Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 761–769.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 883–892.
Original Russian Text Copyright © 2005 by Chesnokov.

ATOMS, MOLECULES, OPTICS

Dependence of the Differential Cross Sections for Radiative–Collisional Excitation of Metastable Atomic States on the Polarization of Radiation

E. A. Chesnokov

Institute of Physics, St. Petersburg State University, Peterhof, St. Petersburg, 198904 Russia
e-mail: tchesn@ec8174.spb.edu

Received June 6, 2005

Abstract—We calculate the total and differential cross sections for radiative–collisional excitation of the metastable 2^1S state of He atoms at collisions with Ne atoms in external radiation fields of various frequencies and polarizations. The calculations are performed for a thermal collision energy of $E = 10^{-3}$ atomic units and light intensity of $I = 1$ MW cm^{-2}, which corresponds to a single photon absorption by a quasi-molecule during the collision. Both the differential and total cross sections are shown to depend strongly on the relative orientation of the radiation polarization vector and the initial relative velocity vector of the colliding atoms. We analyze the azimuthal scattering asymmetry related to the orientation of the angular momentum of the absorbed photon. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In recent years, considerable interest has been shown in analyzing the angular distributions of the atoms produced during experiments with crossed atomic beams in a laser radiation field. In particular, the excitation reaction of Na atoms at collisions with rare-gas atoms in a quasi-resonant laser radiation field was used as an example to show that studying the change in the structure of the differential cross sections with radiation frequency, polarization, and the collision energy allows one to judge the geometry of radiative–collisional processes [1, 2], to analyze the nonadiabatic quasi-molecular transitions [3], and to reproduce the interaction potentials of atoms in the ground and excited states with a fairly high accuracy [4, 5].

At the same time, the differential cross sections for the radiative–collisional processes that lead to the collisionally induced radiative excitation of metastable states in one of the colliding atoms remain essentially unexplored. These processes are difficult to observe experimentally, because the dipole moment of the transition induced by the interaction between atoms and, accordingly, the radiative–collisional excitation cross sections are small. Nevertheless, a number of works on experimental observation of the spectral distributions formed during the processes of this kind can be noted [6–9]. Attempts to give an adequate theoretical description also run into a number of difficulties. In particular, the widely used close coupling method (see, e.g., [2]) proves to be inefficient when there are no good initial data on the atomic interaction potentials and the transition dipole moment. At the same time, determining the latter is a complex quantum-chemical problem, especially for the interaction between two compact atoms in the ^1S states, since it requires simultaneously allowance for the motions of all electrons in the quasi-molecule. Another method commonly used to analyze optical collisions is based on the Condon approximation. However, as was shown in [10], the Condon approximation is valid only for a smoothly changing interaction between quasi-molecular states. In contrast, for an interaction that depends strongly on the internuclear distance R, which is peculiar to collisionally induced radiative transitions (with a zero dipole moment at large R), the validity range of the Condon approximation is very limited.

In [11], based on a model for the interaction between two Coulomb terms that depends exponentially on R [12], which leads to a closed analytical expression for the scattering amplitude, the authors estimated the differential cross sections for the reaction of the collisionally induced radiation,

$$\mathrm{He}(1^1S) + \mathrm{Ne} + \hbar\omega \longleftrightarrow \mathrm{He}(2^1S) + \mathrm{Ne}. \qquad (1)$$

Although the results obtained correctly reproduce some of the general patterns of behavior of the differential cross sections, using the Coulomb fits for the ground- and excited-state potentials leads to a definitely incorrect quantitative estimate of the cross sections and makes it impossible to reproduce a number of features in the differential cross sections related to the short-range nature of the interaction between ground-state atoms. In addition, the calculation in [11] was performed in the approximation of identical centrifugal potentials in both reaction channels. Basically, this

implies the neglect of the photon angular momentum and rules out the possibility of studying the polarization-related effects.

The goal of this work is to calculate and analyze the structure of the differential and total cross sections for reaction (1) with absorption as a function of the polarization. The calculation is based on fairly accurate [10] exponential fits for the radiative width and interaction potentials of atoms in the ground and excited $^1\Sigma$ states and suggests a proper allowance for the photon angular momentum.

The mechanism of reaction (1) was described in detail in [10]. The exponential fits for the potentials of the ground $|^1\Sigma, 1^1S\rangle$ and excited $|^1\Sigma, 2^1S\rangle$ quasi-molecular states in the energy range 10–1000 cm^{-1} based on data from [13, 14] are

$$U_g(R) = 2594\exp(-3.439R),$$
$$U_e(R) = 0.404\exp(-0.917R).$$

The radiative width of the excited $|^1\Sigma, 2^1S\rangle$ quasi-molecular state [15] emerges as the atoms approach each other due to the mixing of the wavefunctions for the metastable 2^1S and resonance 2^1P He atomic states. The exponential fit for the width is $\Gamma(R) = 4.84 \times 10^{-5}\exp(-1.84R)$. In what follows, unless stated otherwise, we use the atomic system of units.

2. THEORY OF RADIATIVE TRANSITION BETWEEN THE TWO $^1\Sigma$ TERMS

2.1. Quantum Theory

At low radiation intensities, when the radiative transition probability is much lower than unity both at each stage and during the entire collision, the amplitude and differential cross section for the radiative–collisional process can be determined using the distorted-wave method [16]:

$$f_{fi}(\mathbf{k}_f, \mathbf{k}_i, \mathbf{e}) = -\frac{\mu}{2\pi}\int \Psi^{(-)*}_{\mathbf{k}_f}(\mathbf{R})V(\mathbf{R})\Psi^{(+)}_{\mathbf{k}_i}(\mathbf{R})d^3R, \quad (2)$$

$$\frac{d\sigma}{d\Omega} = \frac{k_f}{k_i}|f_{fi}(\mathbf{k}_f, \mathbf{k}_i, \mathbf{e})|^2, \quad (3)$$

where μ is the reduced mass of the colliding atoms, $\mathbf{k}_{i,f}$ are the relative momenta of the atoms before and after the collision, and $\Psi^{(+)}_{\mathbf{k}_i}(\mathbf{R})$ and $\Psi^{(-)}_{\mathbf{k}_f}(\mathbf{R})$ are the wavefunctions for the relative motion of the atoms in the initial and final channels that satisfy the well-known asymptotics [17].

Since the dipole moment of the transition between the two $^1\Sigma$ quasi-molecular states is parallel to the internuclear axis, $\mathbf{d}_{fi} = d(R)\hat{\mathbf{R}}$, where $\hat{\mathbf{R}}$ is a unit vector along the internuclear axis, the matrix element of the dipole interaction between the quasi-molecule and an external electromagnetic field

$$\mathbf{E} = E_0\mathrm{Re}[\mathbf{e}\exp(-i\omega t)]$$

of intensity $I = cE_0^2/8\pi$ can be represented as a product of the radial and angular parts of the interaction:

$$V(\mathbf{R}) = V(\hat{\mathbf{R}})V(R), \quad (4)$$

where

$$V(\hat{\mathbf{R}}) = -\mathbf{e}\cdot\hat{\mathbf{R}}, \quad V(R) = (E_0/2)d(R).$$

Expanding the functions $\Psi^{(+)}_{\mathbf{k}_i}(\mathbf{R})$ and $\Psi^{(-)}_{\mathbf{k}_f}(\mathbf{R})$ in terms of partial waves allows the scattering amplitude to be expressed in terms of elements of the S-matrix:

$$f(\theta, \varphi, \mathbf{e}) = \mathbf{e}\cdot\mathbf{f}(\theta, \varphi), \quad \mathbf{f}(\theta, \varphi) = \sum_\nu \mathbf{e}_\nu f_\nu(\theta, \varphi), \quad (5)$$

where the angles θ and φ specify the direction of the final relative momentum vector of the colliding atoms, $\mathbf{k}_f$ (the z axis is assumed to be along the initial relative momentum vector $\mathbf{k}_i$). The spherical components of the scattering amplitudes are defined by the formulas

$$f_0(\theta) = (4k_ik_f)^{-1/2} \times \sum_{l=0}^{\infty}[lS_{l-1\to l} - (l+1)S_{l+1\to l}]P_l(\cos\theta),$$
$$f_{\pm1}(\theta, \varphi) = \mp(1/\sqrt{2})\exp(\pm i\varphi)f_{(1)}(\theta), \quad (6)$$
$$f_{(1)}(\theta) = (4k_ik_f)^{-1/2} \times \sum_{l=1}^{\infty}[S_{l-1\to l} + S_{l+1\to l}]P^1_l(\cos\theta)$$

and are related to the Cartesian components by

$$f_z(\theta) = f_0(\theta), \quad f_x(\theta, \varphi) = \cos\varphi f_{(1)}(\theta), \quad f_y(\theta, \varphi) = \sin\varphi f_{(1)}(\theta). \quad (7)$$

The S-matrix elements are

$$S_{l_i\to l_f} = -i2\pi\exp[i(\delta^{l_i}_i + \delta^{l_f}_f)] \times \int_0^{\infty}\Psi^{l_f}_{E_f}(R)V(R)\Psi^{l_i}_{E_i}(R)dR. \quad (8)$$

Here, $\Psi^l_E(R)$ are the radial, regular (at zero) real wavefunctions normalized to the δ function of energy, and δ^l are the elastic scattering phase shifts.

Expressions (6), along with the Legendre polynomials $P_l(\cos\theta)$, contain the associated Legendre functions

$P_l^1(\cos\theta)$ that emerge when calculating the matrix element of the angular part of the dipole interaction $V(\hat{\mathbf{R}})$ between the spherical functions that specify the initial and final rotational quasi-molecular states. Photon absorption causes a change in the relative angular momentum of the atoms, $l \pm 1 \longrightarrow l$, which corresponds the *P* and *R* branches in the theory of molecular radiation.

For an arbitrary linear polarization,

$$\mathbf{e} = \mathbf{e}_z\cos\theta_p + \mathbf{e}_x\sin\theta_p\cos\varphi_p + \mathbf{e}_y\sin\theta_p\sin\varphi_p, \quad (9)$$

the scattering amplitude is

$$\begin{aligned} f(\theta, \varphi, \mathbf{e}) &= \cos\theta_p f_0(\theta) \\ &+ \sin\theta_p\cos(\varphi - \varphi_p) f_{(1)}(\theta). \end{aligned} \quad (10)$$

In this case, the differential cross section is given by the formula

$$\begin{aligned} \frac{d\sigma(\mathbf{e})}{d\Omega} &= \frac{k_f}{k_i}\{\cos^2\theta_p |f_0(\theta)|^2 \\ &+ \sin^2\theta_p\cos^2(\varphi - \varphi_p)|f_{(1)}(\theta)|^2 \\ &+ \sin(2\theta_p)\cos(\varphi - \varphi_p)\mathrm{Re}[f_0(\theta) f_{(1)}^*(\theta)]\}. \end{aligned} \quad (11)$$

Integrating (11) over the solid angle Ω yields an expression for the total cross section:

$$\sigma(\mathbf{e}) = \sigma_\parallel\cos^2\theta_p + \sigma_\perp\sin^2\theta_p, \quad (12)$$

where

$$\begin{aligned} \sigma_\parallel &= 2\pi\frac{k_f}{k_i}\int |f_0(\theta)|^2\sin\theta d\theta, \\ \sigma_\perp &= \pi\frac{k_f}{k_i}\int |f_{(1)}(\theta)|^2\sin\theta d\theta \end{aligned} \quad (13)$$

are the total cross sections for polarizations parallel and perpendicular to the initial relative momentum, respectively. The factor 1/2 appears in the expression for $\sigma_\perp$ when averaged over the azimuthal angle φ. Having integrated the squares of the magnitudes of the amplitudes, $|f_0(\theta)|^2$ and $|f_{(1)}(\theta)|^2$, over the scattering angle θ and taking into account the orthogonality of the Legendre polynomials, we obtain

$$\begin{aligned} \sigma_\parallel &= \frac{\pi}{k_i^2}\sum_{l=0}^{\infty}\frac{1}{2l+1}|lS_{l-1\to l} - (l+1)S_{l+1\to l}|^2, \\ \sigma_\perp &= \frac{\pi}{2k_i^2}\sum_{l=1}^{\infty}\frac{l(l+1)}{2l+1}|S_{l-1\to l} + S_{l+1\to l}|^2. \end{aligned} \quad (14)$$

Averaging the total cross section (12) over the radiation polarizations or over the directions of the initial relative velocity of the colliding atoms under gas-cell conditions yields

$$\begin{aligned} \sigma &= \frac{\sigma_\parallel + 2\sigma_\perp}{3} \\ &= \frac{\pi}{3k_i^2}\sum_{l=0}^{\infty}[(l+1)|S_{l\to l+1}|^2 + l|S_{l\to l-1}|^2], \end{aligned} \quad (15)$$

where the factor 2 of $\sigma_\perp$ corresponds to the two possible perpendicular polarizations.

When the polarization vector **e** lies in the collision plane, we assume the *x* axis to lie in the same half-plane as the polarization vector ($\varphi_p = 0$). In this case, for an observer whose head is oriented along the *y* axis and whose face is directed along the *z* axis, $\varphi = 0$ and π correspond to the leftward and rightward scatterings, respectively. The expressions for the scattering amplitude and differential cross section take the form

$$f(\theta, \theta_p) = \cos\theta_p f_0(\theta) \pm \sin\theta_p f_{(1)}(\theta), \quad (16)$$

$$\begin{aligned} \frac{d\sigma(\theta_p)}{d\Omega} &= \frac{k_f}{k_i}(\cos^2\theta_p |f_0(\theta)|^2 + \sin^2\theta_p |f_{(1)}(\theta)|^2 \\ &\pm \sin(2\theta_p)\mathrm{Re}[f_0(\theta) f_{(1)}^*(\theta)]), \end{aligned} \quad (17)$$

where the upper and lower signs correspond to the leftward and rightward scatterings, respectively.

In the case of a right-hand or left-hand circular polarization in the collision plane,

$$\mathbf{e}_{r,l} = \mp\frac{\mathbf{e}_z \pm i\mathbf{e}_x}{\sqrt{2}}, \quad (18)$$

the scattering amplitude and differential cross section are

$$f_{r,l}(\theta) = \mp\frac{f_0(\theta) \pm i f_{(1)}(\theta)}{\sqrt{2}}, \quad (19)$$

$$\begin{aligned} \frac{d\sigma_{r,l}}{d\Omega} &= \frac{k_f}{k_i}\left(\frac{1}{2}|f_0(\theta)|^2\right. \\ &\left. + \frac{1}{2}|f_{(1)}(\theta)|^2 \pm \mathrm{Im}[f_0(\theta) f_{(1)}^*(\theta)]\right). \end{aligned} \quad (20)$$

The upper and lower signs in Eq. (20) correspond to the leftward and rightward scatterings, respectively, for a right-hand circular polarization and vice versa for a left-hand circular polarization.

2.2. The Condon Approximation

In the Condon approximation (see, e.g., [2, 10, 11]), the radiative transitions occur when the atoms are separated by distance R_C. The radius R_C depends on the

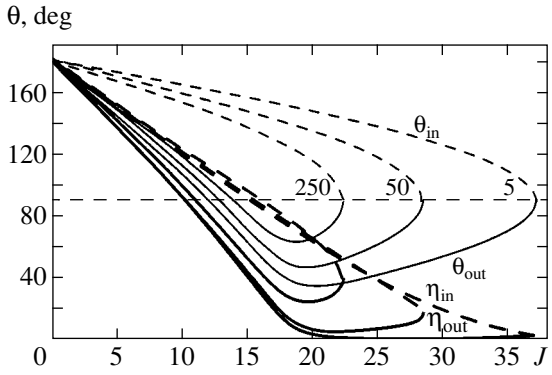


Fig. 1. The deflection functions $\eta_{in,\,out}(J)$ and the angles specifying the directions of the Condon vectors $\theta_{in,\,out}(J)$ calculated for the collision energy $E = 10^{-3}$ atomic units at various frequency detunings: $\Delta\omega = 5$, 50, and 250 cm^{-1}.

radiation frequency and can be determined from the equation

$$\Delta U(R_C) = \Delta\omega, \tag{21}$$

where $\Delta U = U_e - U_g$ is the difference between the excited- and ground-state potentials measured from the energy levels of isolated atoms, and $\Delta\omega$ is the detuning of the radiation frequency ω from the frequency of the forbidden atomic transition ω_0. For reaction (1) under consideration, the difference potential ΔU has a maximum at $R = 4$. In this case, Eq. (21) in the frequency range $\Delta\omega \in$ 0–300 cm^{-1} of interest defines two Condon points: one lies at $R > 6$, and the other lies at $R < 3.5$. It was shown in [10] that at collision energies $E \sim 10^{-3}$ atomic units, the contribution from the second Condon point could be ignored, and the analysis could be restricted to the radiative transitions at $R > 6$. Thus, within the framework of the semiclassical treatment, the scattering through a fixed angle θ is formed via the contribution from two paths with angular momenta $J_{1,2}$, which can be determined from the equations

$$\eta_{in,\,out}(J) = \theta, \tag{22}$$

where the deflection functions

$$\begin{aligned}\eta_{in,\,out}(J) &= \pi - (\chi_e + \chi_g)(\infty, J)\\ &\mp (\chi_e - \chi_g)(R_C, J),\end{aligned} \tag{23}$$

$$\chi(R, J) = J\int_{R_{tp}}^{R}\frac{dR}{R^2 k(R)},$$

$$k(R) = \sqrt{2\mu(E - U(R)) - \frac{J^2}{R^2}},$$

correspond to the paths for which the radiative transition takes place when the atoms approach each other (in) or separate from each other (out).

Figure 1 shows the deflection functions $\eta_{in,\,out}$ for various frequency detunings $\Delta\omega$. At $\theta > \theta_C$, where the angle θ_C corresponds to the maximum angular momentum J_C for which the transition point R_C is classically attainable, the in and out paths contribute to the scattering through θ. At $\theta_r < \theta < \theta_C$, where θ_r is the minimum of the deflection function η_{out}, both paths correspond to the transition during the separation. The range of angles $\theta < \theta_r$ is classically unattainable. Since θ_C and θ_r are close, we assume below that the semiclassical Condon approximation is suitable for analyzing the differential scattering cross sections in the range of angles $\theta > \theta_C$, where the scattering is determined by the sum of the contributions from the in and out paths.

When the linear polarization vector lies in the collision plane, the expressions for the scattering amplitude and differential cross section calculated in the semiclassical Condon approximation are [2]

$$\begin{aligned}f_{FC}(\theta, \theta_p) &= \left(\frac{k_i}{k_f}\right)^{1/2}[\sigma_{in}^{1/2}\cos(\theta_{in} - \theta_p)\exp(i\phi_{in})\\ &+ \sigma_{out}^{1/2}\cos(\theta_{out} - \theta_p)\exp(i\phi_{out})],\end{aligned} \tag{24}$$

$$\begin{aligned}\frac{d\sigma_{FC}(\theta, \theta_p)}{d\Omega} &= \sigma_{in}\cos^2(\theta_{in} - \theta_p)\\ &+ \sigma_{out}\cos^2(\theta_{out} - \theta_p)\\ &+ 2(\sigma_{in}\sigma_{out})^{1/2}\cos(\phi_{in} - \phi_{out})\\ &\times \cos(\theta_{in} - \theta_p)\cos(\theta_{out} - \theta_p).\end{aligned} \tag{25}$$

Here,

$$\sigma_{in,\,out} = \frac{P_{L1}(J_{in,\,out})J_{in,\,out}|dJ_{in,\,out}/d\theta|}{k_i^2\sin\theta}$$

are the single-path scattering cross sections obtained without the spherically asymmetric part of the interaction $V(\hat{\mathbf{R}})$,

$$P_{L1} = \frac{2\pi\mu V_C^2}{k_C\Delta F_C}$$

is the Landau probability calculated for a single passage through the nonadiabaticity region [17], $V_C = V(R_C)$ is the interaction matrix element at the term crossing point R_C,

$$\Delta F_C = (U_g' - U_e')(R_C)$$

is the difference between the forces at R_C, and $k_C = k_e(R_C) = k_g(R_C)$ is the relative momentum of the atoms

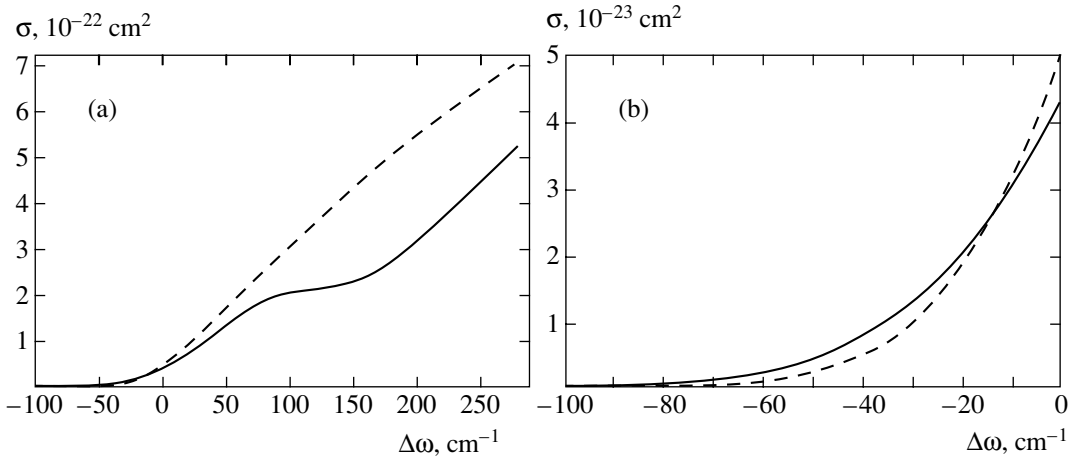


Fig. 2. The total cross sections $\sigma_{\|}$ (solid lines) and $\sigma_{\perp}$ (dashed lines) calculated for the collision energy $E = 10^{-3}$ atomic units: the blue (a) and red (b) wings of the forbidden line.

at R_C. The phases $\phi_{\text{in, out}}$ are defined by the formulas

$$\phi_{\text{in, out}} = 2\delta_{\text{in, out}}(J_{\text{in, out}}) - \theta J_{\text{in, out}} \pm \frac{\pi}{4}, \qquad (26)$$

$$2\delta_{\text{in, out}}(J) = \delta_e^J + \delta_g^J \pm (S_e - S_g)(R_C), \qquad (27)$$

where

$$S(R) = \int_{R_{tp}}^{R} k(R)dR$$

are the classical action functions, and δ^J are the quasi-classical WKB asymptotics for the scattering phase shifts δ^l. The angles $\theta_{\text{in, out}}$ specify the directions of the Condon vectors $\mathbf{R}_C^{\text{in, out}}$, i.e., the directions of the internuclear axis at the times of the radiative transitions,

$$\theta_{\text{in, out}}(J) = \pi - \chi_g(\infty, J) \pm \chi_g(R_C, J). \qquad (28)$$

Figure 1 shows the functions $\theta_{\text{in, out}}(J)$ for various radiation frequencies.

Formulas (24) and (25) correspond to the leftward scattering ($\varphi = 0$); when the scattering into the right half-plane ($\varphi = \pi$) is considered, θ_p should be substituted with $\pi - \theta_p$. The factors $\cos(\theta_{\text{in, out}} - \theta_p)$ represent the angular part of the dipole interaction operator $V(\hat{\mathbf{R}})$.

In the case of a right-hand or left-hand circular polarization in the collision plane, the differential scattering cross section is

$$\frac{d\sigma_{r,l}^{FC}}{d\Omega} = \frac{\sigma_{in}}{2} + \frac{\sigma_{\text{out}}}{2} + (\sigma_{\text{in}}\sigma_{\text{out}})^{1/2}\cos(\phi_{\text{in}} - \phi_{\text{out}} \pm (\theta_{\text{in}} - \theta_{\text{out}})). \qquad (29)$$

The total scattering cross sections for parallel and perpendicular linear polarizations are given by the formulas

$$\sigma_{\|}^{FC} = \frac{2\pi}{k_i^2}\int_0^{J_C} P_{L1}(J)(\cos^2\theta_{\text{in}} + \cos^2\theta_{\text{out}})JdJ,$$
$$\sigma_{\perp}^{FC} = \frac{\pi}{k_i^2}\int_0^{J_C} P_{L1}(J)(\sin^2\theta_{\text{in}} + \sin^2\theta_{\text{out}})JdJ. \qquad (30)$$

Adding $\sigma_{\|}^{FC}/3$ and $2\sigma_{\perp}^{FC}/3$ yields a well-known expression for the total cross section of a nonadiabatic transition (see, e.g., [18]), in which the factor 1/3 implies that, on average, only a third of the quasi-molecules have a dipole moment oriented along the polarization vector under gas-cell conditions.

3. RESULTS OF CALCULATIONS

3.1. Total Scattering Cross Sections

Calculating the differential (11) and total (14) scattering cross sections is reduced to calculating the *S*-matrix elements (8) and summing over the partial cross sections. For the collision energy $E = 10^{-3}$ atomic units under consideration, ~40 partial waves should be included in the summation (see Fig. 1). The interaction potentials of the atoms in both the ground and excited quasi-molecular states are purely repulsive, thereby ruling out the possibility of orbiting effects and the formation of quasi-bound states. For this reason, the calculation of the *S*-matrix elements can be simplified significantly by using uniform Langer approximations (see, e.g., [19]) for the radial wavefunctions of the initial and final states:

$$\Psi_E^l(R) = (2\mu)^{1/2}\left(\frac{\xi(R)}{k^2(R)}\right)^{1/4}\text{Ai}(-\xi(R)), \qquad (31)$$

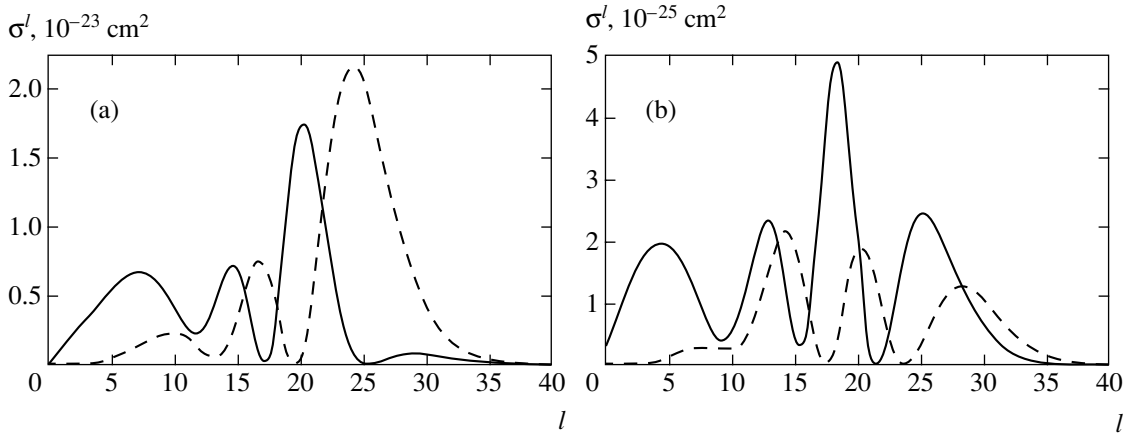


Fig. 3. Partial cross sections $\sigma_{\parallel}^{l}$ (solid lines) and $\sigma_{\perp}^{l}$ (dashed lines) versus angular momentum l calculated for the collision energy $E = 10^{-3}$ atomic units at various frequency detunings: $\Delta\omega = 50$ (a) and -50 cm^{-1} (b).

where

$$\xi(R) = \left(\frac{3}{2}S(R)\right)^{2/3},$$

Ai is the Airy function.

The total cross sections calculated using formulas (14) and (8) are shown in Fig. 2. In the blue wing of the forbidden spectral line, the cross section for a perpendicular polarization dominates over the cross section for a parallel polarization. As the frequency detuning decreases, the difference between the cross sections for parallel and perpendicular polarizations approaches zero, so that the cross section becomes polarization-independent at $\Delta\omega = -14$ cm^{-1}. As the detuning decreases further, the situation changes and the cross section for a parallel polarization begins to dominate over the cross section for a perpendicular polarization.

The relative behavior of the cross sections for parallel and perpendicular polarizations in the blue wing can be understood on the basis of the semiclassical Condon approximation (30). Since the interaction between ground-state atoms is a short-range one, to a first approximation, we can set $U_g \equiv 0$; the directions of the Condon vectors are defined by simple formulas:

$$\theta_{\text{in}} = \pi - \arcsin\frac{J}{k_i R_C}, \quad \theta_{\text{out}} = \arcsin\frac{J}{k_i R_C}. \qquad (32)$$

Substituting $\theta_{\text{in, out}}$ from (32) into Eqs. (30) and integrating over the angular momentum yields $\sigma_{\parallel}^{FC} = \sigma_{\perp}^{FC}$. Thus, in the absence of interaction between ground-state atoms, the total cross section is polarization-independent. Including the short-range repulsion in the ground-state potential causes an increase in the factor $\sin^2\theta_{\text{out}}$ and, accordingly, a decrease in $\cos^2\theta_{\text{out}}$ for small impact parameters, $b < 6$ ($J < 20$), with the cross section $\sigma_{\perp}^{FC}$ becoming larger than $\sigma_{\parallel}^{FC}$. As the frequency detuning $\Delta\omega$ decreases, the Condon radius R_C tends to infinity; the range of impact parameters that contribute to the cross section grows, so that the relative contribution from the range of small impact parameters that actually determines the difference between $\sigma_{\perp}$ and $\sigma_{\parallel}$ becomes progressively less significant.

Figure 3 shows the partial cross sections for parallel and perpendicular polarizations. The following peculiarities can be noted for all radiation frequencies. First, the cross section $\sigma_{\perp}^{l}$ becomes zero at $l = 0$. Second, the cross section $\sigma_{\perp}^{l}$ dominates over $\sigma_{\parallel}^{l}$ for large l, while the reverse is true for small l. Indeed, the radiative transitions for paths with large impact parameters occur when the internuclear axis is nearly perpendicular to the initial relative velocity vector of the colliding atoms, with the dipole interaction being at a maximum for perpendicularly polarized radiation. In contrast, for small impact parameters, the internuclear axis is nearly parallel at the times of the radiative transitions.

3.2. Differential Cross Sections

Figure 4 shows the differential cross sections $\sigma_z(\theta)$ and $\sigma_x(\theta)$ calculated for parallel and perpendicular linear polarizations in the collision plane ($\varphi = \varphi_p$). The figures demonstrate the great difference between the cross sections $\sigma_z(\theta)$ and $\sigma_x(\theta)$ in both the blue and red wings of the forbidden spectral line. A characteristic feature of the cross section $\sigma_x(\theta)$ is that it becomes zero at $\theta = 0$. Indeed, the emission of a photon with the spin projection $m_{\text{ph}} = \pm 1$ for the $^1S \longrightarrow {}^1S$ transitions causes the projection of the relative angular momentum of the atoms m to change by ± 1, which is not possible for the strict forward scattering (for $\theta = 0$, we have $m_i = m_f = 0$). In contrast, the cross section $\sigma_z(\theta)$ has a distinct maxi-

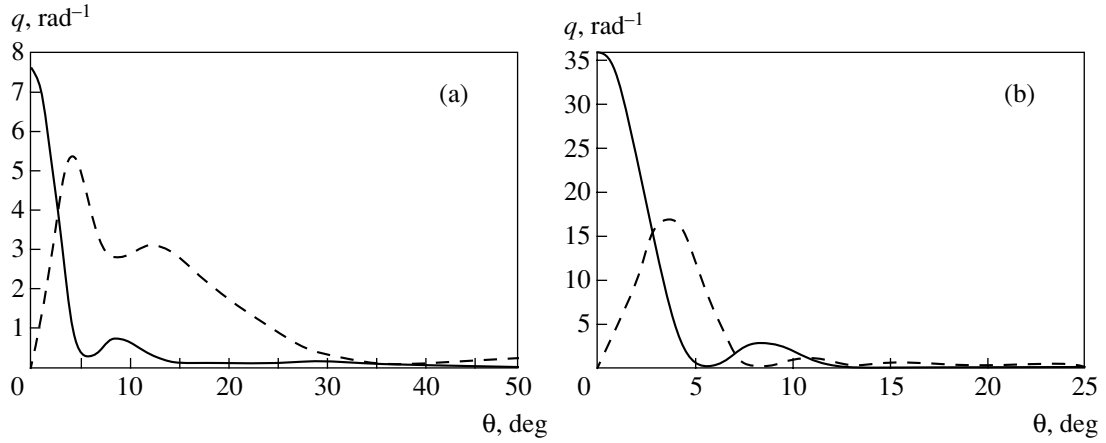


Fig. 4. The differential cross sections $q_z(\theta) = (d\sigma_z/d\Omega)/\sigma$ (solid lines) and $q_x(\theta) = (d\sigma_x/d\Omega)/\sigma$ (dashed lines) calculated for two mutually perpendicular polarizations in the collision plane. The calculation was performed for the collision energy $E = 10^{-3}$ atomic units at various frequency detunings: $\Delta\omega = 50$ (a) and -50 cm^{-1} (b). The cross sections were normalized by the condition $\frac{2\pi}{3}\int_0^\pi (q_x + q_z)\sin\theta d\theta = 1$.

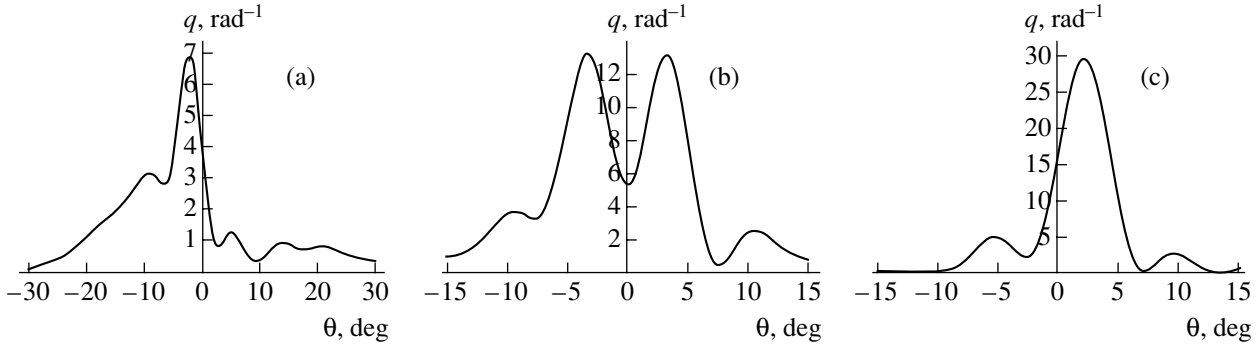


Fig. 5. The differential cross sections $q(\theta, \theta_p) = (d\sigma(\theta_p)/d\Omega)/\sigma$ calculated for a linear polarization, $\theta_p = 45°$, in the collision plane. The calculation was performed for the collision energy $E = 10^{-3}$ atomic units at various frequency detunings: $\Delta\omega = 50$ (a), 0 (b), and -50 cm^{-1} (c). The negative and positive values of θ correspond to the leftward and rightward scatterings, respectively. The cross sections were normalized by the condition $\frac{2\pi}{3}\int_{-\pi}^{\pi} q\sin|\theta| d\theta = 1$.

mum at $\theta = 0$, which is similar in nature to the glory effect. At small frequency detunings $\Delta\omega$, when the rainbow angle θ_r is close to zero (see Fig. 1), almost all of the particles that have passed to an excited state during their separation are scattered through small angles, with the bulk of the cross section σ_z being concentrated near the central maximum.

Figures 5 and 6 demonstrate the left–right scattering asymmetry in the plane formed by the vectors $\mathbf{k}_i$ and $\mathbf{e}$ for linear and circular polarizations. Figure 5 shows the differential cross sections calculated using formula (17) for the polarization angle $\theta_p = 45°$. The left–right scattering asymmetry is attributable to the asymmetric location of the polarization vector in the scattering plane. Thus, for example, in the range of parameters where the Condon approximation is admissible, the scattering into both the left and right half-planes is defined by Eq. (25). However, as was noted above, when the scattering into the right half-plane is considered, the polarization angle θ_p should be substituted with $\pi - \theta_p$. An interesting feature of the differential cross section is the disappearance of the scattering asymmetry for small θ at $\Delta\omega \approx 0$ (Fig. 5b). As more accurate calculations show, the asymmetry disappears for $E = 220$ cm$^{-1} = 10^{-3}$ au at $\Delta\omega = 0.3$ cm^{-1} and for $E = 100$ cm^{-1} at $\Delta\omega = -2.3$ cm^{-1}. The above feature is attributable to the interference between the particle beams scattered into different half-planes and cannot be explained in terms of the classical ideas of scattering. Nevertheless, a qualitative interpretation based on the Condon approximation can be offered.

As follows from Eq. (17), the left–right scattering asymmetry disappears when the real part of the product

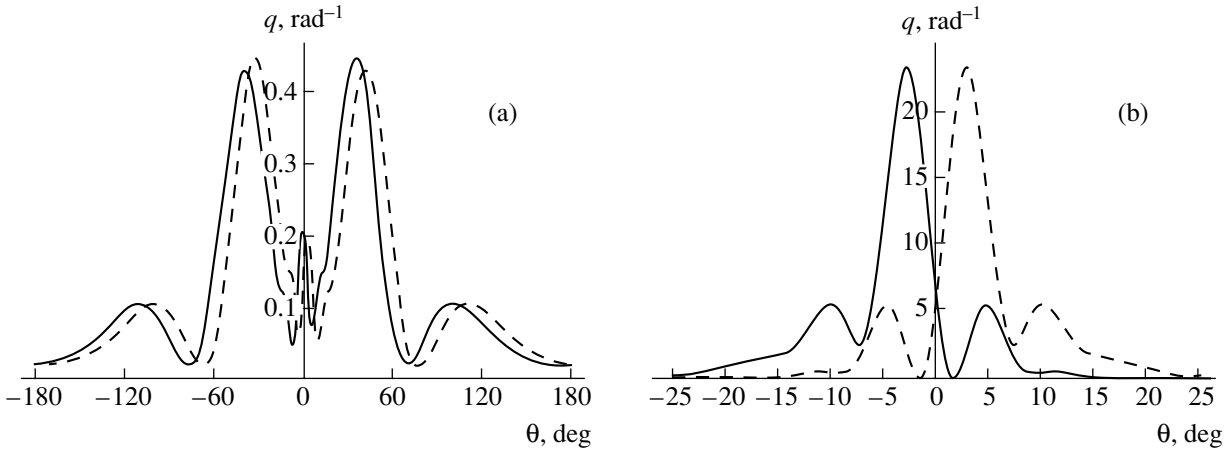


Fig. 6. The differential cross sections $q_{r,l}(\theta) = (d\sigma_{r,l}/d\Omega)/\sigma$ calculated for right-hand (solid lines) and left-hand (dashed lines) circular polarizations in the collision plane. The calculation was performed for the collision energy $E = 10^{-3}$ atomic units at various frequency detunings: $\Delta\omega = 250$ (a) and 0 cm^{-1} (b). The negative and positive values of θ correspond to the leftward and rightward scatterings, respectively. The cross sections were normalized by the condition $\frac{2\pi}{3}\int_{-\pi}^{\pi} q_{r,l}\sin|\theta|d\theta = 1$.

$f_0(\theta) f_1^*(\theta)$ becomes zero. Let us substitute the Legendre functions in Eqs. (6) with their asymptotic expressions at small θ (see, e.g., [20]). To the first order in θ, we have

$$P_l(\cos\theta) \approx 1, \quad P_l^1(\cos\theta) \approx l(l+1)\frac{\theta}{2}.$$

Given also that the small-angle scattering is determined by large l (in our case, $l > 20$; see Fig. 1) and making the standard (for large l) substitution of integration for summation, we obtain

$$f_0(\theta) = (4k_ik_f)^{-1/2}\int_0^\infty (S_{l-1\to l} - S_{l+1\to l})JdJ,$$

$$f_{(1)}(\theta) = (4k_ik_f)^{-1/2}\frac{\theta}{2} \tag{33}$$

$$\times\int_0^\infty (S_{l-1\to l} + S_{l+1\to l})J^2dJ.$$

Using the Condon approximation for the S matrix elements (see, e.g., [11]) and given that

$$2\delta_{\text{in, out}}(J \pm 1 \longrightarrow J)$$

$$= 2\delta_{\text{in, out}}(J \longrightarrow J) \pm \left(\theta_{\text{in, out}}(J) - \frac{\pi}{2}\right)$$

and that the small-angle scattering is formed mainly by the atoms that have passed to an excited state during their separation, we obtain

$$f_0(\theta) = -i\exp\frac{i\pi}{4}(k_ik_f)^{-1/2}$$

$$\times\int_0^{J_C} P_{L1}^{1/2}\exp(i2\delta_{\text{out}})\cos\theta_{\text{out}}JdJ, \tag{34}$$

$$f_{(1)}(\theta) = -\exp\frac{i\pi}{4}(k_ik_f)^{-1/2}\frac{\theta}{2}$$

$$\times\int_0^{J_C} P_{L1}^{1/2}\exp(i2\delta_{\text{out}})\sin\theta_{\text{out}}J^2dJ.$$

As we see from Fig. 1, the deflection function η_{out} for $\Delta\omega \longrightarrow 0$ is almost zero at $J > 20$, so that the scattering phase

$$2\delta_{\text{out}} = \int\eta_{\text{out}}dJ$$

is close to a constant over the entire integration range significant for small θ. Factoring $\exp(i2\delta_{\text{out}})$ outside the integral signs, we find that the complex phases for the scattering amplitudes f_0 and $f_{(1)}$ differ by π/2, with $\text{Re}[f_0 f_{(1)}^*] = 0$. Thus, we conclude that the disappearance of the left–right scattering asymmetry at $\Delta\omega \approx 0$ has the same origin as the spike in the differential cross section $\sigma_z(\theta)$ at $\theta = 0$ and is directly related to the short-range nature of the interaction between ground-state atoms.

Figure 6 shows the differential cross sections calculated using formula (20) for left-hand and right-hand

circular polarizations in the scattering plane. The cross sections demonstrate a displacement of the oscillatory structure to the left for a right-hand circular polarization and to the right for a left-hand circular polarization, in agreement with the results of [2]. Using Eq. (29), we can obtain a simple semiclassical estimate for the displacement $\Delta\theta$ between the oscillatory structures of the cross sections that correspond to right-hand and left-hand circular polarizations:

$$\Delta\theta = 2\frac{\theta_{in} - \theta_{out}}{J_{in} - J_{out}}. \quad (35)$$

As comparative calculations show, this estimate allows the displacement between the maxima of the differential cross sections for right-hand and left-hand circular polarizations to be determined to within 5% for $\theta > \theta_C$.

4. CONCLUSIONS

Using the first order of the distorted-wave method in combination with uniform Langer approximations for the radial wavefunctions of the initial and final quasimolecular states, we were able to calculate the differential and total cross sections for radiative–collisional excitation of a metastable state in one of the colliding atoms at various radiation polarizations over a wide range of radiation frequencies, including both wings and the line center of the forbidden atomic transition. The suggested approach imposes certain constraints on the ranges of admissible collision energies and radiation intensities. Our estimate of the interaction matrix element shows that the upper limit for the radiation intensities admissible for the distorted-wave method is $\sim 10^{11}$ W cm^{-2} at collision energies of $E \sim 10^{-3}$ atomic units. For its part, using the Langer approximations for the wavefunctions excludes the effects related to orbiting and resonant scattering at quasi-discrete levels from the analysis. The latter sets a lower limit for the admissible collision energies that is comparable to the depth of the potential wells attributable to the forces of polarization attraction between atoms; it is ~20 cm^{-1} for reaction (1).

Invoking the Condon approximation made it possible to qualitatively interpret a number of peculiarities of the total and differential cross sections in the blue wing of the forbidden spectral line. At the same time, this approximation does not allow the structure of the cross sections in the red wing to be interpreted correctly. This interpretation may prove to be possible on the basis of a uniform quasi-classical approximation [10–12, 21] generalized to the case of a spherically asymmetric interaction between atoms.

In conclusion, we would like to point out the topicality of the world's first experiment on studying the angular distributions of the atoms produced during reaction (1) or a similar reaction that leads to the excitation of a metastable state in one of the colliding atoms.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education of Russia (project no. PD02-1.2-377) and the administration of St. Petersburg (project no. PD03-1.2-193).

REFERENCES

1. J. Grosser, O. Hoffmann, C. Rakete, and F. Rebentrost, J. Phys. Chem. A **101**, 7627 (1997).
2. F. Rebentrost, S. Klose, and J. Grosser, Eur. Phys. J. D **1**, 277 (1998).
3. J. Grosser, O. Hoffmann, W. F. Schulze, and F. Rebentrost, J. Chem. Phys. **111**, 2853 (1999).
4. J. Grosser, O. Hoffmann, S. Klose, and F. Rebentrost, Europhys. Lett. **39**, 147 (1997).
5. J. Grosser, O. Hoffmann, and F. Rebentrost, J. Phys. B **33**, L577 (2000).
6. P. D. Kleiber and K. M. Sando, Phys. Rev. A **35**, 3715 (1987).
7. P. D. Kleiber, A. K. Fletcher, and K. M. Sando, Phys. Rev. A **37**, 3584 (1988).
8. T. Kurosawa, K. Ohmori, H. Chiba, *et al.*, J. Chem. Phys. **108**, 8101 (1998).
9. E. Bichoutskaya, A. Devdariani, K. Ohmori, *et al.*, J. Phys. B **34**, 2301 (2001).
10. A. Z. Devdariani, A. L. Zagrebin, F. Rebentrost, *et al.*, Zh. Éksp. Teor. Fiz. **122**, 481 (2002) [JETP **95**, 413 (2002)].
11. A. Z. Devdariani and E. A. Chesnokov, in *Laser Studies in the St. Petersburg State University* (NII Rossiĭskiĭ Tsentr Lazernoĭ Fiziki, St. Petersburg, 2001), p. 145 [in Russian].
12. A. Z. Devdariani and E. A. Chesnokov, Khim. Fiz. **22**, 109 (2003).
13. M. Keil, L. J. Danielson, U. Buck, *et al.*, J. Chem. Phys. **89**, 2866 (1988).
14. H. Haberland, W. Konz, and P. Oesterlin, J. Phys. B **15**, 2969 (1982).
15. A. L. Zagrebin and S. I. Tserkovnyĭ, Opt. Spektrosk. **79**, 556 (1995) [Opt. Spectrosc. **79**, 511 (1995)].
16. N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions*, 3rd ed. (Clarendon, Oxford, 1965; Inostrannaya Literatura, Moscow, 1951).
17. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 4th ed. (Nauka, Moscow, 1989; Oxford Univ. Press, Oxford, 1980).
18. E. E. Nikitin and S. Ya. Umanskiĭ, *Nonadiabatic Transitions in Slow Atomic Collisions* (Atomizdat, Moscow, 1979), p. 177 [in Russian].
19. J. N. L. Connor, J. Chem. Phys. **74**, 1047 (1981).
20. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun, 2nd ed. (Dover, New York, 1971; Nauka, Moscow, 1979).
21. A. Z. Devdariani, Zh. Éksp. Teor. Fiz. **96**, 472 (1989) [Sov. Phys. JETP **69**, 266 (1989)].

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 770–778.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 893–903.
Original Russian Text Copyright © 2005 by Blagov, Koval'chuk, Kohn, Lider, Pisarevskiĭ.

ATOMS, MOLECULES, OPTICS

Possibilities of Controlling an X-ray Beam with a Crystal Subjected to Long-Wave Ultrasonic Vibrations

A. E. Blagov[a,*], M. V. Koval'chuk[a,b], V. G. Kohn[b], V. V. Lider[a], and Yu. V. Pisarevskiĭ[a]

[a] *Shubnikov Institute of Crystallography, Russian Academy of Sciences, Leninskii pr. 59, Moscow, 119333 Russia*
**e-mail: aopt@ns.crys.ras.ru*
[b] *Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia*

Received June 18, 2005

Abstract—X-ray diffraction is experimentally studied in the Laue geometry in a germanium crystal carrying a long-wave ultrasonic wave that creates an alternating lattice deformation along the sample surface. Stroboscopic equipment is used to separate different phases and, correspondingly, different profiles of a spatial deformation distribution from the periodic deformation. A uniform deformation is shown to change the angular position of the X-ray beam, and a nonuniform deformation broadens the angular region of reflection and decreases the peak intensity. Ultrasound can be used to compensate for the static deformation at the place where the single-crystal sample and the resonator are glued together. Apart from the fundamental long-wave harmonic, the crystal contains a parasitic deformation with a shorter wavelength. A simple theoretical model is developed, and it rather accurately describes the experimental results. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

X-ray diffraction in crystals subjected to elastic deformation is now an actively developing field in solid state physics.

In many works dealing with X-ray-acoustic interactions, researchers attempted to obtain information on the structure of elastic vibrations using X-ray diffraction. This information is important for physical acoustics; moreover, it is widely used to determine the parameters and quality of various piezoelectric and acoustoelectric devices. Some works deal with studying the fundamental features of X-ray-acoustic interaction, such as an X-ray-acoustic resonance [1] resulting in the suppression of the Borrmann effect with ultrasound of a certain frequency.

The possibility of controlling the parameters of an X-ray beam with ultrasound is also of interest; however, the physical foundations of this control depend substantially on the elastic-strain wavelength. We can distinguish two main groups of interaction for different relations between the elastic-strain wavelength Λ and the crystal surface region illuminated by an X-ray beam D.

(1) High frequencies, where $D \gg \Lambda$. In this range, an elastic wave forms a superlattice with a period equal to the ultrasound wavelength. This superlattice results in the formation of additional X-ray diffraction maxima—satellites. This range is studied in most works dealing with X-ray acoustics (e.g., see [2–6] and references therein).

(2) Low and medium frequencies, where $\Lambda \gg D$. This range features an aperiodic (uniform or gradient) lattice deformation across the section (aperture) of the X-ray beam, and this range is poorly understood.

At present, works on controlling the X-ray beam amplitude with ultrasound are being advanced. A large set of works dealt with the modulation of an X-ray beam by ultrasound [2–5]. In essence, the authors of these works propose electronic analogs of a mechanical interrupter—a chopper [6]. Work [7] on modulation by long-wave ultrasound is also of interest.

Ultrasound can also be used to change the angular position of the diffracted X-ray beam. In the case of low frequencies, this control can be reached by a uniform change in the lattice parameter across the X-ray beam aperture due to the mechanical deformation of an ultrasonic wave in the crystal. Long-wave ultrasound can also create a gradient elastic deformation in the crystal, which affects the structure of the X-ray beam.

Controlling the spatial position and structure of the X-ray beam with long-wave ultrasound has only been studied in [8, 9]. The authors of these works showed that fresh opportunities for developing controlled X-ray optics appear in the case of low-frequency bending ultrasonic vibrations for Bragg diffraction.

In this work, we theoretically and experimentally study the effect of long-wave ultrasonic vibrations on the characteristics of an X-ray beam under conditions of dynamical diffraction. We think that our results can serve as a basis for long-wave ultrasonic vibrations to be applied for controlling the angular position and spatial structure of the X-ray beam.

By analyzing the use of bending vibrations in the Bragg geometry, we revealed substantial difficulties in

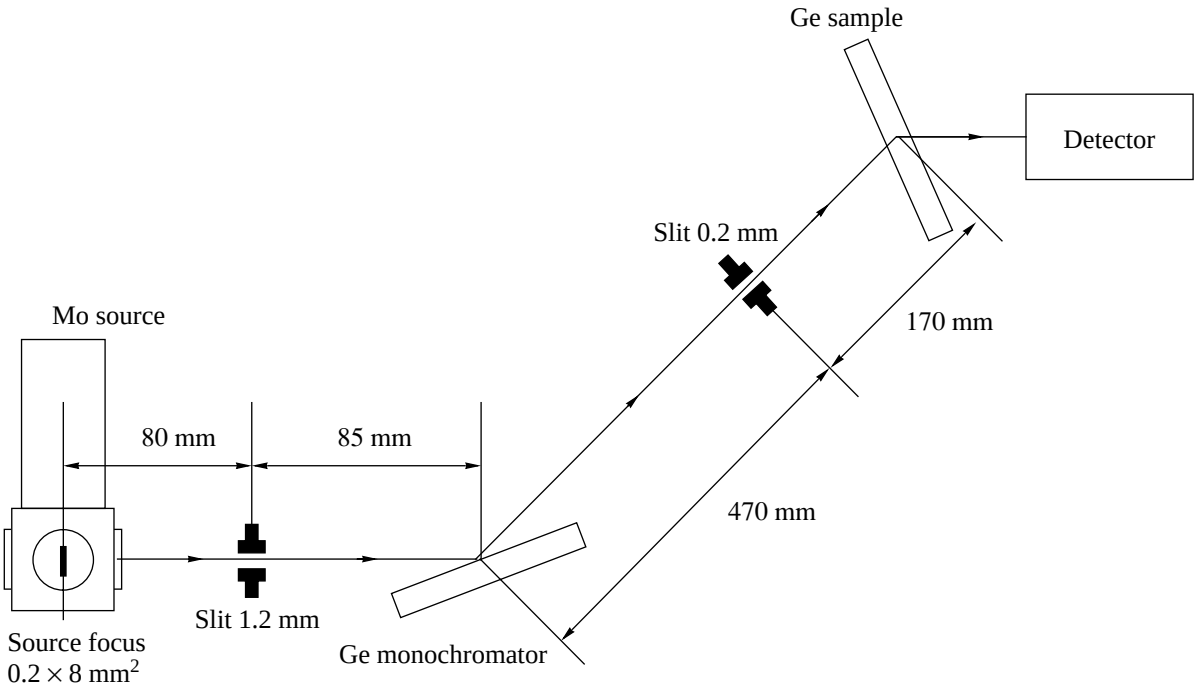


Fig. 1. X-ray optical schematic diagram of the experimental setup.

practical realization of this scheme. Therefore, we designed and actualized a scheme using long-wave ultrasonic vibrations for Laue diffraction. In this work, we present the results of the first stage, where we experimentally confirm and theoretically ground the possibility of controlling the spatial characteristics of the X-ray beam using diffraction in a crystal subjected to long-wave ultrasonic vibrations. In the next section, we describe a designed experimental scheme. In Section 3, we present the time-integrated and stroboscopic experimental results. In Section 4, we develop a theoretical model to analyze the results obtained and show that this model can adequately describe the structure of the appearing deformation from the angular dependence of the intensity of the diffracted X-ray beam.

2. EXPERIMENTAL SCHEME

The experimental setup was based on a TRS-1 X-ray spectrometer [10]. The X-ray optical scheme of the experiment is shown in Fig. 1. We used a double-crystal dispersion-free X-ray diffraction scheme, Mo$K_{\alpha 1}$ radiation, and a 0.2×8-mm^2 radiation-source focus. The angle of the main goniometer was set to an accuracy of $0.1''$. After a crystal monochromator, the collimated beam passes through a 0.2-mm slit in the diffraction plane and falls on the crystal to be studied, where it is subjected to periodic ultrasonic vibrations. The intensity of the diffracted X-ray beam is measured by a BDS scintillation detector.

As the monochromator and samples, we used single-crystal [110] and [$1\bar{1}1$] germanium plates, respectively. In both cases, we generate the symmetrical (220) reflection: the monochromator was in the Bragg diffraction position, whereas the sample was in the Laue diffraction position.

2.1. Scheme for the Excitation of Ultrasonic Vibrations

We used the resonance vibrations of an elastic longitudinal wave along the sample in the Laue diffraction geometry. When vibrations are excited in the crystal, a standing wave with the spatial deformation-amplitude distribution shown in Fig. 2a forms in the ideal case (pure mode). The deformation distribution along the crystal is seen to have the shape of a half-sinusoid with nodes at the sample ends and an antinode at its center. At the center of the crystal, the deformation distribution is quasi-uniform; at its periphery, a near-linear deformation gradient appears just as a deformation is created in statically bent gradient X-ray monochromators. The difference from this static case consists in a periodic change in the deformation in time.

The samples were $19.5 \times 10 \times 0.4$ mm^3 in size, and their working surface was elongated in the [110] axis. They were part of a composite resonator consisting of a sample and a piezoelectric crystal resonator glued together (Fig. 3). An alternating electromagnetic signal was applied to the lateral faces of the crystal resonator to create longitudinal elastic vibrations along the crystal. To this end, a conducting platinum or nickel layer was deposited on the lateral surfaces by cathode sputtering.

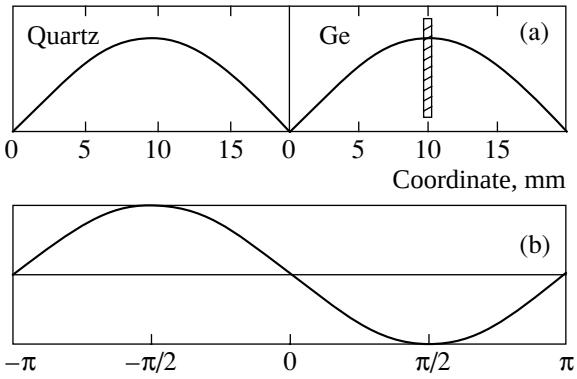


Fig. 2. (a) Estimated spatial distribution of the deformation amplitude in the resonator and sample and (b) the time variation of the deformation. The hatched region demonstrates the area illuminated by the beam of the vibrating crystal.

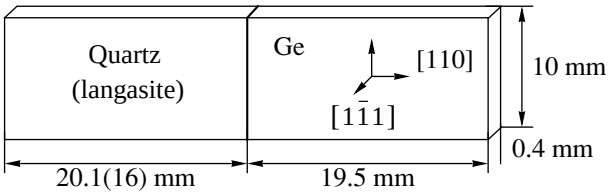


Fig. 3. Schematic diagram of the composite resonator.

The piezoelectric resonators were made of crystalline quartz with the (XYtwl-18.5°/0°/0°) cut or langasite with the (XYtwl-0°/0°/0°) cut. Such resonators are the best to provide longitudinal tension–compression vibrations almost without parasitic excitation of other vibration types. For effective ultrasound excitation, the resonance frequencies of the sample and the exciting piezoelectric plate must be coincident. In this case, when an alternating electromagnetic signal with a frequency equal to the resonance frequency is applied, high-Q vibrations appear in the crystal–piezoelectric element system. Then, an elastic half-wavelength with a deformation maximum at the center of the crystal is along the length of each plate (Fig. 2a). If parasitic vibrations are not excited, the deformation amplitude is a simple sinusoidal function in space and time (see Fig. 2b). The interface contains a deformation node; therefore, we can retain the high-Q state of the resonance system and rather simply generate high vibration amplitudes.

The resonance frequency of the germanium plate was 126 kHz. The sizes of the piezoelectric elements were chosen so that the resonance frequencies were equal, since vibration excitation was most effective in this case. To generate a pure vibration mode, the width of the piezoelectric element was several times smaller than its length. Using this experimental scheme, we can measure a rocking curve averaged over the ultrasound period, i.e., the angular dependence of the intensity of the diffracted X-ray beam for a small rotation of the sample about the X-ray beam.

2.2. Reflected-Beam Detecting Unit

The unit for detecting the intensity of the diffracted X-ray beam consisted of a scintillation detector of X-ray quanta, an amplifier with a discriminator, and a stroboscopic system (Fig. 4). This scheme allowed us to count X-ray quanta reflected by the crystal under study both continuously and periodically. The counting system processed and converted signals from the detector, and the stroboscopic system differentiated them with respect to time (i.e., it presented them as a function of the crystal vibration phase).

In the stroboscopic experiments, a synchronizing signal from the sinusoidal-signal generator that excited the crystal resonator was applied to a pulse generator, whose main purpose was to form a pulse displaced in phase with respect to the synchronizing signal. The pulse generator generated pulses with the crystal-resonator vibration frequency and a controlled phase shift with respect to crystal-resonator vibrations (the vibration phase was set to an accuracy of 10%). According to each incoming pulse, the coincidence circuit allowed the counting system to count the X-ray quanta reflected by the sample. The counting time (the time of recording signals from the detector) was set in the coincidence circuit and was one-tenth of the resonator vibration period. Using this scheme, we could detect the diffracted X-ray beam only at a given resonator vibration phase.

3. EXPERIMENTAL RESULTS

We experimentally studied X-ray-acoustic interaction during the excitation of long-wave ultrasound with and without stroboscopic analysis. As noted above, in the absence of excited parasitic vibrations, the deformation amplitude should be a simple sinusoidal function in space and time. If the crystal length is much larger than the X-ray beam width, we can change the distribution (gradient) of the ultrasonic-deformation amplitude within the X-ray beam width by moving the crystal with respect to the X-ray beam.

