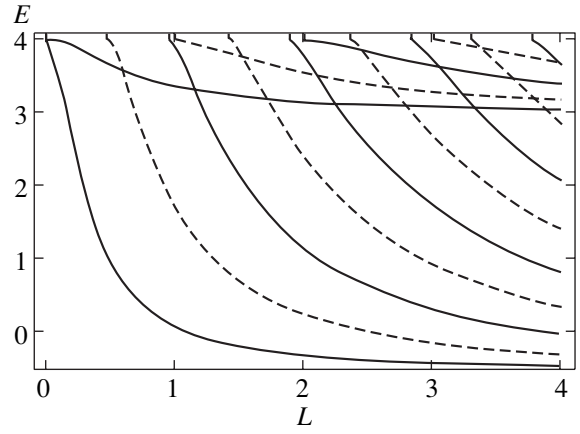

Fig. 6. Graphic solution of Eqs. (84) and (85) for the two-parametric problem. The intersection of solid curves gives critical parameters of symmetric states, while the intersection of dashed curves gives critical parameters of asymmetric states of the system.

D. Resonances in a Waveguide with Asymmetric Impurity

Let us now consider the case when the impurity potential does not possess parity relative to the transformation $x \longrightarrow -x$: $V(x, y) \neq V(-x, y)$. For the sake of def-

initeness, we consider a 2D two-component impurity (see Fig. 4d) whose potential field will be chosen in the form

$$V(x, y - Y_s) = \begin{cases} -V_a\vartheta(L_a + x)\vartheta\left(\frac{W_{\text{att}}}{2} - |y - Y_s|\right), & x < 0, \\ -V_b\vartheta(L_b - x)\vartheta\left(\frac{W_{\text{att}}}{2} - |y - Y_s|\right), & x > 0. \end{cases} \quad (86)$$

The position of energy levels in the well $V_{22}(x)$ in this case is determined by the relation between the quantities V_a and V_b in expression (86) (for the sake of definiteness, we assume that $|V_a| > |V_b|$). The energy levels can be determined from the equation

$$(1 + \beta\tan\theta_{2a})(\alpha - \tan\theta_{2b}) = (1 + \alpha\tan\theta_{2b})(\gamma\tan\theta_{2a} - \alpha), \quad (87)$$

where $\alpha = \kappa/q_{2b}$, $\beta = \kappa/q_{2a}$, $\gamma = q_{2a}/q_{2b}$, $\theta_{2a} = q_{2a}L_a$, $\theta_{2b} = q_{2b}L_b$, and $L = L_a + L_b$. The coefficient of transmission of a particle above the well $V_{11}(x)$ has the form

$$t = \frac{1}{\cos\theta_{1a}\left(\cos\theta_{1b} - \frac{i}{2}(\tilde{\beta} + \tilde{\beta}^{-1})\sin\theta_{1b}\right) - \frac{i}{2}\sin\theta_{1a}(\cos\theta_{1b}(\tilde{\alpha} + \tilde{\alpha}^{-1}) - i(\tilde{\gamma} + \tilde{\gamma}^{-1})\sin\theta_{1b})}, \quad (88)$$

where $\tilde{\alpha} = k_1/q_{1b}$, $\tilde{\beta} = k_1/q_{1a}$, $\tilde{\gamma} = q_{1a}/q_{1b}$, $\theta_{1a} = q_{1a}L_a$, $\theta_{1b} = q_{1b}L_b$, and $q_{1a,b} = \sqrt{2m(E - E_1 + |V_{11,a,b}|)}/\hbar$. It should be noted that, in the case of scattering in an asymmetric field $V_{11}(x)$, the transmission associated with potential scattering (88) also displays a series of alternating maxima and minima, but in contrast to the symmetric problem, the values of transmission amplitude at the peaks may be smaller than unity.