3.1. Time-Integrated Measurements

Information on an actual deformation distribution along the crystal can be extracted from the evolution of rocking curve measured when the position of the X-ray beam is scanned from the free end of the sample to the place of gluing with the piezoelectric transducer without stroboscopy. The sample thickness meets the conditions of the Borrmann effect with an absorption factor $\mu t \approx 12$ (where μ is the linear absorption factor and t is the sample thickness). The samples, i.e., the germanium plates, have a high quality: the FWHM is close to the theoretical value and does not exceed 6″ along the

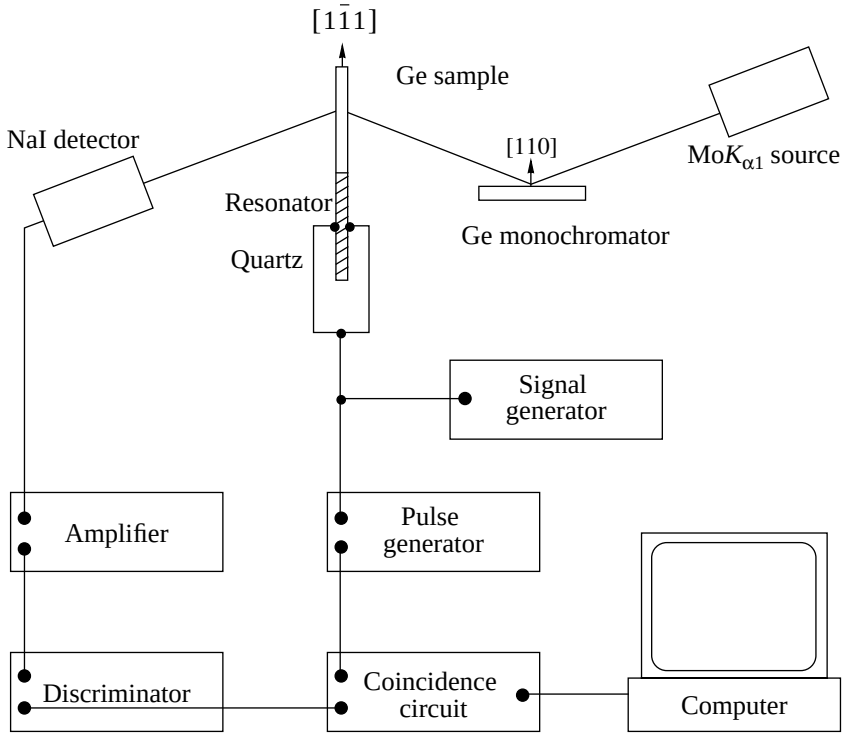


Fig. 4. General experimental scheme for studying the effect of an ultrasonic wave on X-ray diffraction in the crystal.

whole plate (the angle measurement error was less than half a percent, and the intensity measurement error was less than one percent). The only exception is a region about 3 mm wide next to the place of gluing with the piezoelectric transducer, since it excites ultrasonic deformation. In this region, the rocking curve differs from a Gaussian shape and has a strongly broadened asymmetric shape caused by a nonuniform static mechanical deformation in the germanium crystal. When ultrasound is turned on, the rocking curve width increases and the integrated intensity remains the same. The rocking curve is found to change significantly depending on the X-ray beam position on the crystal. We measured the dependence of the FWHM on the X-ray beam coordinate on the sample (Fig. 5). As follows from Fig. 5, short-wave modulation is superimposed on the pure deformation mode, when the half-wavelength of excited ultrasound is along the sample length. This finding demonstrates the excitation of additional parasitic vibrations and allows us to make preliminary conclusions regarding a real deformation distribution in the sample.

As follows from Fig. 5, the wavelength of the parasitic harmonic is 2.5 mm. This distance is much larger than the X-ray beam width on the crystal (0.2 mm). Note that this modulation substantially increases the deformation gradient in some areas of the crystal across the X-ray beam width. On the other hand, using this modulation, we can create a situation where the acoustic deformation compensates for the static deformation and where the total deformation becomes virtually uniform in a certain portion of the sample.

When ultrasound is turned on, the rocking curves in regions close to deformation maxima are strongly broadened (by almost an order of magnitude at the ultrasound amplitudes used) (Fig. 6). In regions with a strong linear ultrasound-deformation gradient, the rocking curve shape is asymmetric.

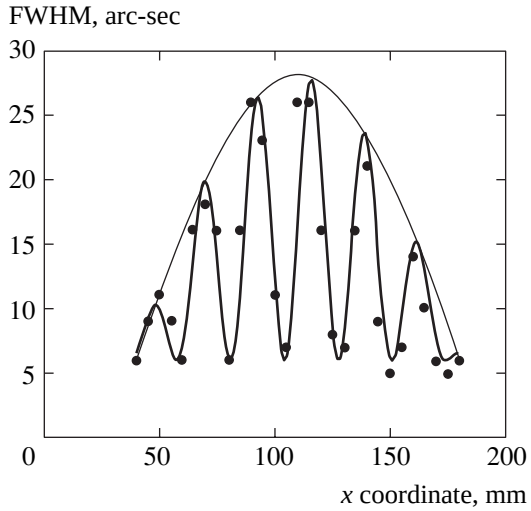


Fig. 5. Dependence of the FWHM on the position of an illuminated site on the crystal and the assumed deformation distribution in the case of a pure deformation mode.

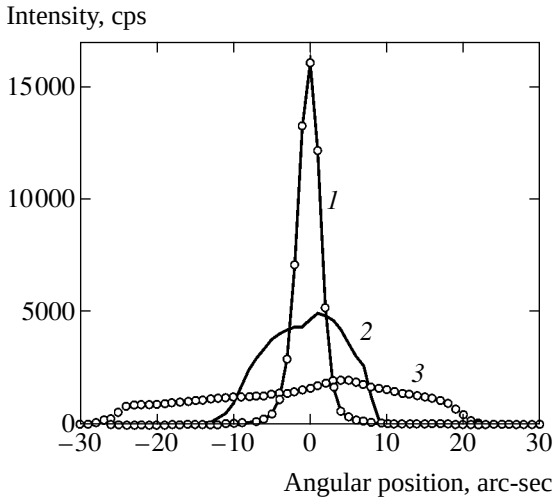


Fig. 6. Rocking curve for Ge(220) at different ultrasound powers. The ultrasound intensity is indicated in percent of the maximum power: (*1*) without ultrasound, (*2*) 40% power, and (*3*) 100% power.

We detected a rather interesting effect in a statically stressed region (near the place of gluing with the piezoelectric transducer). In this region, the rocking curve has a triangular shape with a large half-width when ultrasound is turned off. If ultrasound is turned on, the rocking curve changes its shape from an asymmetric triangular with a half-width of 50″ to an almost Gaussian shape with a half-width of 15″ (Fig. 7).

3.2. Stroboscopic Measurements

We performed time-resolved (stroboscopic) measurements in several crystal regions with characteristic deformation distributions. The first is a central region with a quasi-uniform deformation distribution across the beam aperture (region 1). Two other chosen regions have a gradient deformation distribution. One of them is near the free end (region 2), where the FWHM without ultrasound is close to the theoretical half-width. The other region (region 3) is chosen near the place of gluing to study the effect of interaction between static and dynamic deformations. In each region, we measured a series of rocking curves as a function of the resonator vibration phase. The equipment used allowed us to perform stroboscopic measurements with a given delay.

Region 1. The rocking curves measured when an X-ray beam passes through the center of the germanium plate at various resonator vibration phases are slightly broadened (Fig. 8). The rocking curves measured at phases $\varphi = -\pi/2$ and $\pi/2$ are shifted –20″ and +20″ with respect to the rocking curve recorded at $\varphi = 0$. These shifts can explain the broadening of the rocking curve recorded without stroboscopy.

In the germanium crystal, the Bragg condition (more specifically, the Bragg angle) changes because of a periodic change in the lattice parameter, and the center of rocking curve shifts with respect to the normal (zero) position. In this case, the maximum shift at the center of rocking curve is 40″; the corresponding change in the lattice parameter is 0.0023 Å; and the relative change is 0.11%.

Region 2. We also carried out measurements at three characteristic values of the vibration phase in the region near the free end of the crystal, where a linear deformation gradient is assumed to occur. The results are shown in Fig. 9. The rocking curve corresponding to a zero vibration phase is the slightly broadened rocking curve of the germanium crystal in the absence of ultrasound vibrations. The curves corresponding to vibration phases $\varphi = -\pi/2$ and $\pi/2$ were recorded at the

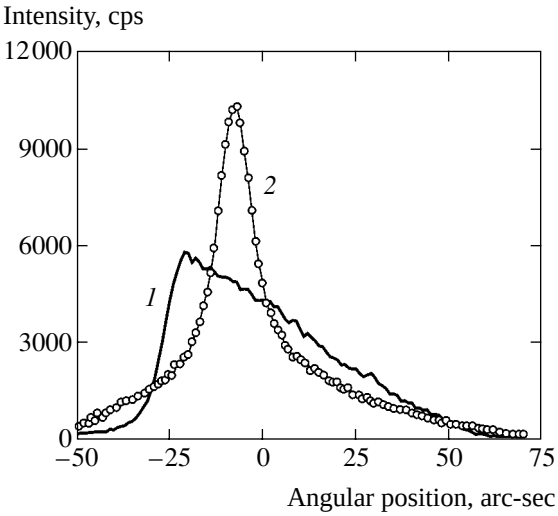


Fig. 7. Effect of ultrasound on the rocking curve shape of the Ge crystal at the place of gluing with the crystal resonator: (*1*) without and (*2*) with ultrasound.

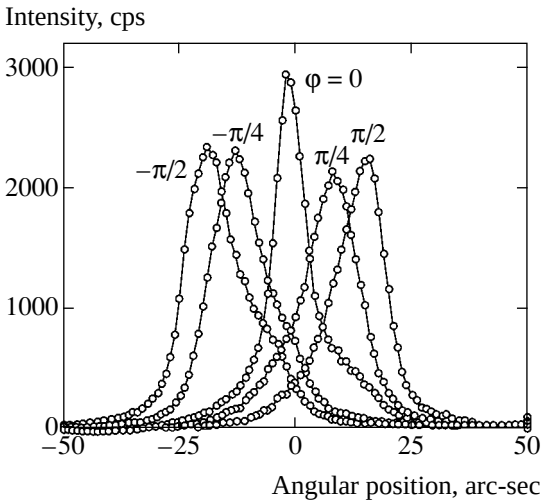


Fig. 8. Rocking curves measured at different resonator vibration phases. Points stand for experimental data.

instant of maximum deformation. They have a complex asymmetric shape that is close to a distorted asymmetric triangle. The FWHM is several times that of the ideal germanium crystal. An analysis of the rocking curve suggests that, within the X-ray beam width, the deformation is nonuniform and the gradient changes its sign as the sign of the generator vibration phase changes.

Region 3. Figure 10 shows rocking curves for the same vibration phases but in the region near the boundary with the piezoelectric element, where the rocking curve of the unexcited crystal was strongly and asymmetrically broadened because of gluing-induced stresses. In this region, as in the previous case, the ultrasound-induced deformation gradient is almost linear. The curve measured at $\varphi = 0$, as in the previous cases, agrees well with the curve measured without stroboscopy. It has a double-humped shape with a small dip at its center and maxima spaced approximately 100″ apart. The rocking curve measured at $\varphi = -\pi/2$ has a similar shape, but the intermaximum distance increases approximately to 140″. The amplitude of the higher maximum decreases, and the amplitude of the lower maximum increases by a factor of 1.5. The most interesting effect is observed at a vibration phase $\varphi = \pi/2$, where the amplitude of the higher maximum increases by about an order of magnitude, and the lower maximum almost disappears. The intermaximum distance decreases to 60″. An analysis of these curves suggests that, at different signs of deformation gradients, the ultrasound-induced dynamic deformation partly compensates for the static deformation and that the dynamic deformation enhances the static deformation if their signs coincide.

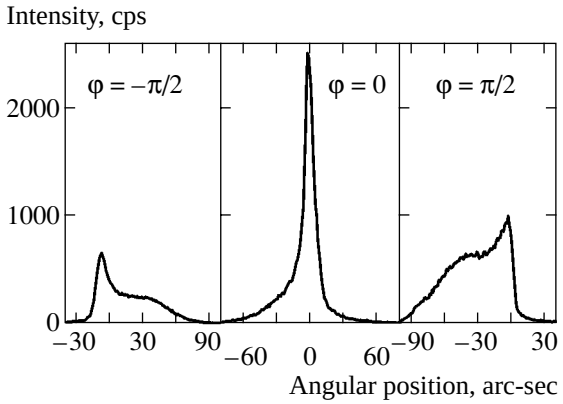


Fig. 9. Rocking curve measured for Ge in the Laue geometry at three characteristic vibration phases in the crystal region where the ultrasound-induced deformation gradient is almost linear.

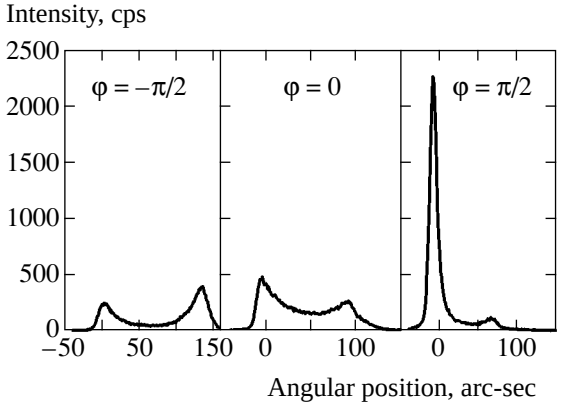


Fig. 10. Rocking curve measured for Ge at three characteristic vibration phases in the statically stressed crystal region.

4. THEORY

In our experiment, the deformation of the crystal substantially breaks the space uniformity in the direction normal to the X-ray beam propagation direction; therefore, the plane wave method, which is usually applied to calculate the angular dependences of the X-ray intensity in the case of diffraction in ideal crystals, is invalid in this case. It should be replaced by the general scheme developed for the calculation of X-ray topograms or phase-contrast images (see [11, 12]). In this scheme, a coherent radiation component is separated in the first stage; it is the monochromatic component of the spherical wave emitted by individual atoms in the anode of an X-ray tube or by orbital electrons in the case of synchrotron radiation. The propagation of this wave along a preferred trajectory (optical axis) is described by Kirchhoff's equations in space and by the Takagi equations in crystals. At the detector, the electric field (amplitude) of this wave and the local intensity (the amplitude modulus squared) depend on the coordinates, and this dependence can be measured with a photographic film or a position-sensitive detector. If the detector counts all photons, the recorded dimensionless intensity (the number of photons per measurement time) is an integral of many parameters with respect to the coherent intensity. The differential intensity should be integrated with respect to all wavelengths, to the source size with allowance for the transverse position of each atom, and to the detector window size. Moreover, we can use the Huygens–Fresnel principle and introduce a propagator for not only space but also for a crystal to describe the diffraction of a point source by the crystal surface.

We have actualized this calculation scheme, and we will use it in our subsequent works. In this work, deformation in the sample changes very slowly over transverse distances comparable with the region of diffraction scattering of a point source located on the sample surface (the so-called Borrmann delta). Moreover, due to sufficiently strong absorption, the transverse size of the diffraction region becomes even narrower. Therefore, by making allowance for the relatively large sizes of the source and the slit that limits the beam in front of the crystal, we can use the so-called "ray" approximation (not to be confused with geometrical optics used

for coherent radiation) to ensure a reasonable accuracy. In this approximation, independent trajectories of a coherent X-ray beam in space are considered for each wavelength and each transverse coordinate (point) in the source; when this beam enters the crystal, it is diffracted by a local region in the crystal as a plane wave.

The result of diffraction of such a local coherent X-ray beam that meets a crystal monochromator and a sample is a function $R(\Delta\theta(s))$ that describes the rocking curve. This function is dimensionless and presents the convolution of the reflection curves of the monochromator and sample with allowance for the local misfit of the lattice parameters and the angular positions of these two crystals. It is important that this function is independent of the incident-radiation frequency for a dispersion-free scheme. More specifically, the argument is the difference in the Bragg angles in the sample and monochromator, which depends on the coordinate of the incident beam on the sample, since the Bragg angle depends on the coordinate along the plate surface due to a deformation in the sample. In this approach, all the coherent properties of the radiation enter in this function, and various trajectories of the "rays" that correspond to different source points, different frequencies, and different detector coordinates are assumed to be incoherent.

For simplicity, instead of a coordinate x on the crystal, we use a coordinate s on the slit located in front of the crystal. These coordinates are related by a ray trajectory and are at the opposite ends of the trajectory from the slit to the crystal. Since the slit is usually located normal to the optical axis, we have

$$s = C_1 x \cos\theta_B,$$

where C_1 is the coefficient that is slightly lower than unity and is equal to the ratio of the source–slit to source–crystal distance along the ray trajectory. Thus, we obtain

$$\Delta\theta(s) = \Delta\theta_0 - (\Delta d(s)/d)\tan\theta_B,$$

where $\Delta\theta_0$ is the angle of rotation of the sample about the monochromator; in other words, this is a parameter that changes in experiment. On the other hand, $\Delta d(s)$ is the change in the interplanar spacing in the sample as compared to the monochromator. This parameter depends on the crystal deformation, and it is not known in advance.

We introduce a dimensionless normalized function $B(\delta_E)$ to describe the spectral line of the X-ray source as a function of $\delta_E = \Delta E/E$, where E is the photon energy. For simplicity, we approximate the slit and source by rectangles of widths S and P, respectively, located normal to the trajectory, and the coordinate on the source is denoted by p. Each trajectory begins at a point p on the source and ends in a point s on the slit. In the middle section, the trajectory changes its direction upon reflection by the monochromator, and the point of reflection by the monochromator depends on the photon frequency. Reflection occurs in a very narrow region, which serves as a basis for applying the ray approximation. Having passed through the slit, the beam broadens substantially after Laue diffraction in the crystal; however, this is not important for us, since the detector records the integrated intensity. The only important parameter is the trajectory coordinate in the crystal. When considering all possible coordinates for the signal measured experimentally, we obtain the formula

$$I(\Delta\theta_0) = I_0 \int_{-P/2}^{P/2} dp \int_{-S/2}^{S/2} ds \int_{-\infty}^{\infty} d\delta_E B(\delta_E) \times \delta(s - p - L_B\delta_E) R(\Delta\theta(s)). \quad (1)$$

Here, the left-hand side contains the dimensionless integrated intensity (the number of photons); $\delta(x)$ is the delta function; $L_B = L\tan\theta_B$, where L is the source–sample distance; and I_0 is the differential intensity of the incident radiation per unit transverse source length for the whole spectral line width. Like the delta function, this integrated intensity has the dimension of reciprocal length. Note that, in the absence of deformation, all integrals only give the coefficient of proportionality, and the result is equal to $I_0 PR(\Delta\theta_0)$.

In our case, a certain simplification can be made, since the $R(\Delta\theta(s))$ function is independent of both the source coordinate and the photon energy. Due to the presence of the delta function, one integral is easily computed, and the result can be written in the form

$$I(\Delta\theta_0) = I_0 L_B^{-1} \int_{-S/2}^{S/2} ds F(s) R(\Delta\theta(s)),$$
$$F(s) = \int_{-P/2}^{P/2} dp B((s-p)/L_B). \quad (2)$$

Formulas (1) and (2) have common features, and the result depends substantially on the source and slit sizes via the $F(s)$ weight function.

In our case, $P/2 = S/2 = 100$ μm and $L_B = 11.4$ cm. Thus, for the center of the slit ($s = 0$), the maximum argument of the function under the integral taken over the source coordinates is equal to 10^{-3}. This value should be compared with the half-width of the Mo$K_{\alpha 1}$ spectral line used in the experiment, which is equal to 3×10^{-4} [13]. It is obvious that, at these parameters, the $F(s)$ function is equal to the integrated value of the spectral function virtually throughout the whole slit width, and it only halves at the slit edges. To a first approximation, we may neglect its shape and change it into unity; then, we have

$$I(\Delta\theta_0) = I_0 \int_{-S/2}^{S/2} ds R(\Delta\theta(s)). \quad (3)$$

With this formula, we can rather simply interpret the experimental curves. The reflection intensity upon diffraction is independent of both the radiation wavelength and the position of a point source, and it is only specified by the local difference in the Bragg angles in the crystal and monochromator. The summation of the local reflection coefficients does not depend on the trajectory of a certain ray, and it is only determined by the total illumination of each point inside the slit (and, hence, on the crystal surface). This illumination is virtually uniform inside the space limited by the slit. If the deformation is uniform inside the beam width, the experimental curve should be equal to the curve of the ideal crystal and the peak position can shift if the deformation is nonzero. This experimental result seems to be obvious. If the deformation is nonuniform inside the beam width, we have to integrate the reflection intensity with respect to the beam width by making allowance for the local difference in the Bragg angles at each point.

The experiment shows that some curves have a width well above the width of the reflection curve of the undeformed crystal. In terms of the theoretical model proposed, this means that the deformation in the sample changes sufficiently strongly inside the beam width, so that the region of a changed Bragg angle is significantly larger than the width of the $R(\Delta\theta)$ function. When analyzing such curves, we may neglect the width of the $R(\Delta\theta)$ function and replace it by the $R_0\delta(\Delta\theta)$ function, where R_0 is the integral of the reflection curve of the undeformed crystal. From a physical standpoint, this means that only the part of the beam width having the corresponding Bragg angles is reflected rather than the whole beam width. The real intensity for every angle of crystal rotation is determined by the width of this part. Assuming that the deformation gradient inside the beam width does not change its sign and using this approximation, we obtain the simple formulas

$$t = (\Delta d/d)\tan\theta_B = \Delta\theta_0, \quad ds/dt = CI(t), \qquad (4)$$

where C is a normalizing constant to be easily determined from the sizes of the angular and spatial regions.

The t parameter specifies the local deformation, and the $t(s)$ function describes the desired deformation profile within the beam width. In real practice, we first determine

$$s(t) = -S/2 + C\int_{t_0}^{t} dxI(x),$$
$$C = S\left(\int_{t_0}^{t_1} dxI(x)\right)^{-1}, \qquad (5)$$

where t_0 and t_1 are the boundaries of the angular region with a noticeable intensity between two pronounced slopes. The desired $t(s)$ function can readily be determined graphically. It directly demonstrates an interplanar-spacing profile in the beam width.

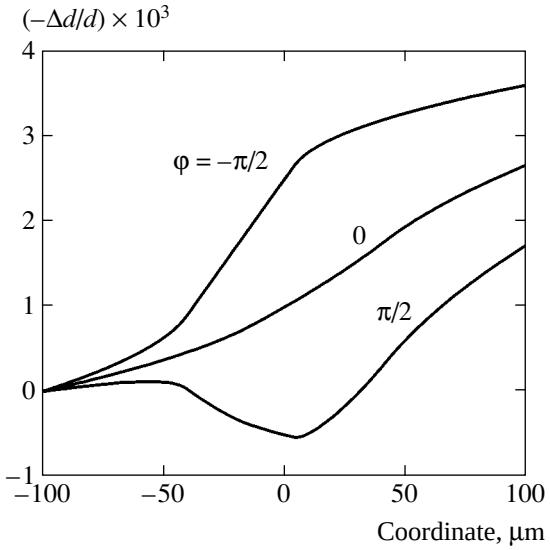


Fig. 11. Deformation profiles calculated from the experimental data for the region near the place of gluing with the resonator (see Fig. 10).

This simple model is found to reliably reconstruct a deformation profile within the beam width from an experimental curve. The experimental curves obtained in region 3 (Fig. 10) satisfy the conditions of its applicability best of all. Figure 11 shows three curves demonstrating such deformation profiles. The deformation was determined from the angular position of the crystal, and the zero mark was set at the same accuracy. The argument is the position of a point in the illuminated area. The upper curve was calculated by Eqs. (4) and (5) from the experimental curve for a phase $\varphi = -\pi/2$, and the middle curve, from the curve for a phase $\varphi = 0$. It is obvious that, in the second case, the crystal has only a static deformation, whereas, in the first case, an ultrasonic deformation is superimposed on the static deformation. Both deformations have the same sign, which increases the lattice deformation.

Note that the purely ultrasonic deformation can be obtained by the subtraction of the lower curve from the upper curve. For a phase $\varphi = \pi/2$, the result obtained should be subtracted from the purely static deformation. The lower curve in Fig. 11 is obtained as a result of this subtraction. It is seen to have an almost flat long segment in the region of zero deformation. Therefore, this region should correspond to a sharp reflection peak with a width close to the width of the reflection curve of the undeformed crystal. This behavior was detected in the experiment (see Fig. 10). Thus, the compensation of static and dynamic deformations can be described in terms of the model proposed despite the fact that the assumption of a weakly changing deformation within a dynamic diffraction region lies on the boundary of applicability.

Theoretical analysis supports the assumption that, apart from the fundamental long-wave harmonic, a parasitic harmonic with a smaller spatial period is excited in the crystal.

5. CONCLUSIONS

We were the first to experimentally show the possibility of both uniform and gradient periodic time modulation of the lattice parameter by long-wave ultrasound. This possibility allowed us to electronically control the angular position and the spatial structure of an X-ray beam. When ultrasound is excited in a composite resonator with a thin sample to be measured, we detected a parasitic deformation with a significantly smaller period. This issue has to be studied in more detail. The static deformation was shown to be compensated for by an ultrasound-induced dynamic deformation in the crystal.

We developed a simple theoretical model for X-ray diffraction by a crystal with a spatially nonuniform deformation that is induced by a long-wave ultrasonic wave along the sample surface. Using this model, we described the experimental curves even in the case where deformation changes relatively rapidly.

ACKNOWLEDGMENTS

We thank O.P. Aleshko-Ozhevskiĭ†, A.A. Lomov, and E.Kh. Mukhamedzhanov for their assistance in the experimental work and for helpful discussions of the results.

This work was supported by the Russian Foundation for Basic Research, project no. 04-02-0817 ofi-a.

REFERENCES

1. I. R. Éntin, Pis'ma Zh. Éksp. Teor. Fiz. **26**, 392 (1977) [JETP Lett. **26**, 269 (1977)].
2. A. Hauer and S. J. Burns, Appl. Phys. Lett. **27**, 524 (1975).
3. D. V. Roshchupkin, R. Tucoulou, M. Brunel, and V. V. Shchelokov, in *Proceedings of All-Russian Meeting on X-ray Optics* (Nizhni Novgorod, 1999), Vol. 1, p. 83.
4. D. Shilo, E. Lakin, E. Zolotoyabko, *et al.*, Synchrotron Radiat. News **15**, 17 (2002).
5. E. Zolotoyabko and J. P. Quintana, Rev. Sci. Instrum. **75**, 699 (2004).
6. R. Tucoulou, D. V. Roshupkin, O. Mathon, *et al.*, J. Synchrotron Radiat. **5**, 1357 (1998).
7. M. A. Navasardyan, J. Appl. Crystallogr. **34**, 763 (2001).
8. V. L. Nosik and M. V. Kovalchuk, Nucl. Instrum. Methods Phys. Res. A **405**, 480 (1998).
9. V. L. Nosik and M. V. Kovalchuk, Poverkhnost, No. 1, 91 (2000).
10. M. V. Koval'chuk, É. K. Kov'ev, and Z. G. Pinsker, Kristallografiya **20**, 142 (1975) [Sov. Phys. Crystallogr. **20**, 81 (1975)].
11. V. G. Kohn, I. Snigireva, and A. Snigirev, Phys. Status Solidi B **222**, 407 (2000).
12. V. G. Kohn, Zh. Éksp. Teor. Fiz. **124**, 224 (2003) [JETP **97**, 204 (2003)].
13. M. A. Blokhin and I. G. Shveĭtser, *X-ray Spectrum Handbook* (Nauka, Moscow, 1982), p. 376 [in Russian].

† Deceased.

Translated by K. Shakhlevich

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 779–787.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 904–912.
Original Russian Text Copyright © 2005 by Alshits, Lyubimov.

ATOMS, MOLECULES, OPTICS

Dispersion Polaritons on Metallized Surfaces of Optically Uniaxial Crystals

V. I. Alshits and V. N. Lyubimov

Shubnikov Institute of Crystallography, Russian Academy of Sciences, Moscow, 119333 Russia
e-mail: alshits@ns.crys.ras.ru

Received June 21, 2005

Abstract—We have constructed a theory of dispersion polaritons (localized electromagnetic waves) on arbitrarily oriented metallized surfaces of optically uniaxial crystals. The domain of existence of polaritons is defined by the following inequalities for permittivities ε_o and ε_e of the crystal and the angle θ between the optical axis and the surface: $-\varepsilon_e \tan^2\theta < \varepsilon_o < 0$. Thus, polaritons exist only in the range of wave frequencies ω ensuring negative values of $\varepsilon_o(\omega)$ for $\varepsilon_e > 0$. The frequency boundaries of this region are specified for the case when the $\varepsilon_o(\omega)$ dependence corresponds to the model of a single polar excitation. The azimuthal orientation φ of the optical axis projection onto the surface does not appear in the criterion for polariton existence, but affects (together with angle θ) its main dispersion characteristics, such as the refractive index and partial wave localization parameters. This effect is analytically described in detail. Anomalies in the behavior of polariton parameters are studied in the vicinity of the boundaries of the domain of its existence, where the wave fields are especially sensitive to variations in the angles θ and φ. It is shown that a polariton in the plane of propagation (sagittal plane) passing through the optical axis is transformed into a one-partial bulk wave satisfying the boundary conditions. Accordingly, the wave branch under investigation for close orientations (when the optical axis forms a small angle with the sagittal plane) describes deeply penetrating (quasi-bulk) polaritons. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

It is well known [1–6] that localized electromagnetic waves (polaritons) can propagate under certain conditions along certain directions on the surface of a crystal in contact with an isotropic dielectric. Such modes appear due to strong frequency dispersion of the crystal permittivity tensor $\hat{\varepsilon}$ in the vicinity of certain resonance states [1–3], for which the tensor components of $\hat{\varepsilon}(\omega)$ can be negative. It was shown in [4–6], however, that surface polaritons can also exist in crystals (owing to dielectric anisotropy) in an ordinary dispersion-free version, when tensor $\hat{\varepsilon}$ is positive definite and weakly depends on the wave frequency ω.

A metal coating deposited on the crystal surface serves as a reflecting screen confining the electromagnetic field in the crystal. In accordance with the general theory [7], dispersion-free polaritons in principle cannot exist in this case. We will demonstrate here that this prohibition on the existence of polaritons at a metallized surface can be removed if we do not impose the condition of positive definiteness of tensor $\hat{\varepsilon}$ (i.e., we consider the situation when components of $\hat{\varepsilon}(\omega)$ can assume negative values).

The theory of localized electromagnetic waves on the "open" surface of a crystal in contact with a dielectric is usually quite cumbersome even for uniaxial crystal and permits analytic solutions only for preferred symmetric orientations of the surface and directions of propagation (see, for example, [3, 5, 6]). It turned out that the theory of surface polaritons at a metallized boundary of a uniaxial crystal is simpler from the mathematical point of view and can be constructed in a general analytic form for an arbitrary geometry of the problem.

In this paper, we determine the conditions for propagation of such polaritons in optically uniaxial crystals and establish the domain of their existence, which is defined by certain relations between the components of $\hat{\varepsilon}$ and the angle θ between the optical axis and the surface. It is important that the azimuthal orientation φ of the optical axis projection onto the surface plane does not appear in the criterion for the emergence of such polaritons. Nevertheless, the main parameters of a polariton naturally depend on the azimuth. The orientation dependences of polariton characteristics, as well as the properties of polaritons near the boundaries of the domain of their existence, will be studied analytically in detail. The boundaries of this domain will be specified for the case when the dispersion branch $\varepsilon_o(\omega)$ of the ordinary wave corresponds to the model of a single polar excitation.

It will be shown that the polariton considered here is transformed into an exceptional bulk wave in the special case when the optical axis of the crystal is parallel to the sagittal plane defined by the set **m** and **n** of the

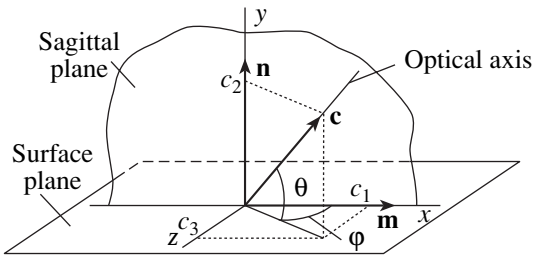


Fig. 1. System of coordinates *xyz* and orientation of the optical axis *c* of the crystal relative to this system.

direction of propagation and the normal to the surface. In the vicinity of this orientation, when the angle between the optical axis and the sagittal plane is small, weak localization of the wave is observed; i.e., the polariton becomes a quasi-bulk mode. Analogous transformations of exceptional bulk waves into quasi-bulk waves are well known in crystal acoustics [8, 9].

2. FORMULATION OF THE PROBLEM

Let us consider a semi-infinite optically uniaxial medium with an arbitrarily oriented metallized boundary. We choose the Cartesian system of coordinates with the *y* axis directed along the inward normal **n** to the surface and the *x* axis directed along the propagation vector **m**. In this case, the *xy* plane coincides with the sagittal plane of the wave and the *xz* plane coincides with the crystal surface (Fig. 1). In this system of coordinates, the orientation of the optical axis defined by unit vector **c** is defined by two angles (θ and φ).

The wave fields studied here can be represented as a superposition of two partial (ordinary and extraordinary) components. Subscripts "o" and "e" mark the corresponding wave parameters. In the general case, the structure of such fields has the form

$$\begin{pmatrix} \mathbf{E}(x,y,t) \\ \mathbf{H}(x,y,t) \end{pmatrix} = \left\{ C_\mathrm{o} \begin{pmatrix} \mathbf{E}_\mathrm{o}(y) \\ \mathbf{H}_\mathrm{o}(y) \end{pmatrix} + C_\mathrm{e} \begin{pmatrix} \mathbf{E}_\mathrm{e}(y) \\ \mathbf{H}_\mathrm{e}(y) \end{pmatrix} \right\} \times \exp[ik(x - vt)], \tag{1}$$

$$\begin{pmatrix} \mathbf{E}_\mathrm{o}(y) \\ \mathbf{H}_\mathrm{o}(y) \end{pmatrix} = \begin{pmatrix} \mathbf{E}_\mathrm{o}^{(0)} \\ \mathbf{H}_\mathrm{o}^{(0)} \end{pmatrix} \exp(-q_\mathrm{o} ky), \quad \begin{pmatrix} \mathbf{E}_\mathrm{e}(y) \\ \mathbf{H}_\mathrm{e}(y) \end{pmatrix} = \begin{pmatrix} \mathbf{E}_\mathrm{e}^{(0)} \\ \mathbf{H}_\mathrm{e}^{(0)} \end{pmatrix} \exp[(ip - q_\mathrm{e})ky]. \tag{2}$$

Here, **E** and **H** are the electric and magnetic field strengths, respectively; *k* is the common *x* component of the wavevectors of the ordinary and extraordinary partial waves ($k = \mathbf{k}_\mathrm{o} \cdot \mathbf{m} = \mathbf{k}_\mathrm{e} \cdot \mathbf{m}$); $v = \omega/k$ is the reduced phase velocity of the wave; and C_o and C_e are the amplitude factors determined from the boundary conditions. Complex wavevectors $\mathbf{k}_\mathrm{o}$ and $\mathbf{k}_\mathrm{e}$ appearing in formulas (1) and (2) differ only in the components normal to the surface:

$$\mathbf{k}_\mathrm{o} = k(1, iq_\mathrm{o}, 0), \quad \mathbf{k}_\mathrm{e} = k(1, p + iq_\mathrm{e}, 0). \tag{3}$$

For our purposes, it is more convenient to use the corresponding dimensionless refractive vectors

$$\mathbf{n}_\mathrm{o} = \mathbf{k}_\mathrm{o}/k_0 = n(1, iq_\mathrm{o}, 0), \quad \mathbf{n}_\mathrm{e} = \mathbf{k}_\mathrm{e}/k_0 = n(1, p + iq_\mathrm{e}, 0). \tag{4}$$

Here, $k_0 = \omega/c$, where *c* is the velocity of light in vacuum, and $n = k/k_0 = c/v$ is the dimensionless slowness of the wave, which is also known as the refractive index.

Parameters q_o, q_e, and *p* appearing in Eqs. (2)–(4) depend both on the material characteristics of the crystal (ε_o and ε_e) and on the orientation of the unit vector of optical axis $\mathbf{c} = (c_1, c_2, c_3)$ relative to the surface. Using general relations in the optics of uniaxial crystals [10–12], we can easily obtain these dependences in explicit form,

$$q_\mathrm{o}^2 = 1 - \frac{\varepsilon_\mathrm{o}}{n^2}, \quad q_\mathrm{e}^2 = \left(\frac{B}{\varepsilon_\mathrm{o} A} - \frac{1}{n^2}\right)\frac{\varepsilon_\mathrm{e}}{A}, \quad p = \frac{(\varepsilon_\mathrm{o} - \varepsilon_\mathrm{e})c_1 c_2}{\varepsilon_\mathrm{o} A}, \tag{5}$$

where the following notation has been introduced:

$$A = 1 - c_2^2\left(1 + \frac{1}{\kappa}\right), \quad B = 1 - c_3^2(1 + \kappa), \quad \kappa = -\frac{\varepsilon_\mathrm{o}}{\varepsilon_\mathrm{e}}. \tag{6}$$

The value of refractive index *n* appearing in expressions (5) must be determined from the boundary conditions. According to [10], a specific feature of the boundary conditions for a crystal with a metallized surface is that the tangential electric field components $\mathbf{E}_t$ vanish at the surface:

$$\mathbf{E}_t\big|_\mathrm{surf} = 0. \tag{7}$$

This relation automatically implies that the normal components of magnetic field **H** and Poynting's vector $\mathbf{P} = \mathbf{E} \times \mathbf{H}$ also vanish at the surface:

$$\mathbf{H}_n\big|_\mathrm{surf} = 0, \quad \mathbf{P}_n\big|_\mathrm{surf} = 0. \tag{8}$$

3. DISPERSION RELATION AND DOMAIN OF EXISTENCE OF POLARITONS

For an arbitrary orientation of the optical axis relative to the crystal surface, the vector amplitudes of electric and magnetic fields can be written in the following specific form [10–12]:

$$\begin{pmatrix} \mathbf{E}_o^{(0)} \\ \mathbf{H}_o^{(0)} \end{pmatrix} = -\frac{1}{n}\begin{pmatrix} \mathbf{n}_o \times \mathbf{c} \\ \mathbf{n}_o \times (\mathbf{n}_o \times \mathbf{c}) \end{pmatrix},$$
$$\begin{pmatrix} \mathbf{E}_e^{(0)} \\ \mathbf{H}_e^{(0)} \end{pmatrix} = \begin{pmatrix} \mathbf{c} - \mathbf{n}_e(\mathbf{n}_e \cdot \mathbf{c})/\varepsilon_o \\ \mathbf{n}_e \times \mathbf{c} \end{pmatrix}. \quad (9)$$

Taking into account these relations, we can reduce boundary conditions (7) and (8) for superposition (1) and (2) of the ordinary and extraordinary waves under investigation to the equation

$$\begin{pmatrix} iq_o c_1 - c_2 & c_3 \\ iq_o c_3/n^2 & gc_1 + iq_e c_2/\varepsilon_o \end{pmatrix}\begin{pmatrix} C_o \\ C_e \end{pmatrix} = 0, \quad (10)$$

where the function

$$g(n) = \frac{1}{\varepsilon_o A} - \frac{1}{n^2}$$

has been introduced. Nonzero amplitudes $C_{o,e}$,

$$C_e = \frac{-iq_o c_1 + c_2}{c_3} C_o, \quad (12)$$

exist if the determinant of the matrix in Eq. (10) vanishes:

$$f(n) = \left(g(n)c_1 + \frac{iq_e(n)c_2}{\varepsilon_o}\right) \times (c_2 - iq_o(n)c_1)n^2 + iq_o(n)c_3^2 = 0. \quad (13)$$

In the complex dispersion equation (13), the real and imaginary parts must vanish simultaneously:

$$s - 1/A - q_o(s)q_e(s) = 0,$$
$$[s(1-c_2^2) - c_1^2/A]q_o(s) + c_2^2 = 0. \quad (14)$$

For brevity, we introduced the notation $s = \varepsilon_0/n^2$. After cumbersome calculations, we can prove that the problem is not overdetermined and each equation in (14) can be reduced to the same real quadratic equation in unknown s. An analogous result has been obtained in the general theory of surface polaritons in crystals with positive definite $\hat{\varepsilon}$ [7, 13, 14]. The roots of the resultant equation have the form

$$s_\pm = \frac{A + c_1^2 \pm r}{2A(1-c_2^2)}. \quad (15)$$

In this expression, radical $r > 0$ is defined by the relation

$$r^2 = (A + c_1^2)^2 - 4A(1-c_2^2)(1-c_3^2) = (A - c_1^2)^2 - 4Ac_2^2c_3^2. \quad (16)$$

The sought localized wave fields can correspond to only those roots $s_\pm$ in relation (15) which ensure positive values of parameters n^2, q_o, and q_e. For such roots, the initial system of equations (14) is solvable if the inequalities

$$\frac{1}{A} < s < \frac{1}{A}\frac{c_1^2}{c_1^2 + c_3^2}, \quad (17)$$

which are compatible only for $A < 0$, hold. It can be seen from formula (16) that $|A + c_1^2| < r$ in this case also; consequently, only root s_+ from the two roots (15) is negative and can satisfy system of inequalities (17). Hence, the dispersion of a polariton in the domain of its existence is defined by the equation

$$n^2 = \frac{\varepsilon_o}{s_+} = \frac{\varepsilon_0(A + c_1^2 - r)}{2(1-c_3^2)}. \quad (18)$$

To find this domain, we note that the negative value of the root $s_+ \equiv \varepsilon_0/n^2 < 0$ leads to the necessary condition for the existence of a polariton:

$$\varepsilon_o < 0. \quad (19)$$

On the other hand, taking into account relations (6), we note that the inequality $A < 0$ is equivalent to the requirement

$$0 < \kappa \equiv -\frac{\varepsilon_o}{\varepsilon_e} < \frac{c_2^2}{c_1^2 + c_3^2} \equiv \tan^2\theta, \quad (20)$$

where θ is the angle of inclination of the optical axis (see Fig. 1). It follows from relation (20) that, in addition to necessary negativeness of component ε_o (19), positiveness of ε_e should also be ensured for the existence of a polariton:

$$\varepsilon_e > 0. \quad (21)$$

The system of inequalities (19)–(21) defining the domain of existence of polaritons can be represented in

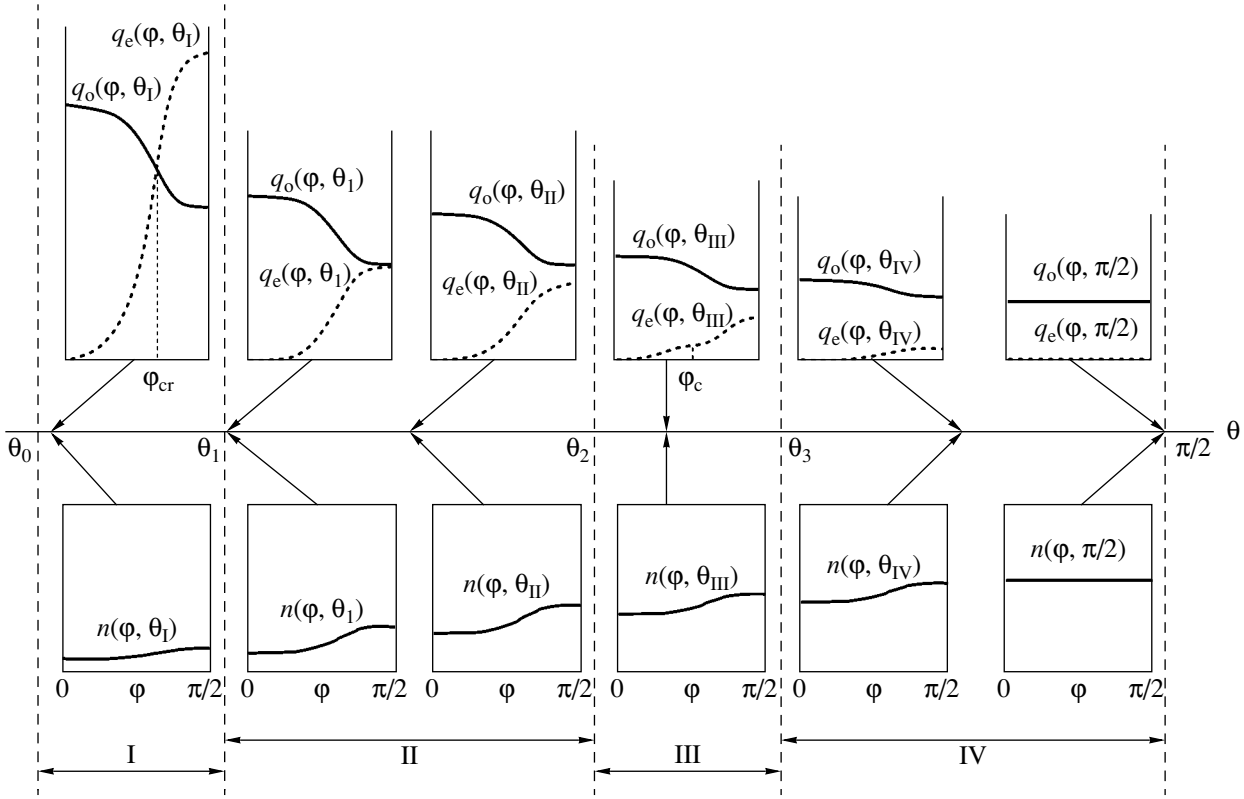


Fig. 2. Schematic diagram of dispersion curves $q_o(\varphi, \theta)$, $q_e(\varphi, \theta)$, and $n(\varphi, \theta)$ for fixed values of θ (marked by arrows) in four regions of the polariton domain for $\theta_0 < \theta \le \pi/2$.

a more compact form:

$$-\varepsilon_e \tan^2\theta < \varepsilon_o < 0. \tag{22}$$

4. DEPENDENCE OF POLARITON PARAMETERS ON THE OPTICAL AXIS ORIENTATION

It is interesting to note that domain (22) of admissible variations of the parameters of the medium depends only on angle θ between the optical axis and the surface and not on its azimuth φ (see Fig. 1). This naturally does not rule out the azimuthal dependence of the main polariton characteristics such as refractive index n (18) and localization parameters q_o and q_e (5):

$$q_o^2 = 1 - \frac{2(1-c_3^2)}{A + c_1^2 - r} = 1 - \frac{A + c_1^2 + r}{2A(1-c_2^2)},$$
$$q_e^2 = \frac{1}{\kappa A^2}\left(\frac{A + c_1^2 + r}{2(1-c_2^2)} - B\right). \tag{23}$$

Taking into account (6) and the explicit relation of the components of vector $\mathbf{c}$ with spherical angles θ and φ (see Fig. 1)

$$\mathbf{c} = (c_1, c_2, c_3) = (\cos\theta\cos\varphi, \sin\theta, \cos\theta\sin\varphi), \tag{24}$$

it can easily be verified that all functions $n(\varphi, \theta)$, $q_o(\varphi, \theta)$, and $q_e(\varphi, \theta)$, when plotted as polar diagrams in φ for a fixed θ, are symmetric about straight lines $\varphi = 0$ and $\varphi = \pi/2$; consequently, the complete pattern of the behavior of these functions in any cross sections θ = const is fully characterized by an interval of $0 \le \varphi \le \pi/2$. If we use conventional plots of these functions of argument φ in the given interval instead of polar diagrams for mapping the above-mentioned azimuthal dependences, these plots must have horizontal tangents at the ends of this interval since their first derivatives with respect to φ are proportional to $\sin 2\varphi$. For any fixed angle θ, function $n(\varphi, \theta)$ increases monotonically, while function $q_o(\varphi, \theta)$ decreases monotonically in the interval $0 < \varphi < \pi/2$ (see Fig. 2), remaining, however, greater than unity (see Eqs. (5)):

$$q_o \ge 1. \tag{25}$$

The behavior of the function $q_e(\varphi, \theta)$ is as follows, for $\varphi = 0$, irrespective of the value of θ, we have

$$q_e(0, \theta) = 0. \tag{26}$$

In the same interval $0 \le \varphi \le \pi/2$, function $q_e(\varphi, \theta)$ at $\theta =$ const can exhibit different modes of behavior depending on the range in which angle θ falls. In the domain of polariton existence (22), we must distinguish between four such regions (Fig. 2):

$$\begin{aligned} &\text{(I) } \theta_0 < \theta < \theta_1, \quad \text{(II) } \theta_1 < \theta < \theta_2, \\ &\text{(III) } \theta_2 < \theta < \theta_3, \quad \text{(IV) } \theta_3 < \theta < \pi/2, \end{aligned} \tag{27}$$

whose boundaries are determined by the relations

$$\begin{aligned} \theta_0 &= \arctan\sqrt{\kappa}, \quad \theta_1 = \arctan\sqrt{\kappa_1}, \\ \theta_2 &= \arctan\sqrt{\kappa_2}, \quad \theta_3 = \arctan\sqrt{2\kappa + 1}, \end{aligned} \tag{28}$$

where

$$\begin{aligned} \kappa_1 &= \frac{1}{4}(3\kappa - 1 + \sqrt{9\kappa^2 + 10\kappa + 1}), \\ \kappa_2 &= \frac{1}{4}(3\kappa + 2 + \sqrt{9\kappa^2 + 4\kappa + 4}). \end{aligned} \tag{29}$$

For angles of inclination θ belonging to regions I, II, and IV, function $q_e(\varphi, \theta)$ increases monotonically with angle φ, while for values of θ fixed in region III, this function only does not decrease monotonically, having a "cubic" point of zero slope for the azimuth $\varphi_c(\theta)$ defined by the equation

$$\sin\varphi_c = 2\sin\theta[1 + (2\kappa)^{-1}(1 - \tan^2\theta)]^{1/2}. \tag{30}$$

In other words, two derivatives of function $q_e(\varphi, \theta)$ with respect to φ must vanish simultaneously at point φ_c (i.e., $q_e'(\varphi_c, \theta) = q_e''(\varphi_c, \theta) = 0$), and the function must exhibit a very low sensitivity to variations of φ in the vicinity of angle φ_c (see Fig. 2):

$$q_e(\varphi, \theta) \approx q_e(\varphi_c, \theta) + \lambda(\varphi - \varphi_c)^3, \quad \lambda > 0. \tag{31}$$

With increasing angle θ, the position of point of inflection φ_c (30) in region III shifts from the right to the left end of the interval $0 \le \varphi \le \pi/2$.

It should be noted that curves $q_o(\varphi)$ and $q_e(\varphi)$ intersect only for angles θ fixed in region I (i.e., for $\varphi = \varphi_{cr}(\theta)$; see Fig. 2):

$$\cos 2\varphi_{cr} = (1 + \sin^2\theta)(1 - \kappa^{-1}\tan^2\theta). \tag{32}$$

As the value of angle θ increases in interval I, the position of the azimuth of intersection φ_{cr} (32) changes from $\varphi_{cr} = \pi/4$ to $\pi/2$.

Let us introduce the total azimuthal dispersions for polariton localization parameters $q_o(\varphi, \theta)$ and $q_e(\varphi, \theta)$ for a fixed value of θ:

$$\Delta q_o(\theta) = q_o(0, \theta) - q_o\left(\frac{\pi}{2}, \theta\right),$$

$$\Delta q_e(\theta) = q_e\left(\frac{\pi}{2}, \theta\right) - q_e(0, \theta) = q_e\left(\frac{\pi}{2}, \theta\right).$$

On the basis of Eqs. (23), we can easily verify that all functions $q_o(0, \theta)$, $q_o(\pi/2, \theta)$, and $\Delta q_o(\theta)$, as well as $\Delta q_e(\theta) = q_e(\pi/2, \theta)$, decrease monotonically with increasing angle θ in the domain of polariton existence. In the same interval, functions $n(0, \theta)$ and $n(\pi/2, \theta)$ increase monotonically in accordance with Eqs. (5), while the difference $\Delta n(\theta) = n(\pi/2, \theta) - n(0, \theta)$ does not exhibit monotonicity since it vanishes both at the left and right ends of the interval $\theta_0 \le \theta \le \pi/2$ (see Fig. 2).

5. SOLUTIONS IN THE VICINITY OF DOMAIN BOUNDARIES

The boundaries of the polariton domain (22) are defined by the relations

$$\varepsilon_o = -\varepsilon_e\tan^2\theta, \quad \varepsilon_o = 0. \tag{33}$$

Let us consider in greater detail the behavior of the main polariton parameters in the vicinity of these boundaries.

5.1. Neighborhood of the Lower Boundary

The lower boundary for ε_o in inequality (22) corresponds to the limiting value $A = 0$ which, in accordance with relations (6) and (20), fixes the slope $\theta = \theta_0$,

$$\tan^2\theta_0 = \kappa \equiv -\varepsilon_o/\varepsilon_e, \tag{34}$$

preserving arbitrariness in azimuth φ. Precisely at this boundary, the polariton is obviously absent since the refractive index vanishes ($n = 0$) for $A = 0$ in accordance with relations (16) and (18). In other words, $k = 0$, which "suppresses" the stationary wave field (1) propagating parallel to the surface. However, in accordance with the theory developed here, the polariton must exist in any small neighborhood of boundary (34), albeit with quite peculiar properties.

For a small but nonzero A ($|A| \ll 1$), dispersion equation (18) can be simplified so that it assumes the form

$$n^2 \approx -\varepsilon_o/q_o^2. \tag{35}$$

However, localization parameters q_o and q_e have different forms depending on the additional relation between $|A|$ and c_1^2, i.e., in two limiting cases

$$(1)\ 0 < -A \ll c_1^2, \quad (2)\ 0 < -A \ll 1, \quad c_1^2 \ll 1. \tag{36}$$

(1) This limiting case is equivalent to the condition

$$0 < \tan^2\theta - \kappa \ll \kappa\cos^2\varphi. \tag{37}$$

In such sections of the crystal, a surface polariton is characterized by the parameters

$$\begin{aligned} q_o^2 &\approx \frac{c_1^2}{-A(1-c_2^2)} = \frac{\cos^2\varphi}{-A} \\ &\approx \frac{\kappa(1+\kappa)}{\tan^2\theta - \kappa}\cos^2\varphi \gg 1+\kappa, \end{aligned} \tag{38}$$

$$\begin{aligned} q_e^2 &\approx \frac{c_3^4}{-Ac_1^2(1-c_2^2)} = \frac{1}{-A}\frac{\sin^4\varphi}{\cos^2\varphi} \\ &\approx \frac{\kappa(1+\kappa)}{\tan^2\theta - \kappa\cos^2\varphi}\frac{\sin^4\varphi}{\cos^2\varphi}. \end{aligned} \tag{39}$$

It can be seen from relation (38) that the localization parameter of the ordinary component in the given limiting case (37) is automatically large. In accordance with relations (38) and (39), in this case we have

$$q_e/q_o \approx \tan^2\varphi. \tag{40}$$

Consequently, the localization of the extraordinary partial wave can be smaller or larger than that of the ordinary wave depending on azimuth φ.

When angle φ in formula (40) approaches π/2 without violating condition (37) so that $1 \ll q_o \ll q_e$, the polariton is found to be strongly localized and one-partial almost everywhere and is characterized by parameters of the ordinary component. However, because the product $q_o c_1$ in relation (12) can be either small or large in this case, we can state that the surface structure of the polariton becomes anomalously sensitive to small variations in the orientation of vector **c** in the vicinity of $\theta \approx \theta_0$, $\varphi \approx \pi/2$.

If, however, azimuth φ in relation (40) is close to zero (i.e., the optical axis forms a small angle with the sagittal plane), the opposite situation takes place: $q_o \gg q_e$; in this case, the polariton is mainly determined by the extraordinary partial wave and its localization can be controlled arbitrarily by choosing angle φ. In particular, parameter q_e can be chosen arbitrarily small and even equal to zero, which corresponds to a bulk (nonlocalized) polariton. This case will be considered separately at a later stage.

(2) The other limiting case in (36) is defined by the system of inequalities

$$\begin{aligned} 0 < \frac{\tan^2\theta - \kappa}{\kappa(1+\kappa)} &\ll 1, \\ \frac{\cos^2\varphi}{1+\kappa} &\ll 1. \end{aligned} \tag{41}$$

The corresponding expressions for the localization parameters have the form

$$\begin{aligned} q_o^2 &\approx \frac{c_2^2}{-Ac_3^2} \approx \frac{\kappa}{\sin^2\varphi}\frac{\kappa(1+\kappa)}{\tan^2\theta - \kappa}, \\ q_e^2 &\approx \frac{c_3/c_2}{\sqrt{-A^3}} \approx \frac{\sin\varphi}{\sqrt{\kappa}}\left(\frac{\kappa(1+\kappa)}{\tan^2\theta - \kappa}\right)^{3/2}. \end{aligned} \tag{42}$$

In this case, their ratio has a much more complex form than simple formula (40):

$$\frac{q_e}{q_o} \approx \left(\frac{\sin^2\varphi}{\kappa}\right)^{3/4}\left(\frac{\kappa(1+\kappa)}{\tan^2\theta - \kappa}\right)^{1/4}. \tag{43}$$

It can be seen from relations (41)–(43) that subsequent analysis is determined to a considerable extent by the value of parameter κ. If we disregard for the time being the region of small values of κ corresponding to the vicinity of the upper boundary of the polariton domain (22) ($\varepsilon_o \longrightarrow 0$), which will be considered separately (see Section 5.2), two possibilities remain: $\kappa \sim 1$ and $\kappa \gg 1$.

For $\kappa \sim 1$, the second inequality in system (41) can be ensured only for small values of $\cos^2\varphi$ (i.e., $\sin^2\varphi \sim 1$). Obviously, both localization parameters must be large in this case and, additionally, $q_e \gg q_o$ in the immediate vicinity of the boundary ($\theta \longrightarrow \theta_0$); i.e., the pattern is qualitatively analogous to that observed in the first case for $\varphi \approx \pi/2$.

In the limiting case of $\kappa \gg 1$, the second inequality in (41) is observed for any angle φ; parameter q_o is universally large, while the value of q_e can be either large or small. For $q_e \ll q_o$, the polariton is mainly determined by the extraordinary partial wave and the depth of its penetration can be controlled arbitrarily by the choice of the section of the surface determining azimuth φ. For $q_e \gg q_o \gg 1$, the ordinary component plays the major role; however, it is strongly localized for any angle φ.

5.2. Neighborhood of the Upper Boundary

The upper boundary $\varepsilon_o = 0$ of the polariton domain corresponds to the limiting values

$$\kappa \longrightarrow 0, \quad -A \longrightarrow \infty, \quad \kappa A \longrightarrow -c_2^2. \tag{44}$$

Substituting these relations into Eqs. (18) and (23), we can easily obtain the main parameters of the corresponding limiting polariton:

$$n^2 = \frac{\varepsilon_e}{1 + \cot^2\theta\cos^2\varphi}, \quad q_o = 1, \quad q_e = 0. \tag{45}$$

Thus, the limiting polariton considered here is completely delocalized with respect to one (extraordinary) component for any orientation of the optical axis. Obviously, the corresponding wave branch acquires weak

localization in the vicinity of this boundary, becoming a quasi-bulk wave.

6. EXCEPTIONAL BULK WAVE AND A QUASI-BULK POLARITON

Thus, the above system of dispersion relations permits delocalized solutions: bulk waves with a component characterized by zero localization parameter may in principle propagate along some surfaces and directions.

In accordance with relation (26), the condition of delocalization of the extraordinary component ($q_e = 0$) in the entire region of (22) is automatically satisfied for $\varphi = 0$, which corresponds to the choice of the sagittal plane parallel to the optical axis. In contrast to two-partial solution (45), such a bulk wave is one-partial in principle since, in accordance with relation (12), $C_e \neq 0$ and $C_o = 0$ for $c_3 = \cos\theta\sin\varphi = 0$. For the wave under investigation, we have

$$\mathbf{c} = (\cos\theta, \sin\theta, 0), \quad \mathbf{n}_e = (1, p, 0)n, \tag{46}$$

$$n^2 = \varepsilon_o\cos^2\theta + \varepsilon_e\sin^2\theta, \quad p = \frac{(\varepsilon_o - \varepsilon_e)\sin 2\theta}{2n^2}. \tag{47}$$

The vector amplitude of this wave in terms of relations (1) and (2) has the form

$$\begin{pmatrix} \mathbf{E}_e^{(0)} \\ \mathbf{H}_e^{(0)} \end{pmatrix} = \begin{pmatrix} (0, 1, 0)/n \\ (0, 0, 1) \end{pmatrix}. \tag{48}$$

The bulk solution considered here belongs to the continuous branch of polaritons; it emerges in the limit for $\varphi = 0$. Consequently, for a small perturbation of φ, it obviously must be transformed into a weakly localized quasi-bulk polariton. Indeed, let us introduce a small parameter c_3 ($c_3^2 \ll 1$) that draws the optical axis from the sagittal plane. In this case, the wave field components (48) of a bulk wave change in proportion to c_3, while the initial parameters n and p (47) change in proportion to c_3^2. The initially zero component q_e determining the wave field localization now becomes nonzero:

$$q_e = \frac{c_3^2}{c_2}\frac{\kappa}{n}\sqrt{\varepsilon_e - \varepsilon_o}. \tag{49}$$

In addition, for the above perturbation, the extraordinary wave with a small amplitude ($C_o \sim c_3 C_e$) is supplemented with an ordinary partial wave with the parameters

$$\mathbf{n}_o = (1, iq_o, 0)n, \quad q_o = \frac{c_2}{n}\sqrt{\varepsilon_e - \varepsilon_o}, \tag{50}$$

$$\begin{pmatrix} \mathbf{E}_o^{(0)} \\ \mathbf{H}_o^{(0)} \end{pmatrix} = \frac{n(iq_o c_1 - c_2)}{\varepsilon_e c_2}\begin{pmatrix} (0, 0, 1) \\ (iq_o, -1, 0)n \end{pmatrix}. \tag{51}$$

In the domain of surface polaritons specified by condition (22), when $\varepsilon_o < 0$, extraordinary wave (46)–(48) is a single bulk wave. In this region, the bulk ordinary wave is impossible in principle since $q_o \geq 1$ (25) in all cases.