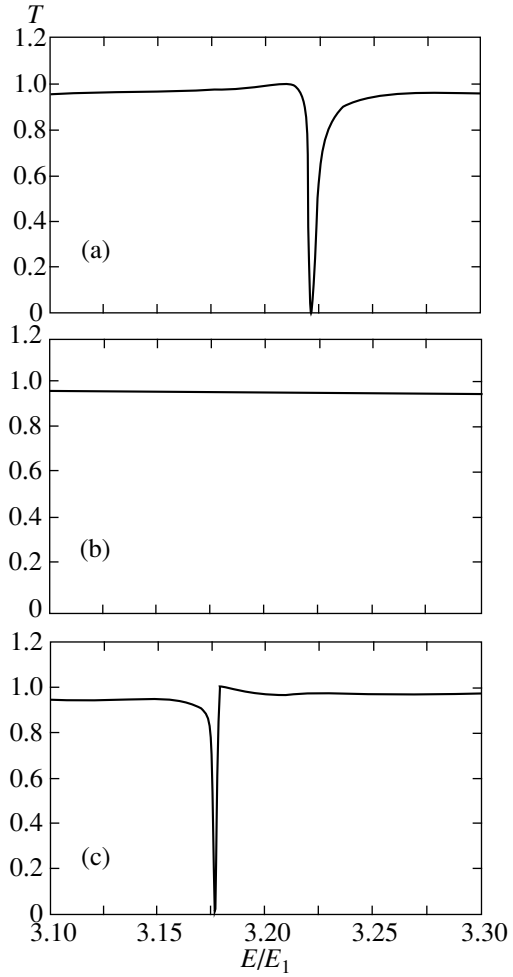

Fig. 7. Swing of a Fano resonance upon the change of zero–pole to pole–zero positions for three values of parameter L: $1.95W$, $L = 2.25116W$, and $2.45W$.

In order to determine the resonant contribution to the transmission amplitude, we must calculate the matrix elements F_j, R_j, and $U_{jj'}$ taking into account the fact that $V_{12}(x)$ is now a step function. For this reason, integration is carried out over the two regions $(-L_a, 0)$ and $(0, L_b)$. The matrix elements $U_{jj'}$ can be written, as before, in the form (33), where $Q_{jj'}$ is now calculated by using the formula

$$Q_{jj'} = -\frac{2m}{k_1\hbar^2}\int_{-L_a}^{0} dx V_{12}(x)u_j(x) \times \left(\int_x^0 dx' V_{12}(x')u_{j'}(x')\sin q_{1a}(x - x') + \int_0^{L_b} dx' V_{12}(x')u_{j'}(x')\sin q_{1b}(x - x')\right) + \frac{2m}{k_1\hbar^2}\int_0^{L_b} dx V_{12}(x)u_j(x)\int_x^{L_b} dx' V_{21}(x')u_{j'}(x')\sin q_{1b}(x - x'). \quad (89)$$

It can be seen that the statement that $\text{Im}\, Q_{jj'} = 0$ is also valid for an asymmetric well; however, in the general case, the matrix $U_{jj'}$ may contain nonzero nondiagonal elements. Let us find out how this circumstance changes the structure of resonances.

In the case when the well $V_{22}(x)$ contains one energy level, the expression for the channel transmission formally coincides with formula (42), but the resonance

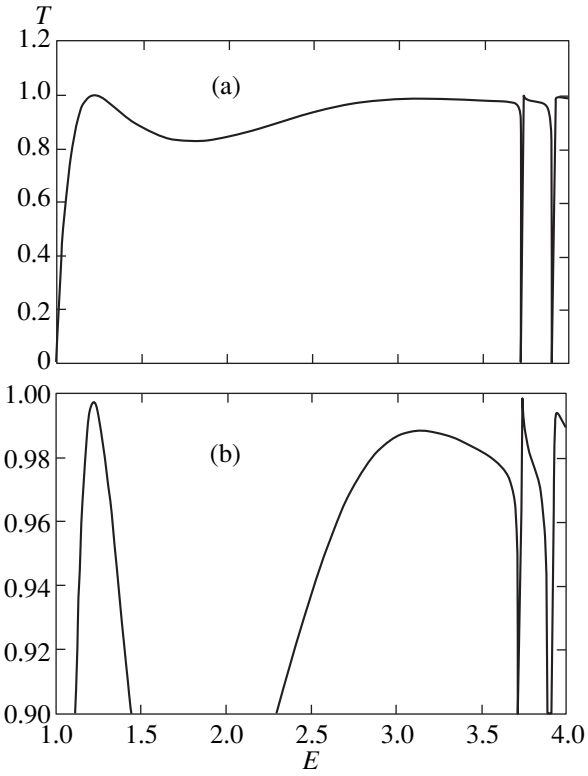

Fig. 8. Transmission of a channel with a two-component impurity (the impurity size $L = L_a + L_b = 1.7W$, $V_b/V_a = 0.5$). The remaining parameters are the same as in Fig. 4.