At the same time, in the region outside domain (22), where $\varepsilon_o > 0$, the bulk extraordinary wave (46)–(48) continues to exist. In addition, another bulk wave, viz., an ordinary wave, exists outside region (22). This is a one-partial wave satisfying the boundary conditions at the metallized surface in the case when the optical axis of the crystal is parallel to this surface: $\theta = 0$. In this case, the optical axis can form an arbitrary angle φ with the direction of propagation. For this wave, we have

$$\mathbf{c} = (\cos\varphi, 0, \sin\varphi), \quad \mathbf{n}_o = (1, 0, 0)n, \quad n^2 = \varepsilon_o, \tag{52}$$

and the polarization is determined by relation (48), in which the indices should be changed (e $\longrightarrow$ o). It is important that this bulk mode does not transform into a "quasi-bulk" polariton as the optical axis slightly deviates from the surface plane, but either disappears or becomes a wave field component in the reflection problem (depending on the relative values of ε_o and ε_e). In the limit of an isotropic medium ($\varepsilon_e = \varepsilon_o > 0$, $p = 0$), the expressions for $\mathbf{n}_e$ (46) and n^2 (47) coincide with the corresponding relations (52). This means that the bulk wave described by formulas (48) and (52) can propagate in an arbitrary direction along the plane surface of an isotropic solid.

7. FREQUENCY DISPERSION IN MODEL DESCRIPTION

In the following analysis, we will use a simple model of an isolated polar excitation [2, 15], assuming that function $\varepsilon_o(\omega)$ is described by the formula

$$\varepsilon_o = \varepsilon_\infty + \frac{a\omega_{TO}^2}{\omega_{TO}^2 - \omega^2}. \tag{53}$$

Here, a is the oscillator force and ω_{TO} is the frequency of transverse optical (TO) phonons, for which

$$\varepsilon_o(\omega_{TO} \pm 0) = \mp\infty.$$

For simplicity, we assume that $\varepsilon_e = \text{const} > 0$. In the model under investigation, the frequency domain of

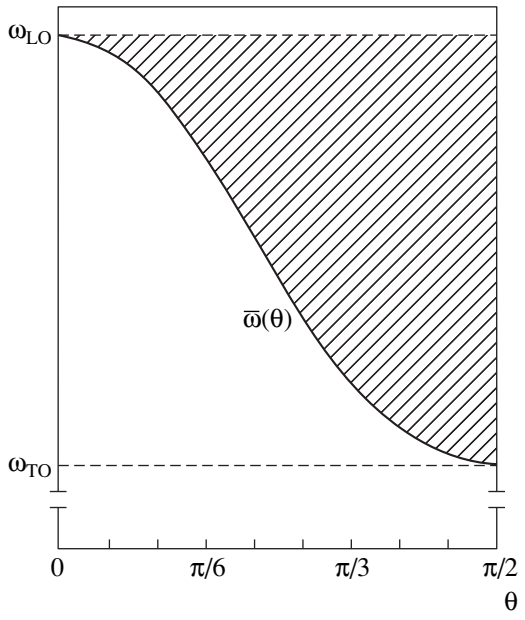


Fig. 3. Domain of surface polaritons in the ω vs. θ coordinates (hatched) in the model of a single polar excitation (we assume that $\varepsilon_\infty/a = 1.2$ and $\varepsilon_o/a = 1.8$).

polaritons can easily be determined. Substituting relation (53) into (22), we obtain

$$\bar{\omega}(\theta) \equiv \omega_{TO}\left(1 + \frac{a}{\varepsilon_\infty + \varepsilon_e \tan^2\theta}\right)^{1/2} < \omega < \omega_{TO}\left(1 + \frac{a}{\varepsilon_\infty}\right)^{1/2} \equiv \omega_{LO}. \quad (54)$$

Here, ω_{LO} is the frequency of longitudinal optical (LO) phonons, for which $\varepsilon_o(\omega_{LO}) = 0$. This frequency corresponds to the upper boundary of the region of negative values of ε_o. It is essential that the lower frequency boundary $\bar{\omega}$ of the polariton domain (54) depend on the angle θ formed by the optical axis with the surface (Fig. 3).

For $\theta \longrightarrow 0$, expression for $\bar{\omega}(0)$ coincides with the LO phonon frequency ($\bar{\omega} = \omega_{LO}$) so that the lower frequency boundary of the polariton domain approaches the upper boundary upon a decrease in the angle between the optical axis and the surface, and the frequency interval in which it can propagate vanishes. On the contrary, for an optical axis oriented orthogonally to the surface ($\theta = \pi/2$), the frequency domain of polaritons is the broadest ($\omega_{TO} < \omega < \omega_{LO}$). In other words, the larger the angle of inclination θ, the broader the frequency domain of polaritons (see Fig. 3). On the other hand, the same figure shows that the higher the frequency ω in the interval (54), the larger the admissible range of variation of angle θ.

Let us consider typical examples of characteristic frequencies determining the domain (54) of polaritons in optically uniaxial media. Thus, in accordance with [2], $\omega_{TO} = 769$ cm^{-1} and $\omega_{LO} = 832$ cm^{-1} for α-$LiIO_3$ crystals and $\omega_{TO} = 431$ cm^{-1} and $\omega_{LO} = 450$ cm^{-1} for $LiNbO_3$ crystals.

8. DISCUSSION

The above analysis was based on the assumption that metallization of the surface ensures complete confinement of electromagnetic fields in the crystal. This is so as long as the thickness h of the metallic coating considerably exceeds the characteristic depth d of wave field penetration in the crystal ($h \gg d$). It is well known that depth d is the smaller, the larger the imaginary part of the refractive index of the metal (corresponding estimates for many metals are given in [16]). For example, the penetration depth for copper is $d = 6.2 \times 10^{-8}$ cm for a wavelength of $\lambda = 10^{-5}$ cm (ultraviolet range), while $d = 6.2 \times 10^{-7}$ cm for $\lambda = 10^{-3}$ (infrared range); i.e., $d \ll \lambda$. Thus, the condition $h \gg d$ can easily be realized. At the same time, if the thickness of the coating is comparable to the penetration depth ($h \sim d$), such a coating becomes transparent for wave fields (specific features of such a situation are considered in [17]).

In the absence of metallization of the surface, the geometry and condition for propagation of localized wave fields radically differ from the situation considered above. In this case, a surface polariton in the crystal is accompanied by a localized wave in the contacting medium on the other side of the interface. It was noted in the Introduction that the theory of such polaritons becomes much more cumbersome and, in contrast to the case considered here, does not permit simple analytic solutions for general position orientations. This naturally does not imply that the corresponding conditions for the existence of localized natural waves in such media are more stringent than for crystals with metallized surfaces. On the contrary, dispersionless localized solutions [5, 6] can exist in uniaxial (and even biaxial) crystals along with dispersion polaritons [3] for positive components of the crystal permittivity exhibiting a weak dependence on frequency. For example, according to [5], an entire sector of allowed directions of propagation of dispersionless surface waves exists on the surface of a uniaxial crystal, which is parallel to the optical axis ($\theta = 0$) provided that $\varepsilon_e > \varepsilon > \varepsilon_o > 0$ (ε is the permittivity of the contacting medium). It was shown above that such solutions do not exist on a metallized surface.

ACKNOWLEDGMENTS

The authors are grateful to L.M. Barkovsky and A.N. Furs for information on the results of investigations of related problems.

This study was carried out in cooperation between the Institute of Crystallography, Russian Academy of Sciences, and Kielce University of Technology (Poland).

V.N.L. is grateful to the Russian Foundation for Basic Research (project no. 03-02-16871), and V.I.A. thanks the Polish–Japanese Institute of Information Technologies, Warsaw (grant no. PJ/MKT/02/2005) for financial support.

REFERENCES

1. *Surface Polaritons: Electromagnetic Waves at Surfaces and Interfaces*, Ed. by V. M. Agranovich and D. L. Mills (North-Holland, Amsterdam, 1982; Nauka, Moscow, 1985).
2. N. L. Dmitruk, V. G. Litovchenko, and V. L. Strizhevskiĭ, *Surface Polaritons in Semiconductors and Dielectrics* (Naukova Dumka, Kiev, 1989) [in Russian].
3. V. I. Alshits, V. N. Lyubimov, and L. A. Shuvalov, Fiz. Tverd. Tela (St. Petersburg) **43**, 1322 (2001) [Phys. Solid State **43**, 1377 (2001)].
4. F. N. Marchevskiĭ, V. L. Strizhevskiĭ, and S. V. Strizhevskiĭ, Fiz. Tverd. Tela (Leningrad) **26**, 1501 (1984) [Sov. Phys. Solid State **26**, 911 (1984)].
5. M. I. D'yakonov, Zh. Éksp. Teor. Fiz. **94** (4), 119 (1988) [Sov. Phys. JETP **67**, 714 (1988)].
6. V. I. Alshits and V. N. Lyubimov, Fiz. Tverd. Tela (St. Petersburg) **44**, 371 (2002) [Phys. Solid State **44**, 386 (2002)]; Fiz. Tverd. Tela (St. Petersburg) **44**, 1895 (2002) [Phys. Solid State **44**, 1988 (2002)].
7. A. N. Furs and L. M. Barkovsky, J. Opt. A: Pure Appl. Opt. **1**, 109 (1999).
8. V. N. Lyubimov and D. G. Sannikov, Fiz. Tverd. Tela (Leningrad) **15**, 1851 (1973) [Sov. Phys. Solid State **15**, 1234 (1973)]; **17**, 478 (1975) [**17**, 300 (1975)].
9. J. Lothe and V. I. Alshits, Kristallografiya **22**, 906 (1977) [Sov. Phys. Crystallogr. **22**, 519 (1977)].
10. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 3rd ed. (Nauka, Moscow, 1992; Pergamon, New York, 1984).
11. F. I. Fedorov and V. V. Filippov, *Reflection and Refraction of Light by Transparent Crystals* (Nauka i Tekhnika, Minsk, 1976) [in Russian].
12. F. I. Fedorov, *Theory of Gyrotropy* (Nauka i Tekhnika, Minsk, 1976) [in Russian].
13. V. M. Galynsky, A. N. Furs, and L. M. Barkovsky, J. Phys. A: Math. Gen. **37**, 5083 (2004).
14. A. N. Furs and L. M. Barkovsky, Microwave Opt. Technol. Lett. **14**, 301 (1997).
15. V. M. Agranovich and V. L. Ginzburg, *Crystal Optics with Spatial Dispersion, and Excitons*, 2nd ed. (Nauka, Moscow, 1979; Springer, New York, 1984).
16. M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon, Oxford, 1969; Nauka, Moscow, 1970).
17. K. P. Lyshenyuk and F. N. Marchevskiĭ, Fiz. Tverd. Tela (St. Petersburg) **37**, 211 (1995) [Phys. Solid State **37**, 116 (1995)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 788–794.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 913–921.
Original Russian Text Copyright © 2005 by Romanenko, Yatsenko.

ATOMS, MOLECULES, OPTICS

Stimulated Raman Adiabatic Passage in Fields with Stochastic Amplitudes

V. I. Romanenko* and L. P. Yatsenko

Institute of Physics, National Academy of Sciences of Ukraine, Kiev, 03028 Ukraine
**e-mail: vr@iop.kiev.ua; victor_romanenko@list.ru*

Received June 23, 2005

Abstract—A theoretical analysis is presented of the effect of correlation between fluctuations of laser pulse amplitudes on population transfer between the states of a three-level atom coupled by the laser field. The carrier frequencies of the pulses are tuned to resonance with the transitions between the ground and excited states, $|1\rangle$ and $|2\rangle$, and the excited and metastable states, $|2\rangle$ and $|3\rangle$, in a lambda-type configuration. The laser pulses are timed so that population transfer between states $|1\rangle$ and $|3\rangle$ is made possible by stimulated Raman adiabatic passage (STIRAP) in the absence of fluctuations. STIRAP does not occur when the laser fields are not correlated. When the fluctuations of one pulse amplitude duplicate those of the other, STIRAP can be observed for pulse amplitudes larger than those required in the absence of fluctuations. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

One promising method for control of internal states of atoms and molecules is stimulated Raman adiabatic passage (STIRAP), which has been the subject of numerous theoretical and experimental studies (e.g., see the review in [1]). This phenomenon is observed in an atom or molecule interacting with two temporally overlapping laser pulses. (In what follows, an atom or molecule is referred to as an atom.) In the simplest case, STIRAP can be described by analyzing the three-level atom schematized in Fig. 1 (lambda-configuration). Initially, the atom is in a stable or metastable state $|1\rangle$. The pump pulse couples this state to an excited state $|2\rangle$, and the latter is then coupled to a metastable state $|3\rangle$ by the Stokes pulse. Population transfer from $|1\rangle$ to $|3\rangle$ is observed if the interaction of the atom with the Stokes pulse precedes its interaction with the pump pulse. It is important that the simultaneous interaction of the atom with both pulses during a certain time interval is followed by its interaction with the pump pulse only.

Population transfer in a three-level atom interacting with two pulses is explained by the fact that one of the eigenstates of the corresponding Hamiltonian (dark state) is a linear combination of the initial and final states, $|1\rangle$ and $|3\rangle$. If the field variation is sufficiently slow, then transition from state $|1\rangle$ to state $|3\rangle$ occurs in an adiabatic process of atom–field interaction that does not involve significant population of the excited state $|2\rangle$. Thus, population transfer via STIRAP is almost insensitive to spontaneous decay of the excited state. This is important for experiments on atomic beams, which are generally characterized by atom–field interaction times much longer than the excited-state lifetimes.

The efficiency of STIRAP transfer depends on how close it is to a perfectly adiabatic process and on the precision of maintaining the two-photon resonance condition in which the dark state is an eigenstate of the Hamiltonian of an atom interacting with electromagnetic field. The dependence of population-transfer efficiency on the carrier-frequency detuning from two-photon resonance (two-photon line shape) was analyzed in [2–5]. Uncontrollable detuning from two-photon resonance is caused by fluctuations of laser frequencies. Their effect on population-transfer efficiency was studied in [5–7]. The effect of field amplitude fluctuations on population transfer via STIRAP has never been investigated. This problem is addressed in the present study.

2. BASIC EQUATIONS

To separate the effect of noise on the population-transfer efficiency from the effect of carrier-frequency

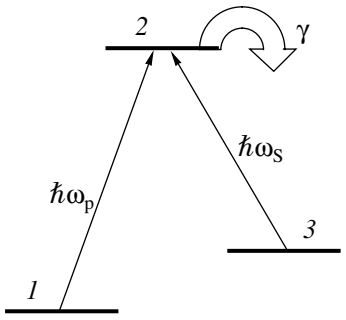


Fig. 1. Schematic diagram of interaction between three-level atom and laser pulses. The atom interacts first with a Stokes pulse characterized by carrier frequency ω_S.

1063-7761/05/10105-0788$26.00 © 2005 Pleiades Publishing, Inc.

detuning (which was analyzed in [2–5]), we consider the case when the pump and Stokes carrier frequencies, ω_p and ω_S, are resonant with the $|1\rangle \longrightarrow |2\rangle$ and $|3\rangle \longrightarrow |2\rangle$ transition frequencies, respectively. The electric field interacting with the atom can be written as

$$E = E_p(t)\exp(-i\omega_p t) + E_S(t)\exp(-i\omega_S t) + \text{c.c.} \quad (1)$$

Assuming that the probability of transition from state $|2\rangle$ to state $|1\rangle$ or $|3\rangle$ via spontaneous decay is negligible, we can represent the wavefunction of the atom as

$$\Psi = [c_1(t), c_2(t), c_3(t)]^T.$$

The time variation of the probability amplitude $c_k(t)$ of state $|k\rangle$ (k = 1, 2, 3) is described by the Schrödinger equation with the Hamiltonian

$$H = \frac{\hbar}{2}\begin{bmatrix} 0 & \Omega_p(t) & 0 \\ \Omega_p^*(t) & -i\gamma & \Omega_S^*(t) \\ 0 & \Omega_S(t) & 0 \end{bmatrix}, \quad (2)$$

where

$$\Omega_p(t) = -d_{12}E_p(t)/\hbar, \quad \Omega_S(t) = -d_{32}E_S(t)/\hbar$$

are the Rabi frequencies characterizing the interactions between the atom and the pump and Stokes pulses, respectively, and γ is the rate of spontaneous decay of the excited state into states other than $|1\rangle$ and $|3\rangle$.

3. PULSE SHAPES

Analytical calculations and numerical simulations are performed for two combinations of laser pulse shapes:

$$\Omega_p(t) = \begin{cases} \Omega_0 f_p(t)\sin(\pi t/\tau), & 0 \le t \le \tau, \\ 0, & \text{otherwise}, \end{cases}$$
$$\Omega_S(t) = \begin{cases} \Omega_0 f_S(t)\cos(\pi t/\tau), & -\frac{1}{2}\tau \le t \le \frac{1}{2}\tau, \\ 0, & \text{otherwise}, \end{cases} \quad (3)$$

and

$$\Omega_p(t) = \Omega_0 f_p(t)\exp\left(-\frac{(t-t_d/2)^2}{\tau^2}\right),$$
$$\Omega_S(t) = \Omega_0 f_S(t)\exp\left(-\frac{(t+t_d/2)^2}{\tau^2}\right). \quad (4)$$

Here, Ω_0 is the maximum pulse amplitude in the absence of fluctuations, τ is pulse duration, t_d is the delay of the pump pulse relative to the Stokes pulse ($\tau/2$ for pulse shape (3)), and $f_p(t)$ and $f_S(t)$ represent amplitude fluctuations. In the absence of fluctuations,

$$f_p(t) = f_S(t) = 1.$$

We consider two models of envelope amplitude fluctuations of the laser pulses [8]: a Gaussian model with real $f_p(t)$ and $f_S(t)$ characterized by correlation functions

$$\langle f_n(t)\rangle = 0,$$
$$\langle f_n(t) f_n(t')\rangle = \exp(-G|t-t'|) \quad (5)$$

(n = p, S) and a chaotic field model with $f_n(t)$ having fluctuating real and imaginary parts,

$$\langle f_n(t)\rangle = 0, \quad \langle \mathrm{Re}\, f_n(t)\,\mathrm{Im}\, f_n(t')\rangle = 0,$$
$$\langle \mathrm{Re}\, f_n(t)\,\mathrm{Re}\, f_n(t')\rangle = \frac{1}{2}\exp(-G|t-t'|), \quad (6)$$
$$\langle \mathrm{Im}\, f_n(t)\,\mathrm{Im}\, f_n(t')\rangle = \frac{1}{2}\exp(-G|t-t'|),$$

where angle brackets denote ensemble averages. The factor 1/2 on the right-hand side of (6) is introduced to ensure that the average values of $|f_n(t)|^2$ in models (5) and (6) are equal. Chaotic field model (6) is a realistic representation of a multimode laser field [8].

The correlation between the Stokes and pump fields is assumed to be such that $f_p(t)$ is a time-delayed copy of $f_S(t)$:

$$f_p(t) = f_S(t - t_N). \quad (7)$$

The limit cases of $t_N = 0$ and $t_N = \infty$ correspond to perfectly correlated and mutually independent pulse fluctuations.

4. IMPLEMENTATION OF CORRELATED FLUCTUATIONS

As an example of Stokes and pump fields with correlated fluctuations, we consider two circularly polarized pulses generated by the same laser that couple an atomic state with total angular momentum $J = 1$ and magnetic quantum number $M = 1$ to the excited state with $J = 0$ and the excited state to the state with $J = 1$ and $M = -1$. The atom moves across parallel laser beams, first interacting with the Stokes pulse, which induces $J = 0 \longleftrightarrow J = 1$, $M = -1$ transitions. This interaction facilitates population transfer from the state with $J = 1$ and $M = 1$ to the state with $J = 1$ and $M = -1$, which must be depopulated by optical pumping before the interaction. Any desired delay t_N is easy to implement

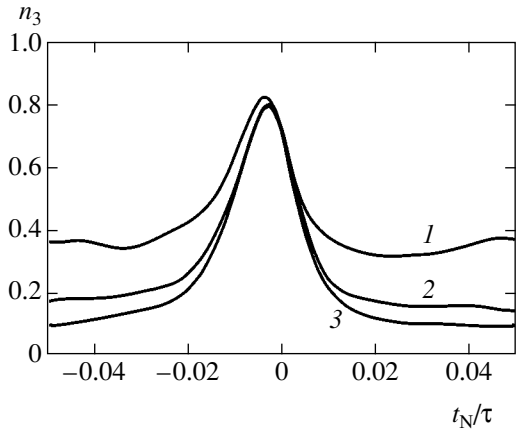


Fig. 2. Population of target state $|3\rangle$ vs. normalized delay t_N between pump and Stokes pulse fluctuations obtained by solving the Schrödinger equation for pulse shape (3) and averaging over 100 realizations of Gaussian amplitude fluctuations. Parameters: $\Omega_0\tau = 100$ and $G\tau = 50$ for all curves; $\gamma\tau = 0$, 10, and 20 for curves *1*, *2*, and *3*, respectively.

by adjusting the path difference between the beams prior to the atom–field interaction.

5. NUMERICAL SIMULATION OF FLUCTUATIONS

Sequences of values $\xi(t_j)$ of a random variable were generated to simulate either the real part of $f_S(t_j)$ (in the Gaussian model) or its real and imaginary parts (in the chaotic field model) at instants $t_j = t_{j-1} + \Delta t$ by using the colored-noise algorithm proposed in [9, 10]:

$$\xi(t_{j+1}) = \xi(t_j)\exp(-G\Delta t) + h(t_j), \tag{8}$$

where $h(t_j)$ has a zero-mean Gaussian distribution with

$$\langle h(t_j)^2\rangle = (1 - e^{-2G\Delta t}) \tag{9}$$

in the Gaussian model and

$$\langle h(t_j)^2\rangle = \frac{1}{2}(1 - e^{-2G\Delta t}) \tag{10}$$

in the chaotic field model. The sequence $h(t_j)$ is generated by using the standard Matlab function *randn*.

5.1. Numerical Example

Figure 2 shows the population of the target state $|3\rangle$ versus the delay t_N between pump and Stokes pulse fluctuations computed by using Gaussian model (5), in which the amplitude of the field fluctuates while its phase remains constant. When $|t_N|$ exceeds the correlation time $1/G$, the population-transfer efficiency (i.e., the population n_3 of state $|3\rangle$) substantially decreases, and this decrease is more pronounced for higher rates of spontaneous decay. Note that the maximum population-transfer efficiency corresponds to a delay much smaller than $1/G$, rather than to fully correlated pulse fluctuations, and the maximum values of n_3 are almost equal for different γ. This implies insignificant population of the excited state during the atom–field interaction time, which is characteristic of STIRAP processes.

6. GAUSSIAN MODEL OF AMPLITUDE FLUCTUATIONS

Since the phases of both fields remain constant according to (5), we can set them to zero. Assuming that both pump and Stokes field fluctuations are perfectly correlated ($t_N = 0$), we can find an expression for the efficiency of population transfer from $|1\rangle$ to $|3\rangle$ in the case when the decay time of state $|2\rangle$ is much shorter than the atom–field interaction time. The resulting efficiency is lower than the maximum efficiency attained by adjusting t_N. To simplify analysis, we represent the Rabi frequencies $\Omega_p(t)$ and $\Omega_S(t)$ in terms of the rms Rabi frequency $\Omega_m(t)$ and the mixing angle $\theta(t)$:

$$\begin{aligned}\Omega_p(t) &= \Omega_m(t)\sin\theta(t),\\ \Omega_S(t) &= \Omega_m(t)\cos\theta(t).\end{aligned} \tag{11}$$

Both $\Omega_m(t)$ and $\theta(t)$ are real quantities. Changing from the bare-state basis to the dressed-state basis consisting of the excited state $\varphi_2 = |2\rangle$, the bright state φ_b, and the dark state φ_d, where

$$\begin{aligned}\varphi_b &= \sin\theta(t)|1\rangle + \cos\theta(t)|3\rangle,\\ \varphi_d &= \cos\theta(t)|1\rangle - \sin\theta(t)|3\rangle,\end{aligned} \tag{12}$$

we represent the dressed wavefunction as

$$\Psi = [B_b(t), B_2(t), B_d(t)]^T,$$

where $B_k(t)$ (k = b, 2, d) is the probability amplitude of the state φ_k. In this basis, the evolution of the atom is governed by the Hamiltonian

$$H_d = \frac{\hbar}{2}\begin{bmatrix} 0 & \Omega_m(t) & 2i\dot\theta(t)\\ \Omega_m(t) & -i\gamma & 0\\ 2i\dot\theta(t) & 0 & 0\end{bmatrix}. \tag{13}$$

Before the interaction with the laser beams, the atom is assumed to be in state $|1\rangle$. As the atom successively interacts with the Stokes pulse, the combination of the Stokes and pump pulses, and the pump pulse alone in the absence of amplitude fluctuations, the mixing angle monotonically varies with time from 0 to $\pi/2$:

$$\theta(t) = \arctan(\Omega_p(t)/\Omega_S(t)).$$

Before the interaction, only the dark state φ_d (i.e., $|1\rangle$) is populated. When $\Omega_0\tau \gg 1$ (the process is nearly adiabatic in the absence of amplitude fluctuations), the time

derivative in (13) is negligible, and the population of the dark state φ_d remains constant during the atom–field interaction time. After the atom–field interaction, φ_d reduces to $|3\rangle$; i.e., 100% of the population is transferred from $|1\rangle$ to $|3\rangle$. When the derivative of $\theta(t)$ is taken into account in Hamiltonian (13), the states $|2\rangle$ and $|b\rangle$ are populated; i.e., the population-transfer efficiency is lower. A theory of population transfer driven by nonfluctuating fields can be developed by using the series expansion of the wavefunction in terms of the small parameter $\dot{\theta}(t)\tau$. A perturbation theory in terms of $\dot{\theta}(t)\tau$ can also be developed when the amplitude fluctuations are perfectly correlated ($f_p(t) = f_S(t)$) and $\theta(t)$ is independent of fluctuating parameters.

Hereinafter, we consider the case of fully correlated fluctuations. Accordingly, the derivative $\dot{\theta}(t)$ can be neglected in Hamiltonian (13), and its eigenvalues λ_n and eigenfunctions χ_n are

$$\lambda_1 = \frac{1}{2}\Omega_m(t), \quad \chi_1 = \frac{1}{\sqrt{2}}(\varphi_2 + \varphi_b),$$
$$\lambda_2 = -\frac{1}{2}\Omega_m(t), \quad \chi_2 = \frac{1}{\sqrt{2}}(\varphi_2 - \varphi_b), \qquad (14)$$
$$\lambda_3 = 0, \quad \chi_3 = \varphi_d.$$

Owing to the perfect correlation between the Stokes and pump field fluctuations, the current rms Rabi frequency $\Omega_m(t)$ has a zero-mean Gaussian distribution. Therefore, the adiabaticity criterion

$$|\lambda_3 - \lambda_n|\tau_{corr} \gg 1, \quad n = 1, 2,$$

cannot be satisfied during the atom–field interaction time ($\tau_{corr} = 1/G$ is the autocorrelation time of field fluctuations). This leads to loss in the dark-state population and ensuing decrease in the population-transfer efficiency. When the pulse fields fluctuate independently, the value of $|\lambda_3 - \lambda_n|\tau_{corr}$ is much larger, because the probability that two independent variables nearly vanish simultaneously is much lower than the analogous probability for one variable. However, since $\theta(t)$ fluctuates with a characteristic time τ_{corr} in this case, the dark-state population also decreases. Figure 2 demonstrates that the ensuing population leakage is larger than that in the case of fully correlated fluctuations. A small delay ($t_N \ll \tau_{corr}$) can play a positive role: it causes $|\lambda_3 - \lambda_n|$ to increase, while the fluctuations of $\theta(t)$ are insignificant (except for the time intervals when $\Omega_p(t)$ or $\Omega_S(t)$ approaches zero). This may explain the shift of the peaks relative to the origin in Fig. 2.

We define

$$\Phi(t) = \frac{1}{2}\int_{-\infty}^{t}\Omega_m(t')dt' \qquad (15)$$

and

$$a_d(t) = \ln(B_d(t)),$$
$$a_2(t) = B_2(t)/B_d(t), \qquad (16)$$
$$a_b(t) = B_b(t)/B_d(t).$$

Then, introducing a small parameter ε to reflect the slow variation of $\theta(t)$, we obtain equations for $a_b(t)$, $a_2(t)$, and $a_d(t)$:

$$\dot{a}_b(t) = \varepsilon\dot{\theta}(t) - \dot{a}_d(t)a_b(t) - i\dot{\Phi}(t)a_2(t),$$
$$\dot{a}_2(t) = -\dot{a}_d(t)a_2(t) - i\dot{\Phi}(t)a_b(t) - \frac{\gamma}{2}a_2(t), \qquad (17)$$
$$\dot{a}_d(t) = -\varepsilon\dot{\theta}(t)a_b(t).$$

These equations can be solved perturbatively by representing $a_b(t)$, $a_2(t)$, and $a_d(t)$ as series expansions in ε and using the fact that only the dark state is populated before the atom–field interaction. For arbitrary pulse shapes, as well as for a two-level atom interacting with a pulse with resonant carrier frequency [11], a solution can be found only when spontaneous decay of the excited state is negligible. Setting $\gamma = 0$, we solve (17) to second order in ε to obtain

$$\dot{a}_d(t) = -\varepsilon^2\dot{\theta}(t)\int_{-\infty}^{t}\dot{\theta}(t')\cos(\Phi(t) - \Phi(t'))dt'. \qquad (18)$$

We assume that the correlation time $1/G$ of amplitude fluctuations is small as compared to the pulse duration τ. The opposite case ($1/G \gg \tau$), when the amplitudes do not fluctuate during the atom–field interaction time, but fluctuate from pulse to pulse, and the problem is solved by averaging the population-transfer efficiency over the pulse-amplitude distribution, is not considered here.

Since

$$\langle \exp(i\xi(t))\rangle = \exp\left(-\frac{1}{2}\langle\xi(t)\rangle^2\right) \qquad (19)$$

for any zero-mean Gaussian process $\xi(t)$ (see [12]), we can use (5) and the assumption that $G\tau \gg 1$ to write

$$\langle a_d(t)\rangle = -\int_{-\infty}^{t}\dot{\theta}(t')\int_{-\infty}^{t'}\dot{\theta}(t'')$$
$$\times \exp\left(-\frac{1}{8G}\int_{t''}^{t'}\langle\Omega_m(t''')^2\rangle\right. \qquad (20)$$
$$\left.\times[2 - \exp(G(t'' - t''')) - \exp(G(t''' - t'))]dt'''\right)dt''dt',$$

setting $\varepsilon = 1$. Since the dark-state population equals the population of state $|3\rangle$ after the atom–field interaction,

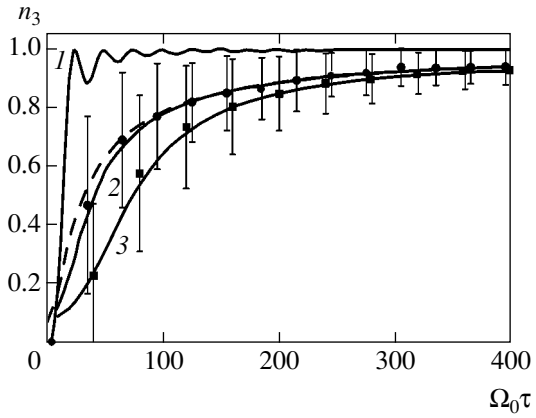


Fig. 3. Population of target state $|3\rangle$ vs. pulse fluence represented by $\Omega_0\tau$ obtained by solving the Schrödinger equation for pulse shape (3) and averaging over 100 realizations of Gaussian amplitude fluctuations (circles and squares) and by calculating formula (22) (curves *2* and *3*). Error bars represent rms deviations of transfer efficiencies from their mean values. Curve *1* corresponds to nonfluctuating amplitudes. Parameters: $\gamma = 0$ for all curves; $G\tau = 20$ and 50 for curves *2* (circles) and *3* (squares), respectively. Dashed curve corresponds to Eq. (24).

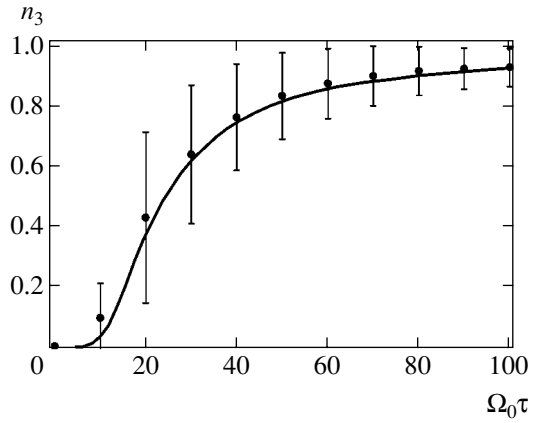


Fig. 4. Population of target state $|3\rangle$ vs. pulse fluence represented by $\Omega_0\tau$ obtained by solving the Schrödinger equation for pulse shape (4) and averaging over 100 realizations of Gaussian amplitude fluctuations (circles) and by calculating formulas (21) and (26) (solid curve). Error bars represent rms deviations of transfer efficiencies from their mean values. Parameters: $\gamma = 0$, $G\tau = 50$, and $\tau_d = \tau$.

the efficiency of population transfer from $|1\rangle$ to $|3\rangle$ is expressed as

$$\eta = \langle \exp(2a_d(\infty)) \rangle.$$

Invoking the expansion

$$\langle \exp(X) \rangle = \exp(\langle X \rangle)\left(1 + \frac{1}{2}\langle (X - \langle X \rangle)^2 \rangle + \ldots\right),$$

we calculate η as

$$\eta = \exp(2\langle a_d(\infty) \rangle). \quad (21)$$

This approximation is accurate when the rms deviation of $a_d(\infty)$ from its mean value is relatively small. It was used in [7] to analyze the effect of frequency fluctuations on population transfer via STIRAP. Expressions (20) and (21) can be used to find the efficiency of population transfer driven by pulses with fluctuating amplitudes and arbitrary time-dependent ensemble-averaged intensities in the Gaussian model of amplitude fluctuations.

For pulse shape (3), expression (20) can be substantially simplified. Performing the integrals, taking into account the assumption that $G\tau \gg 1$ used to obtain this expression, and substituting $a_d(t)$ into (21), we obtain

$$\eta = \exp\left\{-\frac{\pi^2}{(G\tau)^2}\left[\frac{e^z}{z}(G\tau - 2)(\Gamma(z) - \Gamma(z, z)) + \frac{2}{z^2}(e^{z(1 - G\tau/2)} - 1)\right]\right\}, \quad (22)$$

where $z = \Omega_0^2/(2G)^2$ and

$$\Gamma(x, y) = \int_y^\infty \exp(-t)t^{x-1}dt \quad (23)$$

is the upper incomplete gamma function. In the limit case of $\Omega_0 \gg G$, we have

$$\eta = \exp\left[\frac{8\pi^2}{\Omega_0^2\tau^2} - \frac{\pi^2\sqrt{2\pi}}{\Omega_0\tau}\right], \quad (24)$$

because the dominant contributions to expression (20) in this limit correspond to the values of t''' such that

$$G(t''' - t'') \ll 1, \quad G(t' - t''') \ll 1.$$

In the absence of amplitude fluctuations (when $f_p(t) = f_S(t) = 1$), the population-transfer efficiency is found by using (18):

$$\eta_0 = \exp\left[-\frac{8\pi^2}{\Omega_0^2\tau^2}\left(1 - \cos\frac{\Omega_0\tau}{4}\right)\right]. \quad (25)$$

Comparing (25) with (24), we see that the limit behavior of the logarithm of the population-transfer efficiency at large $\Omega_0\tau$ changes from $(\Omega_0\tau)^{-2}$ to $(\Omega_0\tau)^{-1}$ with increasing frequency of pulse-amplitude fluctuations.

Figure 3 compares the populations of state $|3\rangle$ versus $\Omega_0\tau$ obtained by solving the Schrödinger equation for

pulse shape (3) and averaging over 100 realizations for several values of G^{-1} in the Gaussian model with those predicted by calculating (22). Curve *1* is obtained in the absence of pulse amplitude fluctuations. Its shape is in almost perfect agreement with (25) at $\Omega_0\tau$ above $\Omega_0\tau \approx 20$. The results of Monte Carlo computations are consistent with predictions based on formula (22). It is clear that amplitude fluctuations substantially reduce the population-transfer efficiency as compared to that in the absence of fluctuations, but this negative effect can be compensated for by increasing the pulse intensity. As the pulse intensity increases, the transfer efficiency approaches the dashed curve predicted by (24) and follows this curve at $\Omega_0 > 4G$.

For pulse shapes (4), direct application of expression (20) involves extensive computations, which are required to evaluate an integral in three variables. However, the integral in the exponent of (20) can be calculated by noting that the dominant contributions correspond to values of t', t'', and t''' that differ by intervals on the order of $1/G$:

$$\langle a_d(t)\rangle = -\int_{-\infty}^{t}\dot{\theta}(t')^2\int_{-\infty}^{t'}\exp\left\{-\frac{1}{4G^2}\langle\Omega_m(t')^2\rangle\right.$$
$$\left.\times[-1+G(t'-t'')-\exp(-G(t'-t''))]\right\}dt''dt'. \quad (26)$$

This expression is not valid for small values of $\Omega_0^2\tau/G$, but this case is of little importance, because the corresponding transfer efficiency is low.

Figure 4 compares the population of state $|3\rangle$ versus $\Omega_0\tau$ obtained by solving the Schrödinger equation for Gaussian pulse shapes (4) in the Gaussian model with those calculated by using (21) and (26). The figure demonstrates that the population-transfer efficiencies predicted by formulas (21) and (26) are in good agreement with the results of Monte Carlo computations. This suggests that formulas (21) and (26) can also be used to evaluate the population-transfer efficiency for non-Gaussian pulses with smooth ensemble-averaged envelopes.

7. CHAOTIC FIELD MODEL

In the chaotic field model, when the real and imaginary parts of the field amplitude fluctuate independently, a theory analogous to that presented above for the Gaussian model cannot be developed, mainly because of the difficulty of ensemble averaging of two independent Gaussian random variables.

Figure 5 demonstrates that population-transfer efficiency higher than 90% can also be achieved for correlated Stokes and pump fields. The figure shows the

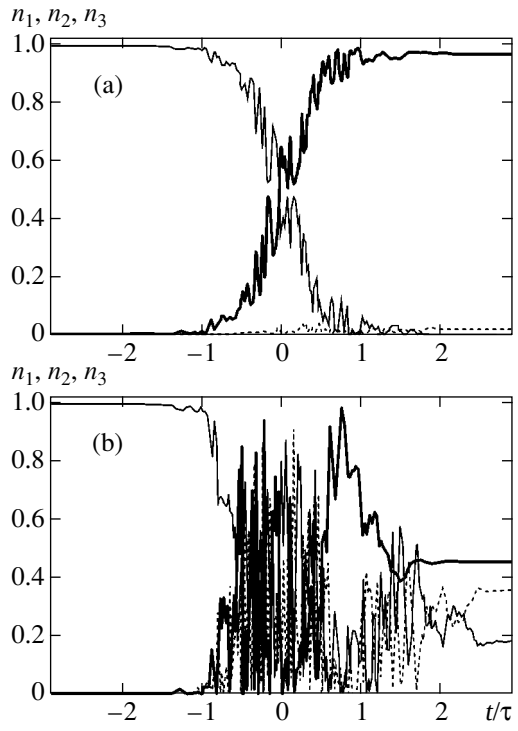


Fig. 5. Populations states $|1\rangle$ (thin curve), $|2\rangle$ (dotted curve), and $|3\rangle$ (thick curve) vs. $\Omega_0\tau$ obtained by solving the Schrödinger equation for pulse shape (4) in a single realization of the chaotic field model. Parameters: $\gamma = 0$, $t_d = \tau$, $G\tau = 50$, and $\Omega_0\tau = 100$; $t_N = 0$ (a) and $1/G$ (b).

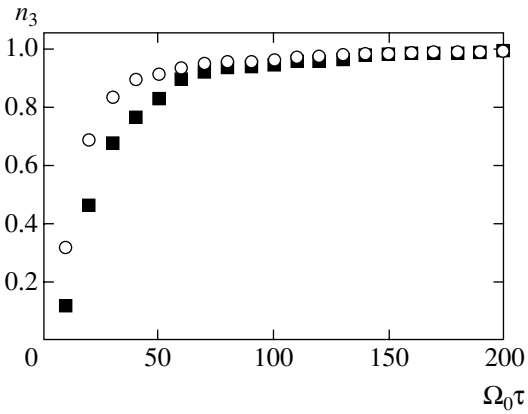


Fig. 6. Population of target state $|3\rangle$ vs. $\Omega_0\tau$ obtained by solving the Schrödinger equation for pulse shape (4) and averaging over 100 realizations in the chaotic field model. Parameters: $\gamma = 0$, $t_d = \tau$; $G\tau = 20$ (circles) and 50 (squares).

respective populations n_1, n_2, and n_3 of states $|1\rangle$, $|2\rangle$, and $|3\rangle$ predicted for fully correlated fluctuations (with $t_N = 0$) and for Stokes amplitude fluctuations shifted by the correlation time $1/G$ relative to pump fluctuations. It is obvious that the populations vary irregularly in the latter case, whereas the evolution of population in the

case of $t_N = 0$ is similar to STIRAP transfer in the absence of fluctuations [1]. Moreover, the excited-state population is very low when the field fluctuations are perfectly correlated, as in STIRAP processes driven by nonfluctuating fields, which makes STIRAP attractive as a mechanism of population transfer between metastable states.

Figure 6 shows the population n_3 of state $|3\rangle$ versus $\Omega_0\tau$ obtained by solving the Schrödinger equation for pulse shapes (4) and averaging over 100 realizations for fully correlated Stokes and pump field fluctuations in the chaotic field model and by calculating formula (22). As in the Gaussian model, the population-transfer efficiency substantially increases with the pulse fluence.

8. CONCLUSIONS

It is shown that laser amplitude fluctuations do not preclude population transfer between metastable states of atoms and molecules via STIRAP when the Stokes and pump fluctuations are perfectly correlated.

In the Gaussian model of amplitude fluctuations, the target state population is expressed in terms of pulse parameters when the atom–field interaction time is shorter than the spontaneous decay time. The predicted results are in good agreement with those of Monte Carlo computations.

In the chaotic field model (with independently fluctuating real and imaginary parts of the complex electric field amplitude), population transfer is analyzed numerically. The calculated curves of population-transfer efficiency versus pulse fluence are similar to those obtained in the Gaussian model of amplitude fluctuations.

ACKNOWLEDGMENTS

We thank K. Bergmann for fruitful discussion of the problem.

This work was supported by the National Academy of Sciences of Ukraine, project nos. VTs 93/24, V/112, and F7/445.

REFERENCES

1. K. Bergmann, H. Theur, and B. W. Shore, Rev. Mod. Phys. **70**, 1003 (1998).
2. M. V. Danileiko, V. I. Romanenko, and L. P. Yatsenko, Opt. Commun. **109**, 462 (1994).
3. V. I. Romanenko and L. P. Yatsenko, Opt. Commun. **140**, 231 (1997).
4. F. Renzoni, A. Lindner, and E. Arimondo, Phys. Rev. A **60**, 450 (1999).
5. V. I. Romanenko, L. P. Yatsenko, and L. Bergmann, Ukr. Fiz. Zh. **48**, 533 (2003).
6. A. Kuhn, S. Schiemann, G. Z. He, *et al.*, J. Chem. Phys. **96**, 4215 (1992).
7. L. P. Yatsenko, V. I. Romanenko, B. W. Shore, and K. Bergmann, Phys. Rev. A **65**, 043409 (2002).
8. A. T. Georges, Phys. Rev. A **21**, 2034 (1980).
9. R. F. Fox, I. R. Gatland, R. Roy, and G. Vemuri, Phys. Rev. A **38**, 5938 (1988).
10. G. Vemuri and R. Roy, Opt. Commun. **77**, 318 (1990).
11. B. W. Shore, *The Theory of Coherent Atomic Excitation* (Wiley, New York, 1990).
12. S. A. Akhmanov, Yu. E. D'yakov, and A. S. Chirkin, *Introduction to Statistical Radio Physics and Optics* (Nauka, Moscow, 1981) [in Russian].

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 795–810.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 922–937.
Original Russian Text Copyright © 2005 by Elesin.

ATOMS, MOLECULES, OPTICS

A Tunable Terahertz-Band Oscillator Based on a Two-Well Nanostructure with a Coherent Electron Subsystem

V. F. Elesin

Moscow State Engineering Physics Institute, Kashirskoe sh. 31, Moscow, 115409 Russia
e-mail: VEF@supercon.mephi.ru

Received June 27, 2005

Abstract—A theory of coherent resonance tunneling of electrons in a two-well nanostructure (TWNS) in the presence of a strong electromagnetic field is developed. The TWNS consists of two identical tunnel-coupled quantum wells to which a dc electric field is applied. Radiative transitions occur between two levels that arise due to the interwell interference and the dc electric field. The wavefunctions and polarization currents in the TWNS are found in the case of a strong electromagnetic field, and the oscillation power is determined as a function of the coherent pumping current and the parameters of the structure. It is shown that oscillations are possible in the relevant terahertz band, with fine frequency tuning by a dc field. It is found that the interference of electrons between quantum wells plays a crucial role. This interference significantly suppresses the effect of the electromagnetic field on the resonance tunneling and enhances the oscillation up to the highest possible level. It is proved that there exists an optimal regime of strong-field oscillations without inverse population and saturation, which are inherent in conventional lasers. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In 1971, Kazarinov and Suris [1] proposed a new type of a semiconductor laser in which radiative transitions occur between resonance levels (subbands) of a quantum well. The pumping of electrons to the upper level and their extraction from the lower level are performed by means of resonance tunneling. These lasers (which were called cascade lasers) were developed in 1994 [2].

In 1997, it was demonstrated in [3] that coherent resonance tunneling (coherent pumping) allows one to develop an essentially new type of a laser (following [3], we call it briefly a coherent laser) in which the electron subsystem, like the photon subsystem, is coherent. In such a laser, oscillations may occur without dissipative processes and are described by pure states of quantum mechanics. The coherent laser is an example of a quantum device in which the interference of electrons plays a fundamental role. The coherent laser has specific features and offers advantages over conventional lasers; these are oscillations without inverse population, high efficiency, the absence of the amplification-loop expansion due to a strong field, stability with respect to multimode oscillations and to the Coulomb interaction [3], and others.

The reason for the remarkable properties of the coherent laser is that electrons from the emitter are supplied by resonance tunneling (coherent pumping) into an optimal energy interval with optimal phases and are coherently taken to the collector. Under these conditions, each electron brought to the upper level of the quantum well makes a single radiative transition and is taken away to the collector from the lower level. Therefore, this laser is free of the saturation effect, which consists in the leveling of the overpopulation due to multiple transitions of electrons between energy levels and leads to a steady-state field of a laser. The saturation is responsible for many unwanted phenomena such as burning through a "hole" in the overpopulation, multimode oscillations, broadening of the laser linewidth, and the power limitation [4].

In coherent lasers, a steady-state field is obtained in a fundamentally different way, namely, due to the influence of the electromagnetic field on resonance tunneling. The field reduces the rate of resonance tunneling by changing the energy of the resonance levels. The absence of saturation removes the above-listed drawbacks of conventional lasers.

The main problem in the development of coherent lasers consists in the realization of the condition under which the electron subsystem in nanostructures is coherent. The current state of the art in technology can guarantee coherence over sufficiently large lengths (up to dozens of quantum wells (see [5])). It should also be emphasized that a simple condition of coherent tunneling $\tau_\Gamma < \tau_{ph}$ (τ_Γ is time during which an electron stays in a well and τ_{ph} is the decoherence time) may be too stringent. Indeed, according to the detailed theory, under certain conditions, the electron–phonon interaction does not influence the decay of Bloch oscillations [6], while the resonance tunneling remains coherent even for $\tau_\Gamma \gg \tau_{ph}$ [7]. Moreover, the situation strongly

1063-7761/05/10105-0795$26.00 © 2005 Pleiades Publishing, Inc.

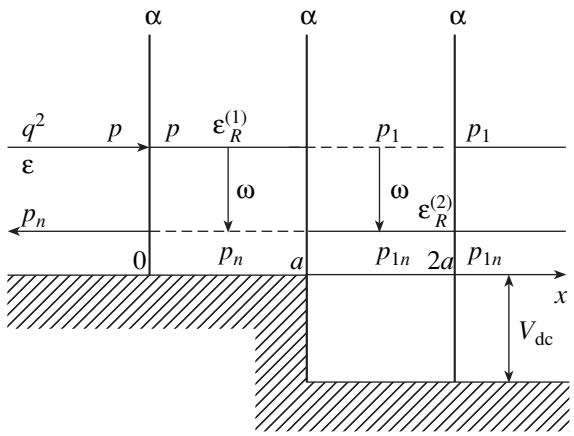


Figure.

depends on the frequency band of oscillations and on a specific nanostructure.

In the present paper, we study a two-well nanostructure (TWNS) that consists of two identical tunnel-coupled wells to which a dc electric voltage V_{dc} is applied. Radiative transitions occur between two levels of the TWNS that arise due to the interwell interference (the splitting of a level of an isolated well) and the dc electric voltage. Such a structure supports electromagnetic oscillations in the relevant terahertz band (which can hardly be attained by cascade lasers) with fine frequency tuning and facilitates the fulfillment of coherence conditions (since the oscillation frequency is less than the frequency of the optical phonon); moreover, it still offers the above-mentioned advantages that are inherent in the coherent laser.

In addition, the TWNS is also of interest as the first step from a single-well structure (which is usually called a resonance tunneling diode (RTD)) to the Bloch single-band superlattice.

In this paper, we develop a consistent theory of oscillations in the TWNS and find the wavefunctions and the polarization currents for a strong electromagnetic field, as well as the oscillation power as a function of the coherent pumping current and the parameters of the structure. We show that there exists an optimal regime of strong-field oscillations in which each electron emits a single photon (the efficiency equals one). The high efficiency at sufficiently low frequencies is attributed to the phenomenon, discovered in the present study, of significant suppression of the electromagnetic field's influence on the resonance tunneling due to the interwell interference and the dc voltage V_{dc}.

The possibility of oscillations in the TWNS (with different wells and $V_{dc} = 0$) was first pointed out in [8]. In [9], an analytic theory of a TWNS consisting of two identical wells and with $V_{dc} \neq 0$ was developed for a weak electromagnetic field, and it was shown that a linear amplification is much greater than the amplification in an RTD and that the TWNS allows for frequency tuning. The results of the numerical solution of the Schrödinger equation for this structure are in good agreement with the results of [9] that were published in [10].

2. STATEMENT OF THE PROBLEM AND BASIC EQUATIONS

We will study the oscillation of an electromagnetic field in a two-well nanostructure within the model of [9]. We consider a one-dimensional structure with δ-shaped barriers at the points $x = 0$, a, and $2a$ (see figure). A dc electric field V_{dc} is applied to the second well. A steady stream of electrons proportional to q^2 and with energy ε approximately equal to the energy $\varepsilon_R^{(1)}$ of the first well is incident on the first well from the left ($x \longrightarrow -\infty$).

An electromagnetic field

$$E_z(z, t) = E \sin kz \cos \omega t \tag{1}$$

with amplitude E, wavevector k, and frequency ω acts in the TWNS region. This field is polarized perpendicular to the plane of the well, while the wavevector is directed along the plane (along the z axis).

The wavefunction $\Psi(x, t)$ of the structure satisfies the Schrödinger equation

$$i\frac{\partial \Psi}{\partial t} = -\frac{\partial^2 \Psi}{\partial x^2} + \alpha[\delta(x) + \delta(x-a) + \delta(x-2a)]\Psi - V_{dc}\theta(x-a)\Psi + \hat{V}\Psi. \tag{2}$$

Here, α is the "power" of the barriers (the product of the barrier height multiplied by its width), $\theta(x)$ is the unit function, $\hbar/2m^* = 1$, and $c = 1$. The last term in (2),

$$\hat{V}(x, t)\Psi = i2eA_x(t)\frac{\partial \Psi}{\partial x} = V(e^{i\omega t} - e^{-i\omega t})\frac{\partial \Psi}{\partial x}, \quad V = -\frac{eE}{\omega}, \tag{3}$$

describes the interaction of electrons with the electromagnetic field; $A_x(t)$ is the vector potential in the Coulomb gauge.

In this paper, following the approach of [3], we use the interaction in form (3) (with a vector potential) in contrast to [9], where we used the interaction in the form "$-xE$." This allows us to find an analytic solution for a strong electromagnetic field and construct an oscillation theory. We will show that, in the case of a weak field, the results for different gauges coincide. Note that, in (3), just as in [3] (see also [4]), the term quadratic in $A(t)$ is dropped, which is valid if

$$V/p = eE/p\omega \ll 1$$

(p is the characteristic momentum of an electron); henceforth, we assume that this condition is fulfilled.

The steady-state equation for the amplitude E of the electromagnetic field is expressed as (see, for example, [3])

$$\frac{E}{2\tau_0} = \frac{2\pi}{\kappa} J_c, \tag{4}$$

$$J_c = \frac{1}{a}\left[\int_0^a dx J_{1c}(x) + \int_a^{2a} dx J_{2c}(x)\right] \equiv \frac{J_{1c} + J_{2c}}{2}, \tag{5}$$

where J_{lc} is the active polarization current that describes transitions with emission (absorption) of photons in the first ($l = 1$) and second ($l = 2$) wells.

As is shown in [9], the energy spectrum of the TWNS consists of two levels that arise due to the interwell interference (splitting of levels) and the dc electric field V_{dc}. The parameters of the structure are chosen so that the energies of the levels $\varepsilon_R^{(1)}$ and $\varepsilon_R^{(2)}$ are displaced with respect to the energy ε_R of the lower resonance level of an isolated well by a quantity that is small compared with ε_R.

Obviously, the transitions between energy levels are most intense when the frequency of the electromagnetic field is approximately equal to the difference $\varepsilon_R^{(1)} - \varepsilon_R^{(2)}$. In the resonance approximation, we can seek steady-state solution (2) in the form [9]

$$\Psi_1(x, t) = e^{-i\varepsilon t}[\Psi_{10}(x) + e^{i\omega t}\Psi_{11}(x)], \quad 0 \le x \le a, \tag{6}$$

$$\Psi_2(x, t) = e^{-i\varepsilon t}[\Psi_{20}(x) + e^{i\omega t}\Psi_{21}(x)], \quad a \le x \le 2a. \tag{7}$$

The wavefunctions Ψ_{ln} ($l = 1, 2$ is the well number, and $n = 0, 1$ is the level number) correspond to the states with quasi-energies ε and $\varepsilon - \omega$ and satisfy the system of equations

$$p^2\Psi_{10} + \Psi''_{10} = V\Psi'_{11}, \quad p_n^2\Psi_{11} + \Psi''_{11} = -V\Psi'_{10}, \quad p^2 = \varepsilon, \quad p_n^2 = \varepsilon - \omega, \tag{8}$$

$$p_1^2\Psi_{20} + \Psi''_{20} = V\Psi'_{21}, \quad p_{1n}^2\Psi_{21} + \Psi''_{21} = -V\Psi'_{20}, \quad p_1^2 = \varepsilon + V_{dc}, \quad p_{1n}^2 = p_1^2 - \omega \tag{9}$$

subject to the boundary conditions (see [3, 9])

$$\begin{gathered}
\Psi_{10}(0)(1-\beta) + \Psi'_{10}(0)/ip = 2q, \quad \beta = \alpha/ip, \\
\Psi_{20}(2a)(1-\beta_1) - \Psi'_{20}(2a)/ip_1 = 0, \quad \beta_1 = \alpha/ip_1, \\
\Psi_{10}(a) = \Psi_{20}(\alpha), \quad \Psi'_{20}(a) - \Psi'_{10}(a) = \alpha\Psi_{10}(a), \\
\Psi_{11}(0)(1-\beta_n) + \Psi'_{11}(0)/ip_n = 0, \quad \beta_n = \alpha/ip_n, \\
\Psi_{21}(2a)(1-\beta_{1n}) - \Psi'_{21}(2a)/ip_{1n} = 0, \quad \beta_{1n} = \alpha/ip_{1n}, \\
\Psi_{11}(a) = \Psi_{21}(a), \quad \Psi'_{21}(a) - \Psi'_{11}(a) = \alpha\Psi_{11}(a).
\end{gathered} \tag{10}$$

The boundary conditions describe an incident stream of electrons from $x = -\infty$, their reflection, and their traveling away into the domain $x > 2a$, as well as the continuity and jump of the derivatives of the wavefunctions Ψ_{ln} at $x = 0, a, 2a$.

The current $J_c(x)$ can be expressed in terms of the functions $\Psi_{ln}(x)$:

$$J_{lc} = -ie[(\Psi_{l0}^*\Psi'_{l1} + \Psi_{l1}^*\Psi'_{l0}) - \text{c.c.}]. \tag{11}$$

The system of equations (8)–(11) allows one to calculate the currents J_{lc} and the oscillation power in the TWNS as a function of the pumping current and the parameters of the structure.

3. A GENERAL SOLUTION TO THE SCHRÖDINGER EQUATION FOR THE TWNS

Following [3], we will seek a general solution to Eqs. (8) and (9) in the form

$$\Psi_{1n}(x) = \sum_j A_{1n}^j \exp(\gamma_j x), \tag{12}$$

$$\Psi_{2n}(x) = \sum_j A_{2n}^j \exp(\bar{\gamma}_j x). \tag{13}$$

The complex eigenvalues γ_j and $\bar{\gamma}_j$ satisfy the equations

$$\begin{gathered}
\gamma^4 + 2\gamma^2\left(p_n^2 + \frac{V^2 + \omega}{2}\right) + p^2 p_n^2 = 0, \\
\bar{\gamma}^4 + 2\bar{\gamma}^2\left(p_{1n}^2 + \frac{V^2 + \omega}{2}\right) + p_1^2 p_{1n}^2 = 0,
\end{gathered} \tag{14}$$

and the coefficients A_{ln}^j are related by the formulas

$$\begin{gathered}
A_{11}^j = \varepsilon_j A_{10}^j, \quad A_{21}^j = \bar{\varepsilon}_j A_{20}^j, \quad A_{l0}^j \equiv A_{lj}, \\
\varepsilon_j = -\frac{\gamma_j V}{p_n^2 + \gamma_j^2}, \quad \bar{\varepsilon}_j = -\frac{\bar{\gamma}_j V}{p_{1n}^2 + \bar{\gamma}_j^2}.
\end{gathered} \tag{15}$$

The eigenvalues possess the following properties:

$$\gamma_1 + \gamma_2 = 0, \quad \gamma_3 + \gamma_4 = 0, \quad \bar{\gamma}_1 + \bar{\gamma}_2 = 0, \quad \bar{\gamma}_3 + \bar{\gamma}_4 = 0.$$

Substituting $\Psi_{ln}(x)$ from (12) and (13) into boundary conditions (10), we arrive at the following system of algebraic equations in the coefficients:

$$\tilde{A}_{1j} = A_{1j}e^{\gamma_j a}, \quad \tilde{A}_{2j} = A_{2j}e^{\tilde{\gamma}_j a},$$

which can be rewritten as the matrix equation

$$\mathbf{EA} = \mathbf{Q}, \tag{16}$$

where

$$\mathbf{E} = \begin{pmatrix} E_1 & E_2 & E_3 & E_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \tilde{\bar{E}}_1 & \tilde{\bar{E}}_2 & \tilde{\bar{E}}_3 & \tilde{\bar{E}}_4 \\ \varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \varepsilon_4 & \bar{\varepsilon}_1 & \bar{\varepsilon}_2 & \bar{\varepsilon}_3 & \bar{\varepsilon}_4 \\ \varepsilon_1 n_1 & \varepsilon_2 n_2 & \varepsilon_3 n_3 & \varepsilon_4 n_4 & \bar{\varepsilon}_1\bar{\gamma}_1 & \bar{\varepsilon}_2\bar{\gamma}_2 & \bar{\varepsilon}_3\bar{\gamma}_3 & \bar{\varepsilon}_4\bar{\gamma}_4 \\ M_1 & M_2 & M_3 & M_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \tilde{\bar{M}}_1 & \tilde{\bar{M}}_2 & \tilde{\bar{M}}_3 & \tilde{\bar{M}}_4 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ n_1 & n_2 & n_3 & n_4 & \bar{\gamma}_1 & \bar{\gamma}_2 & \bar{\gamma}_3 & \bar{\gamma}_4 \end{pmatrix},$$

$$\mathbf{A} = \begin{pmatrix} \tilde{A}_{11} \\ \tilde{A}_{12} \\ \tilde{A}_{13} \\ \tilde{A}_{14} \\ \tilde{A}_{21} \\ \tilde{A}_{22} \\ \tilde{A}_{23} \\ \tilde{A}_{24} \end{pmatrix}, \text{ and } \mathbf{Q} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 2q \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Here, we introduced the following notation:

$$\left.\begin{aligned} E_j &= \varepsilon_j M_{jn}, \quad \tilde{\bar{E}}_j = \tilde{\bar{M}}_{jn}\bar{\varepsilon}_j \\ M_{jn} &= m_{jn}e^{-\gamma_j}, \quad \tilde{\bar{M}}_{jn} = -\tilde{\bar{m}}_{jn}e^{\bar{\gamma}_j}, \\ M_j &= m_j e^{-\gamma_j}, \quad \tilde{\bar{M}}_j = -\tilde{\bar{m}}_j e^{\bar{\gamma}_j}, \end{aligned}\right\} \tag{17}$$

$$\left.\begin{aligned} m_{jn} &= 1 - \frac{\alpha - \gamma_j}{ip_n}, \quad \tilde{m}_{jn} = 1 - \frac{\alpha + \gamma_j}{ip_n}, \\ \tilde{\bar{m}}_{jn} &= 1 - \frac{\alpha + \bar{\gamma}_j}{ip_n}, \quad m_j = 1 - \frac{\alpha - \gamma_j}{ip}, \\ \tilde{m}_j &= 1 - \frac{\alpha + \gamma_j}{ip}, \\ \tilde{\bar{m}}_j &= 1 - \frac{\alpha + \bar{\gamma}_j}{ip_1}, \quad n_j = \alpha + \gamma_j. \end{aligned}\right\} \tag{18}$$

Taking into account that the parameter V/p is small, we can represent the roots of Eqs. (14) as

$$\begin{aligned} \gamma_{1,2} &= \pm ip_n\left(1 - \frac{V^2}{\omega}\right)^{1/2}, \\ \gamma_{3,4} &= \pm ip\left(1 + \frac{V^2}{\omega}\right)^{1/2}, \\ \bar{\gamma}_{1,2} &= \pm ip_{1n}\left(1 - \frac{V^2}{\omega}\right)^{1/2}, \\ \bar{\gamma}_{3,4} &= \pm ip_1\left(1 + \frac{V^2}{\omega}\right)^{1/2}. \end{aligned} \tag{19}$$

Accordingly, the quantities ε_j and $\bar{\varepsilon}_j$ take the form

$$\begin{aligned} \varepsilon_{3,4} &= \pm i\tilde{V}, \quad \varepsilon_{1,2} = \mp\frac{i}{\tilde{V}}\frac{p}{p_n}, \\ \bar{\varepsilon}_{3,4} &= \pm i\tilde{V}\frac{p_1}{p}, \quad \bar{\varepsilon}_{1,2} = \mp\frac{i}{\tilde{V}}\frac{p}{p_{1n}}, \end{aligned} \tag{20}$$

$$\tilde{V} = \frac{Vp}{\omega}. \tag{21}$$

Further, following [3], we will consider the interval of fields in which the dimensionless parameter $\tilde{V}$ can be assumed to be small. This fact significantly simplifies the computations, although is not fundamentally necessary. When $\tilde{V} \ll 1$, the quantities ε_3 and ε_4 are small, whereas ε_1 and ε_2 are large.

4. DETERMINANT OF THE SYSTEM

A key quantity is the determinant $\Delta(\lambda)$ of system (16), which describes the resonance properties of the TWNS in an electromagnetic field, in particular, the influence of a strong field on the resonance tunneling. To avoid cumbersome calculations, we will consider in detail a situation in the absence of dc bias voltage V_{dc} and then generalize it to the case of $V_{dc} \neq 0$.

If

$$V_{dc} = 0,$$

then

$$\begin{aligned} p_1 &= p, \quad p_{1n} = p_n, \quad \bar{\varepsilon}_j = \varepsilon_j, \quad \bar{\gamma}_j = \gamma_j, \\ \tilde{\bar{E}}_j &= \tilde{E}_j, \quad \tilde{\bar{M}}_j = \tilde{M}_j. \end{aligned}$$

The determinant $\Delta(\lambda)$ of the system of equations (16) can be expressed as a sum of terms that represent a product of matrices Π_{iklm}, Π^n_{iklm}, K_{iklm}, $\tilde{K}_{iklm}$, K^n_{iklm}, and $\tilde{K}^n_{iklm}$, which are given by

$$\Pi_{iklm} = \begin{vmatrix} M_i & M_k & 0 & 0 \\ 0 & 0 & \tilde{M}_l & \tilde{M}_m \\ 1 & 1 & 1 & 1 \\ N_I & N_k & \gamma_l & \gamma_m \end{vmatrix}, \quad (22)$$

$$\Pi^n_{iklm} = \begin{vmatrix} M_{in} & M_{kn} & 0 & 0 \\ 0 & 0 & \tilde{M}_{ln} & \tilde{M}_{mn} \\ 1 & 1 & 1 & 1 \\ n_i & n_k & \gamma_l & \gamma_m \end{vmatrix}, \quad (23)$$

$$K_{iklm} = \begin{vmatrix} M_i & 0 & 0 & 0 \\ 0 & \tilde{M}_k & \tilde{M}_l & \tilde{M}_m \\ 1 & 1 & 1 & 1 \\ n_i & \gamma_k & \gamma_l & \gamma_m \end{vmatrix}, \quad (24)$$

$$\tilde{K}_{iklm} = \begin{vmatrix} M_i & M_k & M_l & 0 \\ 0 & 0 & 0 & \tilde{M}_m \\ 1 & 1 & 1 & 1 \\ n_i & n_k & n_l & \gamma_m \end{vmatrix}, \quad (25)$$

$$K^n_{iklm} = \begin{vmatrix} M_{in} & 0 & 0 & 0 \\ 0 & \tilde{M}_{kn} & \tilde{M}_{ln} & \tilde{M}_{mn} \\ 1 & 1 & 1 & 1 \\ n_i & \gamma_k & \gamma_l & \gamma_m \end{vmatrix}, \quad (26)$$

$$\tilde{K}^n_{iklm} = \begin{vmatrix} M_{in} & M_{kn} & M_{ln} & 0 \\ 0 & 0 & 0 & \tilde{M}_{mn} \\ 1 & 1 & 1 & 1 \\ n_i & n_k & n_l & \gamma_m \end{vmatrix}, \quad (27)$$

multiplied by the coefficients ε_j.

In relatively weak fields, when the dimensionless field $\tilde{V}$ is small (see (21)), we can restrict the analysis to two types of terms. Other terms make a small contribution proportional to $\tilde{V}^4$, $\tilde{V}^6$, and $\tilde{V}^8$. As a result, we arrive at the following expression for $\Delta(\lambda)$:

$$\Delta(\lambda) \approx \varepsilon_1^2\varepsilon_2^2\left\{\Pi_{3434}\Pi^n_{1212} - \frac{\varepsilon_3}{\varepsilon_2}[\Pi + K]\right\}, \quad (28)$$

$$\begin{aligned}\Pi = [&\Pi_{3424}\Pi^n_{1213} + \Pi_{3423}\Pi^n_{1214} + \Pi_{2434}\Pi^n_{1312} \\ &+ \Pi_{2334}\Pi^n_{1412} + \Pi_{3414}\Pi^n_{1213} + \Pi_{3413}\Pi^n_{1234} \\ &+ \Pi_{1434}\Pi^n_{2312} + \Pi_{1334}\Pi^n_{2412}],\end{aligned} \quad (29)$$

$$\begin{aligned}K = [&K_{4234}\tilde{K}^n_{1231} + K_{3234}\tilde{K}^n_{1241} + \tilde{K}_{2344}K^n_{1123} \\ &+ \tilde{K}_{2343}K^n_{1124} + K_{4134}\tilde{K}^n_{1232} + K_{3134}\tilde{K}^n_{1242} \\ &+ \tilde{K}_{1344}K^n_{2123} + \tilde{K}_{1343}K^n_{2124}].\end{aligned} \quad (30)$$

The symmetric matrices Π_{3434} and Π^n_{1212} represent the determinants of the TWNS in the absence of an electromagnetic field for the upper level with the energy $\varepsilon = p^2$ and for the lower level, with the energy $\varepsilon - \omega = p_n^2$, respectively:

$$\begin{aligned}\Pi_{3434} &= (-ip)e^{-2ipa}\Delta_0(2), \\ \Pi^n_{1212} &= (-ip_n)e^{-2ip_na}\Delta_n(2), \\ \Delta_0(2) &= (2-\beta)^3 - 2\beta^2(2-\beta)e^{2ipa} \\ &\quad - \beta^2(2+\beta)e^{4ipa}, \\ \Delta_n(2) &= (2-\beta_n)^3 - 2\beta_n^2(2-\beta)e^{2ip_na} \\ &\quad - \beta_n^2(2+\beta_n)e^{4ip_na}.\end{aligned} \quad (31)$$

The properties of the determinants $\Delta_0(2)$ and $\Delta_n(2)$ were studied in detail in [9], where it was shown that they can be represented as

$$\Delta_0(2) \approx \frac{8i}{\Gamma t_0}[\varepsilon - \varepsilon_R^{(1)} + i\Gamma][\varepsilon - \varepsilon_R^{(2)} + i\Gamma], \quad (32)$$

$$\Delta_n(2) \approx \frac{8i}{\Gamma t_0}[\varepsilon - \omega - \varepsilon_R^{(1)} + i\Gamma][\varepsilon - \omega - \varepsilon_R^{(2)} + i\Gamma] \quad (33)$$

in the neighborhood of the resonance. Here, $\varepsilon_R^{(1)}$ and $\varepsilon_R^{(2)}$ are the energies of the levels in the TWNS that

arise due to the splitting of the resonance level of an isolated well:

$$\varepsilon_R^{(1)} = p_0^2 - \frac{2p_0\delta}{a}, \quad \varepsilon_R^{(2)} = p_0^2 - \frac{6p_0\delta}{a}, \qquad p_0 = \frac{\pi}{a}, \quad \delta = \frac{p_0}{\alpha}, \tag{34}$$

t_0 is the value of splitting:

$$t_0 = \varepsilon_R^{(1)} - \varepsilon_R^{(2)} = \frac{4p_0^2}{\alpha a}, \tag{35}$$

and Γ is the width of the resonance levels (which is the same for both levels):

$$\Gamma = \frac{2p_0^3}{\alpha^2 a}. \tag{36}$$

Note that, for simplicity, we omitted the quadratic (in δ) corrections in $\varepsilon_R^{(1,2)}$ and t_0 (see [9]). Formulas (32) and (33) are valid in the vicinity of a resonance when the energy of the electrons coming from the emitter is close to $\varepsilon_R^{(1)}$ or $\varepsilon_R^{(2)}$ ($\varepsilon - \varepsilon_R^{(1)} \ll \varepsilon_R^{(1)}$). In what follows, we assume that this condition is satisfied.

Moreover, it is usually assumed [3, 9] that the width Γ is much less than the energies ε_R; i.e., $\Gamma/\varepsilon_R = 2\delta^2/\pi \ll 1$. It is these conditions under which quantum wells manifest their remarkable properties. It can easily be seen that the value of splitting $t_0/\varepsilon_R \approx \delta \ll 1$ is also small compared with ε_R.

The terms Π describe the contribution of electron transitions between energy levels in the first and the second wells, and the terms K describe the contribution of transitions from the first well to the second. If the dimensionless field $\tilde{V}$ (21) is small, then the ratio

$$\frac{\varepsilon_3}{\varepsilon_2} \approx \tilde{V}^2 \ll 1 \tag{37}$$

is also small, and it suffices to restrict the analysis to two terms (28) and (30) in the determinant. The calculation of Π and K is associated with certain difficulties because, as we will see below, the leading terms in the parameter α/p compensate each other due to the interference between the contributions of the two wells.

Let us start the calculations from the term K. The matrices K_{iklm}, $\tilde{K}_{iklm}$, K^n_{iklm}, and $\tilde{K}^n_{iklm}$ can be expressed as

$$K_{iklm} = M_i\tilde{L}_{klm}, \quad \tilde{K}_{iklm} = \tilde{M}_m L_{ikl}, \qquad K^n_{iklm} = M_{in}\tilde{L}^n_{klm}, \quad \tilde{K}^n_{iklm} = \tilde{M}_{mn}L^n_{ikl}, \tag{38}$$

where the matrices L_{klm} are given by

$$\tilde{L}_{234} = -L_{134} = ip\left\{-2Z_n(2-\beta)e^{-ipa} + \frac{2x}{p}e^{-ip_n a} - \frac{x}{p}e^{-ip_n}\psi_0\right\},$$

$$\tilde{L}_{134} = -L_{234} = ip\left\{2Z_n\beta e^{ip_n a} + \frac{2x}{p}e^{ip_n a} + \frac{x}{p}e^{-ip}\psi_0\right\},$$

$$\tilde{L}^n_{123} = -L^n_{124} = ip_n\left\{-2Z_n\beta_n e^{ip_n a} - \frac{2x}{p_n}e^{ipa} + \frac{x}{p_n}e^{-ip_n}\psi_n\right\},$$

$$\tilde{L}^n_{124} = -L^n_{123} = ip_n\left\{-2Z_n(2-\beta_n)e^{-ipa} + \frac{2x}{p_n}e^{-ipa} - \frac{x}{p_n}e^{-ip_n}\psi_n\right\}. \tag{39}$$

Here, we introduced the following notation:

$$\psi_n = [2 + \beta_n(e^{2ip_n a} - 1)], \qquad \psi_0 = [2 + \beta(e^{2ipa} - 1)], \tag{40}$$

$$Z_n = e^{i(p-p_n)a} - 1, \quad x = p - p_n. \tag{41}$$

Substituting (38) and (39) into (30), after a series of transformations, we obtain

$$K = (ipip_n)2\left\{4Z_n^2\varphi^2 - \frac{2x^2}{pp_n}Z_n\varphi(4 - \tilde{\psi}_0\tilde{\psi}_n) - \frac{x^2}{pp_n}[4 - \tilde{\psi}_0\tilde{\psi}_n]^2\right\}, \tag{42}$$

$$\varphi = -\beta_n\left[(2-\beta)Z_n^* + \frac{x}{p}\right], \qquad \tilde{\psi}_n = e^{-ip_n a}\psi_n, \quad \tilde{\psi}_0 = e^{-ipa}\psi_0. \tag{43}$$

Let us evaluate the contribution of the three terms in (42). First of all, consider the important factor Z_n (41), which describes the intensity of radiative transitions between levels. If the transitions occur between

resonance levels of space quantization of the same well (like in the coherent laser [3]), then $p \approx 2\pi/a$, $p_n \approx \pi/a$, and the factor $Z_n \approx -2$ takes the maximum possible value. In the case of the TWNS, when the splitting t_0 and, hence, the resonance frequency $\omega \approx t_0$, are small compared with ε_R, the factor Z_n and the parameter x are equal to

$$\begin{aligned} Z_n &\approx i(p - p_n)a \approx i\frac{\omega a}{2p} \ll 1, \\ x &= p - p_n = \frac{\omega}{2p}, \quad x/p \ll 1, \end{aligned} \tag{44}$$

respectively. That is, Z_n and x/p are small quantities on the order of $\delta = p/\alpha$. Hence, it is clear that the main contribution to K is made by the first term in braces,

$$4Z_n^2\varphi^2 \approx 4Z_n^2\beta_n^2\beta^2 Z_n^{*2} = 4|Z_n|^4\beta_n^2\beta^2,$$

so that, using the condition $\omega = t_0$, we obtain

$$K = (ipip_n)\cdot 64 \cdot 2. \tag{45}$$

Note that each term in K (see (30)) is on the order of

$$ip\left(\frac{\alpha}{p}\right)^4 4Z_n^2 \approx ip\left(\frac{\alpha}{p}\right)^2 \cdot 16;$$

i.e., each term is $(\alpha/p)^2$ times greater than $K/8$. Hence, the leading (in α/p) terms in K compensate each other due to the interference in the TWNS. A similar situation occurs with the contribution of Π_{iklm} to Π.

This result is of fundamental importance because, due to the compensation, the effect of the ac field on the resonance tunneling is sharply reduced and, as a consequence, the oscillation efficiency increases.

Let us pass to the calculation of Π. First of all, we can show that the last four terms in Π exactly coincide with the first four terms. The matrices Π_{iklm} and Π^n_{iklm} entering these terms can be represented as

$$\begin{aligned} \Pi_{3424} &= ipe^{-2ipa}\left\{-Z_n(2-\beta)\Pi_0(1)\right. \\ &\left. + \frac{x}{p}[\Pi_0(1)e^{i(p-p_n)a} - (2-\beta)\psi_0]\right\}, \end{aligned} \tag{46}$$

$$\begin{aligned} \Pi_{3423} &= ip\left\{\Delta_0(2)e^{-2ipa} + Z_n(2-\beta)\Delta_0(1)\right. \\ &\left. + \frac{x}{p}[-\Delta_0(1)e^{i(p-p_n)a} + \beta\psi_0]\right\}, \end{aligned} \tag{47}$$

$$\begin{aligned} \Pi_{2434} &= ip\left\{Z_n^*\beta\Delta_0(1)e^{2ipa}\right. \\ &\left. + \frac{x}{p}[-\Delta_0(1)e^{i(p+p_n)a} + \beta\psi_0]\right\}, \end{aligned} \tag{48}$$

$$\begin{aligned} \Pi_{2334} &= ip\left\{\Delta_0(2)e^{-2ipa} - Z_n^*\beta\Pi_0(1)\right. \\ &\left. + \frac{x}{p}[\Pi_0(1)e^{ip_na} - (2-\beta)\tilde{\psi}_0]\right\}, \end{aligned} \tag{49}$$

$$\begin{aligned} \Pi^n_{1213} &= ip_n\left\{Z_n\beta_n\Delta_n(1)e^{2ip_na}\right. \\ &\left. + \frac{x}{p_n}[\Delta_n(1)e^{i(p+p_n)a} - \beta_n\psi_n]\right\}, \end{aligned} \tag{50}$$

$$\begin{aligned} \Pi^n_{1214} &= ip_n\left\{-\Delta_n(2)e^{-2ip_na} - Z_n^*(2-\beta_n)\Delta_n(1)\right. \\ &\left. + \frac{x}{p_n}[-\Delta_n(1)e^{i(p_n-p)a} + \beta_n\psi_n]\right\}, \end{aligned} \tag{51}$$

$$\begin{aligned} \Pi^n_{1312} &= ip_n\left\{-Z_n^*(2-\beta_n)\Pi_n(1)e^{-2ip_na}\right. \\ &\left. + \frac{x}{p_n}e^{-2ip_na}[-\Pi_n(1)e^{i(p_n-p)a} + (2-\beta_n)\psi_n]\right\}, \end{aligned} \tag{52}$$

$$\begin{aligned} \Pi^n_{1412} &= ip_n\left\{-\Delta_n(2)e^{-2ip_na} + Z_n\beta_n\Pi_n(1)\right. \\ &\left. + \frac{x}{p_n}e^{-ip_na}[\Pi_n(1)e^{ipa} - (2-\beta_n)\tilde{\psi}_n]\right\}. \end{aligned} \tag{53}$$

Here, we used the following notation:

$$\Delta_0(1) = (2-\beta)^2e^{-2ipa} - \beta^2, \tag{54}$$

$$\Pi_0(1) = \beta[(2-\beta) + (2+\beta)e^{2ipa}], \tag{55}$$

$$\Delta_n(1) = (2-\beta_n)^2e^{-2ip_na} - \beta_n^2, \tag{56}$$

$$\Pi_n(1) = \beta_n[(2-\beta_n) + (2+\beta_n)e^{2ip_na}]. \tag{57}$$

Substituting (46)–(53) into the expression for Π (29) and neglecting the terms that are small compared with unity, after rather tedious transformations, we obtain

$$\begin{aligned}\Pi = -2ip\cdot ip_n\{|Z|^4[\beta_n(2-\beta)\Pi_0(1)\Delta_n(1)\\ +\beta(2-\beta_n)\Pi_n(1)\Delta_0(1)]\\ +2\Delta_0(2)\Delta_n(2)e^{-2i(p+p_n)a}\\ +e^{-2ipa}\Delta_n(2)2Z_n[\beta\Pi_0 e^{2i(p-p_n)a}+\beta_n\Pi_n]\}.\end{aligned} \tag{58}$$

For the one-dimensional determinants $\Delta_n(1)$ and $\Delta_0(1)$, as well as for $\Pi_n(1)$ and $\Pi_0(1)$, we can obtain the following approximations with respect to the parameter p/α (cf. [9]):

$$\Pi_0(p^{(1)})\approx 2\beta\left[1+\frac{1}{\alpha a}(1-i\pi)\right],$$

$$\Pi_n(p^{(1)}) = \Pi_0(p^{(2)}) = -2\beta\left[1+\frac{3}{\alpha a}(1-i\pi)\right], \tag{59}$$

$$\Delta_0(p^{(1)})\approx -2\beta,\quad \Delta_n(p^{(1)}) = \Delta_0(p^{(2)}) = 2\beta.$$

Substituting (59) into (58) and applying (32) and (33), we arrive at the following result:

$$\Pi = ip\cdot ip_n\cdot 64\cdot 2\left[1-\frac{2i}{\pi}\right]. \tag{60}$$

Finally, we have

$$\Pi + K = ip\cdot ip_n\cdot 64\cdot 4\left(1-\frac{i}{\pi}\right). \tag{61}$$

Combining the results, we find the determinant of the system:

$$\Delta(\lambda) = \varepsilon_1^2\varepsilon_2^2 pp_n\frac{64}{\Gamma^2}\tilde{\Delta}(\lambda), \tag{62}$$

$$\begin{aligned}\tilde{\Delta}(\lambda) = \Big[(\varepsilon-\varepsilon_R^{(1)}+i\Gamma)(\varepsilon-\omega-\varepsilon_R^{(2)}+i\Gamma)\\ -\lambda^2\left(1-\frac{i}{\pi}\right)\Big],\end{aligned} \tag{63}$$

$$\lambda^2 = 4\tilde{V}^2\Gamma^2 = \left(16\frac{p^2}{a^2}\tilde{V}^2\right)\left(\frac{p}{\alpha}\right)^4. \tag{64}$$

Expression (63) for the determinant of the TWNS is formally analogous to a similar expression for the coherent laser, in which radiative transitions occur between the resonance levels of the same well [3]. The difference is that the definition of λ^2 involves an additional small factor $(p/\alpha)^4$. It is associated with the above-mentioned compensation of leading (in α/p) terms due to the interference and the smallness of the frequency $\omega \ll \varepsilon_R$. The second difference lies in the term $\lambda^2 i/\pi$. This term leads to a certain asymmetry in the tuning mode (see Section 5 below).

The term λ^2 in (63) describes nonlinear properties and reflects the influence of an electromagnetic field on the resonance tunneling [3], which becomes substantial when

$$\lambda > \Gamma. \tag{65}$$

Such fields are called strong fields [3]. The fact that λ sharply decreases due to the compensation indicates that the influence of the field is weakened. This fact is of fundamental importance for increasing the efficiency of the field generation. Note also that the applicability scope of the approximation $\tilde{V}^2 < 1$ sharply decreases due to the factor $(p/\alpha)^4$. Further, we will restrict ourselves to the fields $\tilde{V}^2 \lesssim 1/3$ because, for $\tilde{V}^2 > 1$, the calculations become extremely cumbersome. The analysis of the results carried out in a wide range of fields by numerical methods [10] shows that this approximation is applicable. Moreover, in the most interesting case, when the bias voltage is greater than the splitting, $V_{dc} \gg t_0$, the domain $\tilde{V}^2 < 1$ again expands (see Section 5 below).

It is of interest to compare the values of λ^2 (64) for the TWNS and for the coherent laser [3],

$$\lambda_l^2 = \frac{16pp_n}{a^2}\tilde{V}^2. \tag{66}$$

As pointed out above, these values differ by $(p/\alpha)^4$; however, the frequencies entering $\tilde{V}^2$ are also significantly different. For instance, for the TWNS, one should use a small frequency of

$$\omega = t_0 = \frac{4p^2}{\alpha a};$$

for a coherent laser, a frequency of

$$\omega = \frac{3\pi^2}{a^2} = \frac{3}{4}p^2.$$

As a result, we obtain

$$\begin{aligned}\lambda = \frac{eEa}{4} = \frac{V_{ac}}{8},\quad \lambda_l = \frac{8eEa}{9\pi^2} = V_{ac}^l\frac{8}{9\pi^2},\\ V_{ac} = 2eEa,\quad V_{ac}^l = eEa.\end{aligned} \tag{67}$$

The strong-field criteria for the TWNS and the coherent laser are expressed as

$$V_{ac} > 8\Gamma,\quad V_{ac}^l > \frac{9\pi^2}{8}\Gamma. \tag{68}$$

Hence, the field becomes strong if the energy accumulated by an electron over the length of the structure is comparable with the linewidth. We can see that the difference between V_{ac} and V_{ac}^{l} is only numerical.

5. WAVEFUNCTION AND POLARIZATION CURRENT OF THE FIRST WELL

Let us determine the wavefunction and the current in the first well only, because the currents in the first and second wells coincide. Indeed, for a weak field, this was proved in [9]. In the case of a strong field, the difference is mainly associated with the determinant $\Delta(\lambda)$ of the system, which is common for the two wells. Moreover, the numerical calculation of currents [10] also confirms that these determinants are equal.

Restricting the consideration to the leading terms of expansion in $\tilde{V}$, after quite tedious calculations, we obtain from expression (16) the following expressions for the coefficients A_{1j} (15) of the wavefunction (12):

$$\varepsilon_1 A_{11} \equiv A_1 = A_1^{(1)} + A_1^{(2)} = \frac{2q\varepsilon_1^2\varepsilon_2^2\varepsilon_3}{\Delta(\lambda)} ip \cdot ip_n\{\Delta_0(1)\Pi_{1212} - \beta_n[M_1 - \Phi_1]\}, \quad (69)$$

$$\varepsilon_2 A_{12} \equiv B_1 = B_1^{(1)} + B_1^{(2)} = \frac{2q\varepsilon_1^2\varepsilon_2^2\varepsilon_3}{\Delta(\lambda)} \times ip \cdot ip_n\{-\Pi_0(1)\Pi_{1212} - (2-\beta_n)[M_1 - \Phi_1]\}, \quad (70)$$

$$A_{13}\frac{\Delta(\lambda)}{2q} = -ip\Delta_0(1)\Pi_{1212}\varepsilon_1^2\varepsilon_2^2, \quad (71)$$

$$A_{14}\frac{\Delta(\lambda)}{2q} = -ip\Pi_0(1)\Pi_{1212}\varepsilon_1^2\varepsilon_2^2, \quad (72)$$

$$M_1 = [Z_n\Delta_0(1)\Pi_n(1) + Z_n^*\Pi_0(1)\Delta_n(1)],$$
$$\Phi_1 = 4[Z_n\beta(2-\beta_n) + Z_n^*\beta_n(2-\beta)]. \quad (73)$$

In the limiting case of a weak field $\lambda \ll \Gamma$, the coefficients A_{1j} in expressions (69)–(72) coincide with those obtained earlier in [9]. The coefficients A_1 and B_1 are represented as a sum of two terms. The first term $A_1^{(1)}$ ($B_1^{(1)}$) is divergent as $\omega \longrightarrow 0$ (see [9] for details). The second term $A_1^{(2)}$ ($B_1^{(2)}$) is finite as $\omega \longrightarrow 0$, and only this term makes a contribution to the polarization current.

Let us determine the wavefunctions Ψ_{10} and Ψ_{11} (12). One should retain only the contributions of A_{13} and A_{14} in the function Ψ_{10} and omit the small corrections due to A_{11} and A_{12} that are quadratic in $\tilde{V}$:

$$\Psi_{10}(x) \approx A_{13}e^{\gamma_3 x} + A_{14}e^{\gamma_4 x}.$$

In the same approximation, one can represent the eigenvalues γ_j (19) as

$$\gamma_3 = -\gamma_4 = ip, \quad \gamma_1 = -\gamma_2 = ip_n. \quad (74)$$

Taking into account that, according to (59),

$$\Delta_0(1) \approx -\Pi_0(1), \quad A_{13} \approx -A_{14},$$

for $\alpha/p \gg 1$, we find that the wavefunction $\Psi_{10}(x)$ is

$$\Psi_{10}(x) \approx 2iA_{13}\sin px. \quad (75)$$

All four terms in the function Ψ_{11} are of the same order in $\tilde{V}$. It can easily be shown that, up to the small terms ω/p^2, the coefficients at e^{ipx} and e^{-ipx} compensate the coefficients $A_1^{(1)}$ at $e^{ip_n x}$ and $B_1^{(1)}$ at $e^{-ip_n x}$. This procedure is analogous to the elimination of infrared divergence in a weak field [9]. Taking into account that $A_1^{(2)} \approx -B_1^{(2)}$ in the leading-order approximation in (α/p), we can represent the wavefunction Ψ_{11} as

$$\Psi_{11}(x) = 2iA_1^{(2)}\sin p_n x. \quad (76)$$

Thus, the wavefunction $\Psi_{11}(x)$ depends only on the coefficient $A_1^{(2)}$, which is finite as $\omega \longrightarrow 0$ (see (80)).

Let us find M_1 and Φ_1. Assuming, as usual, that the energy of electrons incident to the TWNS is close to the energy $\varepsilon_R^{(1)}$ of the upper level (i.e., $p \approx p^{(1)}$) and using formula (59) for $\Delta_0(1)$, $\Pi_0(1)$, $\Delta_n(1)$, and $\Pi_n(1)$, we obtain

$$M_1 \approx \frac{\alpha^2 a^2}{p^4}\omega^2. \quad (77)$$

Accordingly, for Φ_1, we have

$$\Phi_1 \approx -\frac{\alpha^2 a^2}{p^4}\omega^2, \quad (78)$$

so that the difference of interest is

$$M_1 - \Phi_1 = 2\frac{\alpha^2 a^2}{p^4}\omega^2. \quad (79)$$

Substituting the difference $M_1 - \Phi_1$ into (69), we find $A_1^{(2)}$:

$$A_1^{(2)} = -\frac{Ua(16q)\alpha}{\Delta(\lambda)\Gamma p}. \quad (80)$$

Notice that the frequency is canceled in the denominator of the coefficient $A_1^{(2)}$, which becomes finite as $\omega \longrightarrow 0$.

In the limit of a weak field, the expression for $A_1^{(2)}$ (80) exactly coincides with the corresponding expression in [9]. Here, it is important to note that we use different forms of interaction with the electromagnetic field in [9] and in the present study. Indeed, interaction (3) is described by a vector potential, whereas, in [9], it is described as "$-xE$."

The wavefunctions $\Psi_{10}(x)$ and $\Psi_{11}(x)$ prove to be similar to the corresponding functions of the coherent laser based on a single-well structure [3]; this fact allows us to determine the current following the method of [3]. Substituting $\Psi_{10}(x)$ and $\Psi_{11}(x)$ into (11), we obtain the active component of the reduced polarization current:

$$J_{1c} = -i4eM_{12}[A_1^{(2)}A_{13}^* - \text{c.c.}]. \quad (81)$$

We calculate the matrix element M_{12} of the transition taking into account the fact that the resonance frequency is small compared with ε_R:

$$M_{12} = \frac{1}{a}\int_0^a dx[p\sin p_n x\cos px - p_n\sin px\cos p_n x] \approx \frac{\omega a}{4}. \quad (82)$$

Substituting $A_1^{(2)}$ and A_{13} into (81), we arrive at the final expression for the current of the first well:

$$J_{1c} = Ua(ep4q^2)\frac{64(\alpha/p)(\Delta_n(2) + \Delta_n^*(2))}{\Gamma|\tilde{\Delta}(\lambda)|^2}, \quad (83)$$

which coincides with the corresponding current obtained in [9] in the weak-field limit. In the general case, it is convenient to represent current (83) in a form analogous to that of the active current of the coherent laser [3]:

$$J_{1c} = -\frac{E\Gamma^2 Q\eta}{|\tilde{\Delta}(\lambda)|^2}, \quad Q = pq^2, \quad \eta = \frac{e^2p^2}{\alpha}, \quad (84)$$

$$|\tilde{\Delta}(\lambda)|^2 = \left[(\lambda^2 + \Gamma^2 - \xi^2)^2 + \left(2\Gamma\xi + \frac{\lambda^2}{\pi}\right)^2\right]. \quad (85)$$

When deriving (84), we used the relation

$$\Delta_n(2) + \Delta_n^*(2) = 16.$$

Formula (84) shows that the TWNS can be considered as a coherent laser in which radiative transitions occur between the energy levels $\varepsilon_R^{(1)}$ and $\varepsilon_R^{(2)}$. Since the photon frequency is small, an additional small factor p/α appears in current (84). In the determinant $\tilde{\Delta}(\lambda)$, there appear a factor $(p/\alpha)^4$ at $\tilde{V}^2$, which plays an essential role, and a correction λ^2/π (see formula (64)).

The current J_{1c} describes nonlinear properties of amplification and allows us to construct an oscillation theory that is presented in Section 7 below. It can easily be shown that, for a fixed energy ξ, the current first increases with λ, reaches its maximum for $\lambda = \lambda_0$, and then decreases in proportion to $1/\lambda^3$. In particular, for $\xi = 0$, we have

$$\lambda_0^2 \approx \frac{\Gamma^2}{3}\left(1 - \frac{1}{4\pi^2}\right),$$

which corresponds to the value $V_{ac}^{(0)} \approx 4.6\Gamma$.

Following [3], to calculate the efficiency of the radiation of photons, we introduce the quantity

$$P = \frac{1}{2\pi}\int_0^{2\pi/\omega} dt\int_0^{2a} dx J_c(x,t)E(t) = \frac{J_{1c}Ea}{\omega}, \quad (86)$$

which makes the sense of the number of photons emitted by the TWNS per unit time. Obviously, the highest possible efficiency is achieved when each of $2pq^2$ electrons incident to the TWNS emits a photon; i.e.,

$$P_{max} = 2pq^2.$$

Then, it is natural to characterize the efficiency by the parameter

$$\tilde{P} = \frac{P}{2pq^2} = \frac{2\lambda^2\Gamma^2}{|\tilde{\Delta}(\lambda)|^2}. \quad (87)$$

When $\xi = 0$, the efficiency $\tilde{P}$ increases with λ, reaches its maximal value $\tilde{P} = 1/2$ at $\lambda_0 = \Gamma$, and then decreases. Here, we neglected the small correction $1/\pi^2$. The situation considered is analogous to the coherent laser [3], the only difference being that the efficiency of the laser is equal to unity. In [3], it was shown that one can attain a regime with the highest efficiency for any λ by varying the energy ξ_0 as λ increases. Based on the maximality condition

$$\frac{dP}{d\xi} = 0,$$

we obtain the following equation for ξ_0:

$$\xi_0^3 + \xi_0^2(\Gamma^2 - \lambda^2) + \frac{\lambda^2\Gamma}{\pi} = 0. \quad (88)$$

If we neglect the last term, then solutions (88) have the form

$$\xi_{01} = 0, \quad \xi_0^2 = \lambda^2 - \Gamma^2, \quad \lambda > \Gamma. \tag{89}$$

The first solution yields a maximum for $\lambda = \Gamma$. The symmetric solutions

$$\xi_{02}, \xi_{03} = \pm\sqrt{\lambda^2 + \Gamma^2}$$

are realized for $\lambda > \Gamma$ and ensure the efficiency

$$\tilde{P} = \frac{2\lambda^2\Gamma^2}{4\lambda^2\Gamma^2} = \frac{1}{2} \tag{90}$$

for any λ. Looking ahead, note that an application of a dc voltage $V_{dc} > t_0$ increases the efficiency up to $\tilde{P} = 1$ (see Section 6).

Taking into account the term $\lambda_0^2\Gamma/\pi$ in expression (88) leads to a slight asymmetry between the second and the third solutions and to the shift ξ_{01} from the zero value.

6. WAVEFUNCTION AND POLARIZATION CURRENT OF TWNS IN THE PRESENCE OF VOLTAGE

Let us generalize the results obtained. Let a dc voltage be applied to the TWNS (see figure). We will seek the wavefunctions of the first well in form (12) with the coefficients A_{lj} satisfying system of equations (16) with $p_1 \neq p$ and $p_{1n} \neq p_n$.

The determinant $\tilde{\Delta}(\lambda)$ is given by expressions (28)–(30) in which one should make the following changes:

$$\Pi_{iklm} \longrightarrow \bar{\Pi}_{iklm}, \quad \Pi^n_{iklm} \longrightarrow \bar{\Pi}^n_{iklm}, \quad K_{iklm} \longrightarrow \bar{K}_{iklm},$$

$$\tilde{K}_{iklm} \longrightarrow \tilde{\bar{K}}_{iklm}, \quad K^n_{iklm} \longrightarrow \bar{K}^n_{iklm}, \quad \tilde{K}^n_{iklm} \longrightarrow \tilde{\bar{K}}^n_{iklm}.$$

The new matrices (supplied with a bar) are given by

$$\bar{\Pi}_{iklm} = \begin{vmatrix} M_i & M_k & 0 & 0 \\ 0 & 0 & \tilde{\bar{M}}_l & \tilde{\bar{M}}_m \\ 1 & 1 & 1 & 1 \\ n_i & n_k & \bar{\gamma}_l & \bar{\gamma}_m \end{vmatrix}, \tag{91}$$

$$\bar{\Pi}^n_{iklm} = \begin{vmatrix} M_{in} & M_{kn} & 0 & 0 \\ 0 & 0 & \tilde{\bar{M}}_{ln} & \tilde{\bar{M}}_{mn} \\ 1 & 1 & 1 & 1 \\ n_i & n_k & \bar{\gamma}_l & \bar{\gamma}_m \end{vmatrix}, \tag{92}$$

$$\bar{K}_{iklm} = \begin{vmatrix} M_i & 0 & 0 & 0 \\ 0 & \tilde{\bar{M}}_k & \tilde{\bar{M}}_l & \tilde{\bar{M}}_m \\ 1 & 1 & 1 & 1 \\ n_i & \bar{\gamma}_k & \bar{\gamma}_l & \bar{\gamma}_m \end{vmatrix}, \tag{93}$$

$$\tilde{\bar{K}}_{iklm} = \begin{vmatrix} M_i & M_k & M_l & 0 \\ 0 & 0 & 0 & \tilde{\bar{M}}_m \\ 1 & 1 & 1 & 1 \\ n_i & n_k & n_l & \bar{\gamma}_m \end{vmatrix}, \tag{94}$$

$$\bar{K}^n_{iklm} = \begin{vmatrix} M_{in} & 0 & 0 & 0 \\ 0 & \tilde{\bar{M}}_{kn} & \tilde{\bar{M}}_{ln} & \tilde{\bar{M}}_{mn} \\ 1 & 1 & 1 & 1 \\ n_i & \bar{\gamma}_k & \bar{\gamma}_l & \bar{\gamma}_m \end{vmatrix}, \tag{95}$$

$$\tilde{\bar{K}}^n_{iklm} = \begin{vmatrix} M_{in} & M_{kn} & M_{ln} & 0 \\ 0 & 0 & 0 & \tilde{\bar{M}}_{mn} \\ 1 & 1 & 1 & 1 \\ n_i & n_k & n_l & \bar{\gamma}_m \end{vmatrix}. \tag{96}$$

The symmetric matrices $\bar{\Pi}_{3434}$ and $\bar{\Pi}_{1212}$ represent the determinants of the TWNS in the absence of an electromagnetic field:

$$\bar{\Pi}_{3434} = (-ip_1)e^{-2ip_1a}\Delta_{10}(2), \tag{97}$$

$$\bar{\Pi}_{1212} = (-ip_{1n})e^{-2ip_{1n}a}\Delta_{1n}(2). \tag{98}$$

The properties of $\Delta_{10}(2)$ and $\Delta_{1n}(2)$ were studied in detail in [9], where it was shown that, near the resonance and for $V_{dc} \ll \varepsilon_R$, these determinants can be represented as

$$\Delta_{10}(2) \approx \frac{8i}{\Gamma t_0}[\varepsilon - \varepsilon_R^{(1)}(1) + i\Gamma][\varepsilon - \varepsilon_R^{(2)}(1) + i\Gamma], \tag{99}$$

$$\begin{aligned}\Delta_{1n}(2) &= \frac{8i}{\Gamma t_0}[\varepsilon - \omega - \varepsilon_R^{(1)}(1) + i\Gamma] \\ &\times [\varepsilon - \omega - \varepsilon_R^{(2)}(1) + i\Gamma].\end{aligned} \tag{100}$$

Here, $\varepsilon_R^{(1)}(1)$ and $\varepsilon_R^{(2)}(1)$ are the energies of levels in the TWNS that arise due to the splitting of the reso-

nance level of an isolated well and due to the applied voltage V_{dc},

$$\varepsilon_R^{(1,2)}(1) = p_0^2 + \frac{2p_0 x_1^{(1,2)}}{a}, \quad (101)$$

$$x_1^{(1,2)} = \delta\left[-(2+V_0) \pm \sqrt{1+V_0^2}\right], \quad V_0 = \frac{V_{dc}}{t_0}. \quad (102)$$

The value of energy splitting

$$t = \varepsilon_R^{(1)}(1) - \varepsilon_R^{(2)}(2) = t_0\sqrt{1+V_0^2} \quad (103)$$

increases with V_{dc}. For $V_{dc} \gg t_0$ ($V_0 \gg 1$), the splitting is determined only by the dc voltage; i.e., $t \approx V_{dc}$. The width Γ of levels does not depend on V_{dc} and is given by formula (36).

Note that, in expressions (101), (102), and (103), the terms on the order of δ^2, $(V_{dc}/\varepsilon_R)^2$, and $(V_{dc}/\varepsilon_R)\delta$ are dropped due to their awkwardness. They do not play any essential role and can again be included if necessary.

In the resonance case of interest, the frequency grows with V_{dc}:

$$\omega = t = t_0\sqrt{1+V_0^2}. \quad (104)$$

Assuming, as usual, that the energy of pumping electrons (that are brought to the structure from the emitter) is close to the energy of the upper level ($\xi = \varepsilon - \varepsilon_R^{(1)}(1) \ll t$), we can represent the determinants $\Delta_{10}(2)$ and $\Delta_{1n}(2)$ as

$$\Delta_{10}(2) \approx \frac{8i}{\Gamma}(\xi + i\Gamma)\sqrt{1+V_0^2}, \quad (105)$$

$$\Delta_{1n}(2) \approx -\frac{8i}{\Gamma}(\xi + i\Gamma)\sqrt{1+V_0^2}. \quad (106)$$

It follows from (105) and (106) that $\Delta_{10}(2)$ and $\Delta_{1n}(2)$ increase with V_{dc} due to the detuning of the exact resonance. This, in particular, reduces the transmission coefficient through the TWNS:

$$T(\varepsilon_R^{(1)}) = \frac{t_0^2}{t_0^2 + V_{dc}^2} = \frac{1}{1+V_0^2}. \quad (107)$$

The nonsymmetric matrices $\overline{\Pi}_{iklm}$ and $\overline{\Pi}^n_{iklm}$ describe electron transitions under an ac field between the energy levels within the first and second wells, whereas the matrices $\overline{K}_{iklm}$, $\tilde{K}_{iklm}$, $\overline{K}^n_{iklm}$, and $\tilde{\overline{K}}^n_{iklm}$ describe such transitions between the wells.

We begin with the calculation of the matrices K, which can be represented as

$$\begin{aligned} \overline{K}_{iklm} &= M_i\tilde{\overline{L}}_{klm}, \quad \tilde{\overline{K}}_{iklm} = \tilde{\overline{M}}_m L_{ikl}, \\ \overline{K}^n_{iklm} &= M_{in}\tilde{\overline{L}}^n_{klm}, \quad \tilde{\overline{K}}^n_{iklm} = \tilde{\overline{M}}_{mn} L^n_{ikl}. \end{aligned} \quad (108)$$

If we make the following changes:

$$p \longrightarrow p_1, \quad p_n \longrightarrow p_{1n}, \quad Z_n \longrightarrow Z_{1n},$$

where

$$Z_{1n} = e^{i(p_1 - p_{1n})a} - 1, \quad (109)$$

then $\tilde{\overline{L}}_{ikl}$ and $\tilde{\overline{L}}^n_{ikl}$ satisfy relations (39). Substituting (108) into (30), after certain transformations, we obtain

$$\begin{aligned} \overline{K} = {} & [M_4 L^n_{123} + M_3 L^n_{124}][\tilde{\overline{M}}_{14}\tilde{\overline{L}}_{234} + \tilde{\overline{M}}_{24}\tilde{\overline{L}}_{134}] \\ & + [M_{14}L_{234} + M_{24}L_{134}][\tilde{\overline{M}}_3\tilde{\overline{L}}^n_{124} + \tilde{\overline{M}}_4\tilde{\overline{L}}_{123}]. \end{aligned} \quad (110)$$

Taking into account the symmetry properties

$$\begin{aligned} \tilde{\overline{L}}_{234} &= -\overline{L}_{134}, \quad \tilde{\overline{L}}_{134} = -\overline{L}_{234}, \\ \tilde{\overline{L}}^n_{124} &= -\overline{L}^n_{123}, \quad \tilde{\overline{L}}^n_{123} = -\overline{L}^n_{124}, \\ \overline{M}_{14} &= -\tilde{\overline{M}}_{24}, \quad \overline{M}_{24} = -\tilde{\overline{M}}_{14}, \\ \overline{M}_3 &= -\tilde{\overline{M}}_4, \quad \overline{M}_4 = -\tilde{\overline{M}}_3, \end{aligned} \quad (111)$$

we arrive at the expression

$$\overline{K} = L_1\overline{L}_2 + \overline{L}_1 L_2, \quad (112)$$

where

$$\begin{aligned} L_1 &= M_4 L^n_{123} + M_3 L^n_{124}, \\ L_2 &= M_{14}L_{234} + M_{24}L_{134}, \\ \overline{L}_1 &= \overline{M}_4\overline{L}^n_{123} + \overline{M}_3\overline{L}^n_{124}, \\ \overline{L}_2 &= \overline{M}_{14}\overline{L}_{234} + \overline{M}_{24}\overline{L}_{134}. \end{aligned} \quad (113)$$

In the absence of a bias voltage ($V_{dc} = 0$), $\overline{L}_1 = L_1$, $\overline{L}_2 = L_2$, and $\overline{K}$ coincides with (42).

The calculation of L_i and $\overline{L}_i$ yields

$$\begin{aligned} L_1 = 2ip_n\Bigg\{ & |Z_n|^2\beta_n(2-\beta) + 2Z_n(\beta_n - \beta) \\ & + \frac{2x}{p_n} - \frac{x}{2p_n}e^{-i(p+p_n)a}\Psi_n\Psi_0 \Bigg\}, \end{aligned} \quad (114)$$

$$\bar{L}_1 = 2ip_{1n}\left\{|Z_{1n}|^2\beta_{1n}(2-\beta_1) + 2Z_{1n}(\beta_{1n}-\beta_1) + \frac{2x_1}{p_{1n}} - \frac{x_1}{2p_{1n}}e^{-i(p_1+p_{1n})a}\Psi_{1n}\Psi_{10}\right\}, \quad (115)$$

$$L_2 = 2ip\left\{|Z_n|^2\beta_n(2-\beta) + 2Z_n(\beta_n-\beta) - \frac{2x}{p} + \frac{x}{2p}e^{-i(p+p_n)a}\Psi_n\Psi_0\right\}, \quad (116)$$

$$\bar{L}_2 = 2ip_1\left\{|Z_{1n}|^2\beta_{1n}(2-\beta_1) + 2Z_{1n}(\beta_n-\beta) - \frac{2x_1}{p_1} + \frac{2x_1}{p_1}e^{-i(p_1+p_{1n})a}\Psi_{1n}\Psi_{10}\right\}. \quad (117)$$

Taking into account that

$$Z_{1n} \approx i\frac{\omega a}{2p_1}$$

and

$$\frac{x_1}{p_1} \approx \frac{\omega}{2p_1^2}$$

are small, we can easily show that the main contribution to (114)–(117) is made by the leading terms. Neglecting the small terms, we obtain

$$\bar{K} \approx 8|Z_n|^2|Z_{1n}|^2 ip_1 \cdot ip_{1n}\beta\beta_1\beta_n\beta_{1n}. \quad (118)$$

Again (like in the case of $V_{dc} = 0$), a number of terms are compensated due to the interference in the TWNS. In contrast to K, $\bar{K}$ involves frequency, which grows with V_{dc}:

$$Z_n = i\frac{at_0}{2p_n}\sqrt{1+V_0^2}, \quad Z_{1n} = i\frac{at_0}{2p_{1n}}\sqrt{1+V_0^2}. \quad (119)$$

Therefore, compared with K, $\bar{K}$ contains an additional term $(1 + V_0^2)^2$:

$$\bar{K} \approx ip_1 \cdot ip_{1n} \cdot 64 \cdot 2(1+V_0^2)^2. \quad (120)$$

After very tedious calculations, we can obtain the following expression for $\bar{\Pi}$:

$$\bar{\Pi} = ip_1 \cdot ip_{1n} \cdot 64 \cdot 2(1+V_0^2)^2. \quad (121)$$

To simplify the formulas, we dropped the small term i/π.

Thus, the total contribution to the determinant $\bar{\Delta}(\lambda)$ (see (28)),

$$\bar{K} + \bar{\Pi} = ip_1 \cdot ip_{1n} \cdot 64 \cdot 4(1+V_0^2)^2, \quad (122)$$

increases with the voltage V_{dc} faster than the contribution of $\Delta_{10}(2)\Delta_{1n}(2)$. However, this factor is canceled out by the factor that arises due to ω in the denominator of the expression for $\tilde{V}^2$ (see (37) and (21)). Hence, the voltage V_{dc} does not influence the contribution of the transitions between levels to $\bar{\Delta}(\lambda)$. This fact is consistent with the result of [9] that the matrix elements of the transitions are independent of V_{dc} (see formulas (134) below). Combining the results, we find the determinant of the TWNS for $V_{dc} \neq 0$:

$$\bar{\Delta}(\lambda) = \varepsilon_1^2\varepsilon_2^2\frac{64}{\Gamma^2}(1+V_0^2)\tilde{\bar{\Delta}}(\lambda), \quad (123)$$

$$\tilde{\bar{\Delta}}(\lambda) = [(\varepsilon - \varepsilon_R^{(1)}(1) + i\Gamma) \times (\varepsilon - \omega - \varepsilon_R^{(2)}(1) + i\Gamma) - \bar{\lambda}^2], \quad (124)$$

$$\bar{\lambda}^2 = 4\Gamma^2\tilde{V}^2(1+V_0^2) = \frac{V_{ac}^2}{64(1+V_0^2)}. \quad (125)$$

Comparing $\bar{\Delta}(\lambda)$ with $\Delta(\lambda)$, we can see that, in $\bar{\Delta}(\lambda)$, a factor $(1 + V_0^2)$ appears, whereas, in $\tilde{\bar{\Delta}}(\lambda)$, V_{ac}^2 is replaced by $V_{ac}^2/(1 + V_0^2)$. Thus, the growth of the product $\Delta_{10}(2)\Delta_{1n}(2)$ of resonance determinants weakens the effect of the ac field on the resonance tunneling. This result is of fundamental importance and leads to an increase in the oscillation efficiency.

In addition, when $V_{dc} \neq 0$, the applicability domain of the approximation $\tilde{V}^2 < 1$ (for $V_0^2 \gg 1$) expands because the strong-field criteria take the form

$$\bar{\lambda}^2 \geq \Gamma^2, \quad \tilde{V}^2 \geq \frac{1}{4(1+V_0^2)}, \quad \frac{V_{ac}}{\Gamma} \geq 8\sqrt{1+V_0^2}. \quad (126)$$

Let us determine the wavefunction of the first well for $V_{dc} \neq 0$. The coefficients A_{1j} are given by expressions (69)–(73), in which one should make the following substitutions:

$$\Pi_{iklm} \longrightarrow \bar{\Pi}_{iklm}, \ \Delta_0(1) \longrightarrow \Delta_{10}(1), \ \Delta_n(1) \longrightarrow \Delta_{1n}(1),$$

$$\Pi_0(1) \longrightarrow \Pi_{10}(1), \ \Pi_n(1) \longrightarrow \Pi_{1n}(1), \ M_1 \longrightarrow \bar{M}_1,$$

$$\Phi_1 \longrightarrow \bar{\Phi}_1, \ Z_n \longrightarrow Z_{1n},$$

where

$$\bar{M}_1 = [\Delta_{10}(1)\Pi_{1n}Z_{1n} + \Pi_{10}(1)\Delta_{1n}(1)Z_{1n}^*], \quad (127)$$

$$\bar{\Phi}_1 = 4[\beta_1(2-\beta_{1n})Z_{1n} + \beta_{1n}(2-\beta_1)Z_{1n}^*], \quad (128)$$

$$\Delta_{10}(1) = (2-\beta_1)\left(1+\frac{p_1}{p}-\beta\right)e^{-2ip_1a} - \beta_1\left(\beta+\frac{p_1}{p}-1\right), \quad (129)$$

$$\Delta_{1n}(1) = (2-\beta_{1n})\left(1+\frac{p_{1n}}{p_n}-\beta_n\right)e^{-2ip_{1n}a} - \beta_{1n}\left(\beta_n+\frac{p_{1n}}{p_n}-1\right), \quad (130)$$

$$\Pi_{10}(1) = (2-\beta_1)\left(\beta+1-\frac{p_1}{p}\right) + \beta_1\left(\beta+\frac{p_1}{p}+1\right)e^{2ip_1a}, \quad (131)$$

$$\Pi_{1n}(1) = \left(1+\beta_n-\frac{p_{1n}}{p_n}\right)(2-\beta_{1n}) + \beta_{1n}\left(1+\frac{p_{1n}}{p_n}+\beta_1\right)e^{2ip_{1n}a}. \quad (132)$$

We omit cumbersome calculations of the current; they are analogous to the case of $V_{dc} = 0$. We just note a few important moments.

Earlier [9], we showed that

$$\Delta_{1n}(1) \approx -\Pi_{1n}(1) = \frac{2i\alpha}{p_n}[V_0 - \sqrt{1+V_0^2}],$$
$$\Delta_{10}(1) \approx -\Pi_{10}(1) = \frac{2i\alpha}{p_0}[V_0 + \sqrt{1+V_0^2}] \quad (133)$$

in the leading-order approximation in p/α (for $p \approx p^{(1)}$). Using (133), we can easily verify that the field V_{dc} is dropped from the products $\Delta_{10}(1)\Pi_{1n}(1)$ and $\Pi_{10}(1)\Delta_{1n}(1)$, so that

$$\bar{M}_1 = M_1, \quad \bar{\Phi}_1 = \Phi_1. \quad (134)$$

Moreover, the coefficient A_{13} contains $\Delta_{10}(1)$, which, according to (133), depends on V_0. Taking into account the relation

$$\Delta_{1n}(2) + \Delta_{1n}^*(2) = 16\sqrt{1+V_0^2},$$

we arrive at the final expression for the current of the first well:

$$\bar{J}_{1c} = -\frac{EQ\Gamma^2\bar{\eta}}{\left|\tilde{\bar{\Delta}}(\lambda)\right|^2}, \quad \bar{\eta} = \frac{e^2p^2}{\alpha}\frac{[V_0+\sqrt{1+V_0^2}]}{(1+V_0^2)}, \quad (135)$$

$$\left|\tilde{\bar{\Delta}}(\lambda)\right|^2 = [(\xi+i\Gamma)(\xi+i\Gamma)-\bar{\lambda}^2], \quad (136)$$

$$\bar{\lambda} = \frac{V_{ac}}{8\sqrt{1+V_0^2}}. \quad (137)$$

In a weak field, $\bar{\lambda} \ll \Gamma$, the expression for the current $\bar{J}_{1c}$ coincides with the corresponding expression in [9]. The comparison of formulas (135)–(137) with the expression for the current of the coherent laser on a single-well structure [11] shows that a TWNS-based oscillator can be considered as a coherent laser in which radiative transitions occur between split levels of the TWNS in the presence of a dc voltage V_{dc}.

Based on this analogy, we can expect that, as shown in [3], there exists an optimal oscillation regime when the energy ξ_0 of electrons incident to the TWNS satisfies the condition

$$\xi_0 = 0, \quad \bar{\lambda} = \Gamma,$$
$$\xi_0^2 = \bar{\lambda}^2 - \Gamma^2, \quad \bar{\lambda} > \Gamma, \quad (138)$$
$$\frac{V_{ac}}{\Gamma} > 8\sqrt{1+V_0^2}.$$

Now, let us consider the radiation efficiency of photons as this was done in Section 5:

$$\tilde{\bar{P}} = \frac{\bar{P}}{2epq^2} = \frac{2\bar{\lambda}^2\Gamma^2[V_0+\sqrt{1+V_0^2}]}{\left|\tilde{\bar{\Delta}}(\lambda)\right|^2[1+V_0^2]}. \quad (139)$$

The expression for $\tilde{\bar{P}}$ differs from the expression for $\tilde{P}$ in that the former expression involves $\bar{\lambda}$ instead of λ and contains an additional factor

$$\frac{V_0+\sqrt{1+V_0^2}}{1+V_0^2}.$$

It is obvious that, for $\xi_0 = 0$ ($\bar{\lambda} = \Gamma$) and $\xi_0^2 = \bar{\lambda}^2 - \Gamma^2$ ($\bar{\lambda} > \Gamma$), the expression for the efficiency $\tilde{\bar{P}}$ reduces to

$$\tilde{\bar{P}} = \frac{1}{2}\frac{(V_0+\sqrt{1+V_0^2})}{\sqrt{1+V_0^2}}. \quad (140)$$

When $V_0 = 0$, the value $\tilde{\bar{P}} = 1/2$ coincides with (90), while, when $V_0 \gg 1$, we have $\tilde{\bar{P}} = 1$; i.e., the efficiency

attains its highest possible value when each electron brought to the TWNS emits a photon and passes through the structure without reflection.

This result is quite nontrivial. Indeed, the transition frequency ω is small compared with the energy of the resonance level of a quantum well, so that the parameter Z_{1n}, which characterizes the frequency resonance, is small: $Z_{1n} \ll 1$. For comparison, in the coherent laser [3], where $p = 2\pi/a$ and $p_n = \pi/a$, we have $Z_n = -2$.

The high efficiency of the TWNS is associated with the suppression of the influence of the electromagnetic field on the resonance tunneling due to the compensation of a number of terms, which is attributed to the interwell interference (to the coefficient p/α multiplying $\tilde{V}^2$ in the determinant). For example, the voltage V_{dc} additionally reduces the influence of the ac field due to the increase of the determinants $\Delta_{10}(2)$ and $\Delta_{1n}(2)$.

The voltage V_{dc} also results in an increase in the maximal value of the current for $\xi = 0$:

$$\bar{J}_{1c}^m = J_{1c}^m \frac{[V_0 + \sqrt{1 + V_0^2}]}{\sqrt{1 + V_0^2}}, \quad (141)$$

which is attained at

$$\frac{V_{ac}^m}{\Gamma} = \frac{8}{\sqrt{3}}\sqrt{1 + V_0^2}. \quad (142)$$

When $V_0 \gg 1$, the maximal value is doubled.

7. ELECTROMAGNETIC FIELD OSCILLATIONS IN THE TWNS

Taking into account the analogy between the polarization currents $\bar{J}_{1c}$ in the TWNS and in the coherent laser, we apply the theory of oscillations in the coherent laser [3] to describe the electromagnetic field oscillations in the TWNS. An equation for the field amplitude is obtained upon the substitution of the expression for the current $\bar{J}_{1c}$ (135) into (4):

$$1 = \frac{\Gamma^2 \tilde{Q}}{\left|\tilde{\bar{\Delta}}(\lambda)\right|^2}, \quad \tilde{Q} = \frac{4\pi\tau_0 Q\bar{\eta}}{\kappa}. \quad (143)$$

Generally speaking, we need an equation for the oscillation frequency. However, as is shown in [3], under the condition that the stabilization factor is small ($S = 1/\Gamma\tau_0 \ll 1$), the oscillation frequency coincides with the frequency of the resonator, $\omega = \Omega$. Since Ω can be varied, we will assume that ω is a free parameter.

Let us determine $\bar{\lambda}^2$ as a function of the pumping current $\tilde{Q}$ from Eq. (143):

$$\tilde{\bar{\lambda}}^2 = \xi^2 + \Gamma[\sqrt{\tilde{Q} - 4\xi^2} - \Gamma]. \quad (144)$$

The minimal threshold current Q_{th} is obtained if we set $\tilde{\bar{\lambda}} = 0$ and $\xi = 0$:

$$\tilde{Q}_{th} = \Gamma^2,$$
$$Q_{th} = \left[\frac{\kappa\Gamma^2}{4\pi\tau_0\eta_0}\right]\left[\frac{\alpha(1 + V_0^2)}{p(V_0 + \sqrt{1 + V_0^2})}\right], \quad (145)$$
$$\eta_0 = e^2 p.$$

The analysis of Q_{th} is carried out in Section 8 below.

Following [3] (see also Section 5), we can show that there exists an optimal oscillation regime when the energy of the electrons brought to the TWNS satisfies the condition

$$\xi_0^2 = \bar{\lambda}^2 - \Gamma^2, \quad \bar{\lambda} > \Gamma. \quad (146)$$

In this regime, $\bar{\lambda}^2$ is a linear function of the pumping current:

$$\bar{\lambda}^2 = \tilde{Q}/4 \quad (147)$$

and the efficiency is maximal. If we express $\bar{\lambda}^2$ in terms of the number N of photons of frequency ω that are emitted by the structure per unit time, then from (147) we obtain

$$N = \bar{P} = \frac{J_{1c}E2a}{2\omega} = Q\frac{V_0 + \sqrt{1 + V_0^2}}{\sqrt{1 + V_0^2}}. \quad (148)$$

Thus, for $V_0 \gg 1$, the oscillation efficiency is 100%.

8. CONCLUSIONS

It is of interest to compare the results obtained with the corresponding results for the RTD and the coherent laser. According to [11], the threshold current is given by

$$Q_{th}^{RTD} = \frac{\kappa}{\pi e^2 \tau_0 a} \times \frac{[(\varepsilon - \varepsilon_R - \omega)^2 + \Gamma^2][(\varepsilon - \varepsilon_R + \omega)^2 + \Gamma^2]}{(\varepsilon - \varepsilon_R)\Gamma^2}. \quad (149)$$

In the low-frequency regime $\omega \ll \Gamma$ and for $\varepsilon = \varepsilon_R + \Gamma$, this current is α/p times greater than Q_{th} for the TWNS (typical values of α/p vary from 20 to 100). In the high-frequency "quantum regime" [11] $\varepsilon - \varepsilon_R = \omega \gg \Gamma$, the ratio

$$\frac{Q_{th}^{RTD}}{Q_{th}^{TWNS}} \approx \frac{\alpha}{p}\frac{\omega}{\Gamma}$$

becomes still greater and reaches a value of $(\alpha/p)(\omega/\Gamma)^4$ in the classical regime ($\varepsilon - \varepsilon_R \sim \Gamma$ and $\omega \gg \Gamma$).