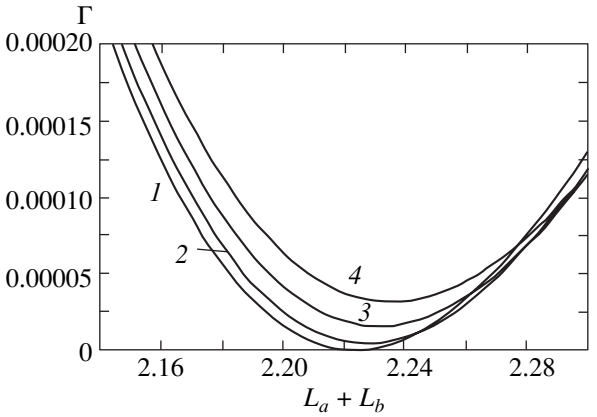

Fig. 9. Dependence of the Fano resonance width on parameter $L = L_a + L_b$ for different values of the impurity asymmetry parameter V_b/V_a: $V_b = V_a$ (*1*), $V_b = 0.95V_a$ (*2*), $V_b = 0.9V_a$ (*3*), and $V_b = 0.85V_a$ (*4*).

parameters depend on the type of well. If the well $V_{22}(x)$ contains two energy levels, we have

$$t_{11}(E) = t(E)\frac{(E-\mathscr{E}_1^0)(E-\mathscr{E}_2^0)-Q_{12}Q_{21}}{(E-\mathscr{E}_1-U_{11})(E-\mathscr{E}_2-U_{22})-U_{12}U_{12}}, \quad (90)$$

where

$$\mathscr{E}_1^0 = \mathscr{E}_1 + Q_{11}, \quad \mathscr{E}_2^0 = \mathscr{E}_2 + Q_{22}.$$

It can easily be seen that zeros of the resonance also lie on the real energy axis. It was noted above that the denominator of expression (90) can be presented in the form $\mathscr{D}(E) = (E - \tilde{\mathscr{E}}_1)(E - \tilde{\mathscr{E}}_2)$, where the positions of the poles $\tilde{\mathscr{E}}_1$ and $\tilde{\mathscr{E}}_2$ are determined by the solutions to the equation

$$(\tilde{\mathscr{E}} - \mathscr{E}_1 - U_{11})(\tilde{\mathscr{E}} - \mathscr{E}_2 - U_{22}) - U_{12}U_{12} = 0.$$

If we assume that the asymmetry parameter of the potential well $V(x, y)$ is small ($V_b/V_a \approx 1$), the position of the poles will be determined in accordance with

$$\tilde{\mathscr{E}}_1 \approx \mathscr{E}_1 + U_{11} - \frac{U_{12}^2}{\mathscr{E}_2 - \mathscr{E}_1},$$

$$\tilde{\mathscr{E}}_2 \approx \mathscr{E}_2 + U_{22} + \frac{U_{12}^2}{\mathscr{E}_2 - \mathscr{E}_1}.$$

This means that, in the present case, we are dealing with a direct interaction of Fano resonances, which is determined by the matrix element U_{12}. Figure 8 shows the transmission of a channel with a two-component impurity (the impurity size is $L = L_a + L_b = 1.7W$, $V_b/V_a = 0.5$). The potential scattering of a particle above the well $V_{11}(x)$ results in a series of maxima and minima on the dependence of T on E (Breit–Wigner resonances), while Fano resonances are connected with the impurity levels split from the E_2 band. The examples considered above show that, in the absence of symmetry, the structure of Fano resonances remains unchanged: transmission has zeros on the real axis and poles corresponding to transmission peaks. However, the peaks of Fano resonances as well as of Breit–Wigner resonances become smaller than unity due to the loss of coherence in the reflection from the edges of an asymmetric well.