A comparison with the single-well coherent laser [3], in which radiative transitions occur between the resonance levels of the well ($p - p_n = \pm\pi/a$) and the frequency is $\omega \approx 3\pi^2/a^2$, shows that the threshold $Q_{\text{th}}^{\text{TWNS}}$ is greater by a factor of α/p. This is associated with the fact that the frequency ω is small compared with ε_R, because the parameter Z_{1n} (see (109)), which characterizes the intensity of radiative transitions $Z_n \sim \omega$, is small. However, it is this frequency that corresponds to the relevant terahertz band. At the same time, in the optimal oscillation regime of the TWNS, the oscillation efficiency reaches 100% even at this small terahertz-band frequency. The reason for this is quite nontrivial and is explained in Section 6.

There are other advantages of TWNS oscillators. First, they allow for the fine tuning of the oscillation frequency by a dc bias voltage V_{dc}, which separates the resonance levels. Second, the use of TWNS oscillators facilitates the fulfillment of the coherence condition because, since the oscillation frequency ω is small compared with the frequency ω_0 of an optical phonon, the most critical mechanism of relaxation is removed. Third, one can easily achieve an optimal oscillation regime because the widths of resonance levels are equal.

At the same time, the TWNS retains the advantages of the coherent laser, such as a high efficiency without inverse population, a narrow amplification loop, the absence of saturation phenomena, etc. This allows one to predict the prospectiveness of tunable high-frequency TWNS oscillators.

We can make one more prediction. As pointed out in the Introduction, the TWNS represents a simple structure on the way from the RTD toward the single-band Bloch superlattice. Therefore, one may expect that the phenomenon of compensation due to the interference will also be inherent in a coherent oscillator based on a single-band Bloch superlattice (the Stark ladder) and, hence, will lead to a high oscillation efficiency. For a two-band Stark ladder, the possibility of an optimal oscillation regime was proved in [12].

ACKNOWLEDGMENTS

This work was supported by the Federal Agency for Science and Innovations.

I am grateful to Yu.V. Kopaev and I.Yu. Kateev for fruitful discussions.

REFERENCES

1. A. Kazarinov and R. Suris, Sov. Phys. Semicond. **5**, 207 (1971).
2. J. Faist, F. Capasso, D. Sivco, *et al.*, Science **264**, 553 (1994).
3. V. F. Elesin, Zh. Éksp. Teor. Fiz. **112**, 483 (1997) [JETP **85**, 264 (1997)]; **122**, 131 (2002) [**95**, 114 (2002)].
4. V. M. Galitskiĭ and V. F. Elesin, *Resonance Interaction of Electromagnetic Fields with Semiconductors* (Énergoatomizdat, Moscow, 1986) [in Russian].
5. S. Haas, T. Stroucken, M. Hübner, *et al.*, Phys. Rev. B **57**, 14860 (1998).
6. I. A. Dmitriev and R. A. Suris, Fiz. Tekh. Poluprovodn. (St. Petersburg) **36**, 1460 (2002) [Semiconductors **36**, 1375 (2002)].
7. V. F. Elesin, JETP **96**, 966 (2003).
8. V. F. Elesin, V. V. Kapaev, Yu. V. Kopaev, and A. V. Tsukanov, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 709 (1997) [JETP Lett. **66**, 742 (1997)].
9. V. F. Elesin, Zh. Éksp. Teor. Fiz. **127**, 131 (2005) [JETP **100**, 116 (2005)].
10. V. F. Elesin and I. Yu. Kateev, Fiz. Tekh. Poluprovodn. (St. Petersburg) (in press).
11. V. F. Elesin, Zh. Éksp. Teor. Fiz. **116**, 704 (1999) [JETP **89**, 377 (1999)].
12. V. F. Elesin and Yu. V. Kopaev, Zh. Éksp. Teor. Fiz. **123**, 1308 (2003) [JETP **96**, 1149 (2003)].

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 811–816.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 938–943.
Original English Text Copyright © 2005 by Kotikov, Lipatov, Zotov.

NUCLEI, PARTICLES, FIELDS, GRAVITATION, AND ASTROPHYSICS

Longitudinal Structure Function F_L: Perturbative QCD and k_T-Factorization versus Experimental Data at Fixed W¶

A. V. Kotikov[a], A. V. Lipatov[b], and N. P. Zotov[b]

[a]*Institut für Theoretische Teilchenphysik, Universität Karlsruhe D-76128, Karlsruhe, Germany*
e-mail: kotikov@particle.uni-karlsruhe.de
[b]*Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119992 Russia*

Received June 13, 2005

Abstract—We use the results for the structure function F_L for a gluon target with a nonzero transverse momentum squared at the order α_s, obtained in our previous paper, for comparison with recent H1 experimental data for F_L at fixed W values and with collinear GRV predictions in the leading-order and next-to-leading-order approximations. © *2005 Pleiades Publishing, Inc.*

1. The longitudinal structure function $F_L(x, Q^2)$ is a very sensitive QCD characteristic and is directly related to the gluon content of the proton. It is equal to zero in the parton model with spin-1/2 partons and acquires nonzero values in the framework of perturbative QCD. The perturbative QCD, however, leads to quite controversial results. In the leading-order (LO) approximation, F_L amounts to about 10–20% of the corresponding F_2 values at large Q^2 and, thus, acquires quite large contributions at low x. The next-to-leading-order (NLO) corrections to the longitudinal coefficient function are large and negative at small x [1–5] and can lead to negative F_L values at low x and low Q^2 values (see [5, 6]). Negative F_L values demonstrate limitations of the applicability of the perturbation theory and the necessity of a resummation procedure that leads to the coupling constant scale higher than Q^2 (see [5, 7–9]).

The experimental extraction of the F_L data requires a rather cumbersome procedure, especially at small values of x (e.g., see [10]). However, new precise preliminary H1 data [11] on the longitudinal structure function F_L presented recently have probed the small-x region $10^{-5} \le x \le 10^{-2}$.

In this paper, the standard perturbative QCD formulas and the so-called k_T-factorization approach [12] based on the Balitsky–Fadin–Kuraev–Lipatov (BFKL) dynamics [13] (see also [14] and references therein) are used for the analysis of the above data. The perturbative QCD approach is hereafter called the collinear approximation and is applied at the LO and NLO levels using Gluck–Reya–Vogt (GRV) parametrizations for parton densities (see [15]). The corresponding coefficient functions are taken from papers [1–3].

In the framework of the k_T-factorization approach, which we primary consider in this paper, the longitudinal structure function F_L was first studied in [16], where the small-x asymptotics of F_L was obtained analytically using the BFKL results for the Mellin transform of the unintegrated gluon distribution, and the longitudinal Wilson coefficient functions for the full perturbative series were calculated at asymptotically small x values. In this paper, we follow a more phenomenological approach in [17], where we analyzed the F_L data in a broader range at small x; we thus use parametrizations of the unintegrated gluon distribution function $\Phi_g(x, k_\perp^2)$ (see [14]).

A similar study has already been done in our paper [17] using previous H1 data [18].[1] The recent H1 preliminary experimental data [11] is essentially more precise, which stimulates the present additional study.

2. The unintegrated gluon distribution $\Phi_g(x, k_\perp^2)$ (where f_g is the integrated gluon distribution in the proton multiplied by x and $k_\perp$ is the transverse part of the gluon 4-momentum k^μ),

$$f_g(x, Q^2) = \int^{Q^2} dk_\perp^2 \Phi_g(x, k_\perp^2) \qquad (1)$$

¶ The text was submitted by the authors in English.

[1] We note that the F_L structure function has also been studied in the framework of the k_T-factorization in [19, 20].

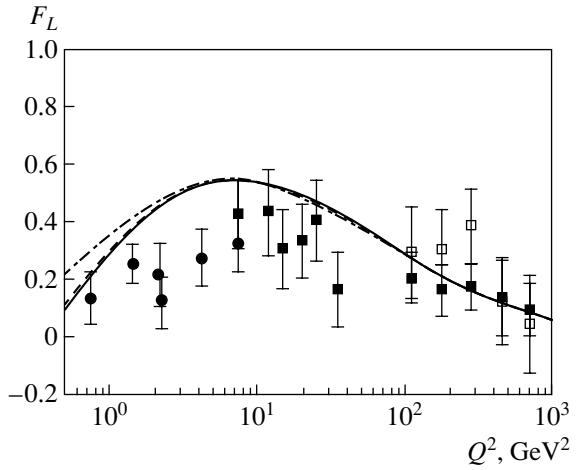


Fig. 1. Q^2 dependence of $F_L(x, Q^2)$ (at fixed $W = 276$ GeV). The H1 preliminary e^+p and e^-p experimental data are shown as black points and black and white squares, respectively (see [11]). Theoretical curves are obtained in the k_T-factorization approach with the JB unintegrated gluon distribution: the solid curve corresponds to a "frozen" coupling constant; the dashed curve, to the analytical coupling constant; and the dash-dotted curve, to the frozen argument in the coupling constant and the unintegrated gluon distribution function.

(herein, $k_2 = -k_\perp^2$) is the basic dynamical quantity in the k_T-factorization approach.[2] It satisfies the BFKL equation [13].

In the k_T-factorization approach, the structure functions $F_{2,L}(x, Q^2)$ are driven at small x primarily by gluons and are related to the unintegrated distribution $\Phi_g(x, k_\perp^2)$ as

$$F_{2,L}(x, Q^2) = \int_x^1 \frac{dz}{z} \int^{Q^2} dk_\perp^2 \times \sum_{i = u, d, s, c} e_i^2 \hat{C}_{2,L}^g(x/z, Q^2, m_i^2, k_\perp^2) \Phi_g(z, k_\perp^2), \quad (2)$$

where e_i are charges of active quarks.

The functions $\hat{C}_{2,L}^g(x, Q^2, m_i^2, k_\perp^2)$ can be regarded as structure functions of the off-shell gluons with virtuality $k_\perp^2$ (hereafter, we call them hard structure functions in analogy to similar relations between cross sections and hard cross sections). They are described by the sum of the quark box (and crossed box) diagram contribution to the photon–gluon interaction (e.g., see Fig. 1 in [17] and [21]).

3. We note that the $k_\perp^2$-integral in Eqs. (1) and (2) can be divergent at the lower limit, at least for some parametrizations of $\Phi_g(x, k_\perp^2)$. To overcome this problem, we change the low Q^2 asymptotics of the QCD coupling constant within hard structure functions. We here apply two models: the "freezing" procedure and the Shirkov–Solovtsov analytization.

The "freezing" of the strong coupling constant is a very popular phenomenological model for the infrared behavior of $\alpha_s(Q^2)$ (e.g., see [22]). The "freezing" can be done in the hard way and in the soft way.

In the hard case (e.g., see [23]), the strong coupling constant is itself modified: it is taken to be constant at all Q^2 values less than some Q_0^2, i.e.,

$$\alpha_s(Q^2) = \alpha_s(Q_0^2) \text{ if } Q^2 \le Q_0^2.$$

In the soft case (e.g., see [20]), the subject of the modification is the argument of the strong coupling constant. It contains the shift $Q^2 \longrightarrow Q^2 + M^2$, where M is an additional scale, which strongly modifies the infrared α_s properties. For massless produced quarks, the ρ-meson mass m_ρ is usually taken as the M value, i.e., $M = m_\rho$. In the case of massive quarks with a mass m_i, the $M = 2m_i$ value is typically used. Below, we use the soft version of the "freezing" procedure.

Shirkov and Solovtsov proposed [24] a procedure of analytization of the strong coupling constant $\alpha_s(Q^2)$, which leads to a new strong analytical coupling constant $a_{an}(Q^2)$ having nonstandard infrared properties. Here, we do not discuss theoretical aspects of the procedure and use only the final formulas for the analytical coupling constant $a_{an}(Q^2)$. They are given by

$$\frac{a_{an}(Q^2)}{4\pi} = \frac{1}{\beta_0}\left[\frac{1}{\ln(Q^2/\Lambda^2)} + \frac{\Lambda^2}{\Lambda^2 - Q^2}\right] \quad (3)$$

in the LO approximation and

$$\frac{a_{an}(Q^2)}{4\pi} = \frac{1}{\beta_0}\left[\frac{1}{\ln(Q^2/\Lambda^2) + b_1 \ln[1 + \ln(Q^2/\Lambda^2)/b_1]} + \frac{1}{2}\frac{\Lambda^2}{\Lambda^2 - Q^2} - \frac{\Lambda^2}{Q^2} C_1\right], \quad (4)$$

in the NLO approximation, where β_0 and β_1 are the first two terms in the α_s-expansion of the β-function and $b_1 = \beta_1/\beta_0^2$. The constant $C_1 = 0.0354$ is very small.

[2] In our previous analysis [21], we have shown that the property $k^2 = -k_\perp^2$ leads to the equality of the Bjorken x value in the standard renormalization-group approach and in the Sudakov approach.

The first terms in the right-hand sides of Eqs. (3) and (4) are the standard LO and NLO representations for $\alpha_s(Q^2)$. The additional terms modify its infrared properties.

We note that numerically, both infrared transformations, the "freezing" procedure and the Shirkov–Solovtsov analytization, lead to very close results (see Fig. 1, as well as [25] and the discussion therein).

4. As was already noted above, the purpose of this paper is to describe new preliminary H1 experimental data for the longitudinal SF $F_L(x, Q^2)$ using our calculations of the hard structure function $\hat{C}^g_{2,L}(x, Q^2, m^2, k_\perp^2)$ given in our previous study [21] and infrared modifications of $\alpha_s(Q^2)$ explained above. For the unintegrated gluon distribution $\Phi(x, k_\perp^2, Q_0^2)$, we use the so-called Blumlein's parameterization (JB) [26]. We note that there are also several other popular parametrizations, which give quite similar results, with a possible exception of contributions from the small-$k_\perp^2$ range $k_\perp^2 \le 1$ GeV^2 (see [14] and the references therein).

The JB form depends strongly on the Pomeron intercept value. In different models, the Pomeron intercept has different values (see [27]). In our calculations, we apply the H1 parameterization [28] based on the corresponding H1 data, which are in good agreement with perturbative QCD (see [28, 29]).

We calculate the structure function F_L as the sum of two types of contributions: that of the charm quark, F_L^c, and of the light quark, F_L^l:

$$F_L = F_L^l + F_L^c. \quad (5)$$

For the F_L^l part, we use the massless limit of the hard structure function (see [17, 21]). We always use $f = 4$ in our fits, because our results depend on the exact f value very weakly (for similar results, see fits of experimental data in [30] and discussions therein). The weak dependence comes from two basic properties. First, the charm part of F_L, F_L^c, is quite small at the considered Q^2 values (see [17] for the F_L^c study). Second, the strong coupling constant depends on f very weakly because of the corresponding relations between Λ values at different f (see [31]).

In Fig. 1, we show the structure function F_L with "frozen" and analytical coupling constants, respectively, as functions of Q^2 for fixed W in comparison with the H1 experimental data sets (see [11]). The results mostly coincide with each other. They are presented as solid and dashed curves, which cannot be actually resolved in the figure.

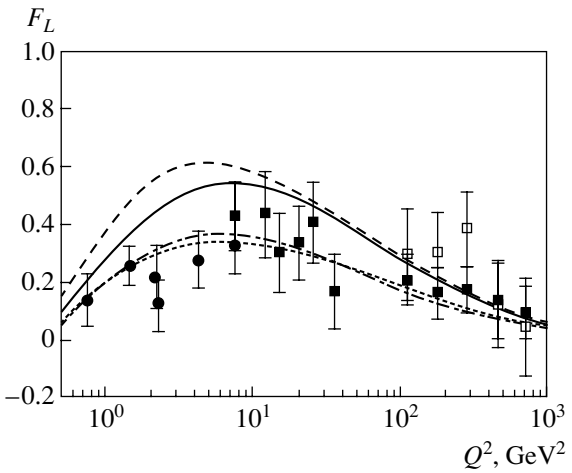


Fig. 2. Q^2 dependence of $F_L(x, Q^2)$ (at fixed $W = 276$ GeV). The experimental points are as in Fig. 1. The solid curve is the result of the k_T-factorization approach with the JB unintegrated gluon distribution and frozen coupling constant, the dashed curve is the GRV LO calculations, the dash-dotted curve is the GRV NLO calculations, and the dotted curve is the result of the GRV LO calculations with $\mu^2 = 127Q^2$.

The dash-dotted curve shows the results obtained with a "frozen" argument also added to the unintegrated gluon density. The difference between the solid and dash-dotted lines is not very big, which demonstrates the unimportance of the infrared modifications of the density argument. Below, we restrict ourselves only to the modification of the argument in the strong coupling constant entering the hard structure function.

Figure 2 contains the same solid curve as Fig. 1 and also shows the collinear results for F_L values. We use the popular GRV parametrizations [15] in the LO and NLO approximations. The k_T-factorization results lie between the collinear ones, which clearly demonstrates the particular resummation of high-order collinear contributions at small x values in the k_T-factorization approach.

We also see excellent agreement between the experimental data and the collinear approach with GRV parton densities in the NLO approximation. The NLO corrections are large and negative and reduce the F_L value by approximately a factor of 2 at $Q^2 < 10$ GeV^2.

In Figs. 1 and 2, our k_T-factorization results are in good agreement with the data for large and small parts of the Q^2 range. We have, however, some disagreement between the data and theoretical predictions at $Q^2 \approx 3$ GeV^2. The disagreement exists in both cases: for the collinear QCD approach in the LO approximation and for the k_T-factorization approach.

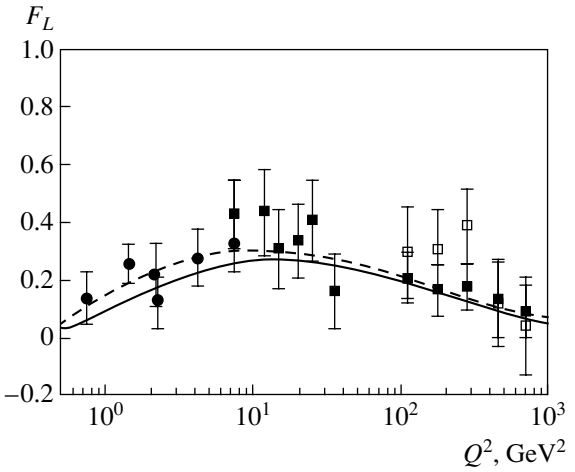


Fig. 3. Q^2 dependence of $F_L(x, Q^2)$ (at fixed $W = 276$ GeV). The experimental points are as in Fig. 1. The solid curve is the result of the k_T-factorization approach with the JB unintegrated gluon distribution and $\mu^2 = M_Z^2$; the dashed curve is the GRV LO calculations at $\mu^2 = M_Z^2$.

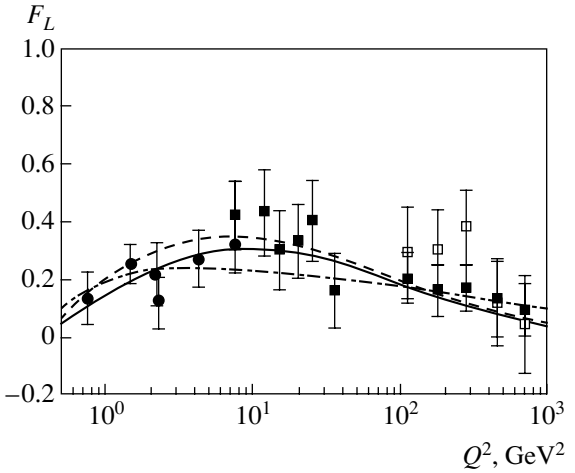


Fig. 4. Q^2 dependence of $F_L(x, Q^2)$ (at fixed W = 276 GeV). The experimental points are as in Fig. 1. The solid curve is the result of the k_T-factorization approach with the JB unintegrated gluon distribution and at $\mu^2 = 127Q^2$, the dashed curve is the GRV LO calculations at $\mu^2 = 127Q^2$, and the dash-dotted curve is from the R_{world}-parametrization.

Comparing these results with Fig. 4 in Lobodzinska's talk in [11], we conclude that the disagreement comes from the use of the LO approximation. Unfortunately, at the moment, only the LO terms are available in the k_T-factorization approach. The calculation of the NLO corrections is a very complicated problem (see [32] and the discussion therein).

A rough estimate of the NLO corrections in the k_T-factorization approach can be performed as follows. We first consider the BFKL approach. A popular resummation of the NLO corrections is done in [8] in some approximation. It is demonstrated in [8] that the basic effect of the NLO corrections is a strong rise in the α_s argument from Q^2 to $Q^2_{\text{eff}} = KQ^2$, where $K = 127$, i.e., $K \gg 1$, which is in agreement with [5, 7, 9].

The use of the effective argument Q^2_{eff} in the DGLAP approach in the LO approximation leads to results that are very close to the ones obtained in the NLO approximation (see the dot-dashed and dotted curves in Fig. 2). Thus, we hope that the effective argument represents the basic effect of the NLO corrections in the k_T-factorization framework, which in a certain sense lies between the DGLAP and BFKL approaches, as was already noted above.

The necessity of large effective arguments is also demonstrated in Fig. 3, where we show the k_T-factorization and collinear results for the nonrunning coupling constant. Its argument is fixed at $Q^2 = M_Z^2$, giving $\alpha_s \approx 0.118$ (see [33]); i.e., the considered argument is larger than the majority of the Q^2-values of the considered experimental data.[3]

The results obtained in the k_T-factorization and collinear approaches based on Q^2_{eff} argument are presented in Fig. 4. In comparison with the ones shown in Fig. 1, they are close to each other because the effective argument is substantially larger than the Q^2 value. There is a very good agreement between the experimental data and both theoretical approaches.

Moreover, in Fig. 4, we also present the F_L results based on the R_{world} parameterization for the $R = \sigma_L/\sigma_T$ ratio (see [34]) (because $F_L = F_2R/(1 + R)$), improved in [35, 36] for low Q^2 values and the parameterization of F_2 data used in our previous paper [17]. The results are in good agreement with other theoretical predictions as well as with experimental data.

5. CONCLUSIONS

In the k_T-factorization framework, we have applied the results of the calculation of the perturbative parts for the structure functions F_L and F_L^c for a gluon target having squared nonzero momentum, in the process of photon–gluon fusion [17, 21], to an analysis of recent H1 preliminary data. The perturbative QCD predictions are also represented in the LO and NLO approximations.

We have found very good agreement between the experimental data and collinear results based on the GRV parameterization in the NLO approximation. The LO collinear and k_T-factorization results show disagreement with the data at some Q^2 values. We argued

[3] The study was also initiated by a conversation with L. Lönnblad, whom we thank.

that the disagreement comes from the absence of NLO corrections in the k_T-factorization approach. Another reason is discussed in [36]. We modeled these NLO corrections by choosing a large effective argument of the strong coupling constant and argued for our choice. The effective corrections significantly improve the agreement with the H1 data under consideration.

ACKNOWLEDGMENTS

We thank S.P. Baranov for careful reading of the manuscript and useful remarks. Our study was supported in part by an a grant from the Russian Foundation for Basic Research. One of the authors (A.V.K.) was supported in part by an Alexander von Humboldt fellowship. A.V.L. was supported in part by INTAS (YSF-2002 grant no. 399) and Dinastiya Foundation. N.P.Z. also acknowledges L. Jönsson for a discussion of the H1 data [11] and support from the Crafoord Foundation (Sweden).

REFERENCES

1. W. L. van Neerven and E. B. Zijlstra, Phys. Lett. B **272**, 127 (1991); E. B. Zijlstra and W. L. van Neerven, Phys. Lett. B **273**, 476 (1991); Nucl. Phys. B **383**, 525 (1992).
2. D. I. Kazakov and A. V. Kotikov, Theor. Math. Phys. **73**, 1264 (1987); Nucl. Phys. B **307**, 721 (1988); Nucl. Phys. B **345**, 299 (1990); Phys. Lett. B **291**, 171 (1992); D. I. Kazakov, A. V. Kotikov, G. Parente, *et al.*, Phys. Rev. Lett. **65**, 1535 (1990).
3. J. Sanchez Guillen, J. Miramontes, M. Miramontes, *et al.*, Nucl. Phys. B **353**, 337 (1991).
4. S. Keller, M. Miramontes, G. Parente, *et al.*, Phys. Lett. B **270**, 61 (1990); L. H. Orr and W. J. Stirling, Phys. Rev. Lett. **66**, 1673 (1991); E. Berger and R. Meng, Phys. Lett. B **304**, 318 (1993).
5. A. V. Kotikov, JETP Lett. **59**, 1 (1994); Phys. Lett. B **338**, 349 (1994).
6. A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, Eur. Phys. J. C **23**, 73 (2002).
7. Yu. L. Dokshitzer and D. V. Shirkov, Z. Phys. C **67**, 449 (1995); W. K. Wong, Phys. Rev. D **54**, 1094 (1996).
8. S. J. Brodsky, V. S. Fadin, V. T. Kim, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 161 (1999) [JETP Lett. **70**, 155 (1999)].
9. M. Ciafaloni, D. Colferai, and G. P. Salam, Phys. Rev. D **60**, 114036 (1999); J. High Energy Phys. **07**, 054 (2000); R. S. Thorne, Phys. Lett. B **474**, 372 (2000); Phys. Rev. D **60**, 054031 (1999); Phys. Rev. D **64**, 074005 (2001); G. Altarelli, R. D. Ball, and S. Forte, Nucl. Phys. B **621**, 359 (2002).
10. A. M. Cooper-Sarkar, G. Ingelman, K. R. Long, *et al.*, Z. Phys. C **39**, 281 (1988); L. Bauerdick, A. Glazov, and M. Klein, in *Proceedings of the International Workshop on Future Physics on HERA* (DESY, Hamburg, 1996), p. 77; hep-ex/9609017.
11. E. M. Lobodzinska, hep-ph/0311180; P. Newman, hep-ex/0312018.
12. S. Catani, M. Ciafaloni, and F. Hautmann, Nucl. Phys. B **366**, 135 (1991); in *Proceedings of the Workshop on Physics at HERA* (Hamburg, 1991), Vol. 2, p. 690; J. C. Collins and R. K. Ellis, Nucl. Phys. B **360**, 3 (1991); E. M. Levin, M. G. Ryskin, Yu. M. Shabelskii, and A. G. Shuvaev, Sov. J. Nucl. Phys. **53**, 657 (1991).
13. L. N. Lipatov, Yad. Fiz. **23**, 642 (1976) [Sov. J. Nucl. Phys. **23**, 338 (1976)]; E. A. Kuraev, L. N. Lipatov, and V. S. Fadin, Zh. Éksp. Teor. Fiz. **71**, 840 (1976) [Sov. Phys. JETP **44**, 443 (1976)]; **72**, 377 (1977) [**45**, 199 (1977)]; Ya. Ya. Balitski and L. N. Lipatov, Yad. Fiz. **28**, 1597 (1978) [Sov. J. Nucl. Phys. **28**, 822 (1978)]; L. N. Lipatov, Zh. Éksp. Teor. Fiz. **90**, 1536 (1986) [Sov. Phys. JETP **63**, 904 (1986)].
14. Bo Andersson *et al.* (Small x Collab.), Eur. Phys. J. C **25**, 77 (2002); J. Andersen *et al.* (Small x Collab.), hep-ph/0312333.
15. M. Gluck, E. Reya, and A. Vogt, Z. Phys. C **67**, 433 (1995).
16. S. Catani and F. Hautmann, Nucl. Phys. B **427**, 475 (1994); S. Catani, Preprint DFF 254-7-96; hep-ph/9608310.
17. A. V. Kotikov, A. V. Lipatov, and N. P. Zotov, Eur. Phys. J. C **27**, 219 (2003); A. V. Kotikov, A. V. Lipatov, G. Parente, and N. P. Zotov, Lect. Notes Phys. **647**, 386 (2004).
18. S. Aid *et al.* (H1 Collab.), Phys. Lett. B **393**, 452 (1997); N. Gogitidze, J. Phys. G **28**, 751 (2002); hep-ph/0201047.
19. J. Blumlein, J. Phys. G **19**, 1623 (1993).
20. B. Badelek, J. Kwiecinski, and A. Stasto, Z. Phys. C **74**, 297 (1997).
21. A. V. Kotikov, A. V. Lipatov, G. Parente, and N. P. Zotov, Eur. Phys. J. C **26**, 51 (2002); in *Proceedings of the XVI International Workshop on High Energy Physics and Quantum Field Theory* (Moscow, 2002), p. 230; hep-ph/0208195.
22. G. Curci, M. Greco, and Y. Srivastava, Phys. Rev. Lett. **43**, 834 (1979); Nucl. Phys. B **159**, 451 (1979); M. Greco, G. Penso, and Y. Srivastava, Phys. Rev. D **21**, 2520 (1980); M. Greco and PLUTO Collaboration, Phys. Lett. B **100B**, 351 (1981).
23. N. N. Nikolaev and B. M. Zakharov, Z. Phys. C **49**, 607 (1991); **53**, 331 (1992).
24. D. V. Shirkov and L. I. Solovtsov, Phys. Rev. Lett. **79**, 1209 (1997); Theor. Math. Phys. **120**, 1220 (1999).
25. A. Yu. Illarionov and A. V. Kotikov, private communication.
26. J. Blumlein, Preprints DESY 95-121; hep-ph/9506403; I. P. Ivanov and N. N. Nikolaev, Phys. Rev. D **65**, 054004 (2002).
27. A. B. Kaidalov, in *At the Frontier of Particle Physics: Handbook of QCD, Boris Ioffe Festschrift*, Ed. by M. Shifman (World Sci., Singapore, 2001), Vol. 1, p. 603; hep-ph/0103011; A. Capella, A. Kaidalov, C. Merino, and J. Tran Thanh Van, Phys. Lett. B **337**, 358 (1994).

28. C. Adloff *et al.* (H1 Collab.), Phys. Lett. B **520**, 183 (2001).
29. A. V. Kotikov and G. Parente, Zh. Éksp. Teor. Fiz. **124**, 963 (2003) [JETP **97**, 859 (2003)].
30. V. G. Krivokhijine and A. V. Kotikov, Preprint JINR E2-2001-190 (Joint Inst. for Nuclear Research, Dubna, 2001); hep-ph/0108224; Acta Phys. Slov. **52**, 227 (2002).
31. W. Marciano, Phys. Rev. D **29**, 580 (1984); K. G. Chetyrkin, B. A. Kniehl, and M. Steinhauser, Phys. Rev. Lett. **79**, 2184 (1997).
32. H. Jung, Nucl. Phys. (Proc. Suppl.) **79**, 429 (1999); hep-ph/9908497; hep-ph/0312066.
33. S. Bethke, J. Phys. C **26**, R27 (2000).
34. L. W. Whitlow *et al.* (SLAC Collab.), Phys. Lett. B **250**, 193 (1990).
35. U. K. Yang *et al.*, J. Phys. G **22**, 775 (1996); A. Bodek, in *Proceedings of the 4th International Workshop on Deep Inelastic Scattering, DIS96* (Rome, 1996), p. 213; A. Bodek, S. Rock, and U. K. Yang, Preprint Univ. Rochester UR-1355 (1995).
36. U. K. Yang *et al.* (CCFR/NuTeV Collab.), Phys. Rev. Lett. **87**, 251802 (2001); A. Bodek (CCFR/NuTeV Collab.), in *Proceedings of the 9th International Workshop on Deep Inelastic Scattering, DIS* (Bologna, 2001); hep-ex/0105067; A. V. Kotikov, A. V. Lipatov, and N. P. Zotov, in *Proc. XVIII Baldin Seminar "Relativistic Nuclear Physics and QCD," JINR, Dubna, Russia* (2004); hep-ph/0503275.

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 817–829.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 944–957.
Original Russian Text Copyright © 2005 by Popov.

NUCLEI, PARTICLES, FIELDS, GRAVITATION, AND ASTROPHYSICS

Feynman Disentangling of Noncommuting Operators in Quantum Mechanics

V. S. Popov

Institute for Theoretical and Experimental Physics, Russian Academy of Sciences, Moscow, 117218 Russia
e-mail: markina@itep.ru

Received June 22, 2005

Abstract—Feynman's disentangling theorem is applied to noncommuting operators in the problem of quantum parametric oscillator, which is mathematically equivalent to the problem of $SU(1, 1)$ pseudospin rotation. The number states of the oscillator correspond to unitary irreducible representations of the $SU(1, 1)$ group. Feynman disentangling is combined with group-theoretic arguments to obtain simple analytical formulas for the matrix elements and transition probabilities between the initial and final states of the oscillator. Feynman disentangling of time evolution operators is also discussed for an atom or ion interacting with a laser field and for a model Hamiltonian possessing the "hidden" symmetry of the hydrogen atom. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Feynman disentangling (FD) of noncommuting operators was applied in [1] to the quantum harmonic oscillator driven by an arbitrary force $f(t)$ and to quantum electrodynamics [1]. The expression for the transition probability between the number states of the oscillator with quantum numbers m and n at $t \longrightarrow -\infty$ and $+\infty$ derived in [1] was considerably simplified in [2]. In [3], FD was used to calculate the time-varying polarization state of a particle with arbitrary spin interacting with an arbitrary applied magnetic field $\mathbf{H}(t)$. However, FD has never been widely used.

In this study, FD is applied to the quantum parametrically driven oscillator

$$\hat{H} = \frac{1}{2}p^2 + \frac{1}{2}\omega^2(t)q^2, \quad \hbar = m = 1, \tag{1.1}$$

to obtain a disentangled expression for the evolution operator $\hat{S}(t, t_0)$. The problem is examined from a group-theoretic perspective, and its relationship to unitary irreducible representations of the group $SU(1, 1)$ is established, which makes it possible to avoid cumbersome algebra in finding analytical expressions for transition probabilities w_{mn} (in terms of Wigner D functions for this group). FD is applied to describe an atom or ion interacting with electric field $\mathscr{E}(t)$ and to disentangle the evolution operator for a model Hamiltonian possessing the "hidden" symmetry of the hydrogen atom. In Appendices A–C, the Wigner D functions corresponding to the representations of interest for this study are obtained in explicit form, exact solutions to the Riccati equation arising in FD applications are considered, and some details of calculations are given.

The results presented here were announced, in part, in [4].

2. FEYNMAN DISENTANGLING AND GROUP THEORY

The Hamiltonian of the harmonic oscillator contains noncommuting operators: $[p, q] = -i$. The frequency $\omega(t)$ can be any function of time satisfying the condition

$$\omega(t) \longrightarrow \begin{cases} \omega_-, & t \longrightarrow -\infty, \\ \omega_+, & t \longrightarrow +\infty, \end{cases} \tag{2.1}$$

which makes it possible to specify initial and final oscillator number states, $|m, \omega_-\rangle$ and $|n, \omega_+\rangle$. Defining

$$\hat{A} = \frac{1}{2}q^2, \quad \hat{B} = \frac{1}{2}p^2, \quad \hat{C} = \frac{1}{4}(pq + qp), \tag{2.2}$$

we have a closed operator algebra, as in the problem examined in [1]:

$$\begin{gathered} \hat{H} = \omega^2(t)\hat{A} + \hat{B}, \\ [\hat{A}, \hat{B}] = 2i\hat{C}, \quad [\hat{A}, \hat{C}] = i\hat{A}, \quad [\hat{B}, \hat{C}] = -i\hat{B}. \end{gathered} \tag{2.3}$$

However, straightforward application of FD to the evolution operator

$$\hat{S}(t, t_0) = T\exp\left\{-i\int_{t_0}^{t}\left(\frac{1}{2}p^2 + \omega^2(t')q^2\right)dt'\right\} \tag{2.4}$$

does not lead to the desired result, because the operator to be disentangled is again contained in the exponential at the next step of the process. This difficulty is elimi-

nated by using a procedure suggested in [3]. Rewriting the integrand as

$$\hat{H} = \frac{1}{2}p^2 + \frac{1}{2}\omega(t)q^2 \equiv \chi\hat{A} + (\omega^2 - \chi)\hat{A} + \hat{B}, \quad (2.5)$$

without specifying the function $\chi(t)$, we apply Feynman's disentangling theorem [1] to the first term:

$$\hat{S}(t, t_0) = \exp(-a(t)\hat{A}) \times T\exp\left\{-i\int_{t_0}^{t}\{(\omega^2 - \chi)\hat{A}' + \hat{B}'\}dt'\right\}, \quad (2.6)$$

where

$$a(t) = i\int_{t_0}^{t}\chi(t')dt', \qquad \hat{O}' \equiv \exp(a\hat{A})\hat{O}\exp(-a\hat{A}). \quad (2.6')$$

Note that t' in both (2.4) and (2.6) is an ordering parameter in the sense of [1]. The operator $\hat{O}'$ is also introduced by following [1]. Differentiating (2.6') with respect to a and using commutation relations (2.3), we obtain

$$\hat{A}' = \hat{A}, \quad \hat{C}' = \hat{C} + ia\hat{A}, \qquad \hat{B}' = \hat{B} + 2ia\hat{C} - a^2\hat{A}. \quad (2.7)$$

The condition for disentangling the exponential operator expression above with respect to the operator $\hat{A}$ is the following Riccati equation for $a(t)$:

$$\dot{a} = i[\omega^2(t) - a^2]. \quad (2.8)$$

After disentangling the operator expression with respect to $\hat{C}$, the operator $\hat{S}(t, t_0)$ is written in the desired completely disentangled form as

$$\hat{S}(t, t_0) = \exp\left(-\frac{a}{2}q^2\right)\exp\left\{\frac{c}{2}(pq + qp)\right\} \times \exp\left(-\frac{b}{2}p^2\right), \quad (2.9)$$

where $b(t)$ and $c(t)$ are expressed in terms of $a(t)$ as quadratures:

$$c(t) = \int_{t_0}^{t}a(\tau)d\tau, \qquad b(t) = i\exp\left[-i\int_{t_0}^{t}c(\tau)d\tau\right]. \quad (2.8')$$

To calculate the transition amplitude between initial and final states $|m, \omega_-\rangle$ and $|n, \omega_+\rangle$,

$$A_{mn} = \langle n, \omega_+|\hat{S}(\infty, -\infty)|m, \omega_-\rangle, \quad (2.10)$$

one should write the initial and final wavefunctions in the p and q representations, respectively, using the fact that the operator $\hat{C}$ has a continuous spectrum of eigenvalues, $-\infty < \lambda < \infty$:

$$\langle x|\lambda\rangle = \pi^{-1/2}x^{-1/2+2i\lambda}, \quad \langle p|\lambda\rangle = \pi^{-1/2}p^{-1/2-2i\lambda}, \quad \langle\lambda|\lambda'\rangle = \delta(\lambda - \lambda').$$

This completes the formal solution, but the resulting integrals are very difficult to calculate analytically. To derive compact expressions for w_{mn}, FD should be combined with group-theoretic arguments, as done in [3].

To do this, let us change from operators (2.2) to their linear combinations

$$\hat{J}_1 = \frac{1}{4}(pq + qp), \quad \hat{J}_2 = \frac{1}{4}(q^2 - p^2), \qquad \hat{J}_0 = \frac{1}{4}(p^2 + q^2), \quad (2.11)$$

which satisfy the commutation relations for the generators of the $SU(1, 1)$ group,

$$[\hat{J}_+, \hat{J}_-] = -2\hat{J}_0, \quad [\hat{J}_0, \hat{J}_\pm] = \pm\hat{J}_\pm, \qquad [\hat{\mathbf{J}}^2, \hat{J}_a] = 0. \quad (2.12)$$

Here, $\hat{J}_\pm = \hat{J}_1 \pm i\hat{J}_2 = \pm i(q \mp ip)^2/4$ and $\hat{\mathbf{J}}^2$ is the Casimir operator (analog of total angular momentum squared), which reduces to a c number for the operators in (2.11):

$$\hat{\mathbf{J}} = \hat{J}_0^2 - \hat{J}_1^2 - \hat{J}_2^2 = -3/16 = j(j + 1). \quad (2.13)$$

This means that the irreducible representations of $SU(1, 1)$ with "weights" $j = -1/4$ and $-3/4$ correspond, respectively, to the even and odd number states, which do not mix under time evolution because the potential $V(q) = \omega^2q^2/2$ is an even function. The spectrum of the operator $\hat{J}_0$ defined in (2.11) consists of $\lambda_n = (2n + 1)/4$ with integer nonnegative n:

$$\lambda_n = k + 1/4 = -j + k, \quad j = -1/4$$

for $n = 2k$, and

$$\lambda_n = -j + k, \quad j = -3/4$$

for $n = 2k + 1$ ($k = 0, 1, 2, \ldots$, where $k = 0$ corresponds to the lowest eigenstate in either representation). Both

representations are unitary (since $S^+S = 1$ for quantum-mechanical evolution operators) and infinite-dimensional.[1] It is well known that all finite-dimensional representations of $SU(1, 1)$ are nonunitary (e.g., see (A.2) in Appendix A).

Thus, the oscillator problem is equivalent to the problem of $SU(1, 1)$ "rotation" of a pseudospin **J**:

$$\hat{S}(t, t_0) = T\exp\left\{-i\int_{t_0}^{t}\omega^a(t')\hat{J}_a dt'\right\},$$

where $\omega^+ = -\omega^- = i(1-\omega^2)/2$ and $\omega^0 = 1 + \omega^2$. By analogy with the Wigner D functions [10, 11], the finite-rotation matrix elements are written as

$$D^{(j)}_{\mu\nu}(g) = \exp(-i\mu\psi)f^{(j)}_{\mu\nu}(\beta)\exp(-i\nu\varphi),$$
$$0 \le \beta < \infty,$$

where $g = g(\psi, \beta, \varphi) \in SU(1, 1)$, ψ and φ denote the Euler angles for Euclidean rotations about the x_0 axis ($0 \le \psi, \varphi < 2\pi$), $f^{(j)}_{\mu\nu}(\beta)$ and the parameter β correspond to a hyperbolic rotation in the (x_1, x_2) plane, and μ and ν are eigenvalues of $\hat{J}_0$ ($\mu = -j$ for the lowest eigenstate in an irreducible representation). According to (2.12), $\hat{J}_+$ and $\hat{J}_-$ are, respectively, the raising and lowering operators that change the eigenvalue of $\hat{J}_0$ by +1 and –1 and the energy eigenvalue by +2 and –2 ($|n\rangle \longrightarrow |n \pm 2\rangle$).

The transition probabilities between oscillator eigenstates are expressed as

$$w_{mn} = |A_{mn}|^2 = \left|f^{(j)}_{k-j, l-j}(\beta)\right|^2, \quad k, l = 0, 1, 2, \ldots, \tag{2.14}$$

where

$$j = -1/4, \quad k = m/2, \quad l = n/2 \tag{2.14'}$$

for even eigenstates;

$$j = -3/4, \quad k = (m-1)/2, \quad l = (n-1)/2 \tag{2.14''}$$

for odd eigenstates; and

$$k - j = \frac{1}{2}\left(m + \frac{1}{2}\right), \quad l - j = \frac{1}{2}\left(n + \frac{1}{2}\right)$$

in either case. These expressions elucidate the group-theoretic aspects of the problem under analysis.

[1] This is a natural consequence of the fact that $SU(1, 1)$ is a noncompact group [5]. An analogy more familiar to physicists is the homogeneous (proper) Lorentz group [6–8]. A detailed representation theory of the Lorentz group can be found in [9].

An expression for $f^{(j)}_{\mu\nu}(\beta)$ is obtained by analytic continuation of the Wigner functions[2] $d^{(j)}_{\mu\nu}(\vartheta)$ for the unitary group $SU(2)$ to the pseudounitary group $SU(1, 1)$ by replacing ϑ with $i\beta$ and $j = 1/2, 1, 3/2, \ldots$ with the values of j given above:

$$\left|f^{(j)}_{k-j, l-j}(\beta)\right|^2 = \frac{k!}{(k-l)!^2 l!}\frac{\Gamma(k-2j)}{\Gamma(l-2j)}\rho^{k-l}(1-\rho)^{-2j} \times\left[F\left(-l, 2j+1-l; k-l+1; -\frac{\rho}{1-\rho}\right)\right]^2, \tag{2.15}$$

$$\rho = \tanh^2\frac{\beta}{2},$$

where $F(\ldots) \equiv {}_2F_1(\ldots)$ is Gauss's hypergeometric function and $k \ge l$ (an analogous expression is valid for $k < l$). In the particular case of $l = 0$, we have

$$\left|f^{(j)}_{k-j, -j}(\beta)\right|^2 = \left|f^{(j)}_{j, j-k}(\beta)\right|^2 = \frac{\Gamma(k-2j)}{k!\Gamma(-2j)}\rho^k(1-\rho)^{-2j}, \tag{2.16}$$

and unitarity is readily verified:

$$\sum_\mu \left|d^{(j)}_{\mu, -j}(\beta)\right|^2 = 1.$$

After performing some algebra (see details in Appendix A), final expressions for the $|m, \omega_-\rangle \longrightarrow |n, \omega_+\rangle$ transition probability are obtained:

$$w_{mn} = \frac{K!}{(K-L)!^2 L!}\frac{\Gamma(K-2j)}{\Gamma(L-2j)}\rho^{K-L}(1-\rho)^{-2j} \times F^2(-L, K-2j; K-L+1; \rho), \tag{2.17}$$

where $K = \max(k, l)$, $L = \min(k, l)$, and $K - L = |m - n|/2$. Since $L \ge 0$ is an integer, the hypergeometric series reduces to a polynomial. Probability (2.17) satisfies the reciprocity condition

$$w_{mn}(\rho) = w_{nm}(\rho), \tag{2.18}$$

which is specific to harmonic oscillator.

3. EXAMPLES

Let us compare the expressions obtained above with results of independent calculations.

[2] These functions were analyzed in detail in [10, 11] and tabulated in [11] for $j \le 5$. However, the tables cannot be used directly to determine the functions $f^{(j)}_{\mu\nu}(\beta)$ derived in Appendix A. The parameter β in (2.14) (or ρ in (2.15)) characterizes the degree of excitation of the oscillator (ρ is discussed in Appendix B, where several exact solutions are considered).

Example 1. Substituting $j = -1/4$ and $-3/4$ into (2.16) yields the transition probabilities from the two lowest oscillator eigenstates:

$$w_{0,2l} = \frac{\Gamma(l+1/2)}{l!\Gamma(1/2)}\rho^l\sqrt{1-\rho},$$
$$w_{1,2l+1} = \frac{\Gamma(l+3/2)}{l!\Gamma(3/2)}\rho^l(1-\rho)^{3/2}, \quad (3.1)$$

in agreement with [12–14]. For transitions from the second excited level,

$$w_{2,2l} = \frac{l\Gamma(l+1/2)}{(l-1)!\Gamma(3/2)}\left(1-\frac{2l+1}{2l}\rho\right)^2 \times \rho^{l-1}\sqrt{1-\rho}, \quad l \geq 1 \quad (3.2)$$

and $w_{20} = \rho\sqrt{1-\rho}/2$ ($l = 0$). Analogous formulas are obtained for transitions from the next level ($m = 3$).

Example 2. For diagonal transitions ($m = n$), we obtain

$$w_{nn}(\rho) = \sqrt{1-\rho}[p_n(\rho)]^2,$$
$$p_n(\rho) = {}_2F_1\left(-\frac{n}{2}, \frac{n+1}{2}, 1; \rho\right) \quad (3.3)$$
$$\equiv \sqrt{1-\rho}\,{}_2F_1\left(\frac{1-n}{2}, \frac{n}{2}+1; 1; \rho\right).$$

As shown in Appendix A, these formulas can be written in the unified form

$$w_{nn} = \sqrt{1-\rho}\,{}_3F_2(-n, n+1, 1/2; 1, 1; \rho) \quad (3.4)$$

by using the generalized hypergeometric series [15]

$${}_3F_2(-n, n+1, 1/2; 1, 1; x) = \sum_{k=0}^{n}\frac{(2k)!}{2^{2k}(k!)^4}\frac{(n+k)!}{(n-k)!}(-x)^k.$$

This representation facilitates analysis when ρ is small (i.e., when the oscillator is in a low excited state):

$$w_{nn}(\rho) = 1 - \frac{1}{2}N\rho + \frac{1}{32}(3N^2-4N-3)\rho^2$$
$$-\frac{1}{576}(5N^3-28N^2+11N+48)\rho^3 + \frac{1}{73728} \quad (3.5)$$
$$\times(35N^4-520N^3+1706N^2+840N-4941)\rho^4 - \ldots,$$

where $N = n^2 + n + 1$ (see detailed calculations in Appendix C). The small parameter in this expansion is $n^2\rho$, and the coefficient of ρ^k is a polynomial of N with the highest degree term $(-1)^k[(2k-1)!!/2^k(k!)^3]N^k$.

For any pair of quantum numbers m and n,

$$w_{mn}(\rho) \propto \rho^{|m-n|/2}, \quad \rho \longrightarrow 0. \quad (3.6)$$

In the adiabatic limit, when $\rho \ll 1$, transitions with $|m-n| \gg 1$ correspond to high-order terms in perturbation series and are therefore strongly suppressed. The adiabatic regime is further discussed in Section 4, where formula (3.6) is refined.

Example 3. In the opposite limit of $\rho \longrightarrow 1$, the probability that the oscillator remains in the initial state as $t \longrightarrow +\infty$ is very low:

$$w_{nn} \approx \begin{cases} c_n(1-\rho)^{1/2} & \text{for even } n, \\ c_n'(1-\rho)^{3/2} & \text{for odd } n, \end{cases} \quad (3.7)$$

where

$$c_n = \left[\Gamma\left(\frac{n+1}{2}\right)/\sqrt{\pi}\Gamma\left(\frac{n}{2}+1\right)\right]^2,$$

i.e., $c_0 = 1$, $c_2 = 1/4$, and $c_n \sim 1/n$ for $n \gg 1$.

Thus, comparison with nontrivial examples confirms the validity of expressions (2.15)–(2.17), which are derived here somewhat heuristically (by analytic continuation of the $SU(2)$ Wigner functions to $SU(1, 1)$). Their relationship with the expressions for w_{mn} obtained in [13] is discussed in Appendix A. Note that neither (2.17) nor (3.4) can be found in available literature, whereas both expressions are suitable for calculational purposes. In particular, they entail adiabatic expansions (3.5) and (4.1).

Example 4. When applied to a D-dimensional oscillator,[3]

$$H = \frac{1}{2}\sum_{i=1}^{D}\{p_i^2 + \omega_i^2(t)q_i^2\}, \quad [p_j, q_k] = -i\delta_{jk}, \quad (3.8)$$

FD results in the product of D operators having the form of (2.9). Since these operators commute, the $|0\rangle \longrightarrow |0\rangle$ transition probability is

$$w_{00}^{(D)} = \sqrt{(1-\rho_1)(1-\rho_2)\ldots(1-\rho_D)}.$$

For an isotropic oscillator,

$$w_{00}^{(D)} = (1-\rho)^{D/2} = (\cosh\beta/2)^{-D}. \quad (3.9)$$

It follows from comparison with (A.5) that

$$j = -D/4, \quad (3.10)$$

in this case, and the $|0\rangle \longrightarrow |2l\rangle$ transition probability is

$$w_{0,2l}^{(D)} = \frac{\Gamma!(l+D/2)}{l!\Gamma(D/2)}\rho^l(1-\rho)^{D/2}. \quad (3.11)$$

[3] See also [16, 17] and further references in [17].

This obviously implies that $\sum_l w_{0,2l}^{(D)} = 1$ for any D (i.e., unitarity holds). The parameter ρ characterizes the probability of transition to the level closest to the ground level: $w_2/w_0 = D\rho/2$.

Since $j = -1/2$ for $D = 2$, the corresponding expression (3.11) is similar to the distribution of n pairs of charged scalar bosons in a state with definite momentum $\mathbf{p}$ created from vacuum by a time-dependent uniform electric field $\mathscr{E}(t)$:

$$W_n(\mathbf{p}) = \left|f^{(-1/2)}_{1/2,\,n+1/2}(\beta)\right|^2 = (1-\rho)\rho^n, \quad n = 0, 1, 2, \ldots, \tag{3.12}$$

where ρ depends both on $\mathscr{E}(t)$ and on $\mathbf{p}$. This distribution was obtained in calculations performed in [18, 19], in particular, in the case of a field with pulse profile $\mathscr{E}(t) = \mathscr{E}_0/\cosh^2\omega t$, for which both Klein–Gordon and Dirac equations can be solved exactly.

In the problem of particle–antiparticle pair creation from vacuum, if the two-mode $SU(1, 1)$ generators $\hat{J}_a$ are defined in terms of the creation and annihilation operators ($\hat{J}_+ = a^\dagger b^\dagger$ and $\hat{J}_- = ba$ for $s = 0$ and 1/2), then probabilities (3.12) can easily be calculated [19]. The group-theoretic aspects of this problem for particles with arbitrary spin s were discussed in [20]. It should be noted here that $\rho \sim \exp(-\pi\mathscr{E}_{cr}/\mathscr{E}_0)$, where $\mathscr{E}_{cr} = m^2c^3/e\hbar$ is the Schwinger critical field in quantum electrodynamics [21–23]. Its numerical value is $\mathscr{E}_{cr} = 1.32 \times 10^{16}$ V/cm for $e^\pm$. Being proportional to m^2, this quantity is on the order of 10^{21} V/cm for π mesons. (Therefore, only the case of $s = 1/2$ is of real interest.)

In recent years, the experimental observability of electron–positron pair creation from vacuum by a focused laser pulse (Schwinger effect) is widely discussed in the literature (e.g., see [24–26] and references therein).

4. ADIABATIC REGIME

When the oscillator frequency is a slowly varying analytic function of t with singular points located at finite distances from the real axis, the Landau–Dykhne adiabatic approximation can be applied [27–30]. In this case, ρ (as well as reflection coefficient $\rho = |R|^2$, where R is the reflected-wave amplitude in the solution to the Schrödinger equation obtained by replacing t with x) is exponentially small (see Appendix B). As $\rho \longrightarrow 0$, expression (2.17) reduces to

$$w_{mn} = \frac{n_>!}{(2^q q!)^2 n_<!}\rho^q \times\left(1 - \frac{mn + (m+n)/2 + 1}{2(q+1)}\rho + \ldots\right), \quad q = |m-n|/2. \tag{4.1}$$

This expansion is valid for arbitrary m and n. The transition probability rapidly decreases with increasing $|m-n|$. In particular,

$$w_{n,\,n\pm2} = d_\pm\rho - d_\pm^2\rho^2 + \ldots, \quad w_{n,\,n\pm4} = h_\pm\rho^2 + \ldots, \quad \rho \ll 1, \tag{4.2}$$

where

$$d_+ = \frac{1}{4}(n^2+3n+2), \quad d_- = \frac{1}{4}(n^2-n), \quad d_+ + d_- = \frac{1}{2}N, \tag{4.2'}$$
$$h_\pm = \frac{1}{4}d_\pm\left[d_\pm + 2 \pm \left(n+\frac{1}{2}\right)\right].$$

The corresponding expansion of the diagonal transition probability is given by (3.5). For $|m-n| \geq 6$, the transition probabilities are

$$w_{n,\,n+2k} = \frac{(n+2k)!}{2^{2k}k!^2 n!}\rho^k + \ldots, \quad w_{n,\,n-2k} = \frac{n(n-1)\ldots(n-2k+1)}{2^{2k}k!^2}\rho^k + \ldots, \quad k \geq 1. \tag{4.3}$$

Hence,

$$\frac{w_{n,\,n-2k}}{w_{n,\,n+2k}} = 1 - \frac{4k^2}{n} + O(n^{-2}), \quad \rho \ll 1, \quad n \longrightarrow \infty.$$

Introducing

$$W_n^{(+)} = \sum_{n'>n} w_{nn'}, \quad W_n^{(-)} = \sum_{n'<n} w_{nn'}, \tag{4.4}$$

and using (4.2), we find

$$W_n^{(+)} - W_n^{(-)} = \left(n+\frac{1}{2}\right)\left[\rho - \frac{1}{4}(n^2+n-1)\rho^2 + \ldots\right]. \tag{4.5}$$

Thus, when ρ is small, the total probability of upward transition from any level exceeds the total probability of downward transition for any $\omega(t)$. According to numerical calculations (see Fig. 1), this is true for $\rho \sim 1$ as well. For comparison, Fig. 2 shows diagonal transition probabilities. All functions $w_{nn}(\rho)$ are not monotonic except for those with $m = 0$ and 1. It can readily be verified that the sum of all expansions in (3.5) and (4.2) is unity up to terms of order ρ^3, which are not written out in (4.2). Thus, the unitarity condition is satisfied to second order in ρ.

Let us compare the expressions obtained here with Dykhne's results for transition amplitudes a_{mn} corre-

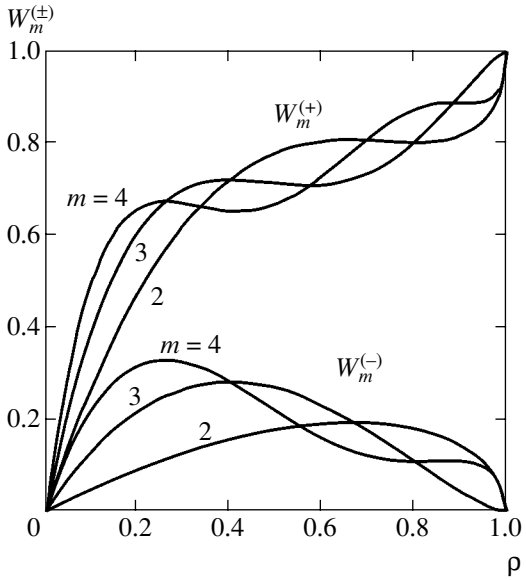


Fig. 1. Probabilities $W_m^{(+)}$ and $W_m^{(-)}$ of upward and downward transitions from the mth oscillator state. Values of m are shown at curves.

sponding to $n \gg 1$ (Eq. (3.28) in [28]):

$$a_{nn} = 1 + i\left(n + \frac{1}{2}\right)\mathrm{Im}R,$$

$$a_{n+2,n} = \frac{1}{2}\left(n - \frac{1}{2}\right)R, \qquad (4.6)$$

$$a_{n-2,n} = -\frac{1}{2}\left(n + \frac{3}{2}\right)R^*.$$

A comparison shows that these formulas with $m \neq n$ agree with (4.2) if $a_{n+2,n}$ and $a_{n-2,n}$ in (4.6) are replaced with $a_{n,n-2}$ and $a_{n,n+2}$, respectively, and the terms of order unity in the coefficients of the powers of ρ in the corresponding expansions of $w_{mn} = |a_{mn}|^2$ are neglected (as compared to n^2 and n).

Note that the unitarity condition can be used to uniquely determine the next term in the expansion of a_{nn} when $n \gg 1$:

$$a_{nn} = 1 + i\left(n + \frac{1}{2}\right)\mathrm{Im}R$$

$$-\frac{1}{4}(n^2 + n)[|R|^2 + 2(\mathrm{Im}R)^2] + \ldots \qquad (4.7)$$

$$= \sqrt{w_{nn}}\exp\left[i\left(n + \frac{1}{2}\right)\mathrm{Im}R + \ldots\right].$$

The value of a_{nn} depends not only on $\rho = |R|^2$, but also on the phase in $R = \sqrt{\rho}\exp(i\varphi)$, which changes with shift in potential or $\omega(t)$ (e.g., see (B.4)), whereas w_{nn} is independent of φ.