If, however, the asymmetry parameter is small, resonances may become narrower, but no resonance collapse takes place in the case of an asymmetric impurity. Expressing F_1R_1 and F_2R_2 explicitly in the regions $(-L_a, 0)$ and $(0, L_b)$ as was done for a symmetric impurity, we can easily verify that the widths of Fano resonances cannot vanish exactly in the absence of parity of states. In this case, the integrands will contain complex-valued functions, and the vanishing of Fano resonances requires that the real and imaginary components of matrix elements become simultaneously equal to zero, which cannot be achieved with a single free parameter (e.g., the longitudinal size of the well). However, as in the symmetric case, a change in the impurity parameters may result in the intersection of the energy levels of

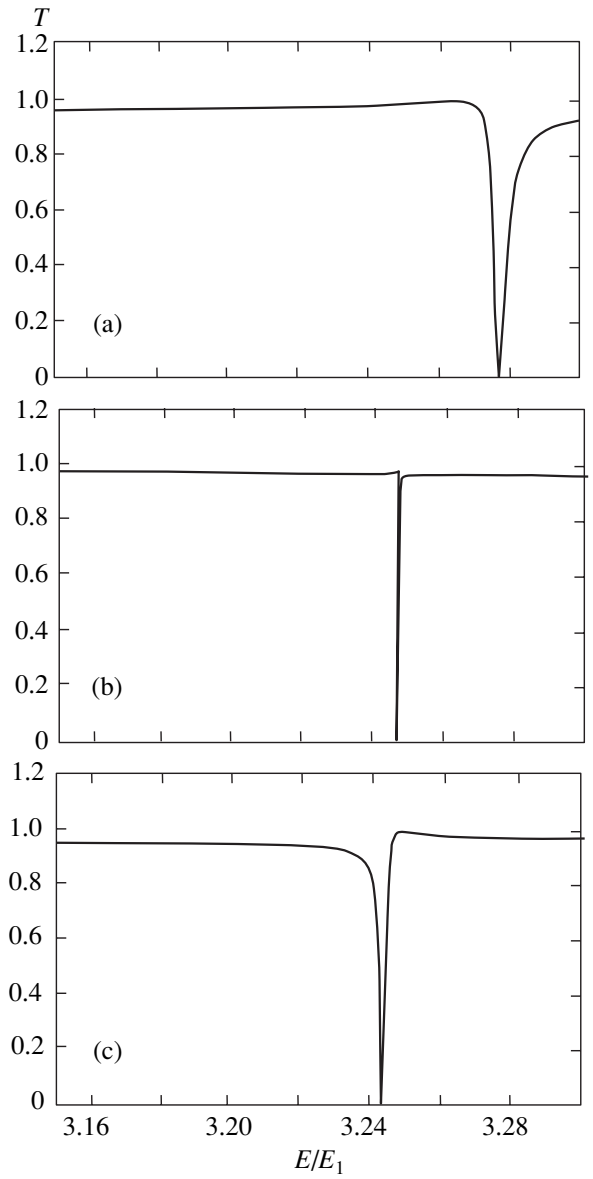

Fig. 10. Swing of a Fano resonance. The impurity asymmetry parameter $V_b/V_a = 0.85$; the parameter $L = L_a + L_b$ assumes the following values: (a) $1.95W$; (b) $2.25W$; and (c) $2.45W$.

effective wells leading to the swing of a Fano resonance. It should be noted that interacting states in the given system have no definite parity. Figure 9 demonstrates the dependences of the width of a Fano resonance on the parameter $L = L_a + L_b$ for various values of the impurity asymmetry parameter. Figure 10 shows the swing of a Fano resonance, when parameter L passes through a value close to the critical value for a symmetric impurity (the asymmetry parameter $V_b/V_a = 0.85$). It can be seen from Fig. 10 that the pole and the zero of an isolated Fano resonance change places upon a change in parameter L, the width of the resonance remaining finite.