5. ATOM IN TIME-DEPENDENT ELECTRIC FIELD

Let us apply FD to the interaction between an atom and a laser field. In the dipole approximation ($a_B = \hbar^2/me^2 \ll \lambda$, where λ is the laser wavelength), the single-electron Hamiltonian is

$$\hat{H} = \frac{1}{2}\mathbf{p}^2 - e\mathscr{E}(t)\mathbf{r} + U(r), \qquad (5.1)$$

where $\mathscr{E}(t) = -c^{-1}\dot{\mathbf{A}}(t)$ is the field of a plane light wave and $U(r)$ is the atomic core potential. Hamiltonian (5.1)

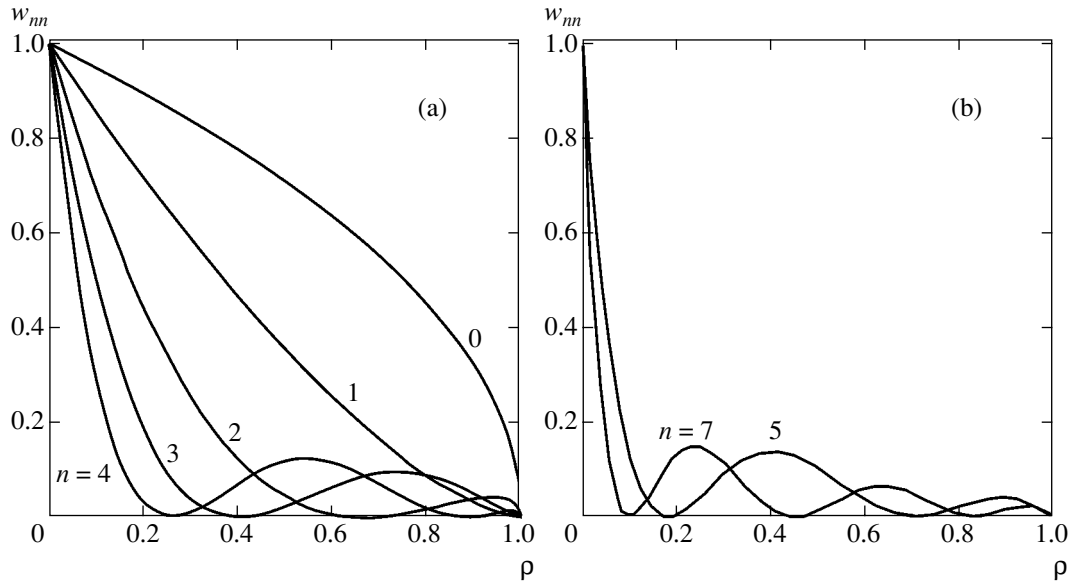


Fig. 2. Probabilities w_{nn} of diagonal transitions vs. ρ: (a) $0 \le n \le 4$; (b) $n = 5, 7$.

contains noncommuting operators (e.g., see [31, 32]). Defining

$$\hat{A} = \frac{1}{2}\mathbf{p}^2, \quad \hat{B} = \mathscr{E}\cdot\mathbf{r}, \quad \hat{C} = \mathscr{E}\cdot\mathbf{p}, \tag{5.2}$$

we have

$$\begin{gathered}[\hat{A},\hat{B}] = -i\hat{C}, \quad [\hat{A},\hat{C}] = 0,\\ [\hat{B},\hat{C}] = i\mathscr{E}^2\hat{I},\end{gathered} \tag{5.3}$$

where $\hat{I}$ is the identity operator. Thus, we have a closed operator algebra,[4] and FD can be applied to obtain

$$\begin{aligned}\hat{S}(t,t_0) &= T\exp\left\{-i\int_{t_0}^{t}\hat{H}(t')dt'\right\}\\ &= \exp\left(-\frac{ie}{c}\mathbf{A}(t)\cdot\mathbf{r}\right)\\ &\times T\exp\left\{-i\int_{t_0}^{t}[\hat{A}' + U(r)]dt'\right\},\end{aligned} \tag{5.4}$$

where

$$\begin{gathered}\hat{A}' = \frac{1}{2}e^{i\mathbf{a}\cdot\mathbf{r}}\mathbf{p}^2e^{-i\mathbf{a}\cdot\mathbf{r}} = \frac{1}{2}(\mathbf{p}-\mathbf{a})^2,\\ \mathbf{a} = \frac{e}{c}\mathbf{A}(t).\end{gathered} \tag{5.5}$$

If $U(r)$ is a short-range potential,[5] then the evolution operator can be disentangled:

$$\begin{aligned}\hat{S}(t,t_0) &= \exp\left[-\frac{ie}{c}\mathbf{A}(t)\cdot\mathbf{r}\right]\\ &\times\exp\left\{-\frac{i}{2}\int_{t_0}^{t}\left[\mathbf{p}-\frac{e}{c}\mathbf{A}(t')\right]^2dt'\right\}.\end{aligned} \tag{5.6}$$

The Volkov wavefunction $\psi_{\mathbf{p}} = \hat{S}\exp(i\mathbf{p}\cdot\mathbf{r})$ of an electron in electromagnetic field has been employed in theoretical calculations since a well-known study by Keldysh [34]. For a nonrelativistic electron,

$$\begin{aligned}\psi_{\mathbf{p}}(\mathbf{r},t) &= (2\pi)^{-3/2}\\ &\times\exp\left\{i\left[\mathbf{P}(t)\cdot\mathbf{r} - \frac{1}{2}\int_{t_0}^{t}\mathbf{P}^2(t')dt'\right]\right\},\end{aligned} \tag{5.7}$$

where $\mathbf{P}(t) = \mathbf{p} - (e/c)\mathbf{A}(t)$. Function (5.7) has been used to obtain various results in the theory of ionization and excitation of atoms and ions by intense laser fields (e.g., see reviews in [35, 36]). Discussion of these results would be inappropriate here, since we only illustrate the disentangling of operator expressions containing (5.1).

6. MODEL WITH A "HIDDEN" SYMMETRY GROUP

The model Hamiltonian

$$\hat{H} = \hat{H}_0 + \boldsymbol{\omega}_1(t)\cdot\mathbf{L} + \boldsymbol{\omega}_2(t)\cdot\mathbf{A}, \tag{6.1}$$

where $\hbar = m = e = 1$,

$$\hat{H}_0 = \frac{1}{2}\mathbf{p}^2 - \frac{1}{r}, \quad \mathbf{A} = \frac{1}{2}(\mathbf{L}\times\mathbf{p} - \mathbf{p}\times\mathbf{L}) + \frac{\mathbf{r}}{r}, \tag{6.2}$$

$\mathbf{L} = \mathbf{r}\times\mathbf{p}$ is the orbital angular momentum, and $\mathbf{A}$ is the Laplace–Runge–Lenz vector [37, 38], can be interpreted as the projection of the Hamiltonian of the hydrogen atom interacting with fields $\mathscr{E}$ and $\mathbf{H}$ onto the subspace spanned by the energy eigenstates (see also remark at the end of this section). It is well known that the energy eigenstates of the hydrogen atom are degenerate. This is explained by a "hidden" symmetry of the Hamiltonian: the bound states (characterized by principal quantum numbers $n = (-2E)^{-1/2} = 1, 2, 3, \ldots$) are invariant under $SO(4)$, and the continuum states are invariant under the Lorentz group $SO(3,1)$ [39–43]. Defining $\mathbf{N} = (-2H_0)^{-1/2}\mathbf{A} = n\mathbf{A}$ for the bound states, we obtain[6]

$$[\mathbf{L},\mathbf{L}] = [\mathbf{N},\mathbf{N}] = i\mathbf{L}, \quad [\mathbf{L},\mathbf{N}] = i\mathbf{N}, \tag{6.3}$$

$$\mathbf{L}^2 + \mathbf{N}^2 = n^2 - 1. \tag{6.4}$$

These relations are analogous to the commutation relations for the $SO(4)$ generators. The eigenvalues of $\mathbf{N}^2$ vary between $n^2 - 1$ for s-states and $\mathbf{N}^2 = n - 1$ for $l = n - 1$ (circular electron orbits). Introducing the operators $\mathbf{I}_{1,2} = (\mathbf{L}\pm\mathbf{N})/2$, we find

$$\begin{gathered}[\mathbf{I}_1,\mathbf{I}_1] = i\mathbf{I}_1, \quad [\mathbf{I}_2,\mathbf{I}_2] = i\mathbf{I}_2, \quad [\mathbf{I}_1,\mathbf{I}_2] = 0,\\ \mathbf{I}_1^2 = \mathbf{I}_2^2 = (n^2-1)/4, \quad [\hat{H}_0,\mathbf{I}_1] = [\hat{H}_0,\mathbf{I}_2] = 0.\end{gathered} \tag{6.5}$$

The change from (6.3) to the operators $\mathbf{I}_{1,2}$ corresponds to the decomposition $SO(4)\sim SU(2)\otimes SU(2)$. Introducing

$$\boldsymbol{\Omega}_{1,2}(t) = \int_0^t\left[\boldsymbol{\omega}_1(t') \pm \frac{1}{n}\boldsymbol{\omega}_2(t')\right]dt' \tag{6.6}$$

[4] Unlike (2.3) or (2.12), it should be a degenerate algebra. However, it can easily be shown to satisfy necessary conditions for a Lie algebra [5].

[5] This is a good approximation in the case of a negative ion, such as H^- or Na^- (e.g., see [33]).

[6] Here, the relation $[\mathbf{A},\mathbf{B}] = i\mathbf{C}$ is interpreted as $[A_j, B_k] = ie_{jkm}C_m$. Note that $[\mathbf{I},\mathbf{I}] = i\mathbf{I}$ holds for the $SU(2)$ generators.

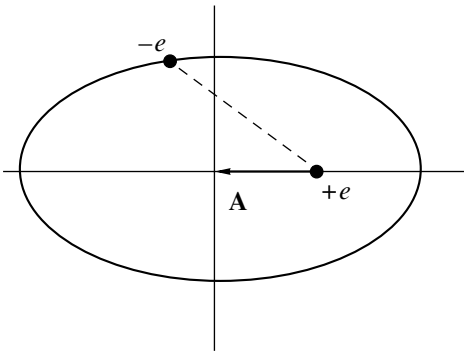


Fig. 3. Position of vector **A** for electron in Kepler motion.

and using the fact that $\mathbf{I}_1$ and $\mathbf{I}_2$ commute, we obtain

$$\hat{S}(t,0) = T\exp\left\{-i\int_0^t \hat{H}(t')dt'\right\} = \exp(-iE_nt)\exp(-i\boldsymbol{\Omega}_1\cdot\mathbf{I}_1)\exp(-i\boldsymbol{\Omega}_2\cdot\mathbf{I}_2). \quad (6.7)$$

Analogously, for continuum states (with energy $E = k^2/2 > 0$), we have

$$\mathbf{N}' = (2H_0)^{-1/2}\mathbf{A} = \frac{1}{k}\mathbf{A}, \quad \mathbf{N}'^2 = l(l+1) + 1 + k^{-2}.$$

Hence, the relations

$$[\mathbf{L},\mathbf{L}] = -[\mathbf{N}',\mathbf{N}'] = i\mathbf{L}, \quad [\mathbf{L},\mathbf{N}'] = i\mathbf{N}',$$

hold for the generators of the Lorentz group instead of (6.3). Defining $\mathbf{J}_{1,2} = (\mathbf{L} \pm i\mathbf{N}')/2$, where $\mathbf{J}_1$ and $\mathbf{J}_2$ satisfy commutation relations similar to (6.5), we write the evolution operator as

$$\hat{S}(t,0) = \exp\left(-i\frac{k^2}{2}t\right)\exp(-i\boldsymbol{\Omega}_1'\cdot\mathbf{J}_1) \times \exp(-i\boldsymbol{\Omega}_2'\cdot\mathbf{J}_2), \quad (6.8)$$

where

$$\boldsymbol{\Omega}_{1,2}'(t) = \int_0^t [\boldsymbol{\omega}_1(t') \mp ik\boldsymbol{\omega}_2(t')]dt'. \quad (6.9)$$

Since the operators $\mathbf{I}_1$ and $\mathbf{I}_2$ (as well as $\mathbf{J}_1$ and $\mathbf{J}_2$) commute, these formulas can be obtained without invoking FD. Then, the exponentials in the evolution operator generated by Hamiltonian (6.1) are disentangled (by introducing I_+, I_0, I_-, etc.) by analogy with the problem of spin precession in a time-dependent magnetic field [3], which completes the application of FD. Even though operator (6.8) is unitary, neither $\exp(-i\boldsymbol{\Omega}_1'\cdot\mathbf{J}_1)$ nor $\exp(-\boldsymbol{\Omega}_2'\cdot\mathbf{J}_2)$ is a unitary operator, because $\mathbf{J}_1$ and $\mathbf{J}_2$ are non-Hermitian operators ($\mathbf{J}_1^+ = \mathbf{J}_2$ and $\mathbf{J}_2^+ = \mathbf{J}_1$).

The matrix elements for transitions between hydrogen atom eigenstates with equal energies E and different orbital angular momenta can easily be written out. For $E < 0$, we have the Wigner functions for the representation $\mathcal{D}(j_1, j_2)$ of the $SO(4)$ group with $j_1 = j_2 = (n-1)/2$. For continuum states, they correspond to the infinite-dimensional[7] unitary principal series representation $\mathcal{D}(0,\rho)$ of the Lorentz group [6–9], with the invariant $\rho = 2/k$ corresponding to the eigenvalue

$$\hat{F} = \frac{1}{2}M_{\mu\nu}M^{\mu\nu} = \mathbf{L}^2 - \mathbf{N}'^2 = -\left(1+\frac{\rho^2}{4}\right) \quad (6.10)$$

of the scalar operator $\hat{F}$, where $M_{\mu\nu} = x_\mu p_\nu - x_\nu p_\mu$ is the 4-momentum operator. (Note that (6.10) is obtained by replacing n in (6.4) with $i\rho/2$.) The pseudoscalar invariant

$$\hat{G} = \frac{i}{8}\varepsilon_{\mu\nu\rho\sigma}M^{\mu\nu}M^{\rho\sigma}$$

of the Lorentz group does not vanish only for nonzero-spin particles. However, the electron spin is neglected in (6.1).

A final remark should be made here about the physical meaning of the term $\boldsymbol{\omega}_2\cdot\mathbf{A}$ in Hamiltonian (6.1). In classical mechanics, $\mathbf{A} = \mathbf{L}\times\mathbf{p} + \alpha\mathbf{r}/r$ for a particle moving in the potential $U(r) = -\alpha/r$. The inner product of this relation with $\mathbf{r}$ yields the orbit equation

$$r = \frac{p}{1-e\cos\varphi}, \quad p = \frac{L^2}{\alpha}, \quad e = \frac{A}{\alpha} \quad (m=1). \quad (6.11)$$

The vector $\mathbf{A}$ is aligned with the major axis of the ellipse, and its magnitude is proportional to its eccentricity e (see Fig. 3). The dipole moment of a particle with charge –1 averaged over the Kepler period is $\langle d\rangle = 3ea/2$, where $a = \alpha/2|E|$ is the semimajor axis of the orbit. In the semiclassical approximation, we therefore obtain

$$\mathbf{A} = \frac{4}{3}|E_n|\langle\mathbf{d}\rangle = \frac{2}{3n^2}\langle\mathbf{d}\rangle_{nn}. \quad (6.12)$$

Thus, for eigenstates with energy E (degenerate with respect to the orbital angular momentum l), the term $\boldsymbol{\omega}_2\cdot\mathbf{A}$ corresponds to the diagonal matrix element of the dipole interaction operator $\boldsymbol{\mathcal{E}}(t)\mathbf{d} \propto (\boldsymbol{\mathcal{E}}\cdot\mathbf{r})$. However, the off-diagonal elements (proportional to $\langle\mathbf{r}\rangle_{nn'}$)

[7] This is clear even from the fact that the orbital angular momentum l is not bounded from above for eigenstates with $E > 0$, whereas $0 \le l \le n-1$ for $E_n = -1/2n^2 < 0$.

are not taken into account here. The reliability of this approximation requires additional analysis.

7. CONCLUSIONS

It is shown that FD can be used to disentangle the evolution operators generated by certain Hamiltonians and find exact solutions to quantum-mechanical problems. However, this approach is not as versatile as the Feynman diagram technique in perturbation theory. This may explain the fact that FD is not widely employed by physicists. However, when FD is applicable, it can be used to obtain a solution for arbitrarily varying parameters (force $f(t)$, oscillator frequency $\omega(t)$, etc.).

The oscillator problem considered in [1] involves only two operators, $\hat{a}$ and $\hat{a}^\dagger$. Since their commutator $[a, a^\dagger] = 1$ is a c number, the application of FD leads directly to the desired result. In the models considered both in [3] and in the present study, a Lie algebra is generated by three operators contained in a Hamiltonian. The corresponding operator expressions are disentangled by using a procedure proposed in [3] in the framework of Feynman's operator calculus and explained by Eqs. (2.5) and (2.6). Once the group-theoretic structure of the evolution operator $\hat{S}$ is elucidated, compact analytical expressions for transition probabilities w_{mn} can be written in terms of matrix elements of unitary irreducible representations of the $SU(1, 1)$ group, which are infinite-dimensional by necessity. Both transition amplitudes A_{mn} and probabilities $w_{mn} = |A_{mn}|^2$ are determined by parameters that can be found by solving the Riccati equations arising in FD applications (see [3] or Eq. (2.8) above).

It would be interesting to explore the applicability of FD to Hamiltonians containing $n > 3$ noncommuting operators and the role played by group theory in the disentangling process.

ACKNOWLEDGMENTS

I thank K.G. Boreskov, O.V. Kancheli, V.D. Mur, Yu.A. Simonov, and participants of the International Conference on Theoretical Physics organized by the Tamm Theory Department of the Lebedev Institute, Russian Academy of Sciences (Moscow, April 2005). I am also grateful to S.G. Pozdnyakov for performing numerical calculations and to M.N. Markina for help in preparing the manuscript.

This work was supported by the Russian Foundation for Basic Research, project no. 04-02-17157.

APPENDIX A

Let us discuss here the relationship between the irreducible representations of the groups $SU(2)$ and $SU(1, 1)$ and derive formulas (2.15)–(2.17).

The spinor representation of the $SU(2)$ group has the form

$$g = \exp\left(-i\frac{\psi}{2}\sigma_z\right)\exp\left(-i\frac{\vartheta}{2}\sigma_x\right)\exp\left(-i\frac{\varphi}{2}\sigma_z\right), \quad g^+ g = 1, \tag{A.1}$$

where ψ, ϑ, and φ are Euler angles ($0 \le \vartheta < \pi$). An analogous representation of $SU(1, 1)$ is obtained by replacing ϑ with $i\beta$:

$$\exp\left(-i\frac{\vartheta}{2}\sigma_x\right) = \begin{pmatrix} \cos(\vartheta/2) & -i\sin(\vartheta/2) \\ -i\sin(\vartheta/2) & \cos(\vartheta/2) \end{pmatrix} \longrightarrow \begin{pmatrix} \cosh(\beta/2) & \sinh(\beta/2) \\ \sinh(\beta/2) & \cosh(\beta/2) \end{pmatrix}. \tag{A.2}$$

The resulting representation is finite-dimensional and nonunitary: $\tilde{g}g = 1$, with

$$\tilde{g} = \sigma_z g^+ \sigma_z, \quad \hat{J}_1 = \frac{i}{2}\sigma_x, \quad \hat{J}_2 = \frac{i}{2}\sigma_y,$$

$$\hat{J}_0 = \frac{1}{2}\sigma_z, \quad \mathbf{J}^2 = 3/4.$$

By assumption, the matrix elements of unitary representations (with $j = -1/4, -3/4$) can be found by analytic continuation of the Wigner D functions [10, 11] of $SU(2)$ with the substitution of

$$\vartheta \longrightarrow i\beta, \quad 0 \le \beta < \infty; \quad j = -1/4 \text{ or } j = -3/4. \tag{A.3}$$

This assumption is substantiated by the following argumentation.

(a) For the $SU(2)$ group, it is known that

$$d^{(j)}_{jj}(\vartheta) = d^{(j)}_{-j,-j}(\vartheta) = \left(\cos\frac{\vartheta}{2}\right)^{2j}, \quad j = 0, \frac{1}{2}, 1, \frac{3}{2}, \ldots. \tag{A.4}$$

Using (A.3), we obtain

$$f^{(j)}_{-j,-j}(\beta) = \left(\cosh\frac{\beta}{2}\right)^{2j} = (1-\rho)^{-j}. \tag{A.5}$$

After substituting $j = -1/4$ and $-3/4$, we have

$$w_{00} = [f^{(-1/4)}_{1/4,\,1/4}(\beta)]^2 = \sqrt{1-\rho}, \quad w_{11} = (1-\rho)^{3/2}, \quad \rho = \tanh^2\frac{\beta}{2}, \tag{A.6}$$

which agrees with the results presented in [12–14].

(b) In the more general case of $m = 0 \longrightarrow n = 2l$ or $m = 1 \longrightarrow n = 2l + 1$ transition,[8]

$$d^{(j)}_{-j,\,l-j}(\vartheta) = C\left(\tan\frac{\vartheta}{2}\right)^{l}\left(\cos\frac{\vartheta}{2}\right)^{2j},$$
$$C = \sqrt{\frac{(2j)!}{l!(2j-l)!}} = i^{\pm l}\sqrt{\frac{\Gamma(l-2j)}{l!\Gamma(-2j)}}. \quad \text{(A.7)}$$

This result entails formulas (3.1), which were obtained independently in [12, 13]. This substantiates the validity of the analytic continuation.

(c) In the general case, the Wigner D functions of $SU(2)$ can be expressed in terms of the hypergeometric function ${}_2F_1$ (see [11]). Invoking formulas (16) and (17) from [11, Section 4.5] with $\mu = l - j$ and $\nu = k - j$, substituting (A.3), and using the representations

$$\frac{(2j-l)!}{(2j-k)!} = \frac{\Gamma(2j+1-l)}{\Gamma(2j+1-k)} = (-1)^{k-l}\frac{\Gamma(k-2j)}{\Gamma(l-2j)},$$

we obtain expression (2.15) for $k \geq l$. Finally, changing from the argument z in ${}_2F_1(\ldots; z)$ to $z/(z-1)$ (see Section 2.9.3 in [15]), we obtain (2.17).

Note that $D^{(j)}_{\mu\nu}(e) = \delta_{\mu\nu}$ if e is the identity of the group. Therefore, the diagonal elements $f^{(j)}_{\mu\mu}(\beta)$ are uniquely determined:

$$f^{(j)}_{k-j,\,k-j}(\beta) = \left(\cosh\frac{\beta}{2}\right)^{2j}$$
$$\times\,{}_2F_1\left(-k, k-2j; 1; \tanh^2\frac{\beta}{2}\right) \quad \text{(A.8)}$$
$$= (1-\rho)^{-j}\sum_{s=0}^{k}\frac{k!\Gamma(k+s-2j)}{s!^2(k-s)!\Gamma(k-2j)}(-\rho)^s,$$

where $j = -1/4$ or $j = -3/4$ ($k = 0, 1, 2, \ldots$). The phase factors corresponding to $\mu \neq \nu$ remain undetermined, but they are not required in calculations of transition probabilities.

Formula (3.4) is obtained by using the identity[9]

$$[{}_2F_1(a, b; a+b+1/2; z)]^2$$
$$= {}_3F_2(2a, 2b, a+b; a+b+1/2, 2(a+b); z),$$

where $a = -n/2$, $b = (n+1)/2$. Analogously, formula (2.17) for $j = -1/4$ can be rewritten as

$$w_{mn} = \frac{K!}{(K-L)!^2L!}\frac{\Gamma(K+1/2)}{\Gamma(L+1/2)}\rho^{K-L}(1-\rho)^{1/2}$$
$$\times\,{}_3F_2(-2L, 2K+1, K-L+1/2;$$
$$K-L+1, 2(K-L)+1; \rho), \quad \text{(A.9)}$$

which yields (3.4) when $m = n$ and $K = L = n/2$.

(d) At first glance, the expression for w_{mn} obtained in [13],

$$w_{mn} = \frac{n_<!}{n_>!}\sqrt{1-\rho}\left|P^{|m-n|/2}_{(m+n)/2}(\sqrt{1-\rho})\right|^2, \quad \text{(A.10)}$$

substantially differs from (2.17) and (A.9). The equivalence of these expressions entails certain relations between special functions. For example, if $m = n$, then

$${}_2F_1(-n/2, (1-n)/2; 1; \sin^2\theta)$$
$$\equiv (\cos\theta)^n P_n(1/\cos\theta). \quad \text{(A.11)}$$

This identity can be verified directly for small n, but it cannot be found in [11, 45]. An analogous identity containing associated Legendre functions holds for arbitrary m and n.

(e) It is clear from (A.2) that the unitarity of a representation may not be preserved after the substitution $\vartheta \longrightarrow i\beta$, requiring that

$$\sum_{\mu}\left|f^{(j)}_{\mu\nu}(\beta)\right|^2 = \sum_{\nu}\left|f^{(j)}_{\mu\nu}(\beta)\right|^2 = 1. \quad \text{(A.12)}$$

These necessary, but not sufficient, conditions are satisfied for (3.1) and (3.11) (i.e., for $m = 0$ and $m = 1$). However, these simplest formulas may not be typical because they correspond to upward transitions. For $m = 2$, unitarity condition (A.12) yields

$$\sqrt{1-\rho}\left\{\frac{1}{2}\rho + \sum_{k=0}^{\infty}\frac{\Gamma(k+3/2)}{(k+1)!\Gamma(3/2)}\right.$$
$$\left.\times\left[k+1-\left(k+\frac{3}{2}\right)\rho\right]^2\rho^k\right\} \equiv 1, \quad \text{(A.13)}$$
$$0 \leq \rho < 1.$$

An analogous condition holds for $m = 3$. These relations were verified numerically for ρ between 0.01 and 0.99 with a precision of at least 10^{-12}.

APPENDIX B

According to [13, 37], its holds that $\rho = |C_2/C_1|^2$, where the constants $C_{1,2}$ are determined by solving the

[8] For details, see [11, 44]. Here, we omit a phase factor, which is not essential for this presentation. Note that the $SU(1, 1)$ group was referred to as $QU(2)$ in [44].

[9] See Section 4.3.1 in [15]. This is the only case when Gauss's hypergeometric function squared can be expressed as the function ${}_3F_2$ of the same argument.

classical equation of motion $\ddot{x} + \omega^2(t)x = 0$ subject to the conditions

$$x(t) = \begin{cases} \exp(i\omega_- t), & t \longrightarrow -\infty, \\ C_1\exp(i\omega_+ t) + C_2\exp(-i\omega_+ t), & t \longrightarrow +\infty. \end{cases} \quad \text{(B.1)}$$

The parameter ρ can be interpreted as a barrier reflection coefficient if t and $x(t)$ are replaced with q and $\psi(q)$, respectively. By changing to $a(t) = -i\dot{x}(t)/x(t)$, equation (2.8) is obtained and it is found that $a(-\infty) = \omega_-$ and

$$\rho = \lim_{t\to\infty}\left|\frac{\omega_+ - a(t)}{\omega_+ + a(t)}\right|^2. \quad \text{(B.2)}$$

Thus, ρ can be calculated by solving the Riccati equation derived by using FD. Consider the following examples.

Suppose that

$$\omega(t) = \omega_-\theta(t_0 - t) + \omega_+\theta(t - t_0),$$

which corresponds to an instant change in oscillator frequency from ω_- to ω_+. The Riccati equation is easily solved: $a(t) = \omega_-$ at $t < t_0$;

$$a(t) = \omega_+ \times \frac{(\omega_+ + \omega_-)\exp(i\omega_+\tau) - (\omega_+ - \omega_-)\exp(-i\omega_+\tau)}{(\omega_+ + \omega_-)\exp(i\omega_+\tau) + (\omega_+ - \omega_-)\exp(-i\omega_+\tau)}, \quad t \geq t_0, \quad \text{(B.3)}$$

where $\tau = t - t_0$. Hence,

$$R = \frac{C_2}{C_1} = \exp(2i\omega_+ t_0)\frac{\omega_+ - \omega_-}{\omega_+ + \omega_-},$$
$$\rho = |R|^2 = \left[\frac{\omega_+ - \omega_-}{\omega_+ + \omega_-}\right]^2, \quad \text{(B.4)}$$
$$\beta = \ln\frac{1+\sqrt{\rho}}{1-\sqrt{\rho}} = \ln\frac{\omega_>}{\omega_<}.$$

In the slightly more complicated case of

$$\omega(t) = \begin{cases} \omega_-, & t < 0, \\ \omega_0, & 0 < t < T, \\ \omega_+, & t > T \end{cases} \quad \text{(B.5)}$$

the result is

$$\rho = \frac{\rho_1 + \rho_2 + 2\sigma_1\sigma_2\cos(2\omega_0 T)}{1 + \rho_1\rho_2 + 2\sigma_1\sigma_2\cos(2\omega_0 T)}, \quad \text{(B.6)}$$

where $\rho_i = \sigma_i^2$, $\sigma_1 = (\omega_0 - \omega_-)/(\omega_0 + \omega_-)$, and $\sigma_2 = (\omega_+ - \omega_0)/(\omega_+ + \omega_0)$. Note that the excitation parameter ρ may vanish ($w_{mn} = \delta_{mn}$); i.e., the oscillator may remain in the initial state, as $t \longrightarrow +\infty$. In this particular example, the corresponding conditions are either $\omega_+ = \omega_-$, $\omega_0 T = n\pi$ or $\omega_0 = \sqrt{\omega_+\omega_-}$, $\omega_0 T = (n + 1/2)\pi$. See also formula (B.11) below, with $\omega_0 = (2n + 1)\omega$.

An exact solution can also be found for

$$\omega^2(t) = \frac{1}{2} \times \left[\omega_+^2 + \omega_-^2 + (\omega_+^2 - \omega_-^2)\tanh\omega t + \frac{\omega_0^2 - \omega^2}{2\cosh^2\omega t}\right], \quad \text{(B.7)}$$

which corresponds to the Eckart potential [46]. In this case, the solution to the Schrödinger equation is expressed in terms of hypergeometric functions [46, 47], and

$$\rho = (\sinh^2\eta + \cos^2\delta)/(\sinh^2\zeta + \cos^2\delta), \quad \text{(B.8)}$$

where

$$\zeta = \frac{\pi(\omega_+ + \omega_-)}{2\omega}, \quad \eta = \frac{\pi(\omega_+ - \omega_-)}{2\omega}, \quad \delta = \frac{\pi\omega_0}{2\omega}.$$

In the adiabatic limit ($\omega \ll \omega_\pm$), the reflection coefficient is exponentially small:

$$\rho \approx \exp(-2\pi\omega_</\omega), \quad \omega_< = \min(\omega_+, \omega_-). \quad \text{(B.9)}$$

According to [48, 49], the preexponential factor here is unity.

The dominant contribution to the value of ρ corresponds to the turning point $t = t_0$ nearest to the real axis, where $\omega(t_0) = 0$. The preexponential factor in (B.9) is unity if there exists only one such point. For example, if $\omega_0 = \omega$, then

$$\omega t_0^{(n)} = \ln\frac{\omega_-}{\omega_+} + i\left(n + \frac{1}{2}\right)\pi \quad \text{(B.10)}$$

and the nearest turning point corresponds to $n = 0$.

In the special case of $\omega_+ = \omega_-$,

$$\rho = \frac{\cos^2\delta}{\sinh^2\zeta + \cos^2\delta}, \quad \zeta = \frac{\pi\omega_\pm}{\omega} \quad \text{(B.11)}$$

and the asymptotic formula

$$\rho \approx 4\cos^2\delta\exp(-2\pi\omega_\pm/\omega), \quad \omega \ll \omega_\pm \quad \text{(B.12)}$$

holds instead of (B.9). In this case, there exist two points t_i at equal distances from the real axis where $\omega(t)$ vanishes:

$$\omega t_{1,2} = i\frac{\pi}{2} \pm \operatorname{arcsinh}\frac{\sqrt{\omega_0^2 - \omega^2}}{2\omega_\pm}. \quad \text{(B.13)}$$

The superposition of their contributions to the transition amplitude yields a preexponential factor other than unity. Note also that $\delta = (k + 1/2)\pi$ when $\omega_0 = (2k + 1)\omega$,

and therefore $\rho \equiv 0$ (which corresponds to a nonreflective potential [27]).

APPENDIX C

The hypergeometric series in (3.4) can be written as

$$
\begin{aligned}
&{}_3F_2\left(-n, n+1, \frac{1}{2}; 1, 1; x\right) \\
&= \sum_{k=0}^{n} \frac{(2k-1)!!}{2^k (k!)^3} b_k (-x)^k,
\end{aligned}
\tag{C.1}
$$

where

$$
\begin{aligned}
&(-1)!! = 1, \quad b_0 = 1, \quad b_1 = n^2 + n = N - 1, \\
&b_2 = N^2 - 4N + 3, \\
&b_k(N) = \prod_{s=0}^{k-1} [N - (s^2 + s + 1)], \\
&N = n^2 + n + 1, \\
&b_k = N^k - \frac{1}{3}(k^3 + 2k) N^{k-1} + \ldots \\
&\quad + (-1)^k 1 \cdot 3 \ldots (k^2 - k + 1)
\end{aligned}
\tag{C.2}
$$

for $k \geq 1$ and $b_k = 0$ for $k = n + 1, n + 2, \ldots$. The product of (C.1) with the known series expansion of $w_{00}(\rho)$,

$$
\begin{aligned}
\sqrt{1-\rho} &= \sum_{k=0}^{\infty} \frac{\Gamma(k - 1/2)}{k! \Gamma(-1/2)} \rho^k \\
&= 1 - \sum_{k=1}^{\infty} \frac{(2k-3)!!}{2^k k!} \rho^k,
\end{aligned}
\tag{C.3}
$$

yields the coefficients in (3.5). This calculation shows that the coefficient of ρ^k in (3.5) is a polynomial of the kth degree in N.

Notes added in proof. (1) Recently, M.A. Trusov rigorously proved that the functions given by (2.17) satisfy the unitarity condition. This justifies the analytic continuation of Wigner functions used here in conjunction with FD. (2) I was advised by K.G. Boreskov that relation (A.11) can be derived by using formulas 15.4.10, 15.4.11, and 8.2.1 in [50] under appropriate choice of parameter values. I thank K.G. Boreskov and M.A. Trusov for their interest in this study and helpful remarks.

REFERENCES

1. R. P. Feynman, Phys. Rev. **84**, 108 (1951).
2. J. Schwinger, Phys. Rev. **91**, 728 (1953).
3. V. S. Popov, Zh. Éksp. Teor. Fiz. **35**, 985 (1958) [Sov. Phys. JETP **8**, 687 (1959)].
4. V. S. Popov, Phys. Lett. A **342**, 281 (2005).
5. C. Fronsdal, in *Theory of Groups and Elementary Particles* (Mir, Moscow, 1967), p. 324.
6. I. S. Shapiro, Dokl. Akad. Nauk SSSR **106**, 647 (1956) [Sov. Phys. Dokl. **1**, 91 (1956)].
7. Chzhou-Guan-chzhao and L. G. Zastavenko, Zh. Éksp. Teor. Fiz. **35**, 1417 (1958) [Sov. Phys. JETP **8**, 990 (1959)].
8. V. S. Popov, Zh. Éksp. Teor. Fiz. **37**, 1116 (1959) [Sov. Phys. JETP **10**, 794 (1960)].
9. M. A. Naĭmark, *Linear Representations of the Lorentz Group* (Fizmatgiz, Moscow, 1958; Pergamon, Oxford, 1964).
10. E. P. Wigner, *Group Theory and Its Applications to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959; Inostrannaya Literatura, Moscow, 1961).
11. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Sci., Singapore, 1988).
12. K. Husimi, Prog. Theor. Phys. **9**, 381 (1953).
13. V. S. Popov and A. M. Perelomov, Zh. Éksp. Teor. Fiz. **56**, 1375 (1969) [Sov. Phys. JETP **29**, 738 (1969)].
14. H. R. Lewis and W. B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969).
15. *Higher Transcendental Functions (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1953; Nauka, Moscow, 1965).
16. I. A. Malkin, V. I. Man'ko, and D. A. Trifonov, Phys. Rev. D **2**, 1371 (1970).
17. V. V. Dodonov, I. A. Malkin, and V. I. Man'ko, Teor. Mat. Fiz. **24**, 164 (1975); V. V. Dodonov and V. I. Man'ko, Tr. Fiz. Inst. im. P. N. Lebedeva, Akad. Nauk SSSR **183**, 71 (1987).
18. N. B. Narozhny and A. I. Nikishov, Yad. Fiz. **11**, 1072 (1970) [Sov. J. Nucl. Phys. **11**, 596 (1970)].
19. V. A. Popov, Zh. Éksp. Teor. Fiz. **62**, 1248 (1972) [Sov. Phys. JETP **35**, 659 (1972)].
20. A. M. Perelomov, Phys. Lett. A **39A**, 165 (1972).
21. F. Sauter, Z. Phys. **69**, 742 (1931).
22. W. Heisenberg and H. Euler, Z. Phys. **98**, 714 (1936).
23. J. Schwinger, Phys. Rev. **82**, 664 (1951).
24. A. Ringwald, Phys. Lett. B **510**, 107 (2001).
25. V. S. Popov, Pis'ma Zh. Éksp. Teor. Fiz. **74**, 151 (2001) [JETP Lett. **74**, 133 (2001)]; Phys. Lett. A **298**, 83 (2002).
26. N. B. Narozhny, S. S. Bulanov, V. D. Mur, and V. S. Popov, Phys. Lett. A **330**, 1 (2004).
27. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed. (Nauka, Moscow, 1974; Pergamon, New York, 1977).
28. A. M. Dykhne, Zh. Éksp. Teor. Fiz. **38**, 570 (1960) [Sov. Phys. JETP **11**, 411 (1960)].
29. E. E. Nikitin and L. P. Pitaevskiĭ, Usp. Fiz. Nauk **163** (9), 101 (1993) [Phys. Usp. **36**, 851 (1993)].

30. N. B. Delone and V. P. Kraĭnov, *Atom in a Strong Light Field* (Énergoatomizdat, Moscow, 1984), Chap. 4 [in Russian].
31. A. M. Perelomov, V. S. Popov, and M. V. Terent'ev, Zh. Éksp. Teor. Fiz. **50**, 1993 (1966) [Sov. Phys. JETP **23**, 924 (1966)]; **51**, 309 (1966) [**24**, 207 (1966)].
32. G. F. Gribakin and M. Yu. Kuchiev, Phys. Rev. A **55**, 3760 (1997).
33. V. D. Mur, S. V. Popruzhenko, S. G. Pozdnyakov, and V. S. Popov, Phys. Lett. A **316**, 226 (2003).
34. L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1964)].
35. T. Brabec and F. Krausz, Rev. Mod. Phys. **72**, 545 (2000).
36. V. S. Popov, Usp. Fiz. Nauk **174**, 921 (2004) [Phys. Usp. **47**, 855 (2004)].
37. A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions and Decays in Nonrelativistic Quantum Mechanics*, 2nd ed. (Nauka, Moscow, 1971; Israel Program for Scientific Translations, Jerusalem, 1966).
38. P. de Laplace, *Traite de mechanique celeste* (Bachelier, Paris, 1829), Vol. 1, Chap. 3.
39. V. A. Fock, Z. Phys. **98**, 145 (1935).
40. V. Bargmann, Z. Phys. **99**, 576 (1935).
41. M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330 (1966).
42. A. M. Perelomov and V. S. Popov, Zh. Éksp. Teor. Fiz. **50**, 179 (1966) [Sov. Phys. JETP **23**, 118 (1966)]; Dokl. Akad. Nauk SSSR **181**, 320 (1968).
43. V. S. Popov, in *Physics of High Energies and Theory of Elementary Particles* (Naukova Dumka, Kiev, 1967), p. 702 [in Russian].
44. N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations* (Nauka, Moscow, 1965; Am. Math. Soc., Providence, 1968).
45. I. S. Gradshteĭn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 4th ed. (Fizmatlit, Moscow, 1962; Academic, New York, 1980).
46. C. Eckart, Phys. Rev. **35**, 1303 (1930).
47. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953; Inostrannaya Literatura, Moscow, 1959), Vol. 2, Sect. 12.3.
48. V. L. Pokrovskiĭ, S. K. Savvinykh, and F. R. Ulinich, Zh. Éksp. Teor. Fiz. **34**, 1272 (1958) [Sov. Phys. JETP **7**, 879 (1958)].
49. V. L. Pokrovskiĭ and I. M. Khalatnikov, Zh. Éksp. Teor. Fiz. **40**, 1713 (1961) [Sov. Phys. JETP **13**, 1207 (1961)].
50. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Ed. by M. Abramowitz and I. A. Stegun (Dover, New York, 1965; Nauka, Moscow, 1979).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 830–843.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 958–973.
Original Russian Text Copyright © 2005 by Anisimov, Popov.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Continuous Spin-Reorientation Phase Transition in the Surface Region of an Inhomogeneous Magnetic Film

A. V. Anisimov and A. P. Popov

Moscow State Engineering Physics Institute, Kashirskoe sh. 31, Moscow, 115409 Russia
e-mail: ap_popov@nm.ru

Received February 21, 2005

Abstract—A continuous spin-reorientation transition from a uniform magnetic state with the in-plane orientation of the moments of all atomic layers to a nonuniform canted state in the surface region is considered. This transition was discovered in experiments on the divergence of magnetic susceptibility in a perpendicular magnetic field at a temperature of about 240 K, which is lower than the Curie point of gadolinium, equal to 292.5 K. These experiments were carried out on an ultrathin iron magnetic film deposited on the (0001) surface of a thin gadolinium film. It is shown that, in the vicinity of the spin-reorientation transition, the thermodynamic potential has a form characteristic of the Landau theory of second-order phase transitions. The orientation angle of the moment of the surface atomic layer with respect to the plane of the film, which is chosen as an order parameter, exhibits anomalous behavior and increases with temperature. Expressions are derived for the magnetic susceptibility of each atomic layer. It is shown that, in the vicinity of the transition, the irregular part of the magnetic susceptibility of each atomic layer exhibits behavior characteristic of the susceptibility in the Landau theory: it is less by a factor of two in the low-symmetry phase and diverges at the transition point. The regular part of the magnetic susceptibility of each atomic layer makes an additional contribution to the asymmetry of the total susceptibility in the vicinity of the transition point; this result follows from the fact that the inhomogeneous magnetic system considered is semi-infinite. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Magnetic properties of films consisting of layers of different materials are of considerable fundamental and applied interest. The phenomenon of giant magnetoresistance [1], spin-reorientation transitions, etc., have been observed in these films. Although there are a large number of experimental studies on the magnetization processes and spin-reorientation transitions in these processes (see, for example, [2, 3]), theoretical models that have been used until recently either deal with an infinite medium or rely on numerical analysis [4]; therefore, these models do not give a complete picture of the behavior of a magnetic structure in an external magnetic field. The main difficulty associated with the theoretical analysis of these systems lies in the consideration of the inhomogeneity of the magnetic structure that is formed near the interface between different media on the scale on the order of the lattice constant due to the change in the chemical composition of the films. As a result, the parameters such as magnetization and magnetic susceptibility, which characterize different magnetic states and phase transitions between them, become essentially dependent on the number of an atomic layer at a distance on the order of the lattice constant. Therefore, considerable attention has recently been paid to the study of the surface and bulk anisotropies and to finding out the role of the exterior surface in the process of reorientation analyzed by the models that take into account the real geometry of a magnet [5–7]. In the present paper, we apply a method analogous to that used in [5–7] to analyze the effect of an external magnetic field perpendicular to the surface of a Fe/Gd two-layer magnetic film and a continuous spin-reorientation transition discovered in this film.

In experiments with a two-layer film consisting of one and a half atomic layers of Fe deposited on the surface of a Gd(0001) thin film, it was established that, as temperature increases, two successive spin-reorientation transitions occur in this film at temperatures below the Curie point of bulk Gd, which is equal to 292.5 K [8]. At low temperatures, a magnetic structure is realized in which the moments of all atomic layers lie in the plane of the film; in this case, the surface moment is antiparallel to those of lower lying Gd layers. An increase in temperature to about 240 K leads to a continuous spin-reorientation transition to a state that we call here a Néel-domain-wall-like canted state. In this state, the surface moment and the moments of lower lying Gd layers deviate from the in-plane orientation; deeper into the crystal, the orientation of the moments of Gd atomic layers gradually approaches the in-plane orientation. A further increase in temperature leads to a slow increase in the deviation angle of the surface moment from the in-plane orientation. At a temperature as high as 280 K, a discontinuous spin-reorientation transition occurs from one canted state to another, in

which the surface moment is virtually perpendicular to the plane of the film.

Both transitions were discovered and investigated by means of the spin-polarized secondary electron emission spectroscopy method. This method is sensitive to the magnetic state of only a few surface atomic layers. Later, this system, denoted here as 1.5Fe/Gd(0001), was investigated by a method based on the magnetooptic Kerr effect [9]. This method allows one to analyze the magnetic state of deeper bulk layers of Gd. In these experiments, it was established that the moments of many Gd layers deviate from the in-plane orientation under spin-reorientation transitions; thus, many Gd layers take part in each spin-reorientation transition.

A continuous transition from the state with a uniform orientation of the moments of atomic layers to a domain-wall-like canted state is characteristic of the films that consist of layers with different chemical compositions. In this sense, this type of transition is unique. Indeed, it does not occur in any magnetic film with homogeneous chemical composition. In 1954, Néel showed that the anisotropy of a surface layer may differ from that of internal layers even in chemically homogeneous ferromagnets such as Fe, Co, Ni [10]. Therefore, one may expect that, when, say, easy-axis anisotropy is realized in the surface layer, while easy-plane anisotropy is realized in the bulk layers, a domain-wall-like canted magnetic state can be formed in the subsurface region. However, the exchange-interaction energy in these ferromagnets is much greater than the anisotropy energy; therefore, the surface moment cannot deviate from the in-plane orientation [11]. Naturally, the redistribution of the electron density between atomic layers in the surface region occurs even in chemically homogeneous films; i.e., the Friedel oscillations of the electron density and the associated phenomenon of interlayer relaxation occur. These factors, together with the difference between the environment symmetry of atoms on the surface and in the bulk of a crystal, lead not only to a difference between the surface and bulk anisotropies but also to a difference between the exchange interactions in the surface region and in the bulk of the crystal. However, experiments with Fe, Co, and Ni films show that the renormalization of these parameters is not sufficient for the formation of a canted domain-wall-like structure in the surface region. A different situation is realized in the two-layer system of 1.5Fe/Gd(0001). Here, the deposition of an ultrathin Fe layer onto the surface of a thin Gd(0001) film with easy plane anisotropy leads to the formation of an amorphous Fe/Gd film in the surface region of a sample up to the Curie point of gadolinium [12]. It is well known that amorphous Fe/Gd films are characterized by easy axis anisotropy, which favors a perpendicular orientation of the magnetic moment with respect to the plane of the film [13]. As a result, an ultrathin Fe/Gd film with anisotropy different from the bulk anisotropy and enhanced exchange interaction is formed on the surface of a Gd film. Due to the latter fact, the Curie point of the surface layer in the 1.5Fe/Gd(0001) system, 350 K, turns out to be higher than the Curie point of Gd, which is 292.5 K [8].

The so-called one-layer approximation provides the simplest explanation for the transition to a canted state in the surface region. Within this approach, it is assumed that only the magnetic state of the topmost atomic layer is different from the magnetic state of bulk layers. In the vicinity of the Curie point of Gd, the energy of exchange interaction $J_{SB}M_SM_B$ between the surface atomic layer and the adjacent subsurface layer, which is assumed to be a bulk layer in the one-layer approximation, decreases to zero, because the magnetization M_B of the bulk layers tends to zero. At the same time, the surface anisotropy energy $K_SM_S^2$ remains finite in this narrow interval of temperatures and is virtually independent of temperature. Therefore, at a certain temperature, the anisotropy energy of the surface becomes comparable to the energy of exchange interaction in the surface region. Hence, a spin-reorientation transition from a state with a uniform orientation of the moments of atomic layers to a canted domain-wall-like state occurs in the 1.5Fe/Gd(0001) system. Naturally, the interpretation of the spin-reorientation transition to a nonuniform canted domain-wall-like magnetic state within the one-layer approximation is not quite correct. In this approximation, the thickness of a domain wall is on the order of the lattice constant, which contradicts both the experimental data of [9] and the available data on the relation between the anisotropy energy and the exchange-interaction energy in Gd. However, this interpretation at least provides a qualitative explanation for the deviation of the surface moment from the in-plane orientation. Below, we describe a continuous spin-reorientation transition to a canted domain-wall-like state in the surface region with regard to the deviation of the magnetic moment in many atomic layers of the film.

In this paper, we consider only the first of the two spin-reorientation transitions: a continuous spin-reorientation transition from a state with a uniform orientation of the moments of atomic layers to a canted domain-wall-like state. This state was discovered by examining a peak of magnetic susceptibility in a perpendicular magnetic field [8]. According to the Landau theory of second-order phase transitions, the divergence of susceptibility at a certain temperature implies that a continuous second-order phase transition occurs at this temperature [14].

The description and the physical interpretation of the continuous spin-reorientation transition to a canted state in the surface region require that one should take into account the deviation of the moments of many atomic layers from the in-plane orientation. This is a consequence of the fact that the value of the bulk

anisotropy is small compared with the energy of interlayer exchange interaction. The consideration of many layers substantially complicates the description of the continuous phase transition due to the complexity of the thermodynamic potential that describes this system,

$$\Phi_0 = -J_{SB}M_S M_B \cos(\theta_1 - \theta_2) - J_{BB}M_B^2 \sum_{n=2}^{\infty} \cos(\theta_n - \theta_{n+1}) + K_S M_S^2 \sin^2\theta_1 + K_B M_B^2 \sum_{n=2}^{\infty} \sin^2\theta_n. \quad (1)$$

Nevertheless, the problem of the description of this transition with regard to many layers, i.e., within model (1), was largely solved. First, based on stability theory, a criterion for the transition form a magnetic state with a uniform orientation of moments to a nonuniform canted state was derived for a semi-infinite ferromagnet. The effect of an external field parallel to the plane of the film on this criterion, the finiteness of the film thickness, the influence of the anisotropy of the substrate onto which the film was deposited, and the inhomogeneity of the chemical composition of the film on this criterion was studied [15]. However, the effect of a field perpendicular to the film surface was investigated only in the limiting case of the infinite constant of bulk anisotropy $K_B \longrightarrow \infty$. Moreover, a ($k_S$, k_B)-phase diagram was constructed that indicates domains where a uniform magnetic state and a nonuniform canted magnetic state in the surface region are realized [15]; k_S and k_B are dimensionless surface and bulk reduced anisotropy constants, respectively, which are defined as follows:

$$k_S = \frac{2K_S M_S^2}{J_{SB}M_S M_B}, \quad k_B = \frac{2K_B M_B^2}{J_{BB}M_B^2}, \quad \gamma = \frac{J_{SB}M_S M_B}{J_{BB}M_B^2}. \quad (2)$$

The parameter γ takes into account, in the simplest approximation, the inhomogeneity of the film, i.e., the difference between the surface and bulk values of the exchange interaction, which is associated with the inhomogeneity of the chemical composition of the 1.5Fe/Gd(0001) film. Later, a criterion for the transition to a canted state in the surface region was again obtained by the method of nonlinear area-preserving maps [16].

In [17], along with a continuous transition to a canted state, which occurs at relatively low temperatures, we also described a discontinuous transition, occurring at a higher temperature, from one canted state to another canted state in which the surface moment is nearly perpendicular to the film plane. In that paper, we actually calculated the moment of an atom in each atomic layer and its orientation as a function of temperature for any set of the model parameters. Then, using the results obtained, we calculated a signal recorded in the experiment based on the magnetooptic Kerr effect with regard to the exponential decrease in the contribution of each atomic layer as the index of a layer increases; the curve obtained was compared with the experimental curve and showed good agreement.

In spite of these achievements, there still remain a number of questions concerning the description of a continuous spin-reorientation phase transition from a uniform magnetic state to a nonuniform canted state in the surface region. First, the derivation of the criterion for the transition to a canted state is based on the expansion of thermodynamic potential (1) only up to quadratic terms in the orientation angles θ_n of atomic layers, regardless of which method is used, [15] or [16]. This approximation allows one to derive the criterion itself but does not allow one to unambiguously judge the kind of the spin-reorientation phase transition described by model (1). Indeed, according to the Landau theory of second-order phase transitions, the thermodynamic potential can be expanded in a series with respect to the order parameter η at the transition point $T = T_C$:

$$\Delta\Phi = a(T - T_C)\eta^2 + B\eta^4. \quad (3)$$

The transition at $T = T_C$ from a low-symmetry state to a high-symmetry state is a second-order phase transition if the coefficient of η^2 changes its sign at the transition point and the coefficient B of η^4 is positive. Intuitively, it is obvious that a smooth increase in the surface anisotropy constant, which favors the perpendicular orientation of the surface moment, should lead to a continuous spin-reorientation transition to a canted state in the surface region. However, based on expression (1) for the thermodynamic potential of a semi-infinite crystal, it is rather difficult to judge the sign of the coefficient multiplying the fourth power of the order parameter. Moreover, it is not quite clear what physical parameter can serve as the order parameter under a continuous transition to a canted state and whether thermodynamic potential (1) in the vicinity of the phase transition has a form characteristic of the Landau thermodynamic potential (3). Therefore, in [17], we determined the kind of the phase transition only for particular values of the model parameters by numerical simulation on a computer. We also established that, in the limiting case of the infinite value of the bulk anisotropy constant, $K_B \longrightarrow \infty$, the boundary that separates the domain with a canted state from the domain with the in-plane orientation of the moments of all layers in the (k_S, k_B) phase diagram corresponds to a second-order

spin-reorientation phase transition. The first problem solved in the present paper is the derivation of an expression for the thermodynamic potential in the form characteristic of the Landau theory (3) on the basis of expression (1) for the thermodynamic potential of a semi-infinite inhomogeneous crystal in the vicinity of the transition point.

The solution of this problem will allow one to make an unambiguous conclusion that the spin-reorientation transition from a magnetic state with a uniform orientation of the moments of atomic layers to a canted state in the surface region, which corresponds to the intersection of a line that separates the relevant domains in the earlier constructed (k_S, k_B) phase diagram, is a second-order phase transition for any values of the model parameters k_S, k_B, and γ. Moreover, the Landau theory implies that the order parameter η that appears in expression (3) for the thermodynamic potential is equal to zero in a high-symmetry phase and is different from zero in a low-symmetry phase. The solution of the first problem formulated above allows one to answer the question of what is the order parameter in the continuous spin-reorientation transition considered and how does it behave with temperature. Finally, the solution of this problem is necessary for solving the next problem, which is the main problem of the present paper: the derivation of an expression for the magnetic susceptibility of an inhomogeneous magnetic film described by thermodynamic potential (1) in a perpendicular field in the vicinity of a continuous spin-reorientation transition. The magnetic susceptibility of a two-layer 1.5Fe/Gd(0001) film is the basic physical quantity that is measured in the experiment. However, the temperature dependence of the magnetic susceptibility of an inhomogeneous 1.5Fe/Gd(0001) film has not been investigated theoretically or discussed. According to the Landau theory, the magnetic susceptibility diverges in the vicinity of a second-order phase transition, and its value in a low-symmetry phase is twice that in a high-symmetry phase. A question of whether these properties follow from model (1), which describes an inhomogeneous magnetic film, is the question that can be answered by solving the second problem considered in this paper. It is also of interest to consider how the chemical inhomogeneity and the bounded geometry of the film influence the above-mentioned features in the behavior of the magnetic susceptibility in the vicinity of a continuous phase transition to a nonuniform domain-wall-like magnetic state in the surface region. The solution of the second problem considered in the present paper will also help to answer the question as to whether it is correct, based on the divergence of the magnetic susceptibility, to classify the discovered transition to a canted state as a second-order phase transition.

One should bear in mind that, in a chemically inhomogeneous semi-infinite film, the magnetic susceptibility is a physical parameter that depends on the index of an atomic layer. Therefore, the derivation of an expression for the magnetic susceptibility meets mathematical difficulties associated with the fact that one should derive a separate expression for the magnetic susceptibility of each atomic layer. To our knowledge, presently there is only one publication, [18], in which the authors consider the influence of a perpendicular field on an inhomogeneous semi-infinite ferromagnet described by model (1). However, these authors investigated only the evolution of magnetization profiles with a perpendicular field within the nonlinear mapping formulation of the mean-field theory; they did not consider the magnetic susceptibility in a perpendicular magnetic field.

In Section 1, we establish a relation between the orientation angles of the moments of atomic layers in a canted state in the surface region of an inhomogeneous semi-infinite ferromagnet and the orientation angles in an imaginary infinite domain wall in a homogeneous magnet ($\gamma = 1$) obtained by an artificial continuation of the semi-infinite magnet beyond its surface. The solution of this technical problem has been stimulated by the fact that, to prove that thermodynamic potential (1) in the vicinity of a continuous spin-reorientation transition has a form characteristic of the Landau theory (3) (which is demonstrated in Section 2), one should express each orientation angle θ_n in terms of the orientation angle θ_1 of the surface layer using the so-called equilibrium conditions. Then, one should substitute the expressions obtained into the expression for thermodynamic potential (1). It turns out that this procedure is significantly simplified if one uses the above-mentioned concept of an imaginary domain wall of an infinite magnet and expresses the orientation angles θ_n of the moments of atomic layers in a real canted magnetic structure in the surface region in terms of fictitious orientation angles Θ_1 of the moments of atomic layers in the imaginary domain wall of an infinite homogeneous magnet.

In Section 3, we derive expressions for the regular and irregular components of magnetic susceptibility in a perpendicular magnetic field as a function of the index of an atomic layer in the system described by model (1). We show that, according to the Landau theory of second-order phase transitions, the magnetic susceptibility of each atomic layer diverges at the transition point and its value in a low-symmetry phase is half that in a high-symmetry phase. The theoretical results obtained are compared with experimental data. The regular component of the magnetic susceptibility is analyzed similarly. To conclude this section, we stress that a direct calculation of the magnetic susceptibility in a perpendicular field proves to be rather difficult. At the same time, the derivation of expressions for the magnetic susceptibility on the basis of the Landau theory and involving the concept of an order parameter substantially simplifies the solution of this problem.

To conclude the Introduction, we make the following two remarks concerning formula (1) for the thermodynamic potential. The transition observed in the experiment occurs at a sufficiently low temperature (at about 230–240 K, which is by 50–60 K below the Curie point of Gd, equal to 292.5 K, and by 110–120 K below the Curie point of the surface, equal to 350 K), the measured value of the magnetization of Gd layers still being close to the saturation value. Therefore, there are few magnons in the system. Thus, one can write out a mean-field thermodynamic potential by introducing the magnetization amplitudes $M_B(T)$. The only possible alternative state for the collinear states to which we restrict our consideration here is the Néel surface wall. Moreover, in view of the statement of the first problem (see above), we only focus on the very beginning of the transition, when the deviation of moments from the in-plane orientation is arbitrarily small. Taking into account that the transition occurs at a low temperature, we can neglect the possible angular dependence of the moment of Gd to a good accuracy. In addition, thermodynamic potential (1) is expressed in a discrete layer-by-layer approximation. The gradient term that enters the well-known Ginzburg–Landau functional, which is used, in particular, for describing domain walls, arises in (1) in the explicit form when passing to a continuum approximation by the formula

$$\cos(\theta_n - \theta_{n+1}) \approx 1 - \frac{a^2}{2}\left[\left|\frac{d\theta(x)}{dx}\right|_{x = na}\right]^2,$$

where a is the interlayer distance. Then, we pass from summation to integration in (1). It can easily be shown that the shape of a wall is described in this approximation by the equation

$$\frac{d^2\theta}{dx^2} = \frac{k_B}{2a^2}\sin(2\theta(x)).$$

Thus, there is no need to include the gradient term in (1) because it is already contained there in the implicit form. The continuum method yields approximate expressions for the second- and fourth-order derivatives of the thermodynamic potential with respect to the orientation angle of the surface layer; these expressions represent the limiting expressions for the exact formulas obtained in this paper by a discrete method. We mean the passage to the limit $\lambda \longrightarrow 1 - \sqrt{k_B}$. The parameter λ is introduced exclusively for convenience, to parameterize the bulk anisotropy constant k_B by the formula $k_B = (1 - \lambda)^2/\lambda$. When λ is varied from zero to unity, k_B varies from zero to infinity. The applicability condition of the continuum method is given by $\sqrt{k_B} \ll 1$ and $\lambda \approx 1$. In this case, if we pass again to the discrete variables, the shape of the domain wall is described by the expression

$$\theta_n = 2\arctan\left\{\tan\left(\frac{\Theta_1}{2}\right)\exp\left(\sqrt{k_B}(n-1)\right)\right\}.$$

As λ decreases, the shape of the wall becomes substantially different from that described by this expression. Therefore, it is essential to verify that the fourth-order derivative at the transition point is positive for any possible value of λ in the interval $0 < \lambda < 1$, i.e., for any value of the reduced bulk anisotropy constant k_B from zero to infinity. Thus, the results obtained below (see Section 2) are of more general character compared to those that would be obtained if one used the continuum method. In spite of this fact, note that the application of the continuum approximation turned out to be rather fruitful, for example, for the description and qualitative explanation of surface magnetism [19–21]. At the same time, the construction of the (k_S, k_B) diagram of the magnetic states of a semi-infinite magnet within the continuum approximation, which was performed in [22, 23], proved to be qualitatively incorrect, which was pointed out in [15]. Here, we focus on finding out the kind and the physical nature of a continuous spin-reorientation phase transition precisely from the viewpoint of the correct (k_S, k_B) phase diagram, which was obtained earlier within the discrete method [15]. This is another reason why we apply a discrete approximation in the present study.

2. DERIVATION OF EXPRESSIONS FOR THE ORIENTATION ANGLES OF THE MOMENTS OF ATOMIC LAYERS IN TERMS OF THE ORIENTATION ANGLE OF THE MOMENT OF THE FIRST LAYER

Using the reduced anisotropy constants k_S and k_B (2) introduced in the previous section, we can write out the following expression for the thermodynamic potential reduced to the dimensionless form:

$$\varphi_0 \equiv \frac{\Phi_0}{J_{BB}M_B^2} = \frac{\gamma k_S}{2}\sin^2\theta_1 - \gamma\cos(\theta_1 - \theta_2) + \sum_{n=2}^{\infty}\left(\frac{k_B}{2}\sin^2\theta_n - \cos(\theta_n - \theta_{n+1})\right). \quad (4)$$

Searching for the state of the system described by this thermodynamic potential, i.e., the minimization of this potential over each orientation angle θ_n, leads to

the following infinite set of equations for the orientation angles θ_n:

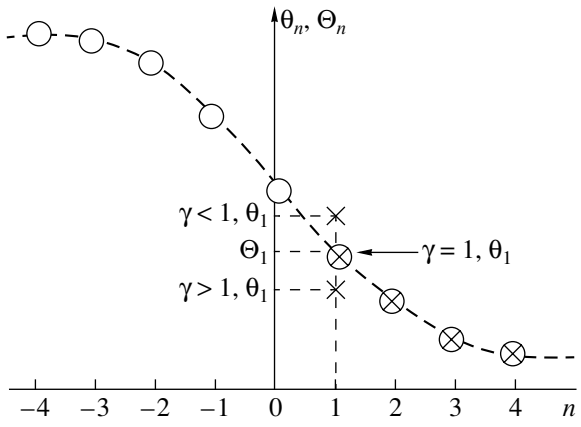
$$\frac{\partial \varphi_0}{\partial \theta_1} = \frac{\gamma k_S}{2}\sin(2\theta_1) + \gamma\sin(\theta_1 - \theta_2) = 0, \quad n = 1, \tag{5.1}$$

$$\frac{\partial \varphi_0}{\partial \theta_2} = \frac{k_B}{2}\sin(2\theta_2) + \sin(\theta_2 - \theta_3) + \gamma\sin(\theta_2 - \theta_1) = 0, \quad n = 2, \tag{5.2}$$

$$\frac{\partial \varphi_n}{\partial \theta_n} = \frac{k_B}{2}\sin(2\theta_n) + \sin(\theta_n - \theta_{n+1}) + \sin(\theta_n - \theta_{n-1}) = 0, \quad n > 2. \tag{5.3}$$

In the Landau theory, thermodynamic potential (3) corresponds to a nonequilibrium state until the order parameter η reaches its equilibrium value; i.e., the order parameter η is a variable quantity. Below, we will show that the orientation angle θ_1 of the surface atomic layer may serve as an order parameter for a continuous spin-reorientation transition to a canted magnetic state in the surface region described by model (4). Therefore, the orientation angle θ_1 of the surface is interpreted here as a variable parameter; i.e., it must not be equal to its equilibrium value. Hence, we will not use the equilibrium condition (5.1). The orientation angles of all the other layers depend on the orientation angle θ_1 of the surface via the recurrence equations (5.2) and (5.3).