5. CONCLUSIONS

Thus, the proposed approach makes it possible to analyze the resonance structure of the amplitude of transmission of a particle through an electronic waveguide containing attracting impurities of an arbitrary shape. It was proved that finite-size impurities create a series of quasi-bound states which are manifested as resonance–antiresonance pairs in transmission. The results presented above demonstrate explicitly that, in all cases considered above, Fano resonances are due to interference of a propagating electron state with quasi-bound states. Fano resonances possess a universal structure and are characterized by three real parameters ($\mathcal{E}^R$, $\mathcal{E}^0$, Γ). Upon a change in the impurity parameters, resonance zeros and poles move in accordance with the change in the position of energy levels in the well. In the case of a symmetric impurity, a pole and a zero may collide, which leads to the collapse and swing of a Fano resonance. The theory developed in this work makes it possible to study the interaction of asymmetric resonances in wells of an arbitrary shape. The predicted effects may be discovered in electronic waveguides with artificial impurities [15].

It should also be noted that the proposed approach makes it possible to consider problems of resonant transformation of waves of any origin, e.g., electromagnetic or acoustic waves.

ACKNOWLEDGMENTS

The authors are grateful to V.A. Volkov, A.A. Gorbatsevich, D.Z. Kwon, and Yu.A. Romanov for their interest in this research and for fruitful discussions.

This work was supported financially by the Russian Foundation for Basic Research (project nos. 01-02-16569 and 02-02-17495). One of the authors (C.S. Kim) is indebted to the Korea Research Foundation for support.

REFERENCES

1. H. Feshbach, Ann. Phys. **5**, 357 (1958).
2. H. Feshbach, Ann. Phys. **19**, 287 (1962).
3. *Quantum Transport in Ultrasmall Devices*, Ed. by D. K. Ferry, H. L. Grubin, C. Jacoboni, and A.-P. Jauho (Plenum, New York, 1995), NATO ASI Ser., Ser. B: Phys., Vol. 342.
4. B. J. van Wees, H. van Houten, C. W. J. Beenakker, *et al.*, Phys. Rev. Lett. **60**, 848 (1988).
5. D. A. Wharam, T. J. Thorton, R. Newbury, *et al.*, J. Phys. C **21**, L209 (1988).
6. U. Fano, Phys. Rev. **124**, 1866 (1961).
7. U. Fano and J. W. Cooper, Phys. Rev. A **137**, 1364 (1965).
8. R. K. Adair, C. K. Bockelman, and R. E. Peterson, Phys. Rev. **76**, 308 (1949).
9. J. A. Simpson and U. Fano, Phys. Rev. Lett. **11**, 158 (1963).

10. F. Cardeira, T. A. Fjeldly, and M. Cardona, Phys. Rev. B **8**, 4734 (1973).
11. J. Feist, F. Capasso, C. Sirtori, *et al.*, Nature (London) **390**, 589 (1997).
12. H. Schmidt, K. L. Campman, A. C. Gossard, and A. Imamoglu, Appl. Phys. Lett. **70**, 3455 (1997).
13. C. P. Holfeld, F. Löser, M. Sudzius, *et al.*, Phys. Rev. Lett. **81**, 874 (1998).
14. O. A. Tkachenko, V. A. Tkachenko, D. G. Baksheev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 366 (2000) [JETP Lett. **71**, 255 (2000)].
15. S. Yamada and M. Yamamoto, J. Appl. Phys. **79**, 8391 (1996).
16. F. H. Mies, Phys. Rev. **175**, 164 (1968).
17. C. S. Chu and R. S. Sorbello, Phys. Rev. B **40**, 5941 (1989).
18. P. F. Bagwell, Phys. Rev. B **41**, 10354 (1990).
19. E. Tekman and S. Ciraci, Phys. Rev. B **42**, 9098 (1990).
20. S. A. Gurvitz and Y. B. Levinson, Phys. Rev. B **47**, 10578 (1993).
21. J. U. Nöckel and A. D. Stone, Phys. Rev. B **50**, 17415 (1994).
22. C. S. Kim and A. M. Satanin, Zh. Éksp. Teor. Fiz. **115**, 211 (1999) [JETP **88**, 118 (1999)].
23. C. S. Kim, A. M. Satanin, Y. S. Joe, and R. M. Cosby, Phys. Rev. B **60**, 10962 (1999).
24. R. Landauer, Phil. Mag. **21**, 863 (1970).
25. M. Buttiker, Phys. Rev. B **35**, 4123 (1987).
26. A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions and Decays in Nonrelativistic Quantum Mechanics* (Nauka, Moscow, 1971, 2nd ed.; Israel Program for Scientific Translations, Jerusalem, 1966).

Translated by N. Wadhwa