In accordance with the formulation of the first problem of this paper, we must express each orientation angle θ_n in terms of the order parameter θ_1 and then substitute the expressions obtained for these angles into the formula for thermodynamic potential (4). However, due to the inhomogeneity of the chemical composition of the film, $\gamma \neq 1$, the equilibrium condition obtained by the differentiation with respect to the orientation angle of the second layer (5.2) does not coincide with similar equilibrium conditions for all the other orientation angles (5.3). This fact substantially complicates the solution of the first problem of the present paper. To avoid this difficulty, we introduce the concept of an imaginary domain wall in an infinite homogeneous crystal in which the orientation angle of each atomic layer satisfies the equilibrium conditions (5.3). In the present section, we find relations between real orientation angles in a canted magnetic state in the surface region and fictitious orientation angles in the imaginary domain wall of an infinite homogeneous crystal, as well as their expressions in terms of the order parameter θ_1 in the vicinity of a continuous spin-reorientation transition to a canted magnetic state.



Fictitious orientation angles of moments (circles) in an imaginary domain wall and real orientation angles of moments (crosses) in a canted magnetic structure in the surface region as functions of the number n of an atomic layer. When $\gamma > 1$, the orientation angle θ_1 of the surface layer is less than the corresponding value of the fictitious angle Θ_1 in the imaginary domain wall. When $\gamma < 1$, the orientation angle θ_1 of the surface layer is greater than the corresponding value of the fictitious angle Θ_1 in the imaginary domain wall. When $\gamma = 1$, the angle θ_1 coincides with Θ_1. For $n = 2$, 3, 4, …, the real orientation angles θ_n of moments coincide with the fictitious orientation angles Θ_n for any value of the parameter γ.

For $\gamma \neq 1$, the profile of magnetization represents a part of a domain wall for all atomic layers except for the first. The equilibrium value of the orientation angle θ_1 of the first layer differs from the corresponding orientation angle in the imaginary domain wall, which is denoted by Θ_1. If the exchange interaction between the surface and the adjacent subsurface layers is less than the exchange interaction between adjacent atomic layers in the bulk of the magnet, $\gamma < 1$, then the real orientation angle θ_1 of the surface is greater than the fictitious orientation angle Θ_1 in the domain wall. If the exchange interaction between the surface and the adjacent subsurface layers is greater than the exchange interaction between adjacent atomic layers in the bulk of the magnet, $\gamma > 1$, then the real orientation angle θ_1 of the surface is less than the fictitious orientation angle Θ_1 in the domain wall. Thus, in an inhomogeneous magnetic film ($\gamma \neq 1$), there is a jump in the equilibrium value of the orientation angle θ_1 of the surface layer with respect to the corresponding value of the fictitious angle Θ_1 in the domain wall of a homogeneous magnet, as illustrated in the figure. Therefore, we use the following notation here: Θ_n are fictitious angles in a domain wall in an infinite homogeneous magnet, and θ_n are real orientation angles in a canted magnetic state in the surface region. For $n > 1$, we have $\Theta_n = \theta_n$, while for

$n = 1$, the relation between Θ_1 and θ_1 will be established below (see (7), (18), and (20)).

The fictitious orientation angles in the imaginary domain wall satisfy the recurrence equation (5.3). Since the angle Θ_1 also belongs to this domain wall, it must obviously satisfy a similar equation

$$\frac{k_B}{2}\sin 2\theta_2 + \sin(\theta_2 - \Theta_1) + \sin(\theta_2 - \theta_3) = 0. \quad (6)$$

A comparison of this equation with Eq. (5.2) shows that the relation between the real orientation angle θ_1 and the corresponding fictitious orientation angle Θ_1 in the domain wall is given by the formula

$$\gamma\sin(\theta_2 - \theta_1) = \sin(\theta_2 - \Theta_1). \quad (7)$$

A transition from the real orientation angle θ_1 of the surface atomic layer to the fictitious orientation angle Θ_1 in the imaginary domain wall allows us to apply, instead of Eq. (5.2), Eq. (6), which is analogous to Eq. (5.3); i.e., the system of recurrence equations (5.2), (5.3) becomes in a sense homogeneous.

To obtain an expansion of thermodynamic potential (4) up to terms of the fourth order in each orientation angle, we must first express the fictitious orientation angles Θ_n in terms of the fictitious orientation angle Θ_1 of the surface in the imaginary domain wall up to the cubic terms. To this end, we must expand the left-hand side of the recurrence equation in a series in terms of fictitious orientation angles in the imaginary domain wall,

$$\begin{gathered}\frac{k_B}{2}\sin 2\Theta_n + \sin(\Theta_n - \Theta_{n-1}) \\ + \sin(\Theta_n - \Theta_{n+1}) = 0, \\ n \geq 2,\end{gathered} \quad (8)$$

up to the cubic terms in each orientation angle Θ_n. For $n \geq 2$, we will seek a solution in the form

$$\Theta_n = \alpha_n\Theta_1 + \frac{\beta_n\Theta_1^3}{6}. \quad (9)$$

Obviously, for $n = 1$, we should set

$$\alpha_1 = 1, \quad \beta_1 = 0 \quad (10)$$

in this formula. Upon substituting (9) into the left-hand side of Eq. (8) expanded up to Θ_1^3 and setting the coefficients at equal powers of Θ_1 to zero, we obtain the following recurrence equations for the coefficients α_n and β_n:

$$(k_B + 2)\alpha_n - \alpha_{n-1} - \alpha_{n+1} = 0, \quad n \geq 2, \quad (11.1)$$

$$(k_B + 2)\beta_n - \beta_{n+1} - \beta_{n-1} = \eta_0\alpha_n^3, \quad n \geq 2. \quad (11.2)$$

Hereinafter, for convenience, we use the parameter λ instead of the reduced bulk anisotropy constant k_B. The relation between λ and k_B is given by

$$k_B = \frac{(1-\lambda)^2}{\lambda}. \quad (12)$$

As λ is increased from 0 to 1, the reduced bulk anisotropy constant k_B decreases from $+\infty$ to 0. Then, the coefficient η_0 can be expressed as

$$\begin{aligned}\eta_0 &= 4k_B - (\lambda - 1)^3 + \left(1 - \frac{1}{\lambda}\right)^3 \\ &= -\frac{(1-\lambda^2)(\lambda^2 - 3\lambda + 1)}{\lambda^3}.\end{aligned} \quad (13)$$

A solution to homogeneous equation (11.1) is given by

$$\alpha_n = \lambda^{n-1}. \quad (14)$$

By analogy with the case of inhomogeneous differential equations, we will seek a solution to the inhomogeneous recurrence equation (11.2) as a sum of the solution to the corresponding homogeneous recurrence equation and the partial solution to the inhomogeneous equation (11.2):

$$\beta_n = \tilde{\beta}\alpha_n + \bar{\beta}\alpha_n^3. \quad (15)$$

Substituting this expression for β_n into the recurrence equation (11.2), we obtain the following formula for the coefficient $\bar{\beta}$:

$$\bar{\beta} = \frac{1 - 3\lambda + \lambda^2}{1 + \lambda^2}. \quad (16)$$

It follows from boundary conditions (10) that $\tilde{\beta} = -\bar{\beta}$. As a result, we obtain the following expression for an arbitrary fictitious orientation angle Θ_n in the imaginary domain wall as a function of the first fictitious orientation angle Θ_1:

$$\Theta_n = \lambda^{n-1}\Theta_1 + \frac{\lambda^2 - 3\lambda + 1}{\lambda^2 + 1}[\lambda^{3(n-1)} - \lambda^{n-1}]\frac{\Theta_1^3}{6}. \quad (17)$$

For small deviations of the moment vector of the surface atomic layer from the in-plane orientation, there exists a one-to-one correspondence between the real orientation angle θ_1 and the fictitious orientation angle Θ_1 introduced by formula (7). Let us find the relation between Θ_1 and θ_1 up to the cubic terms. To this end, we must expand both sides of Eq. (7) up to the cubic terms and apply the expression for Θ_2 in the form (17). A solution is sought in the form

$$\theta_1 = a\Theta_1 + \frac{c\Theta_1^3}{6}. \quad (18)$$

Expressions for the coefficients a and c can be obtained after equating the coefficients at equal powers of Θ_1:

$$a = \frac{1-(1-\gamma)\lambda}{\gamma},$$
$$c = \frac{(1-\lambda)^3}{\gamma}\left(\frac{1}{\gamma^2}-1\right)+\beta_2\left(1-\frac{1}{\gamma}\right), \quad (19)$$
$$\beta_2 = -(\lambda^2-3\lambda+1)\lambda\frac{1-\lambda^2}{1+\lambda^2}.$$

The inverse formula, which expresses the fictitious angle Θ_1 in terms of the real orientation angle θ_1, is obtained analogously:

$$\Theta_1 = \xi\theta_1+\varepsilon\frac{\theta_1^3}{6}, \quad \xi = \frac{1}{a}, \quad \varepsilon = -\frac{c}{a^4}. \quad (20)$$

Finally, we express the fictitious angles Θ_n for $n \geq 2$ in terms of the real orientation angle θ_1 of the surface atomic layer up to the cubic terms with the use of formulas (17) and (20):

$$\Theta_n = \hat{\alpha}_n\theta_1+\frac{\tilde{\beta}_n}{6}\theta_1^3, \quad \hat{\alpha}_n = \frac{1}{a}\lambda^{n-1},$$
$$\hat{\beta}_n = \frac{\bar{\beta}}{a^3}\lambda^{3(n-1)}-\left(\frac{c}{a^4}+\frac{\bar{\beta}}{a^3}\right)\lambda^{n-1}. \quad (21)$$

3. DERIVATION OF EXPRESSION (3) FOR THE THERMODYNAMIC POTENTIAL IN A FORM CHARACTERISTIC OF THE LANDAU THEORY

According to the Landau theory of second-order phase transitions, for a phase transition to be of the second order, it is necessary that the second derivative of the thermodynamic potential (3) with respect to the order parameter η change its sign at the transition point and that the fourth derivative (the coefficient B) be positive. In this section, we will show that the thermodynamic potential (4) of an inhomogeneous magnetic film satisfies these conditions in the vicinity of a spin-reorientation transition from a uniform magnetic state with the in-plane orientation of the moments of all layers to a domain-wall-like canted state in the surface region.

Based on the results of the previous section, we can conclude that, for arbitrary values of the model parameters λ and γ there exists a finite interval of values of the surface orientation angle θ_1 in which there exists a one-to-one correspondence between the orientation angles θ_n and θ_1 in a canted state in the surface region. The fact that the system is discrete is insignificant. Then, in the canted magnetic state, one can expand the thermodynamic potential (4) in a series in θ_1 up to the terms θ_1^4.

For simplicity and convenience, we decompose the expression for the thermodynamic potential (4) into the bulk and surface parts:

$$\varphi_0 = \Delta\tilde{\varphi}_0+\Delta\varphi_{0S},$$
$$\Delta\tilde{\varphi}_0 = \frac{\lambda-1}{2}\sin^2\Theta_1-\cos(\Theta_1-\Theta_2) \quad (22)$$
$$+\frac{k_B}{2}\sum_{n=2}\sin^2\Theta_n-\sum_{n=2}\cos(\Theta_n-\Theta_{n+1}),$$
$$\Delta\varphi_{0S} = \frac{\gamma(k_S-k_{SC})}{2}\sin^2\theta_1$$
$$+\frac{\gamma k_{SC}}{2}\sin^2\theta_1-\frac{\lambda-1}{2}\sin^2\Theta_1 \quad (23)$$
$$-\gamma\cos(\theta_1-\Theta_2)+\cos(\Theta_1-\Theta_2).$$

These formulas show that the bulk part $\Delta\tilde{\varphi}_0$, which corresponds to a domain wall in a homogeneous magnet, does not contain the parameter γ, which takes into account the inhomogeneity of the exchange interaction in an inhomogeneous magnetic film. It neither contains the reduced anisotropy constant k_S of the surface atomic layer, which is different from the reduced anisotropy constant k_B of the bulk layers of the magnet. These model parameters are contained in the surface part $\Delta\varphi_{0S}$ of the thermodynamic potential. Expanding $\Delta\tilde{\varphi}_0$ in a series with respect to each fictitious orientation angle Θ_n, substituting the expressions of Θ_n in terms of the angle Θ_1 for Θ_n by formula (17), and summing the geometric progressions obtained, we derive the following expression for the bulk part of the thermodynamic potential $\Delta\tilde{\varphi}_0$:

$$\Delta\tilde{\varphi}_0 \approx 3\frac{1-\lambda^2}{1+\lambda^2}\frac{\Theta_1^4}{4!}. \quad (24)$$

When deriving (24), we omitted all the terms that do not depend on the orientation angles. Formula (24) shows that the expression for $\Delta\tilde{\varphi}_0$ does not contain quadratic terms in Θ_1. The increment of the total thermodynamic potential can be expanded as follows. First, we should expand the surface part of the thermodynamic potential $\Delta\varphi_{0S}$ in a series with respect to each orientation angle, both real and fictitious, up to the fourth-order terms. Second, we should substitute expression (17) for the fictitious angle Θ_2 in terms of Θ_1. Third, we should substitute expression (20) for the fictitious angle Θ_1 in terms of the real orientation angle θ_1 of the surface atomic layer into (23) and (24). As a result, we obtain the following expression for the increment of the total thermodynamic potential $\Delta\varphi_0$ as a

function of the real orientation angle θ_1 of the surface atomic layer:

$$\Delta\varphi_0 \approx \frac{\gamma}{2}(k_S - k_{SC})\theta_1^2 + B\frac{\theta_1^4}{4!},$$

$$B \equiv \left.\frac{\partial^4\varphi_0}{\partial\theta_1^4}\right|_{\theta_1 = 0} = \frac{1}{a^4}\left[4\gamma(k_{SC} - k_S)a^4 + 3\frac{1-\lambda^2}{1+\lambda^2}\right. \quad (25)$$

$$\left. + 4(1-\lambda)(a^3 - 1) + (1-\lambda)^4\left(1 - \frac{1}{\gamma^3}\right)\right].$$

Here, k_{SC} is the critical value of the reduced constant of anisotropy of the surface atomic layer, which depends on the reduced bulk anisotropy constant k_B and γ (formula (42) in [15]). The graph of the function $k_{SC}(k_B)$ determines the boundary, on the phase diagram (k_S, k_B), between the domain corresponding to a homogeneous magnetic state with the in-plane orientation of all moments and the domain corresponding to a canted state in the surface region. It follows from (25) that the second derivative of the thermodynamic potential with respect to the order parameter θ_1 changes its sign when crossing the boundary between these domains on the (k_S, k_B) phase diagram. If the condition $k_S < k_{SC}$ holds, then the second derivative of the thermodynamic potential is negative; i.e., a canted state in the surface region becomes energetically favorable. When deriving formula (25), we used the following expression for the critical value of the constant k_{SC} of surface anisotropy that enters the criterion for the spin-reorientation transition to a canted state $k_S < k_{SC}$:

$$k_{SC} = -\frac{1-\lambda}{1-(1-\gamma)\lambda}. \quad (26)$$

The expression in square brackets in (25) is positive. Indeed, at the point of spin-reorientation transition, i.e., at $k_S = k_{SC}$, the fourth derivative of the thermodynamic potential with respect to the order parameter θ_1 can be rewritten as

$$B = \left.\frac{d^4\varphi_0}{d\theta_1^4}\right|_{\substack{k_S = k_{SC}\\ \theta_1 = 0}} = \frac{3}{a^4}\left[\frac{\lambda^4(1-\lambda^2)}{1+\lambda^2}\right. \quad (27)$$

$$\left. + \frac{(1-\lambda(1-2\gamma))^2(1-\lambda)^2}{\gamma^3}\right] > 0.$$

For $\gamma = 1$, the expression for the increment of the thermodynamic potential has an especially simple form:

$$\Delta\varphi_0 \approx (k_S - k_{SC})\frac{\theta_1^2}{2!} + 3\frac{1-\lambda^2}{1+\lambda^2}\frac{\theta_1^4}{4!}. \quad (28)$$

If the criterion $k_S < k_{SC}$ for the transition to a canted state is satisfied, then we can derive an equilibrium value of the orientation angle $\theta_1^{(0)}$ of the surface atomic layer from Eq. (25):

$$\theta_1^{(0)} = \sqrt{\frac{6\gamma(k_{SC} - k_S)}{B}}. \quad (29)$$

One can verify that this value of the order parameter θ_1 satisfies the equilibrium condition (5.1) up to the cubic terms. Based on the equilibrium condition (5.1) for the surface atomic layer, one can derive an expression for the coefficient B that coincides with the expression for this coefficient defined by formula (25).

Thus, in the vicinity of a continuous spin-reorientation transition to a canted state in the surface region, the thermodynamic potential (4) of a semi-infinite inhomogeneous magnet has a form characteristic of the Landau theory (3). The positiveness of the coefficient B at θ_1^4 shows that this transition is of the second order. Thus, the spin-reorientation transition from a uniform magnetic state with the in-plane orientation of the moments of all layers to a canted state in the surface region described by model (4) is a physical realization of the second-order phase transitions described by the Landau theory.

Note that, at low temperatures, i.e., for $k_{SC} < k_S$, the order parameter θ_1 vanishes; an increase in temperature leads to a continuous increase in the order parameter. Such behavior of the order parameter with temperature contradicts the vast majority of experimental data on second-order phase transitions. This makes the continuous spin-reorientation transition observed in the 1.5Fe/Gd(0001) system a unique phenomenon. However, according to the Landau theory, a decrease in the order parameter with temperature is not a law of nature [14], because entropy increases all the same.

To conclude this section, we note that one can choose the orientation angle of any atomic layer as an order parameter because the orientation angles of all atomic layers are interdependent. However, the most natural choice of the order parameter is the orientation angle of the surface atomic layer. Obviously, any odd function of the surface orientation angle, say $\sin\theta_1$, may also serve as the order parameter. Similar to the orientation angles, all values of $\sin\theta_n$ are also interdependent; therefore, one can choose the total projection of the moments of all atomic layers onto the normal to the film surface as the order parameter.

4. MAGNETIC SUSCEPTIBILITY OF ATOMIC LAYERS IN AN INHOMOGENEOUS MAGNETIC FILM IN A PERPENDICULAR MAGNETIC FIELD

In the presence of an external magnetic field perpendicular to the plane of the film, the expression for the

thermodynamic potential contains additional terms compared to expression (4) for φ_0:

$$\varphi = \varphi_0 - \mu_1 h_\perp \sin\theta_1 - \mu h_\perp \sum_{n=2}^{\infty} \sin\Theta_n. \quad (30)$$

Here, $h_\perp$ is the reduced value of the external perpendicular magnetic field, which is defined as

$$h_\perp = \frac{M_B H_\perp}{J_{BB} M_B^2}, \quad \mu_1 = \frac{M_S}{M_B}. \quad (31)$$

According to this definition of the parameter $h_\perp$, the coefficient μ in the last term in (30) must be set equal to unity. However, it is more convenient to preserve this coefficient because it helps to follow up the origin of different terms when deriving the final formulas for the susceptibility of each atomic layer in a perpendicular magnetic field. Therefore, below we treat the parameter μ_1 as the moment of the surface atomic layer and μ, as the moment of a bulk atomic layer. Of course, in all the formulas below, we must set $\mu = 1$.

With regard to the notations introduced, the equilibrium conditions for each orientation angle are rewritten as

$$\frac{\partial\varphi}{\partial\theta_1} = \frac{\gamma k_S}{2}\sin 2\theta_1 + \gamma\sin(\theta_1 - \Theta_2) - \mu_1 h_\perp \cos\theta_1 = 0, \quad n = 1, \quad (32.1)$$

$$\frac{\partial\varphi}{\partial\theta_2} = \frac{k_B}{2}\sin 2\Theta_2 + \sin(\Theta_2 - \Theta_3) + \gamma\sin(\Theta_2 - \theta_1) - \mu h_\perp \cos\Theta_2 = 0, \quad n = 2, \quad (32.2)$$

$$\frac{\partial\varphi}{\partial\theta_n} = \frac{k_B}{2}\sin 2\Theta_n + \sin(\Theta_n - \Theta_{n+1}) + \sin(\Theta_n - \Theta_{n-1}) - \mu h_\perp \cos\Theta_n = 0, \quad n > 2. \quad (32.3)$$

The switching on of the external field changes the orientation angle of the moment of each atomic layer compared to the equilibrium value at $h_\perp = 0$. In the bulk of the film, far from the surface, the deviation of the moment from the in-plane orientation is defined by the angle Θ_B, which is independent of the index of an atomic layer:

$$\sin\Theta_B = \frac{\mu h_\perp}{k_B}. \quad (33)$$

Then, the expression for the magnetic susceptibility χ_B of a bulk atomic layer in a perpendicular field is given by the formula

$$\chi_B = \frac{\mu^2}{k_B}. \quad (34)$$

Now, we will find an expression for the magnetic susceptibility of an arbitrary atomic layer in the surface region near the point of a continuous spin-reorientation transition to a canted state. In the canted state and at $h_\perp = 0$, the equilibrium profile with respect to the orientation angles represents a part of a Néel wall. Under the application of an external perpendicular field, the profile of the orientation angles represents a part of a modified Néel wall. Thus, the nonequilibrium thermodynamic potential of the canted state depends not only on the variable orientation angle θ_1 of the surface, the order parameter, but also on the external magnetic field $h_\perp$. In addition, the orientation angles of other layers, $n = 2, 3, \ldots$, are determined by Eqs. (32.2) and (32.3) for a fixed value of the orientation angle of the surface atomic layer:

$$\Delta\varphi(\theta_1, h) = \Delta\varphi_0(\theta_1, h_\perp) - \mu_1 h_\perp \sin\theta_1 - \mu h_\perp \sum_{n=2}^{\infty} \sin\Theta_n, \quad \Theta_n = \Theta_n(\theta_1, h_\perp). \quad (35)$$

Let us show that the first derivative of the total thermodynamic potential $\Delta\varphi_0$ with respect to the external field $h_\perp$ vanishes as the field value tends to zero. Indeed, after the rearrangement of terms, the expression for this derivative for a fixed value of the orientation angle θ_1 of the surface moment has the form

$$\left.\frac{\partial\Delta\varphi_0}{\partial h_\perp}\right|_0 = \gamma\sin(\Theta_2^{(0)} - \theta_1)\left.\frac{\partial\Theta_2}{\partial h_\perp}\right|_0 + \sum_{n=2}^{\infty}\left[\left.\frac{k_B}{2}\sin(2\Theta_n^{(0)})\frac{\partial\Theta_n}{\partial h_\perp}\right|_0 + \sin(\Theta_n^{(0)} - \Theta_{n+1}^{(0)})\left.\frac{\partial(\Theta_n - \Theta_{n+1})}{\partial h_\perp}\right|_0\right]$$

$$= \left[\gamma\sin(\Theta_2^{(0)} - \theta_1) + \frac{k_b}{2}\sin(2\Theta_2^{(0)}) + \sin(\Theta_2^{(0)} - \Theta_3^{(0)})\right]\left.\frac{\partial\Theta_2}{\partial h_\perp}\right|_0 + \sum_{n=3}^{\infty}\left[\frac{k_b}{2}\sin(2\Theta_n^{(0)}) + \sin(\Theta_n^{(0)} - \Theta_{n-1}^{(0)}) + \sin(\Theta_n^{(0)} - \Theta_{n+1}^{(0)})\right]\left.\frac{\partial\Theta_n}{\partial h_\perp}\right|_0 = 0. \quad (36)$$

Here, $\Theta_n^{(0)}$ are the equilibrium values of orientation angles for a fixed value of the orientation angle of the surface moment in zero external field for $n \geq 2$. Each set of square brackets in (36) contains the left-hand side of the equation for the equilibrium values of the orientation angles given by (5.2) and (5.3). Hence, the switching on of a weak perpendicular external field for a fixed value of the orientation angle of the surface moment does not change the total energy of the system, i.e., the sum of the exchange-interaction and anisotropy energies. This means that the first term $\Delta\varphi_0(\theta_1, h_\perp)$ in the thermodynamic potential (35), being expanded in a series with respect to the field $h_\perp$, does not contain a term linear in $h_\perp$.

The magnetic susceptibility of each atomic layer in a canted state in the surface region represents a Laurent series in the parameter $(k_{SC} - k_S) \equiv k_S$:

$$\chi_n = \frac{A_n}{\Delta k_S} + B_n + C_n \Delta k_S + D_n \Delta k_S^2 + \ldots. \qquad (37)$$

Below, we determine only the first two terms in this series: the irregular term $A_n/\Delta k_S$ and the constant B_n. In the previous section, we obtained an expression for the expansion of the thermodynamic potential $\varphi_0(\theta_1, 0)$ in the orientation angle θ_1 of the surface moment up to the fourth-order terms (25). Here, we must generalize this expansion to the case of a nonzero perpendicular external magnetic field $h_\perp$. To find an expression for the susceptibility in a canted state (a low-symmetry phase) with regard to the constant B_n in series (37), we should expand the second and third terms in (35) with respect to the orientation angle θ_1 of the surface moment up to the cubic terms in θ_1. This can be done by formula (17), in which the orientation angles $\Theta_n \equiv \theta_n$ for $n > 1$ are expressed in terms of the fictitious orientation angle Θ_1 of the surface in the imaginary domain wall, as well as by formula (20), in which the fictitious orientation angle Θ_1 of the surface moment is expressed in terms of the real orientation angle θ_1 of the surface moment in a canted state:

$$\sum_{n \geq 2} h_\perp \sin(\theta_n^{(0)}) \approx -\frac{\gamma k_{SC}}{k_b}\theta_1 h_\perp + D\frac{\theta_1^3}{6}h_\perp,$$
$$D = \frac{(\bar{\beta} - 1)\lambda^3}{(1 - \lambda^3)a^3} - \frac{\lambda}{(1 - \lambda)a^3}\left(\frac{c}{a} + \bar{\beta}\right). \qquad (38)$$

As a result, a generalized expression for the expansion of the thermodynamic potential in the orientation angle θ_1 of the surface moment can be represented as

$$\Delta\varphi(\theta_1, h_\perp) \approx \frac{\gamma(k_S - k_{SC})\theta_1^2}{2} + \frac{B\theta_1^4}{4!}$$
$$- \mu_1 h_\perp\left(1 - \frac{\mu\gamma k_{SC}}{\mu_1 k_b}\right)\theta_1 - \mu_1 h_\perp \frac{\theta_1^3}{6}\left(\frac{\mu}{\mu_1}D - 1\right). \qquad (39)$$

The perpendicular component of the moment of the surface atomic layer can be expressed as $\mu_{1\perp} = \mu_1 \sin\theta_1$. This formula implies that the susceptibility χ_1 of the surface atomic layer can be represented as

$$\chi_1 = \left.\frac{\partial \mu_{1\perp}}{\partial h_\perp}\right|_{h_\perp = 0} = \mu_1 \cos(\theta_1^{(0)})\eta_1,$$
$$\eta_1 = \left.\frac{\partial \theta_1^{(0)}}{\partial h_\perp}\right|_{h_\perp = 0}. \qquad (40)$$

The differentiation of thermodynamic potential (39) with respect to the orientation angle θ_1 of the surface moment followed by the differentiation with respect to the field $h_\perp$ leads to an equation in the parameter η_1 that gives an expression for the susceptibility χ_1 of the surface atomic layer. The solution of this equation and the substitution of the obtained expression for η_1 into the formula (40) for χ_1 leads to the following expression for the susceptibility of the surface atomic layer in the vicinity of the spin-reorientation transition to a canted state:

$$\chi_1 = \frac{\mu_1^2(1 - \mu\gamma k_{SC}/\mu_1 k_B)}{\gamma(k_S - k_{SC})}, \quad k_S > k_{SC}, \qquad (41.1)$$

$$\chi_1 = \frac{\mu_1^2(1 - \mu\gamma k_{SC}/\mu_1 k_B)}{2\gamma(k_{SC} - k_S)}$$
$$+ \frac{3}{2}\frac{\mu_1^2}{B}\left[\frac{\mu}{\mu_1}\left(D + \frac{\gamma k_{SC}}{k_B}\right) - 2\right], \qquad (41.2)$$
$$k_S < k_{SC}.$$

When deriving this expression, we used expression (29) for the equilibrium orientation angle of the moment of the surface atomic layer in zero external magnetic field $h_\perp = 0$.

Now, we pass to the determination of the susceptibilities χ_n of other layers, $n \geq 2$. According to the definition of the perpendicular component of the moment of the nth atomic layer, $\mu_{n\perp} = \mu \sin\Theta_n$, the expression for the susceptibility of the nth atomic layer in a perpendicular field $h_\perp$ can be represented as

$$\chi_n = \mu \cos\Theta_n|_{h_\perp = 0}\left.\frac{d\Theta_n}{dh_\perp}\right|_{h_\perp = 0}$$
$$\equiv \mu \cos\Theta_n|_{h_\perp = 0}\eta_n. \qquad (42)$$

According to the Landau approach, which underlies the present study, the orientation angle Θ_n ($n = 2, 3, 4, \ldots$) of the moment of an arbitrary atomic layer is a function of the order parameter, i.e., the orientation angle θ_1 of the surface moment, and the external perpendicular

field $h_\perp$: $\Theta_n = \Theta_n(\theta_1(h_\perp), h_\perp)$. According to this approach, the derivative of the orientation angle Θ_n with respect to the field for $h_\perp = 0$ and the equilibrium value of the order parameter $\theta_1^{(0)}$ can be expressed as

$$\eta_n = \left.\frac{d\Theta_n(\theta_1(h_\perp), h_\perp)}{dh_\perp}\right|_{\substack{h_\perp = 0 \\ \theta_1 = \theta_1^{(0)}}} = \left.\frac{\partial\Theta_n}{\partial\theta_1}\frac{\partial\theta_1}{\partial h_\perp}\right|_{\substack{h_\perp = 0 \\ \theta_1 = \theta_1^{(0)}}} + \left.\frac{\partial\Theta_n}{\partial h_\perp}\right|_{\substack{h_\perp = 0 \\ \theta_1 = \theta_1^{(0)}}} = \nu_n\eta_1 + u_n^{(0)}. \quad (43)$$

According to expansion (21), the parameter ν_n is defined by the formula

$$\nu_n = \left.\frac{\partial\Theta_n}{\partial\theta_1}\right|_{\substack{h_\perp = 0 \\ \theta_1 = \theta_1^{(0)}}} = \hat{\alpha}_n + \frac{\hat{\beta}_n(\theta_1^{(0)})^2}{2}. \quad (44)$$

The partial derivatives $u_n^{(0)}$ satisfy the system of recurrence equations obtained by differentiating Eqs. (32.2) and (32.3) ($n \geq 2$) with respect to the field $h_\perp$ at $h_\perp = 0$ and $\theta_1 = \theta_1^{(0)}$. The superscript in $u_n^{(0)}$ points to the fact that this partial derivative is calculated at the equilibrium value of the order parameter, the orientation angle of the moment of the surface atomic layer:

$$k_b\cos 2\Theta_2 u_2^{(0)} + \cos(\Theta_2 - \Theta_3)(u_2^{(0)} - u_3^{(0)}) + \gamma\cos(\Theta_2 - \theta_1^{(0)})u_2^{(0)} - \mu\cos\Theta_2 = 0, \quad n = 2, \quad (45.1)$$

$$k_b\cos 2\Theta_n u_n^{(0)} + \cos(\Theta_n - \Theta_{n+1})(u_n^{(0)} - u_{n+1}^{(0)}) + \cos(\Theta_n - \Theta_{n-1})(u_n^{(0)} - u_{n-1}^{(0)}) - \mu\cos\Theta_n = 0, \quad n > 2. \quad (45.2)$$

It is important that, when calculating the partial derivative of the orientation angle $\Theta_n = \Theta_n(\theta_1(h_\perp), h_\perp)$ with respect to the field $h_\perp$, one should assume that the second term in (43), the order parameter θ_1, is constant. This is why the system of equations (45) for the partial derivatives $u_n^{(0)}$ does not contain the parameter η_1 in the penultimate term in (45.1). For the same reason, the system of equations (45) does not contain the partial derivative of the first equation (32.1) with respect to the field $h_\perp$.

When determining the susceptibility χ_n in a uniform state with the in-plane orientation of the moments of all layers, $k_S > k_{SC}$, the orientation angle of each atomic layer, including the surface layer, must be set equal to zero, because it is these values of the orientation angles that correspond to the ground state under the condition $k_S > k_{SC}$. In this particular case, the system of equations (45) takes the form

$$(k_b + 1 + \gamma)u_2^{(0)} - u_3^{(0)} = \mu, \quad n = 2,$$
$$(k_b + 2)u_n^{(0)} - (u_{n+1}^{(0)} + u_{n-1}^{(0)}) = \mu, \quad n \geq 3. \quad (46)$$

A solution to this system of recurrence equations is given by

$$u_n^{(0)} = \frac{\mu}{k_b}(1 - \hat{\alpha}_n), \quad n \geq 2. \quad (47)$$

When determining the susceptibility χ_n in a canted state in the surface region ($k_S < k_{SC}$), the orientation angle Θ_n ($n \geq 2$) of the moment of each atomic layer in the system of equations for the parameter $u_n^{(0)}$ can be expressed in terms of the orientation angle of the moment of the surface atomic layer by (21). As a result, the solution to system of equations (45) can be expanded in a series in θ_1. Since Eq. (45) contains only cosines, this expansion contains only even powers of θ_1:

$$u_n^{(0)} \approx a_n + b_n(\theta_1^{(0)})^2 + \ldots.$$

According to (42) and (43), taking into account the quadratic term $b_n(\theta_1^{(0)})^2$ in this expansion implies that this term appears in the expression for the susceptibility. However, in view of (29), $(\theta_1^{(0)})^2 \propto \Delta k_S$. Thus, taking into account the quadratic term in the expansion

$$u_n^{(0)} \approx a_n + b_n(\theta_1^{(0)})^2 + \ldots$$

implies that the term proportional to Δk_S in expression (37) for the susceptibility χ_n should be taken into account. However, as pointed out above, in the present paper, we determine only the first two terms in the Laurent series expansion in Δk_S (37). Therefore, when determining the susceptibility in a canted state in the surface region ($k_S < k_{SC}$), system of equations (45) for the parameters $u_n^{(0)}$ is again transformed to system of equations (46) with a similar solution for $u_n^{(0)}$ defined by formula (47). In other words, in both cases, $k_S < k_{SC}$ and $k_S > k_{SC}$, system of equations (45) is reduced to system of equations (46).

Using (42)–(47) and (29), we obtain the following expressions for the susceptibility of the nth atomic

layer to the left and right of the transition point, respectively:

$$\chi_n = \frac{\mu^2}{k_B}(1-\hat{\alpha}_n) + \frac{\mu\mu_1(1-\mu\gamma k_{SC}/\mu_1 k_B)}{\gamma(k_S - k_{SC})}\hat{\alpha}_n, \quad k_S > k_{SC}, \tag{48.1}$$

$$\chi_n = \frac{\mu^2}{k_B}(1-\hat{\alpha}_n) + \frac{\mu\mu_1(1-\mu\gamma k_{SC}/\mu_1 k_B)}{2\gamma(k_{SC} - k_S)}\hat{\alpha}_n + \frac{3\mu\mu_1}{2B}\left[\left(\frac{\mu}{\mu_1}D - 1\right)\hat{\alpha}_n + \left(1 - \frac{\mu\gamma k_{SC}}{\mu_1 k_B}\right)[\hat{\beta}_n - \hat{\alpha}_n^3]\right], \quad k_S < k_{SC}. \tag{48.2}$$

The first term in (48.1) and (48.2) represents the intrinsic susceptibility of the homogeneous substrate ($n \geq 2$) perturbed by the presence of the surface atomic layer of different nature. The summation of the first term over the layers ($n = 2, 3, \ldots, N$) yields the following result for this part of the susceptibility of the substrate:

$$\chi_{sub} = \frac{\mu^2}{k_b}N + \frac{\mu^2}{k_B}\frac{\gamma k_{SC}}{k_b}. \tag{49}$$

It follows from formula (49) that, in the case of $\gamma = 0$, which means the absence of the surface atomic layer, the susceptibility of each layer is equal to its bulk value ($\hat{\alpha}_n = 0$). Conversely, the presence of the surface atomic layer ($\gamma \neq 0$) with the negative anisotropy constant ($k_S < 0$) reduces the susceptibilities of layers in the subsurface region, because a surface atomic layer with negative anisotropy constant favors the perpendicular orientation of the moments of atomic layers with respect to the plane of the film. This effect is analogous to the application of a local perpendicular magnetic field to the second atomic layer.

The second term in (48.1) and (48.2) is the so-called irregular part of the susceptibility, which diverges at the point of the spin-reorientation transition. The expression for this term shows that the irregular part of susceptibility in a canted state is half the irregular part of susceptibility in the uniform magnetic state with the in-plane orientation of the moments of all atomic layers. This result agrees with the results of the Landau theory, in which the susceptibility of a low-symmetry phase is half the susceptibility of a high-symmetry phase. In both (48.1) and (48.2), the irregular part of susceptibility decreases to zero as λ^n ($0 < \lambda < 1$) as the index of atomic layer increases; this result points to the fact that a continuous spin-reorientation phase transition occurs precisely in the surface region. The summation of the irregular parts of the susceptibility over all layers, including the surface layer, yields the following result:

$$\chi_{irreg}(k_S > k_{SC}) = \sum_{n=1}^{\infty}\chi_{n,\,irreg}(k_S > k_{SC}) = \frac{\mu_1^2(1-\mu\gamma k_{SC}/\mu_1 k_B)^2}{\gamma(k_S - k_{SC})}, \tag{50.1}$$

$$\chi_{irreg}(k_S < k_{SC}) = \sum_{n=1}^{\infty}\chi_{n,\,irreg}(k_S < k_{SC}) = \frac{\mu_1^2(1-\mu\gamma k_{SC}/\mu_1 k_B)^2}{2\gamma(k_{SC} - k_C)}. \tag{50.2}$$

In a real experiment with a 1.5Fe/Gd(0001) film, the measurement of the susceptibility in a perpendicular field actually reduces to the measurement of the spin polarization of secondary electrons that are knocked out from the surface region in a short period after switching off the external field. In the simplest approximation, we can assume that the contribution of individual atomic layers to the resulting signal exponentially decreases with increasing layer number. It is obvious that the summation of the irregular parts of susceptibility (48.1) and (48.2) with an exponential factor gives rise to identical factors on the right-hand sides of (50.1) and (50.2). In other words, this procedure does not remove the divergence at the transition point; neither does it lead to other significant changes in the structure of formulas (50.1) and (50.2). Therefore, the identification of a continuous spin-reorientation transition to a canted state as a second-order phase transition by the divergence of the susceptibility at the transition point [8] is justified.

Finally, the summation of the last terms in (48.2) over the indices of atomic layers of the substrate with regard to a similar term in (41.2) for $k_S < k_{SC}$ yields the following expression for this component of the regular part of susceptibility:

$$\chi_{reg} = \frac{3\mu_1^2}{B}\left(\frac{\mu}{\mu_1}D - 1\right)\left(1 - \frac{\mu\gamma k_{SC}}{\mu_1 k_b}\right), \quad k_S < k_{SC}. \tag{51}$$

This term, just like the subsequent terms in expansion (37) that are proportional to Δk_S, Δk_S^2, …, makes an additional contribution to the asymmetry in the behavior of the susceptibility in the vicinity of the spin-reorientation transition point, in addition to the well-known asymmetry of the irregular part of susceptibility, which is less by a factor of 2 in the low-symmetry case.

ACKNOWLEDGMENTS

This work was supported by NATO, project no. PST.CLG.979374.

We thank D.P. Pappas for help in formulating the problem analyzed in this study and discussion of results.

REFERENCES

1. M. N. Baibich, J. M. Broto, A. Fert, *et al.*, Phys. Rev. Lett. **61**, 2472 (1988).
2. K. Sherifi, C. Dufour, G. Marchal, *et al.*, J. Magn. Magn. Mater. **104–107**, 1833 (1992).
3. M. Nawate, H. Kiriake, K. Doi, *et al.*, J. Magn. Magn. Mater. **104–107**, 1861 (1992).
4. R. E. Camley and D. R. Tilley, Phys. Rev. B **37**, 3413 (1988).
5. K. Zvezdin and S. N. Utochkin, JETP Lett. **57**, 433 (1993).
6. K. Zvezdin and S. N. Utochkin, JETP Lett. **57**, 439 (1993).
7. S. N. Utochkin and A. K. Zvezdin, J. Magn. Magn. Mater. **140–144**, 787 (1995).
8. C. S. Arnold, D. P. Pappas, and A. P. Popov, Phys. Rev. Lett. **83**, 3305 (1999).
9. C. S. Arnold, D. P. Pappas, and A. P. Popov, J. Appl. Phys. **87**, 5478 (2000).
10. L. Néel, J. Phys. Rad. **15**, 225 (1954).
11. L. D. Landau, E. M. Lifshitz, and L. P. Pitaevskii, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 2nd ed. (Nauka, Moscow, 1982; Butterworth Heinemann, Oxford, 1995), pp. 138–140.
12. C. S. Arnold and D. P. Pappas, Phys. Rev. Lett. **85**, 5202 (2000).
13. P. Chaudhari, J. J. Cuomo, and R. J. Gambino, IBM J. Res. Dev. **17**, 66 (1977); Appl. Phys. Lett. **22**, 337 (1973).
14. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon, Oxford, 1980), Part 1.
15. A. P. Popov and D. P. Pappas, Phys. Rev. B **64**, 184401 (2001).
16. M. G. Pini, A. Rettori, D. P. Pappas, *et al.*, J. Magn. Magn. Mater. **272–276**, 1152 (2004).
17. A. P. Popov, A. V. Anisimov, and D. P. Pappas, Phys. Rev. B **67**, 094428 (2003).
18. P. Betti, M. G. Pini, and A. Rettori, Phys. Rev. B **67**, 012402 (2003).
19. M. I. Kaganov and A. N. Omel'yanchuk, Zh. Éksp. Teor. Fiz. **61**, 1679 (1971) [Sov. Phys. JETP **34**, 895 (1972)].
20. M. I. Kaganov, Zh. Éksp. Teor. Fiz. **62**, 1196 (1972) [Sov. Phys. JETP **35**, 631 (1972)].
21. M. I. Kaganov and N. S. Karpinskaya, Zh. Éksp. Teor. Fiz. **76**, 2143 (1979) [Sov. Phys. JETP **49**, 1083 (1979)].
22. D. L. Mills, Phys. Rev. B **39**, 12306 (1989).
23. R. C. O'Handley and J. P. Woods, Phys. Rev. B **42**, 6568 (1990).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 844–855.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 974–986.
Original Russian Text Copyright © 2005 by Ovchinnikov, Shneĭder.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Effective Hamiltonian for HTSC Cuprates Taking into Account Electron–Phonon Interaction in the Strong-Correlation Regime

S. G. Ovchinnikov and E. I. Shneĭder

Kirenskiĭ Institute of Physics, Siberian Division, Russian Academy of Sciences, Akademgorodok, Krasnoyarsk, 660036 Russia

e-mail: shneyder@iph.krasn.ru

Received April 18, 2005

Abstract—Electron–phonon interaction is sequentially derived from a realistic *p–d* multiband model for the cuprates under conditions of strong electron correlations. The electronic structure is described using the representation of the Hubbard *X* operators in a generalized tight-binding method. Dependences of the diagonal and off-diagonal (on lattice sites) matrix elements of electron–phonon interaction on the wavevectors are found for three phonon modes, namely, breathing, apical breathing, and bending modes. The interactions of the breathing and bending modes with electrons are shown to contribute to the formation of kinks in the (0; 0)–(π; π) and (0; 0)–(π; 0) directions, respectively. A low-energy *t–J** model with phonons is developed; apart from electron–phonon interaction, it also includes spin–phonon interaction. The elimination of phonons gives an effective electron–electron interaction that depends on the occupation number of a multielectron term and on the carrier concentration due to strong electron correlations. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Despite significant progress reached in studying high-temperature superconductivity (HTSC) in layered cuprates, the HTSC mechanism is still unclear. Among numerous mechanisms proposed in the initial stage of investigating HTSC, the following two mechanisms have been most often discussed recently: the traditional mechanism of electron–phonon interaction (EPI) and the spin-fluctuation mechanism [1]. The latter is known to be caused by strong electron correlations that result in a long-range antiferromagnetic (AFM) order in undoped dielectric cuprates and to a short-range AFM order in weakly doped cuprates. Interest in electron–phonon interaction, which is present in all substances and can be strong in layered cuprates due to specific features of their crystal structure, has currently quickened because of inflection points (kinks) detected in electron dispersion laws in ARPES (angle-resolved photoemission spectroscopy) measurements [2]. Note that kinks were found in many hole cuprates, but they are absent in the electron cuprates with a *T*' structure ($Nd_{2-x}Ce_xCuO_4$). The kink energy measured from the Fermi level ($\omega_k \approx 70$ meV) is virtually universal, and the effect is most pronounced in the form of a bend in the dispersion law in the diagonal direction $\Gamma \longrightarrow M$, (0; 0)–(π; π) of the Brillouin zone. A kink at an energy $\omega_k \approx 40$ meV was also detected in the vicinity of the *X*((π; 0), (0; π)) and Bi2212 points [3], and it increases sharply as the temperature decreases below T_c (see the review of ARPES data in [4]). The nature of the kink is obviously related to electron–boson interaction; however, the question of what bosons, namely, phonons or spin fluctuations, are responsible for these renormalizations of an electronic spectrum near the Fermi level is a matter of dispute [5]. The kink can result from interaction with optical phonons [6] or with spin fluctuations [7, 8].

Thus, to describe both superconducting pairing and the properties of the normal state in the cuprates, one has to take into account the interactions of electrons with phonons and spin fluctuations. To describe optimally or strongly doped compositions, one can start from ordinary band theory; however, to discuss the entire phase diagram of cuprates, beginning from undoped antiferromagnetic dielectrics, one has to describe electrons in the strong-correlation regime. Various modifications of the one- and multiband Hubbard models led to a low-energy effective *t–J* model that describes electron interaction with spin fluctuations in the Hubbard bands [9, 10]. However, electron–phonon interaction in the strong-correlation regime has been studied to a lesser extent (see recent review [11]). As a rule, researchers consider the *t–J* model with local interaction of electrons with a certain optical mode. At the same time, to discuss the symmetry of the superconducting state and the differences in kinks located in different regions of the Brillouin zone, it is necessary to know an explicit dependence of the matrix elements $g^{(\nu)}(\mathbf{k}, \mathbf{q})$ of electron–phonon interaction on the incoming momentum **k** transferred by **q** and the number ν of the phonon mode. The purpose of this work is to sequentially derive electron–phonon interaction from a

realistic *p–d* multiband model for cuprates in the strong-correlation regime [12] and to realistically describe the phonons that interact most strongly with electrons.

In the general case, we can distinguish diagonal and off-diagonal contributions to the electron–phonon interaction in the nodal representation. Strong electron correlations and the diagonal electron–phonon interaction were simultaneously taken into account in [13–17]. It was found that the following three phonon modes interact most strongly with electrons: the longitudinal breathing mode (oxygen-ion vibrations in the CuO_2 plane that deform the Cu–O bond), the apical breathing mode (vibrations of apical oxygen ions that deform the Cu–O bond along the *c* axis), and the bending mode (oxygen-ion vibrations in the CuO_2 layer normal to the Cu–O bond) (Fig. 1). Neutron-scattering experiments revealed the maximum softening of the breathing mode at the boundary of the Brillouin zone, at the $(\pi/a; 0; 0)$ point [18, 19]. The joint effect of EPI and spin-fluctuation interaction on the superconducting pairing without regard for strong electron correlations was analyzed in [20]. As follows from the results of all these works, the breathing mode interacts most strongly at a phonon quasi-momentum $\mathbf{q} \sim \mathbf{Q} = (\pi/a; \pi/a)$ and breaks pairing with the $d_{x^2-y^2}$ symmetry; the bending mode has maximum interaction at small $\mathbf{q}$; and the apical breathing mode has a matrix interaction element that is independent of the in-plane wavevector $\mathbf{q}$. Oxygen-ion vibrations normal to the CuO_2 plane strongly modulate the ionic component of the chemical bond in the cuprates by changing the Madelung potential; hence, they strongly interact with electrons [11].

Kinks in ARPES experiments at the nodal ($\mathbf{k} = (\pi/2a; \pi/2a)$) and antinodal ($\mathbf{k} = (\pi/a; 0)$) points have different boson frequencies and different temperature dependences; therefore, their analysis requires a detailed description of EPI, in particular, the description of the dependence of the matrix elements $g^{(\nu)}(\mathbf{k}, \mathbf{q})$ on not only the transferred momentum $\mathbf{q}$ but also on the incoming momentum $\mathbf{k}$ [21]. The latter dependence can only be caused by the off-diagonal part of EPI. When the authors of [22, 23] derived EPI, they took into account strong electron correlations within the framework of the three-band *p–d* model and diagonal and off-diagonal EPIs and only considered the breathing mode. As a result, they constructed an effective *t–J* model with EPI. In this work, we investigate the interaction of strongly correlated electrons with all three modes given above and find diagonal and off-diagonal contributions to EPI. By comparing the crystal structures and phonon spectra of LSCO (*T* structure) and NCCO (*T*' structure), we could reveal the contributions to EPI that disappear when passing from the *T* to the *T*' structure and could explain why the EPI in the *T*' structure is significantly weaker than that in the hole cuprates.

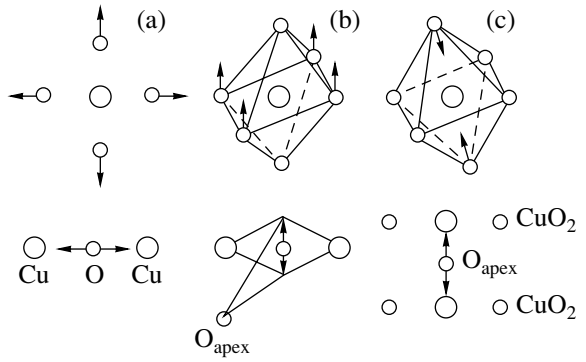


Fig. 1. Schematic diagram for atomic displacements for (a) the breathing mode, (b) the bending mode, and (c) the apical breathing mode.

2. DERIVATION OF ELECTRON–PHONON INTERACTION USING A MULTIBAND *p–d* MODEL FOR LAYERED CUPRATES

The Hamiltonian of the CuO_2 layer in the multiband *p–d* model can be written as follows (in the hole representation) [24]:

$$H_{pd} = \sum_{i\lambda\sigma}(\varepsilon_{i\lambda} - \mu)n_{i\lambda\sigma} + \sum_{ij\sigma}\sum_{\lambda\lambda'} t_{ij}^{\lambda\lambda'} a_{i\lambda\sigma}^{\dagger} a_{j\lambda'\sigma} + \sum_{ij\sigma\sigma'}\sum_{\lambda\lambda'} (V_{ij}^{\lambda\lambda'} n_{i\lambda\sigma} n_{j\lambda'\sigma'} - J_{ij}^{\lambda\lambda'} a_{i\lambda\sigma}^{\dagger} a_{i\lambda\sigma'} a_{j\lambda'\sigma'}^{\dagger} a_{j\lambda'\sigma}). \quad (1)$$

Here, $n_{i\lambda\sigma} = a_{i\lambda\sigma}^{\dagger} a_{j\lambda\sigma}$, $a_{i\lambda\sigma}^{\dagger}$ is the operator of production of a hole on the site $i \equiv R_i$ in the orbital state λ with the spin projection σ and the energy $\varepsilon_{i\lambda}$; μ is the chemical potential; $t_{ij}^{\lambda\lambda'}$ is the matrix element of an atomic jump; and $V_{ij}^{\lambda\lambda'}$ and $J_{ij}^{\lambda\lambda'}$ are the matrix elements of the Coulomb and exchange interactions, respectively. Unlike the three-band *p–d* model [25, 26], the multiband model takes into account both the $d_{x^2-y^2}$ and $d_{3z^2-r^2} \equiv d_{z^2}$ copper orbitals (although the other three t_{2g} orbitals can also be included, they are filled by electrons and their energy levels in the electron valence band are well below the low-energy range ($E \le 1$ eV) to be studied here). For oxygen ions lying in the CuO_2 layer, we take into account the p_x and p_y orbitals, and the p_z orbitals of the apical oxygen (which are present in the *T* structure and are absent in the *T*' structure) are also considered. Important microscopic model parameters are the following: t_{pd} is a hopping between the $d_{x^2-y^2}$ copper and the in-plane oxygen; t_{pp} is a hopping between neighbor-

ing (O_x and O_y) oxygen ions; and t'_{pd} and t'_{pp} are hoppings between copper and the apical oxygen and the in-plane and apical oxygen, respectively. From the Coulomb interactions, we distinguish the intra-atomic interactions of two holes in one (U_d is the Hubbard parameter) and different (V_d) orbitals, the corresponding interactions (U_p and V_p) for oxygen, the Coulomb interaction of neighboring copper and oxygen (V_{pd}), and the interaction of neighboring oxygen ions (V_{pp}). From the exchange interactions, we distinguish intra-atomic (Hund) exchange parameters J_d and J_p. A detailed analysis of Hamiltonian (1) and various matrix elements, as well as a procedure for the calculation of the band structure of quasiparticles with allowance for strong electron correlations using a generalized tight-binding (GTB) method, are given in [12, 27].

After the Wannier functions have been constructed in the framework of the GTB method, Hamiltonian (1) is written as the sum of intracell (H_c) and intercell (H_{cc}) parts [12, 27]

$$H = H_c + H_{cc},$$

$$H_c = \sum_f H_f, \quad H_f = H_f^{(b)} + H_f^{(a)} + H_f^{(ab)}, \qquad (2)$$

$$H_{cc} = \sum_{fg\sigma} (h_{fg}^{(b)} + h_{fg}^{(a)} + h_{fg}^{(ab)}).$$

Here, the f and g sites are only related to the copper sublattice (the cell is the CuO_6 or CuO_4 cluster), since the Wannier functions are centered at the Cu sites. The superscripts a and b indicate the symmetries of the Wannier functions: the $d_{x^2-y^2}$ copper states are hybridized inside the cell with the molecular b_{1g} orbital of the in-plane oxygen, and the d_{z^2} copper states are hybridized with the a_{1g} states of the in-plane oxygen and the p_z states of the apical oxygen. Apart from one-particle p–d and p–p hoppings inside the cell, the $H_f^{(b)}$ and $H_f^{(a)}$ Hamiltonians contain intracell Coulomb interactions. For example, U_p and V_{pd} are involved in all three H_f terms. However, $H_f^{(ab)}$ only contains Coulomb and exchange interactions, since the Wannier functions inside the cell are orthogonal. They are mixed due to hoppings between neighboring cells, and this mixing is contained in the $h_{fg}^{(ab)}$ term.

To take into account strong electron correlations within the framework of the GTB method, we first exactly diagonalize the H_f Hamiltonian and use its complete set of eigenstates $\{|p\rangle\}$ to construct the Hubbard X operators $X_f^{pq} \equiv |p\rangle\langle q|$. In the second stage, the intercell Hamiltonian part H_{cc} is written in the X representation, and it has the same operator structure ($\sim t_{fg} X_f X_g$) as the hopping Hamiltonian in the standard Hubbard model. This fact allows us to find the Green function and a band structure using perturbation theory. A dispersion equation for the calculation of a band structure in the GTB method has the form (in the paramagnetic phase)

$$\det\left\| \frac{E - \Omega_m}{F_m} \delta_{mm'} - \sum_{\lambda\lambda'} \gamma_{\lambda\sigma}(m) T_{\lambda\lambda'}(k) \gamma_{\lambda'\sigma}(m') \right\| = 0, \qquad (3)$$

where the coefficients $\gamma_{\lambda\sigma}(m)$ of the representation of the one-electron operators in terms of the Hubbard X operators,

$$a_{f\lambda\sigma} = \sum_m \gamma_{\lambda\sigma}(m) X_f^m, \quad m \longleftrightarrow (p, q), \qquad (4)$$

are calculated after the exact diagonalization of H_f. The Ω_m energies have a one-particle meaning, and they are defined as resonances between multielectron terms $|p\rangle$ and $|q\rangle$: $\Omega_m = E_p - E_q$. The filling factors $F_m = \langle X^{pp}\rangle + \langle X^{qq}\rangle$, just like the Ω_m energies, are calculated after the exact diagonalization of H_f. Finally, the intercell hopping matrix elements $T_{\lambda\lambda'}(k)$ are defined by different p–d and p–p hoppings. For example, electron production at the bottom of the conduction band of undoped La_2CuO_4 or Nd_2CuO_4 is related to a resonance between the vacuum ($|0\rangle$, the $d^{10}p^6$ configuration) molecular orbitals and the one-hole ($|\sigma\rangle$, $\sigma = \pm 1/2$; a mixture of the d^9p^6 and $d^{10}p^5$ configurations) molecular orbitals. Hole production near the valence band top is related to resonances Ω_S between the one-hole $|\sigma\rangle$ states and the two-hole $|s\rangle$ ^1A singlet that is mixed with the band of triplet excitations Ω_T ($|1, \sigma\rangle \longrightarrow |2, T\rangle$) with the participation of the two-hole ^3B triplet.

In the one-hole sector of the Hilbert space, the blocks of the H_f matrix with the b and a symmetries have the form

$$H_f^{(b)} = \begin{pmatrix} \varepsilon(d_{x^2-y^2}) & -\tau_b \\ -\tau_b & \varepsilon_b \end{pmatrix}, \qquad (5)$$

$$H_f^{(a)} = \begin{pmatrix} \varepsilon(d_{z^2}) & -\tau_a & -\tau'_{pd} \\ -\tau_a & \varepsilon_a & -t'_{pp} \\ -\tau'_{pd} & -t'_{pp} & \varepsilon(p_z) \end{pmatrix}, \qquad (6)$$

where the hopping parameters τ and the energies of the oxygen b and a orbitals are renormalized as compared to the initial atomic values due to the construction of the Wannier functions. The corresponding matrices in the two-hole sector have a large dimension; they are given in an explicit form in [12, 27] and are not discussed here.

After diagonalizing the intracell H_c part and after passing to the X-operator representation, we can write electron Hamiltonian (1) as

$$H_{\rm el} = \sum_{f n\gamma}(E_{n\gamma} - n\mu)X_f^{n\gamma, n\gamma} + \sum_{fg}\sum_{mm'} t_{fg}^{mm'} X_f^{\dagger m} X_g^{m'}. \quad (7)$$

At low energies near the bottom of the conduction band and the top of the valence band, we restrict ourselves to the following set of $|n\gamma\rangle$ terms: $n = 0$, $|0\rangle$; $n = 1$, $|\sigma\rangle$, $\sigma = \pm 1/2$; $n = 2$, singlet $|s\rangle$ and triplet $|T, M\rangle$, $M = 0, \pm 1$ [12]. Indices m, m' denote various hole excitations. For this set of $|n\gamma\rangle$ terms, the disappearance of a hole with a spin σ in Eq. (4) is described by the following quasiparticles: at $m = 0$, $(0, \sigma)$; at $m = 1$, $(\bar{\sigma}, s)$; at $m = 2$, $(\bar{\sigma}, T0)$; and at $m = 3$, $(\sigma, T2\sigma)$.

As usual, when EPI is derived in terms of the GTB method, it is necessary to take into account the modulation of the intra-atomic ($\varepsilon_{i\lambda}$) and interatomic ($t_{ij}^{\lambda\lambda'}$) parameters upon atomic displacements. Moreover, in our case, EPI is contributed by the modulation of the Coulomb interatomic interaction. It is important that the modulation of the one- and two-particle Hamiltonian parameters due to atomic displacements contributes to not only the one-particle but also the two-particle terms; in the general case, it also contributes to the multiparticle terms $E_{n\gamma}$ (where n is the number of electrons, and γ is the set of quantum numbers) that determine the resonance energies Ω in Eq. (3), whence a diagonal contribution to EPI appears. The modulation of various atomic-jump and Coulomb-interaction parameters also causes an off-diagonal contribution to EPI. As a result of atomic displacements, the energies of the $|n\gamma\rangle$ terms become site-dependent:

$$E_{n\gamma}(\mathbf{R}_i) = E_{n\gamma}(\mathbf{R}_{i0} + \mathbf{u}_i) = E_{n\gamma}(0) + \mathbf{g}_{n\gamma}\mathbf{u}_i. \quad (8)$$

Similarly, the hopping and interaction parameters depend on the difference in the sites $\mathbf{R}_i - \mathbf{R}_j = \mathbf{R}_{i0} - \mathbf{R}_{j0} + \mathbf{u}_{ij}$, $\mathbf{u}_{ij} = \mathbf{u}_i - \mathbf{u}_j$. In a linear approximation, we have

$$t_{fg}^{mm'} = t_{fg}^{mm'}(0) + \mathbf{V}^{mm'} \cdot \mathbf{u}_{fg}. \quad (9)$$

Here, a set of the phenomenological $\mathbf{g}_{n\gamma}$ and $\mathbf{V}^{mm'}$ parameters specifies the diagonal and off-diagonal contributions to EPI. As a result, we obtain electron Hamiltonian (7), in which all energies belong to the undisturbed lattice (i.e., $E_{n\gamma}(0)$ and $t_{fg}(0)$), and the EPI Hamiltonian

$$H_{\rm el-ph} = \sum_{f n\gamma}\mathbf{g}_{n\gamma} \cdot \mathbf{u}_f X_f^{n\gamma, n\gamma} + \sum_{fg}\sum_{mm'}\mathbf{V}^{mm'} \cdot \mathbf{u}_{fg} X_f^{\dagger m} X_g^{m'} \quad (10)$$

in the system of strongly correlated electrons.

Using the breathing mode as an example, we consider characteristic displacements and modulation of the corresponding Hamiltonian parameters. Figure 1 shows a fragment of the CuO_2 layer and the directions of oxygen-ion displacements for the three phonon modes under discussion. For the breathing mode, a displacement of the O^{2-} ion along the Cu–O bond (Fig. 1a) changes the Madelung potential and, thus, the d-level energy in the crystal field:

$$\varepsilon_i(d_{x^2-y^2}) = \varepsilon_{d0} + \mathbf{g} \cdot \mathbf{u}_i,$$

$$\mathbf{u}_i = u_x\left(R_i + \frac{a_x}{2}\right) - u_x\left(R_i - \frac{a_x}{2}\right) + u_y\left(R_i + \frac{a_y}{2}\right) - u_y\left(R_i - \frac{a_y}{2}\right). \quad (11)$$

In the approximation that is linear in displacements, modulation of the ε_p level on oxygen is absent because of the symmetry, since the contributions from the left and right copper ions are canceled (Fig. 1a). As a result, we have modulation of the charge transfer energy: $\Delta_i = \varepsilon_p - \varepsilon_i(d_{x^2-y^2}) = \Delta_0 - \mathbf{g} \cdot \mathbf{u}_i$. Analogous linear-in-displacement contributions appear in the parameters of hoppings between copper and the in-plane oxygen ($t_{pd}(i) = t_{pd}(0) + \delta t_{pd}$), between oxygen and the oxygen inside the CuO_2 layer ($t_{pp}(i) = t_{pp}(0) + \delta t_{pp}$), and between the apical oxygen and the in-plane oxygen ($t'_{pp}(i) = t'_{pp}(0) + \delta t'_{pp}$); they also appear in the parameters of the Coulomb interaction of copper with oxygen ($V_{pd}(i) = V_{pd}(0) + \delta V_{pd}$) and of oxygen with oxygen ($V_{pp}(i) = V_{pp}(0) + \delta V_{pp}$).

In [12], the parameters of Hamiltonian (1) were considered as phenomenological and were found from a comparison with ARPES experimental data for undoped $Sr_2CuO_2Cl_2$ oxychlorides. These parameters have recently been calculated using the LDA and LSD + U band-theory methods [28]. All hopping integrals were found to be of the same order of magnitude: $t_{pp} \approx t'_{pp} \approx 0.4$–$0.5t_{pd}$. The displacement dependence of the parameters has not been calculated; therefore, in this work we cannot describe EPI without using fitting parameters. The modulation corrections to the hopping integrals are assumed to be of the same order of magnitude: $\delta t_{pd} \sim \delta t_{pp} \sim \delta t'_{pp} \sim \delta t'_{pd}$.

Apart from the modulation of the Coulomb interactions and the crystal field, all these linear-in-displacement modulations renormalize the energies of the one-hole b_{1g} doublet $|\sigma\rangle$ and the two-hole ^1A singlet and ^3B triplet, which results in the modulation of the Ω_S and Ω_T energies (diagonal contribution to EPI). The off-diagonal contribution results from the modulation of $T_{\lambda\lambda'}$ in dispersion equation (3). Since the distance depen-

dences of all matrix elements of Hamiltonian (1) are unknown, we introduce two parameters of the diagonal and off-diagonal EPIs for each electron band and each phonon mode (ν). Using the completeness condition

$$X_f^{0,0} + \sum_\sigma X_f^{\sigma,\sigma} + X_f^{S,S} + \sum_{M=-1}^{1} X_f^{TM,TM} = 1 \qquad (12)$$

for the multielectron basis of the cell, we can eliminate one parameter of the diagonal EPI and write

$$H_{\text{dia}}^{(\nu)} = \sum_f \mathbf{u}_{f,\nu} \times \left[\sum_\sigma \mathbf{g}_\sigma^{(\nu)} X_f^{\sigma\sigma} + \mathbf{g}_S^{(\nu)} X_f^{SS} + \sum_M \mathbf{g}_T^{(\nu)} X_f^{TM,TM} \right]. \qquad (13)$$

As usual, the displacement vector is represented in the form

$$\mathbf{u}_{f,\nu} = \frac{1}{\sqrt{N}} \sum_{\mathbf{q},\alpha} \frac{\mathbf{e}_{\alpha,\nu}}{\sqrt{2M_\alpha \omega_{\mathbf{q},\nu}}} \varphi_{\mathbf{q},\nu} \exp(i\mathbf{q}(\mathbf{R}_f + \mathbf{R}_\alpha)), \qquad (14)$$

where M_α is the ion α mass, $\mathbf{R}_\alpha$ is the ion radius vector in the $\mathbf{R}_f$ cell, $\mathbf{e}_{\alpha,\nu}$ is the polarization vector, and $\varphi_{\mathbf{q},\nu} = b_{\mathbf{q},\nu} + b^\dagger_{-\mathbf{q},\nu}$ and $b_{\mathbf{q},\nu}(b^\dagger_{-\mathbf{q},\nu})$ are operators of annihilation (production) of phonon ν with the $\mathbf{q}$ wavevector. Neglecting the copper-ion displacements (which are small as compared to the oxygen-ion displacements), for the breathing optical mode (ν = 1) we find

$$\mathbf{u}_{f,1} = \frac{2i}{\sqrt{2M_O N}} \sum_{\mathbf{q}} \frac{\varphi_{\mathbf{q},1}}{\sqrt{\omega_{\mathbf{q},1}}} e^{i\mathbf{q}\cdot\mathbf{R}_f} \times \left[e_x(O_x) \sin\frac{q_x a}{2} + e_y(O_y) \sin\frac{q_y a}{2} \right]. \qquad (15)$$

As a result, the diagonal part of EPI for the ν mode can finally be written as

$$H_{\text{dia}}^{(\nu)} = \frac{1}{\sqrt{N}} \sum_{\mathbf{kq}\nu} \sum_m g_{\text{dia},m}^{(\nu)}(\mathbf{q}) X^{\dagger m}_{\mathbf{k}+\mathbf{q}} X^m_{\mathbf{k}} \varphi_{\mathbf{q},\nu}, \qquad (16)$$

where for the breathing mode we have

$$g_{\text{dia},m}^{(1)}(\mathbf{q}) = \frac{2ig_m^{(1)}}{\sqrt{2M_O\omega_{\mathbf{q},1}}} \times \left[e_x(O_x) \sin\frac{q_x a}{2} + e_y(O_y) \sin\frac{q_y a}{2} \right]. \qquad (17)$$

The off-diagonal part of EPI for this mode is

$$H_{\text{off}}^{(1)} = \sum_{fg} \sum_{mm'} \mathbf{V}_{mm'}^{(1)} \cdot (\mathbf{u}_{f,1} + \mathbf{u}_{g,1}) X_f^{\dagger m} X_g^{m'} \qquad (18)$$

and can be represented as

$$H_{\text{off}}^{(\nu)} = \frac{1}{\sqrt{N}} \sum_{\mathbf{kq}\nu} \sum_{mm'} g_{\text{off},mm'}^{(\nu)}(\mathbf{k},\mathbf{q}) X^{\dagger m}_{\mathbf{k}+\mathbf{q}} X^{m'}_{\mathbf{k}} \varphi_{\mathbf{q},\nu}, \qquad (19)$$

where

$$g_{\text{off},mm'}^{(1)}(\mathbf{k},\mathbf{q}) = \frac{8i\mathbf{V}_{mm'}^{(1)}}{\sqrt{2M_O\omega_{\mathbf{q},1}}} \times \left[e_x(O_x) \sin\frac{q_x a}{2} + e_y(O_y) \sin\frac{q_y a}{2} \right] \times [\gamma(\mathbf{k}) + \gamma(\mathbf{k}+\mathbf{q})] \qquad (20)$$

and $\gamma(\mathbf{q}) = (\cos q_x a + \cos q_y a)/2$.

For the apical breathing mode (ν = 2), the displacements of the apical oxygen ions along the *z* axis modulate the crystal field on copper and, thus, $\delta\varepsilon_d$, the Cu–O_{ap} bond length, the $\delta t'_{pd}$ hoppings, and the $\delta t'_{pp}$ hoppings between the apical and in-plane oxygen (Fig. 1b). All these effects contribute only to the diagonal EPI, since they change the parameters only inside the $\mathbf{R}_f$ cell. Of course coupling between neighboring CuO_2 layers also appear; however, we restrict ourselves to only one-layer cuprates in this work. Strong EPI for this mode and its doping-induced softening were predicted in [29].

For a two-dimensional vector $\mathbf{q} = (q_x, q_y)$, we have $\mathbf{q}\cdot\mathbf{R}_a = 0$; therefore, we can write

$$\mathbf{u}_{f,2} = \frac{1}{\sqrt{N}} \sum_{\mathbf{q}} \frac{e_z(O_{\text{ap}})}{\sqrt{2M_O\omega_{\mathbf{q},2}}} \varphi_{\mathbf{q},2} e^{i\mathbf{q}\cdot\mathbf{R}_f}, \qquad (21)$$

so that

$$g_{\text{dia},m}^{(2)}(\mathbf{q}) = \frac{\mathbf{g}_m^{(2)}}{\sqrt{2M_O\omega_{\mathbf{q},2}}} e_z(O_{\text{ap}}) \qquad (22)$$

depends only weakly on $\mathbf{q}$ through the $\omega_{\mathbf{q},2}$ dispersion. The off-diagonal part of EPI for the apical breathing mode is absent:

$$g_{\text{off}}^{(2)}(\mathbf{k},\mathbf{q}) = 0. \qquad (23)$$

For the bending mode (ν = 3), displacements in the tetragonal phase are transverse to the Cu–O bond (Fig. 1c), and the microscopic nature of EPI for this mode is not so obvious. Indeed, because of the symmetry, the Cu–O bond lengths, the crystal field, and the t_{pd} hopping cannot be modulated in the linear-in-displacement approximation; their modulations are proportional to the displacement squared [13]. Linear contributions appear only in the corrugated CuO_2 layer owing to orthorhombic distortions, and they are small due to

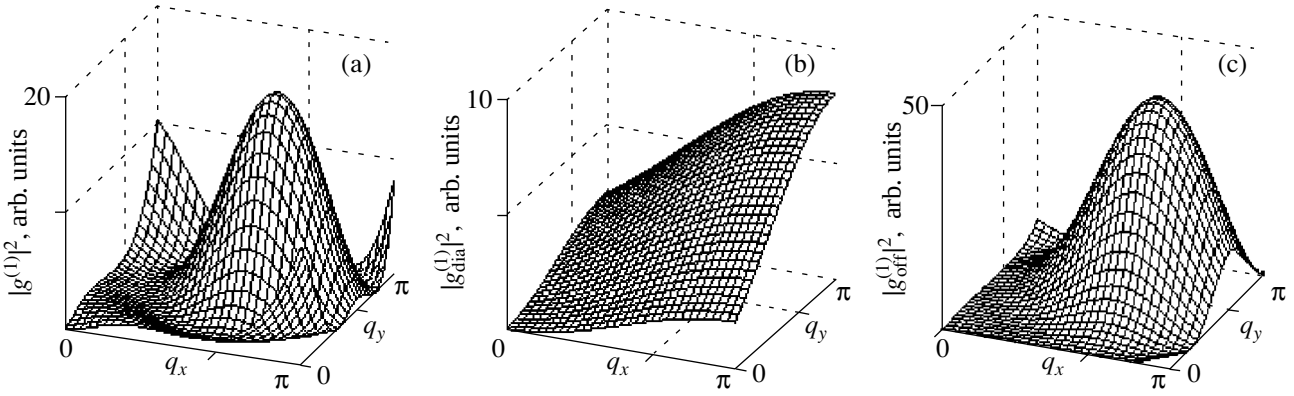


Fig. 2. EPI matrix element for the breathing mode at the $\mathbf{k}_n$ nodal point: (a) the total element, (b) the diagonal part, and (c) the off-diagonal part.

the small angle of corrugation. It should be noted that, in [13] and most related works, EPI is derived from an analysis of displacements using a simplified Hubbard model or the three-band *p–d* model with absent apical-oxygen states. In our *p–d* multiband model, the presence of the apical oxygen leads to the modulation of the distance between the in-plane and apical oxygen (Fig. 1c). As a result, the $\delta t'_{pp}$ and $\delta V'_{pp}$ parameters are modulated in the linear-in-displacement approximation. Moreover, for this mode, the modulation of the Madelung potential (the ionic component of the chemical bond) contributes significantly to EPI [11], since oxygen-ion vibrations transversely to the CuO_2 plane are weakly shielded.

Finally, the diagonal EPI with the bending mode can be written in the form of Eq. (16) with the matrix element

$$g^{(3)}_{\text{dia},\,m}(\mathbf{q}) = \frac{2\mathbf{g}^{(3)}_m}{\sqrt{2M_O\omega_{\mathbf{q},\,3}}} \times\left[e_z(O_x)\cos\frac{q_x a}{2} + \mathrm{e}_z(O_y)\cos\frac{q_y a}{2}\right]. \tag{24}$$

The matrix element of the off-diagonal EPI with the same mode is

$$g^{(3)}_{\text{off},\,mm'}(\mathbf{k},\mathbf{q}) = \frac{2\mathbf{V}^{(3)}_{mm'}}{\sqrt{2M_O\omega_{\mathbf{q},\,3}}} \times\left[e_z(O_x)\cos\left(k_x+\frac{q_x}{2}\right)a + \mathrm{e}_z(O_y)\cos\left(k_y+\frac{q_x}{2}\right)a\right]. \tag{25}$$

By summarizing the results of this section, we write the EPI Hamiltonian as

$$H_{\text{el–ph}} = \sum_{\mathbf{k}\mathbf{q}\nu}\sum_{mm'} g^{(\nu)}_{mm'}(\mathbf{k},\mathbf{q})X^{\dagger m}_{\mathbf{k}+\mathbf{q}}X^{m'}_{\mathbf{k}}(b_{\mathbf{q},\,\nu}+b^{\dagger}_{-\mathbf{q},\,\nu}), \tag{26}$$

$$g^{(\nu)}_{mm'}(\mathbf{k},\mathbf{q}) = \delta_{mm'}g^{(\nu)}_{\text{dia},\,m}(\mathbf{q}) + g^{(\nu)}_{\text{off},\,mm'}(\mathbf{k},\mathbf{q}).$$

The g_{dia} and g_{off} matrix elements for the three phonon modes under study are given by Eqs. (17), (20), (22)–(25).

It should be noted that the introduction of two phenomenological parameters ($\mathbf{g}^{(\nu)}_m$ and $\mathbf{V}^{(\nu)}_{mm'}$) for the diagonal and off-diagonal EPIs for each mode is related to not only a large amount of different microscopic contributions but also to incomplete knowledge of the distance dependences of various parameters. For example, even for the simplest particular case of EPI in the *t–J* model with the breathing mode, two works ([22, 23]) solving similar problems give different results: a large diagonal contribution (~0.25 eV) and a two orders of magnitude smaller off-diagonal contribution in [22] in contrast to virtually the same (~0.03 eV) diagonal and off-diagonal contributions to EPI in [23]. When several EPI mechanisms are taken into account, different contributions begin to interfere; for instance, the contribution of δV_{pd} decreases the contribution of δt_{pd} by approximately 30% [22]. When passing to a realistic model with a large number of contributions to EPI, the estimation errors of matrix elements accumulate; therefore, we think that the decision to restrict ourselves to phenomenological parameters was reasonable.

3. ANALYSIS OF THE SYMMETRY OF ELECTRON–PHONON INTERACTION

It is convenient to consider the dependences of the matrix elements on **k** and **q**, which were obtained by analyzing the atomic displacements in each mode, using maps in which $|g^{(\nu)}(\mathbf{k},\mathbf{q})|^2$ is presented as a function of the phonon momentum **q** at fixed values of the initial electron momentum **k**. The values of **k** were chosen according to ARPES data in which renormalization of the effective electron mass, which indicates interaction between electrons and collective excitations, was detected in the nodal direction for $\mathbf{k}_n = ((1-\delta)\pi/2; (1-\delta)\pi/2)$ and in the nodal direction for $\mathbf{k}_{\text{an}} = (\pi(1-\delta); \delta)$ (where $\delta \sim 0.1$). Figures 2–5 show maps for the diago-

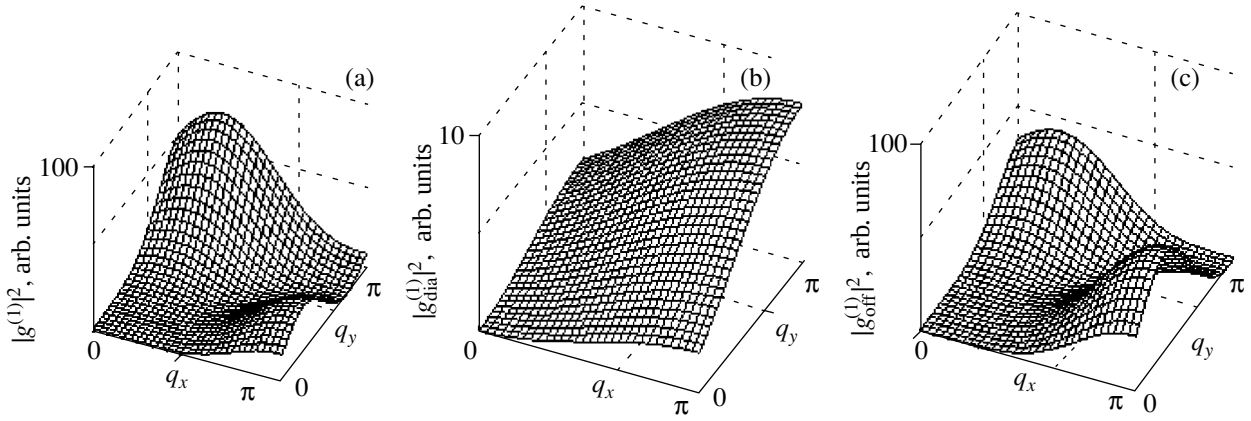


Fig. 3. EPI matrix element for the breathing mode at the $\mathbf{k}_{\rm an}$ antinodal point: (a) the total element, (b) the diagonal part, and (c) the off-diagonal part.

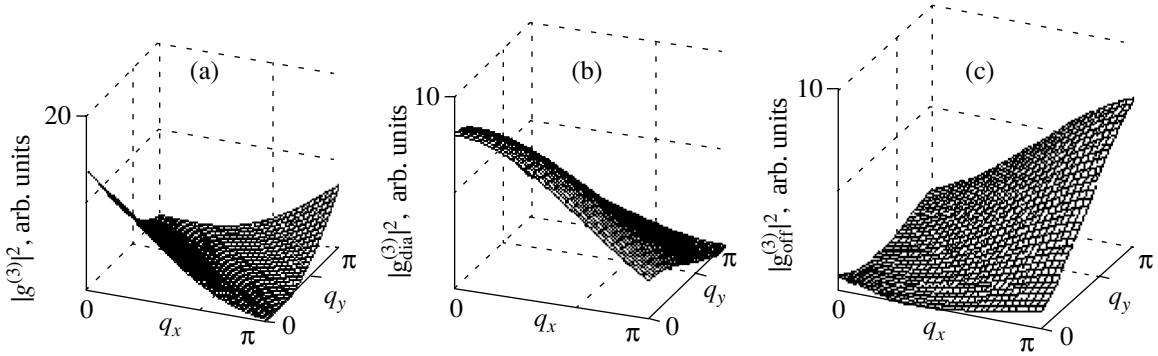


Fig. 4. EPI matrix element for the bending mode at the $\mathbf{k}_n$ nodal point: (a) the total element, (b) the diagonal part, and (c) the off-diagonal part.

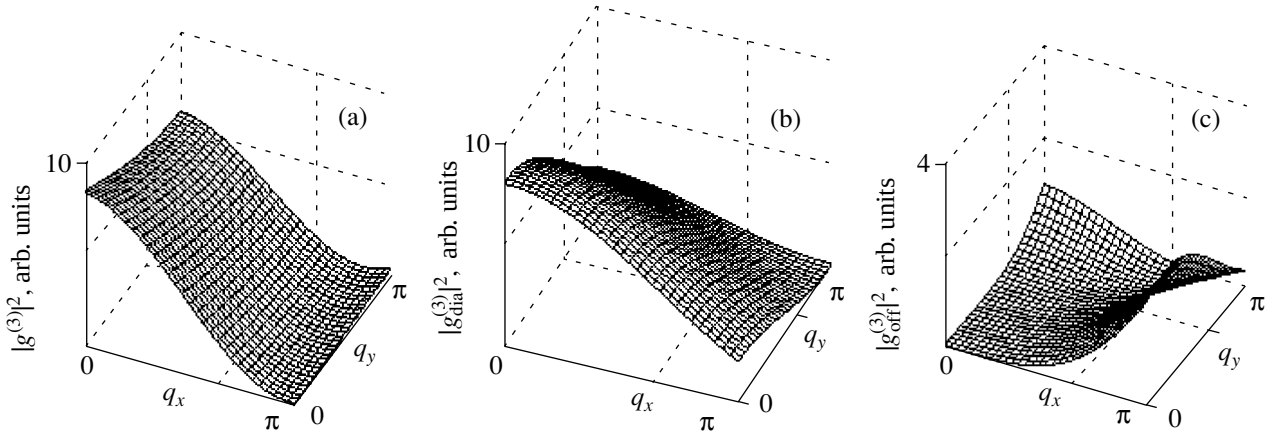


Fig. 5. EPI matrix element for the bending mode at the $\mathbf{k}_{\rm an}$ antinodal point: (a) the total element, (b) the diagonal part, and (c) the off-diagonal part.

nal and off-diagonal matrix elements for EPI with the breathing and bending modes at the nodal and antinodal points. All the maps were plotted for $\mathbf{g}^{(\nu)}_m = \mathbf{V}^{(\nu)}_{mm'} = 1$. The total intensity $|g^{(\nu)}_{{\rm dia},\,m}(\mathbf{q}) + g^{(\nu)}_{{\rm off},\,mm}(\mathbf{k}, \mathbf{q})|^2$ for quasiparticles that are diagonal in the band index m, $g^{(\nu)}_{mm}(\mathbf{k}, \mathbf{q})$ is characterized by interference of the $g_{\rm dia}$ and $g_{\rm off}$ matrix elements. An example of interference is shown in Fig. 2a for the breathing mode, where the peak height at $\mathbf{q} = (3\pi/4, 3\pi/4)$ in the total matrix ele-

ment is smaller than that in the partial $|g_{off}(\mathbf{k}, \mathbf{q})|^2$ contribution (Fig. 2c) for the same $\mathbf{q}$ point.

For the nodal $\mathbf{k}_n$ point, the $|g^{(1)}(\mathbf{k}_n, \mathbf{q})|^2$ maxima of the breathing mode (Fig. 2a) near the $\mathbf{q}$ points equal to $(3\pi/4; 3\pi/4)$, $(\pi; 0)$, and $(0; \pi)$ correspond to an off-diagonal contribution (Fig. 2c), and the diagonal part (Fig. 2b) causes a weak maximum at the $(\pi; \pi)$ point. (It is easy to see that the off-diagonal contribution for $\mathbf{q} = (\pi; \pi)$ becomes zero, just as in [23].) The maximum effective electron–phonon interaction is determined by vibrations with the wavevectors $\mathbf{q}$ located at the edges of the Brillouin zone. It is these phonons that transfer an electron from $\mathbf{k} \sim \mathbf{k}_F$ into states with the final momentum $\mathbf{k}' = \mathbf{k} + \mathbf{q}$ lying at the Fermi surface. The intensity of interaction with electrons for the half-breathing mode $\mathbf{q} = (\pi; 0)$ is higher than that for the full breathing mode $\mathbf{q} = (\pi; \pi)$. This finding agrees well with experiments. As was shown in inelastic neutron scattering experiments, the spectrum renormalization with doping for the $(\pi; 0)$ mode is about 20%, whereas it is only 5% for vibrations with the $(\pi; \pi)$ wavevector [30, 31]. Note that the softening and the line broadening and asymmetry of the half-breathing mode were detected in a number of HTSCs (e.g., in LSCO [32], YBCO [33], BKBO [34]). Moreover, the frequency of this vibration (70–85 meV) falls in the range of the kink energy in the nodal direction. The energy of the full breathing mode is 85–90 meV, which is higher than this value.

The interaction of the breathing mode with electrons having an initial momentum $\mathbf{k}_{an}$ is effectively small. The $|g^{(1)}(\mathbf{k}_{an}, \mathbf{q})|^2$ maximum at the $(0; \pi)$ point (Fig. 3a) corresponds to the scattering of electrons having an initial momentum $\mathbf{k}_{an}$ near the Fermi surface into the final state $\mathbf{k}' = \mathbf{k} + \mathbf{q} \approx (\pi; \pi)$, which is far from the Fermi surface. (Similar considerations are valid for the maximum at the $(\pi; 0)$ point). Note that the diagonal contribution at the maxima is small (Fig. 3b), and the off-diagonal contribution is the main contribution (Fig. 3c).

For the bending mode, the effective interaction is maximal at small values of the phonon momentum in both the nodal (Fig. 4) and antinodal (Fig. 5) directions. In both cases, the result depends on the diagonal contribution to the total matrix element.

For the apical breathing mode, electron–phonon interaction is independent of the $\mathbf{k}$ and $\mathbf{q}$ vectors.

Thus, an analysis of the atomic displacements of the vibrations under study shows the following. The interaction of electrons at the nodal point is maximal for the half-breathing mode with $\mathbf{q} = (\pi; 0)$ and for the bending mode with small values of the wavevector $\mathbf{q}$. The bending mode also strongly interact with electrons at the antinodal point at small values of $\mathbf{q}$. Moreover, the matrix element squared ($|g^{(1)}(\mathbf{k}, \mathbf{q})|^2$) for the half-breathing mode is higher than that for the full breathing mode $\mathbf{q} = (\pi; \pi)$ at any values of the initial electron momentum $\mathbf{k}$.

4. ELECTRON–PHONON INTERACTION IN AN EFFECTIVE LOW-ENERGY MODEL

Intersite hoppings in Hamiltonian (7) contain intraband hoppings in the lower Hubbard hole band ($m = 0$, the conduction band bottom in the electron representation) and in the upper Hubbard hole band ($m = 1$, the valence band top). The corresponding Hubbard operators are $X_f^0 \equiv X_f^{0\sigma}$ and $X_f^1 \equiv X_f^{-\sigma S}$.

Moreover, there are interband hoppings with excitation through a gap with charge transfer Δ that are described by the terms

$$t_{fg}^{01} X_f^{0\dagger} X_g^1 = t_{fg}^{01} X_f^{\sigma 0} X_f^{-\sigma S}. \tag{27}$$

The elimination of the interband hoppings result in the effective low-energy one-band t–J_* model [35]; for the lower Hubbard band (electron doping), it can be written in the form

$$H_{t-J_*} = H_{t-J} + H_{(3)},$$

$$\begin{aligned} H_{t-J} &= \sum_{f\sigma} \varepsilon_1 X_f^{\sigma\sigma} + \sum_{fg\sigma} t_{fg}^{00} X_f^{\sigma 0} X_g^{0\sigma} \\ &+ \sum_{fg} J_{fg}\left(\mathbf{S}_f \cdot \mathbf{S}_g - \frac{1}{4} n_f n_g\right), \end{aligned} \tag{28}$$

$$H_{(3)} = \sum_{fgm\sigma} \frac{t_{fm}^{01} t_{mg}^{01}}{\Delta} (X_f^{\sigma 0} X_m^{\bar\sigma\bar\sigma} X_g^{0\sigma} - X_f^{\sigma 0} X_m^{\bar\sigma\sigma} X_g^{0\bar\sigma}).$$

Here, $J_{fg} = 2(t_{fg}^{01})^2/\Delta$ is the exchange integral; $\mathbf{S}_f$ and n_f are the spin operator and the operator of the number of particles at the site; and $\bar\sigma = -\sigma$.

Off-diagonal EPI processes in Hamiltonian (10) contain intraband processes of the form

$$\mathbf{V}^{01} \mathbf{u}_{fg} X_f^{\sigma 0} X_g^{\bar\sigma S}. \tag{29}$$

Their elimination in the second order in $\mathbf{V}^{01}$ corresponds to the corrections $\delta J_{fg} \sim (\mathbf{V}^{01})^2 \mathbf{u}_{fg}^2/\Delta$ to the exchange integral, and we neglect them, since we restrict ourselves to linear-in-displacement contributions. At the same time, a combination of two perturbations (27) and (29) gives a linear-in-displacement correction to the exchange integral (spin–phonon interaction):

$$\delta J_{fg} = \frac{2 t_{fg}^{01} \mathbf{V}^{01} \cdot \mathbf{u}_{fg}}{\Delta} = \frac{\mathbf{V}^{01} \cdot \mathbf{u}_{fg}}{t_{fg}^{01}} J_{fg}. \tag{30}$$

Since the displacements are small and since $\mathbf{V}^{01}\mathbf{u} \ll t^{01}$ in series (9), we have $\delta J \ll J$. The spin–phonon interac-

tion Hamiltonian is

$$H_{\text{s-ph}} = \sum_{fg\mathbf{q}\nu} A_{fg}(\mathbf{q},\nu)\left(\mathbf{S}_f\cdot\mathbf{S}_g - \frac{1}{4}n_f n_g\right) \times (b_{\mathbf{q},\nu} + b^{\dagger}_{-\mathbf{q},\nu}),$$
$$A_{fg}(\mathbf{q},\nu) = \frac{2it^{01}_{fg}/\Delta}{\sqrt{2M_0\omega_{\mathbf{q},\nu}}} e^{i\mathbf{q}\cdot(\mathbf{R}_f+\mathbf{R}_g)} \times (\mathbf{V}^{(\nu)}_{01,x}\delta_{f,g\pm x} + \mathbf{V}^{(\nu)}_{01,y}\delta_{f,g\pm y}). \quad (31)$$

Analogous linear-in-displacement corrections also appear in the three-center terms:

$$H^{(3)}_{\text{el-ph}} = -\sum_{fgm\sigma} \frac{\mathbf{V}^{01}(t_{fm}\mathbf{u}_{mg} + \mathbf{u}_{fm}t_{mg})}{\Delta} \times (X^{\sigma 0}_f X^{\bar\sigma\bar\sigma}_m X^{0\sigma}_g - X^{\sigma 0}_f X^{\bar\sigma\sigma}_m X^{0\bar\sigma}_g). \quad (32)$$

In (30) and (32), we ignore the corrections that are linear in displacements but small in the $\mathbf{V}^{01}\cdot\mathbf{u}/\Delta$ parameter.

Thus, by making allowance for electron–phonon interaction, we can write the effective low-energy *t–J** model as follows:

$$H_{\text{eff}} = H_{t-J^*} + H^0_{\text{ph}} + H_{\text{el-ph}} + H_{\text{s-ph}} + H^{(3)}_{\text{el-ph}},$$
$$H^0_{\text{ph}} = \sum_{\mathbf{q}\nu}\omega_{\mathbf{q},\nu}b^{\dagger}_{\mathbf{q},\nu}b_{\mathbf{q},\nu}, \quad (33)$$
$$H_{\text{el-ph}} = \sum_{k\mathbf{q}\nu\sigma} g^{(\nu)}_{00}(\mathbf{q},\nu)X^{\sigma 0}_{\mathbf{k}+\mathbf{q}}X^{0\sigma}_{\mathbf{k}}(b_{\mathbf{q},\nu} + b^{\dagger}_{-\mathbf{q},\nu}).$$

The band structure of the *p*-type cuprates is formed with the participation of two-particle $^1A_1$ singlet and $^3B_1$ triplet and is more complex [12]; however, the contribution of the triplet band Ω_T to the dispersion and the density of states manifests itself mainly below (by 0.5 eV) the valence band top, near which the Fermi level is pinned upon doping up to an optimum concentration ($x \le x_{\text{opt}}$). Therefore, to discuss the kinks and superconducting mechanisms, we may neglect the triplet band; then, we obtain an effective hole Hamiltonian that is identical to Hamiltonian (33) in which the operator $X^{0\sigma}$ of the lower Hubbard hole band is replaced by the operator $X^{\bar\sigma 2}$ of the upper band ($m = 0 \longrightarrow m = 1$).

We now consider the simplest EPI contribution to the electron mass operator,

$$\Sigma(\mathbf{k},\varepsilon) = \frac{1}{N}\sum_{\mathbf{q}\nu}\int d\omega\left|g^{(\nu)}_{00}(\mathbf{q},\mathbf{k})\right|^2 \times G(\mathbf{k}-\mathbf{q},\varepsilon-\omega)D(\mathbf{q},\omega). \quad (34)$$

Detailed computation of spectrum renormalizations is beyond the scope of this work, and we only present qualitative notes. First, electrons in the strong-correlation regime in the *t–J* model are described as quasiparticles in the Hubbard subband, and their spectral weight is specified by the filling factor F_m in the numerator of the Green function (see Eq. (3)). For electron doping in $Nd_{2-x}Ce_xCuO_4$, we have

$$F_0 = \langle X^{00}\rangle + \langle X^{\sigma\sigma}\rangle = (1+x)/2,$$

and for hole doping in $La_{2-x}Sr_xCuO_4$ we have

$$F_1 = \langle X^{\sigma\sigma}\rangle + \langle X^{S,S}\rangle = (1+x)/2.$$

This spectral weight of quasiparticles appears in the Hartree–Fock Green function *G* in Eq. (34); as a result, the dimensionless EPI parameter for free electrons, $\lambda_0 = (g^2/\omega_D)N(0)$, decreases by the filling factor: $\lambda = \lambda_0(1+x)/2$.

Second, based on EPI intensity maps and on the energy and momentum conservation laws, we can qualitatively analyze the modes that contribute to the kinks. In this analysis, we assume that the electron energy in the superconducting phase is described by the Bardeen–Cooper–Schrieffer formula

$$E(k) = \pm\sqrt{\varepsilon_k^2 + \Delta_k^2},$$

the Δ_k gap has the $d_{x^2-y^2}$ symmetry, and $\Delta(\mathbf{k}) = \Delta_0(\cos k_x a - \cos k_y a)/2$.

We now consider $\Sigma(\mathbf{k},\varepsilon)$ at the nodal point $\mathbf{k}_n = ((1-\delta)\pi/2; (1-\delta)\pi/2)$, $\delta \le 0.1$. For the breathing mode, we have interaction maxima with transferred momenta $\mathbf{q}_1 = (3\pi/4; 3\pi/4)$ and $\mathbf{q}_2 = (\pi;\pi)$ (Fig. 2). An electron with $\mathbf{k}_n - \mathbf{q}_1$ is far from the Fermi surface, and the state with $\mathbf{k}_n - \mathbf{q}_2$ is near the Fermi level. Here, $E(\mathbf{k}_n - \mathbf{q}_2) \approx \Delta(\mathbf{k}_n - \mathbf{q}_2) = 0$ both above and below T_c; therefore, for a kink energy $\varepsilon(\mathbf{k}_n) = |E(\mathbf{k}_n - \mathbf{q}) - \omega^{(1)}_{\mathbf{q}}|$, we obtain $\varepsilon(\mathbf{k}_n) =$ 70 meV, which corresponds to the breathing-mode energy. An EPI maximum at the point $\mathbf{q}_3 = (\pi; 0)$ (the half-breathing mode) is also visible in Fig. 2. The vector $\mathbf{k}_n - \mathbf{q}_3 \approx (-\pi/2; \pi/2)$ is close to the nodal point; therefore, this mode obeys the energy conservation law. For the bending mode with an energy $\omega \approx 35$ meV, EPI maxima are at the points $\mathbf{q} = (0; 0)$ and $\mathbf{q} = (\pi;\pi)$ (Fig. 4) and the vectors $\mathbf{k}_n - \mathbf{q}$ are close to $\mathbf{k}_n$; however, the energy conservation law with $\varepsilon(\mathbf{k}_n) = 70$ meV does not hold true. Thus, contributions to the electronic-spectrum renormalizations at the nodal point are caused by diagonal EPI with the breathing mode and by off-diagonal EPI with the half-breathing mode, with the kink energy being temperature-independent because of the gap symmetry $\Delta(\mathbf{k})$.

Similarly, for the antinodal point $\mathbf{k}_{\text{an}}$, EPI with the breathing mode has maxima for $\mathbf{q}_2 = (\pi;\pi)$ (from a diagonal matrix element) and for $\mathbf{q}_3 = (\pi; 0)$ (from an

off-diagonal matrix element) (Fig. 3). The state with $\mathbf{k}_{an} - \mathbf{q}_2$ is close to the antinodal point and $E(\mathbf{k}_{an} - \mathbf{q}_2) \approx \Delta(\pi; 0) \sim \Delta_0 \approx 35$ meV (this is true of optimally doped Bi2212, where kinks with energies of 40 meV at T = 100 K and of 70 meV at T = 10 K were detected [4]). The state with $\mathbf{k}_{an} - \mathbf{q}_3$ lies far from the Fermi surface and is of little interest. The contribution from the breathing mode does not obey the energy conservation law at both $T > T_c$ and $T < T_c$. However, EPI with the bending mode at $q = 0$ obeys all conservation laws (Fig. 5). At $T < T_c$, we have

$$\varepsilon(\mathbf{k}_{an}) = 70 \approx \left|\Delta_0 + \omega_{\mathbf{q}}^{(3)}\right|,$$

and at $T > T_c$ we have

$$\varepsilon(\mathbf{k}_{an}) \approx \omega_{\mathbf{q}}^{(3)}.$$

According to [4], a sharp decrease in the kink amplitude at the antinodal point above T_c is caused by two factors: a decrease in the density of states at ε_F in the normal phase as compared to the superconducting phase and temperature-induced smearing. Thus, as in [4], we arrive at the conclusion that the kink at the antinodal point is mainly contributed by the bending mode.

5. DISCUSSION

The consideration of EPI given above implied the $La_{2-x}Sr_xCuO_4$ structure (the T structure). Let us qualitatively discuss changes in the EPI when going to the $Nd_{2-x}Ce_xCuO_4$ structure (the T' structure). The T' structure has no apical oxygen above and below the Cu ions; therefore, the main change is caused by the absence of the apical breathing mode with a strong EPI ($\nu = 2$ in our designations in Eq. (26)). Moreover, $g_{mm'}^{(3)}(\mathbf{k}, \mathbf{q})$ decreases substantially for EPI with the breathing mode, since the $\delta t'_{pp}$ and $\delta V'_{pp}$ contributions, which give linear-in-displacement terms in EPI for this mode (see discussion before Eq. (24)), disappear.

As for EPI with the breathing mode, the basic contribution is generated by oxygen-ion displacements in the Cu–O plane; therefore, we conclude that the values of $g_{mm'}^{(1)}(\mathbf{k}, \mathbf{q})$ for the T and T' structures differ only slightly. Inelastic neutron scattering experiments also indicate that the doping-induced changes in the phonon spectra of the breathing mode in $La_{2-x}Sr_xCuO_4$ and $Nd_{2-x}Ce_xCuO_4$ are similar [19]. A comparison of the EPIs in the T and T' structures makes it clear why the kink in the antinodal direction (0; 0)–(π; 0) is absent in $Nd_{2-x}Ce_xCuO_4$: it is absent because the interaction with the bending mode is smaller than that in $La_{2-x}Sr_xCuO_4$. If the EPI with the breathing mode changes only weakly, it is unclear why the kink in the nodal direction (0; 0)–(π; π) is absent. The ARPES data for $Nd_{2-x}Ce_xCuO_4$ in [36] imply weaker specific features in this direction.

To develop a superconducting theory, the effective Hamiltonian in which EPI is excluded using the Frolich transform is of interest [37]. Until we restrict ourselves to an effective Hamiltonian for the CuO_2 layer, the nonadiabatic effects caused by vibrations with wavevectors q_z along the c axis are insignificant [38]. For the t–J^* model with phonons (33), the Fröhlich transform in terms of the X operators is nontrivial; therefore, we briefly dwell on it. We write a Hamiltonian as

$$H = H_{el} + H_{el-ph}, \tag{35}$$

where H_{el} is the Hamiltonian of the t–J^* model (Eq. (33)) for the conduction band bottom ($m = 0$) or for the valence band top ($m = 1$) and H_{el-ph} is described by Eq. (26). In the canonical transformation $H_S = \exp(-S)H\exp(S)$, the S operator is chosen to be

$$S = \sum_{k\mathbf{q}\sigma}\sum_{mm'}(\alpha_{\mathbf{kq}}^{mm'}b_{-\mathbf{q}}^{\dagger} + \beta_{\mathbf{kq}}^{mm'}b_{\mathbf{q}})X_{\mathbf{k}+\mathbf{q}}^{\dagger m}X_{\mathbf{k}}^{m'}. \tag{36}$$

As usual, we find the α and β coefficients from the condition

$$H_{el-ph} + [H_{el}, S] = 0 \tag{37}$$

and write H_{eff} as

$$H_S = H_{el} + \frac{1}{2}[H_{el-ph}, S]. \tag{38}$$

In solving Eq. (37), we neglect the interband contributions in Eq. (36). First, the EPI interband matrix elements are always smaller than the intraband elements, since the former elements only contain off-diagonal contributions (see Eq. (26)); second, interband excitations go through a large gap Δ between the lower and upper Hubbard subbands, so that the corresponding contributions are $\sim g_{mm'}^{(\nu)}/\Delta \ll 1$. In solving Eq. (37), we also use a Hubbard-I type approximation in the $[H_{el}, S]$ commutator. As a result, for the m band we obtain

$$\begin{aligned}\alpha_{\mathbf{kq},\nu}^{mm} &= g_{mm}^{(\nu)}(\mathbf{k}, \mathbf{q})\\ &\times [(t_m(\mathbf{k}) - t_m(\mathbf{k}+\mathbf{q}))F_m - \omega_{\mathbf{q},\nu}]^{-1},\\ \beta_{\mathbf{kq},\nu}^{mm} &= g_{mm}^{(\nu)}(k, q)\\ &\times [(t_m(\mathbf{k}) - t_m(\mathbf{k}+\mathbf{q}))F_m - \omega_{\mathbf{q},\nu}]^{-1},\end{aligned} \tag{39}$$

where $t_m(\mathbf{k}) = \sum_k t_m(R)\exp(ikR)$. The dependence of α and β on the filling factor F_m and, hence, the dopant concentration appears as the effect of strong correlations. As a result, the effective Hamiltonian can be written as

$$H_{eff} = H_{t-J^*} + H_{el-ph-el}, \tag{40}$$

where H_{t-J^*} is described by Eq. (28) and $H_{\text{el–ph–el}}$ is defined as

$$H_{\text{el–ph–el}} = \sum_{\mathbf{kk'q}\nu}\sum_{m} V^{mm}_{\mathbf{kk'q}} X^{+m}_{\mathbf{k}+\mathbf{q}} X^{+m}_{\mathbf{k'}-\mathbf{q}} X^{m}_{\mathbf{k'}} X^{m}_{\mathbf{k}},$$

$$V^{mm}_{\mathbf{kk'q}} = g^{(\nu)}_{mm}(\mathbf{k}, \mathbf{q}) g^{(\nu)}_{mm}(\mathbf{k'}, -q)\omega_{\mathbf{q},\nu} \quad (41)$$

$$\times [(t_m(\mathbf{k}) - t_m(\mathbf{k}+\mathbf{q}))^2 F_m^2 - w^2_{\mathbf{q},\nu}]^{-1}.$$

Here, $m = 0$ and $m = 1$ give effective Hamiltonians for the cuprates with electron and hole doping types, respectively.

In contrast to the analogous effective interaction of weakly correlated electrons, the effective interaction in the regime of strong electron correlations depends on the occupation numbers and, hence, on the concentration, temperature, and the magnetic field.

6. CONCLUSIONS

We have derived EPI within the framework of a multiband realistic model of cuprates in the regime of strong electron correlations. The number of different microscopic contributions is large, and they are caused by the modulation of all interatomic-distance-dependent parameters upon ionic displacements; these include one-electron parameters (t_{pd} and t_{pp} hoppings between various p and d orbitals and the one-electron energies of p and d orbitals in the crystal field) and two-electron parameters (Coulomb matrix elements). For each vibration mode ν, we combine different microscopic contributions to form two parameters that characterize diagonal EPI (on lattice sites) and off-diagonal EPI. Explicit dependences of the $g_{\text{dia}}(\mathbf{q})$ and $g_{\text{off}}(\mathbf{q})$ matrix elements on the wavevectors were found for three modes, namely, breathing, apical breathing, and bending modes. A symmetrical analysis of these matrix elements allowed the qualitative conclusion that EPI with the breathing mode is involved in the formation of a kink in an electronic spectrum in the nodal direction (0; 0)–(π; π) and that the bending mode is responsible for a kink in the antinodal direction (0; 0)–(π; 0).

It does not follow from our results that the kinks are only caused by EPI. We do not exclude additive EPI contributions and interaction with spin fluctuations. Note that the recent calculation [39] of electronic-spectrum renormalizations by the nonperturbation variational Monte Carlo method, which also includes interaction with spin fluctuations, has not revealed kinks and has detected weaker electronic-spectrum renormalizations.

A comparison of EPIs in the n-type cuprates with the T' lattice and in the p-type cuprates with the T lattice showed a weaker EPI in the T' lattice as compared to the T lattice. However, the EPIs for the breathing mode differ only weakly for these two types of cuprates. An additional origin of electron–hole asymmetry in the cuprates is related to different natures of the carrier bands: in the hole cuprates, carriers are holes moving along the oxygen p_σ orbitals, whereas in the electron cuprates carriers are predominantly electrons of the $d_{x^2-y^2}$ orbital of Cu.

The effects of strong correlations in EPI manifest themselves in the filling factors F_m, which are self-consistently determined via the occupation numbers of multielectron terms and are functions of the carrier concentration, the temperature, and the magnetic field. The same factors specify the nonintegral spectral weight of the Hubbard quasiparticles, i.e., the specific features of a band structure in strongly correlated systems.

ACKNOWLEDGMENTS

We thank E.G. Maksimov, V.V. Val'kov, and M.M. Korshunov for fruitful discussions.

This work was supported by the Russian Foundation for Basic Research (project no. 03-02-16124), the program "Quantum Macrophysics" of the Presidium of the Russian Academy of Sciences, the Foundation in Support for Russian Science, the Foundation of Noncommercial Programs Dynasty, and the International Center of Fundamental Physics in Moscow.

REFERENCES

1. E. G. Maksimov, Usp. Fiz. Nauk **170**, 1033 (2000) [Phys. Usp. **43**, 965 (2000)].
2. A. Damascelli, Z.-X. Shen, and Z. Hussain, Rev. Mod. Phys. **75**, 473 (2003).
3. A. D. Gromko, A. V. Fedorov, Y.-D. Chuang, *et al.*, Phys. Rev. B **68**, 174520 (2003).
4. T. Cuk, D. H. Lu, X. J. Zhou, *et al.*, Phys. Status Solidi B **242**, 11 (2005).
5. E. Schachinger, J. J. Tu, and J. P. Carbotte, Phys. Rev. B **67**, 214508 (2003).
6. A. Lanzara, P. V. Bogdanov, X. J. Zhou, *et al.*, Nature **412**, 510 (2001).
7. A. Kaminski, M. Randeria, J. C. Campuzano, *et al.*, Phys. Rev. Lett. **86**, 1070 (2001).
8. S. V. Borisenko, A. A. Kordyuk, T. K. Kimet, *et al.*, Phys. Rev. Lett. **90**, 207001 (2003).
9. T. M. Rice, Phys. Rev. B **41**, 7243 (1990).
10. V. I. Belinicher and A. L. Chernyshev, Phys. Rev. B **47**, 390 (1993).
11. M. L. Kulic and O. V. Dolgov, Phys. Status Solidi B **242**, 151 (2005).
12. V. A. Gavrichkov, S. G. Ovchinnikov, A. A. Borisov, and E. G. Goryachev, Zh. Éksp. Teor. Fiz. **118**, 422 (2000) [JETP **91**, 369 (2000)].
13. J. Song and J. F. Annett, Phys. Rev. B **51**, 3840 (1995).
14. R. Zeyer and M. L. Kulic, Phys. Rev. B **53**, 2850 (1996).

15. A. Nazarenko and E. Dagotto, Phys. Rev. B **53**, 2987 (1996).
16. T. Dahm, D. Manske, D. Fay, and L. Tewordt, Phys. Rev. B **54**, 12006 (1996).
17. N. Bulut and D. J. Scalapino, Phys. Rev. B **54**, 14971 (1996).
18. L. Pintschovius and M. Braden, Phys. Rev. B **60**, 15039 (1999).
19. L. Pintschovius, Phys. Status Solidi B **242**, 30 (2005).
20. T. S. Nunner, J. Schmailian, and K. N. Bennemann, Phys. Rev. B **59**, 8859 (1999).
21. T. P. Devereaux, T. Cuk, Z.-X. Shen, and N. Nagaosa, Phys. Rev. Lett. **93**, 117004 (2004).
22. O. Rösch and O. Gunnarsson, Phys. Rev. Lett. **92**, 146403 (2004).
23. S. Ishihara and N. Nagaosa, Phys. Rev. B **69**, 144520 (2004).
24. Yu. B. Gaididei and V. M. Loktev, Phys. Status Solidi B **147**, 307 (1988).
25. V. J. Emeri, Phys. Rev. Lett. **58**, 2794 (1987).
26. C. M. Varma, S. Schmitt-Rink, and E. Ebrahams, Solid State Commun. **62**, 681 (1987).
27. V. V. Val'kov and S. G. Ovchinnikov, *Quasi-particles in Strongly Correlated Systems* (Sib. Otd. Ross. Akad. Nauk, Novosibirsk, 2001) [in Russian].
28. M. M. Korshunov, V. A. Gavrichkov, S. G. Ovchinnikov, *et al.*, Zh. Éksp. Teor. Fiz. **126**, 642 (2004) [JETP **99**, 559 (2004)].
29. C. Falter, M. Klenner, and W. Ludwig, Phys. Rev. B **47**, 5390 (1993).
30. R. J. McQueeney, J. L. Sarrao, P. G. Pagliuso, *et al.*, Phys. Rev. Lett. **87**, 077001 (2001).
31. L. Pintschovious and M. Braden, Phys. Rev. B **60**, R15039 (1999).
32. L. Pintschovious, N. Pyka, W. Reichardtet, *et al.*, Physica C (Amsterdam) **185**–**189**, 156 (1991).
33. W. Reichardt, N. Pyka, L. Pintschovius, *et al.*, Physica C (Amsterdam) **162**–**164**, 464 (1989).
34. M. Braden, W. Reichardt, W. Schmidbauer, *et al.*, J. Supercond. **8**, 1 (1995).
35. L. N. Bulaevskiĭ, É. L. Nagaev, and D. I. Khomskiĭ, Zh. Éksp. Teor. Fiz. **54**, 1562 (1968) [Sov. Phys. JETP **27**, 836 (1968)].
36. N. P. Armitage, D. H. Lu, C. Kim, *et al.*, Phys. Rev. B **68**, 064517 (2003).
37. H. Fröhlich, Phys. Rev. **79**, 845 (1950).
38. C. Falter, Phys. Status Solidi B **242**, 78 (2005).
39. S. Yunoki, E. Dagotto, and S. Sorella, Phys. Rev. Lett. **94**, 037001 (2005).

Translated by K. Shakhlevich

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 856–867.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 987–999.
Original Russian Text Copyright © 2005 by Basharov.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Ultrashort Pulse Propagation under Manipulated Resonance Conditions

A. M. Basharov

Russian Research Centre Kurchatov Institute, Moscow, 123182 Russia
e-mail: bash@online.ru; basharov@gmail.com

Received May 31, 2005

Abstract—The propagation of a polarized ultrashort laser pulse is analyzed by the inverse scattering method under initial conditions including a spatial pulse profile, a state of the medium, and a "switched-on" resonant atom–field interaction. Magnetic degeneracy of atomic levels is taken into account. The Maxwell–Bloch equations are rewritten in Hamiltonian form without redefining the spatial and temporal variables. The inverse scattering method is based on an analysis of a new spectral problem. Gelfand–Levitan–Marchenko-type equations are derived, a soliton solution is obtained, and the changes in parameters of two solitons after their collision are calculated. A possible experimental setup for implementing the system under analysis is discussed. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Pulse propagation in resonant media has long since attracted the attention of experimentalists and theorists by the diversity of propagation regimes and related phenomena of practical interest, its amenability to various analytical methods, and a broad scope of numerical analysis. Problems of interest include propagation of single resonant ultrashort pulses (USPs) and simultaneous propagation of such pulses combined with waves having different carrier frequencies or long background pulses. These processes are of interest as feasible regimes of pulse propagation with group velocities much lower than the phase velocity of light and as impressive demonstrations of soliton propagation in optics. In the simplest case of a single-mode pulse propagating in a resonant medium, concurrent absorption and stimulated emission processes sustain a time-independent shape of a pulse propagating with a group velocity that may be much lower than the phase velocity of light in the medium. Such pulses are described by soliton solutions to the Maxwell–Bloch equations. This phenomenon is known as self-induced transparency [1].

To date, the common setting for both experimental and theoretical studies of self-induced transparency has been as follows [2, 3]. At the input end (point $z = 0$) of a resonant medium occupying the half-space $z \geq 0$, the temporal pulse profile (minus the profile of the reflected pulse) is $\mathscr{E}_0(t)$ This profile is used as an initial condition in solving the Cauchy problem for appropriate Maxwell–Bloch equations. The propagation of the pulse into the medium is treated as its evolution in time. Note that the Maxwell–Bloch equations are written in the semi-light-cone gauge in terms of z and $t - z/c$. Previous studies, inverse scattering method, its application to mathematical models of self-induced transparency, and a variety of solutions were reviewed in [4].

In this paper, the problem of USP propagation is considered in the following setting. After an off-resonant USP propagating in a medium has traveled far from the boundary and its spatial profile can be described by a function $\mathscr{E}_0(z)$, resonant interaction between the pulse and the medium is "switched on" at the instant $t = 0$. Thus, the spatial profile $\mathscr{E}_0(z)$ is used as an initial condition in the Cauchy problem for the same Maxwell–Bloch equations written in the original variables z and t, and the profile of the evolving pulse is treated as a function of t.

Mathematically, the proposed and conventional statements of the problem are essentially different, even though the starting equations may be similar, as in the case of boundary value problems for the same operator. In terms of the inverse scattering method used to solve these problems, the difference can be explained as follows. Suppose that Lax operators $\hat{L}$ and $\hat{A}$ define the zero-curvature representation of the original system of equations. Then, different inverse scattering problems for the operators $\hat{L}$ and $\hat{A}$ are to be solved. For waves propagating in a Kerr medium, this difference in statement and solution of the problem was discussed in [5].

An analysis of possible physical implementations of USP propagation in a resonant medium suggests that it can be treated by applying the widely used Stark-pulse technique [6]. However, this technique has previously been used to analyze coherent optical phenomena, such as photon echoes, optical nutation, and optical free-induction decay. The idea behind the Stark-pulse technique as applied in those studies is as follows. Instead of generating a USP in a medium, a cw laser is focused

into the medium placed in a capacitor. The electric field generated by applying voltage across the capacitor for a short time shifts the atomic energy levels in the medium, and atomic transitions are brought into resonance with the laser beam. This is equivalent to irradiation of atoms by a resonant ultrashort pulse. Thus, pulse propagation, interaction with atoms, and their excitation are substituted by briefly switching the capacitor electric field on and off and thus manipulating resonant interaction conditions by means of Stark effect.

In the setting proposed here for analysis of self-induced transparency, the Stark-pulse technique is applied by "switching on" resonant interaction with an ultrashort pulse propagating in a medium, i.e., by manipulating resonance conditions as in studies of other coherent transient processes. As an example, consider a USP propagating in a resonant medium placed in a capacitor where static electric field is already applied. Due to the Stark shift, the USP propagation takes place under off-resonance conditions. Then, conditions for resonant USP propagation can be created by switching off the Stark field.

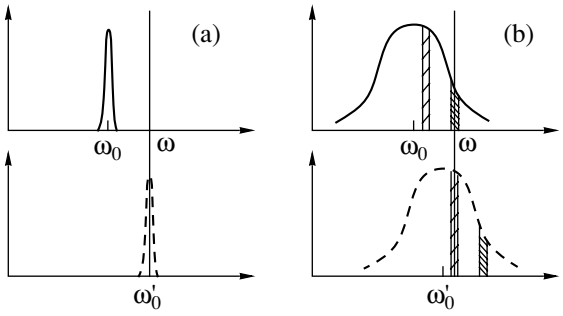


Fig. 1. Manipulation of resonance conditions by Stark shifting for (a) narrow or (b) broad spectral lines. Bell-shaped curves represent the number of atoms vs. transition frequency between two energy levels. The central frequencies ω_0 and ω_0' of these distributions correspond to zero and nonzero Stark field strengths, respectively. The USP frequency is ω_0. Hatched areas correspond to groups of atoms involved in resonant interaction at zero and/or nonzero Stark field strengths. In the reversed setting, ω_0' and ω_0 are the central frequencies of atom distributions in zero and nonzero Stark fields, respectively.

In physical terms, the proposed setting for USP propagation in a resonant medium can be interpreted as USP propagation under manipulated resonance conditions or as a special Stark-pulse technique designed to study self-induced transparency.

The present analysis of USP propagation under manipulated resonance conditions is performed for an arbitrary USP polarization and magnetically degenerate resonant energy levels. The novelty of the mathematical statement of the problem can be explained as follows. According to [7], the generalized Maxwell–Bloch equations taking into account USP polarization and degeneracy of energy levels can be solved by the inverse scattering method. In the conventional approach, this solution involves analysis of the spectral problem formulated by Manakov [8]. In the proposed approach, solution of the Cauchy problem by the inverse scattering method leads to a new spectral problem. In this study, an analysis of the Jost matrices is presented, and the Gelfand–Levitan–Marchenko-type equations required to solve the inverse scattering problem are derived.

It should be noted that the proposed approach can be applied not only to single-mode USP propagation under one-photon resonance conditions, but also to other problems of USP propagation in resonant media. In particular, analogous settings can be discussed for two-frequency USPs of polarized light under conditions of double resonance with degenerate transitions or for USP combined with a long background pulse under double-resonance or Raman-resonance conditions. The present analysis is focused on polarized USP propagation in a two-level medium with magnetically degenerate energy levels with magnetic quantum numbers 0 and 1. An analogous problem was considered in [9] without taking into account energy-level degeneracy and pulse polarization. Note that the proposed setting can also be used to analyze superfluorescence, which is generally considered in the framework of the conventional spectral problem [10].

Finally, the present analysis of a spectral problem can serve as a basis for correct description of quantum solitons of the Maxwell–Bloch lattice without redefining the spatial and temporal variables, because the semi-light-cone quantization [11], strictly speaking, corresponds to a somewhat different problem.

It should also be noted that USP propagation in real resonant media is usually complicated by inhomogeneous broadening of spectral lines. More precisely, as the Stark field is switched on, the electromagnetic wave is detuned from resonance with the initially resonant transitions in one group of atoms and brought into resonance with another group of atoms within an inhomogeneously broadened spectral line (see Fig. 1). Accordingly, when the Stark shift is not sufficiently large, the USP is always in resonance with atomic transitions in a group of atoms, which complicates the pattern of interaction between the USP and the medium under Stark shift. Analysis of an analogous pattern in photon echoes has resulted in the discovery of a new effect [12, 13]. Self-induced transparency is described by the Maxwell–Bloch equations with detuning depending on spatial coordinates and time, for which two different Cauchy problems can also be formulated and solved by an inverse scattering method with spectral parameter depending on time (or coordinate). This approach has been successfully applied to some problems of this kind [14–17], but its applicability to the problem considered here remains an open question. Note that both photon echoes [12, 13] and USP propagation in optically dense

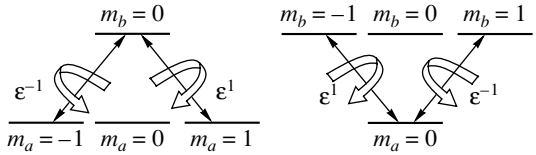


Fig. 2. Transitions between Zeeman sublevels of resonant levels for a two-level system with angular momentum values 0 and 1.

media could be described here by a unified model analogous to that considered in [18]. However, the present analysis is restricted to the simplest conditions illustrated by Fig. 1a, in which case the aforementioned dependence of coefficients in the Maxwell–Bloch equations can be eliminated by the method described in Section 2.

In Section 2, we formulate the problem, derive the generalized Maxwell–Bloch equations, and write out their zero-curvature representation. In Section 3, the Maxwell–Bloch equations are rewritten in Hamiltonian form without redefining the spatial and temporal variables. In Section 4, we determine fundamental Poisson brackets. In Section 5, we analyze a new spectral problem under the simplest boundary conditions, which correspond to a medium in thermodynamic equilibrium, and derive Gelfand–Levitan–Marchenko-type equations. In Section 6, we discuss other regimes of resonant USP propagation for which solutions can be obtained by solving similar spectral problems. We also discuss the problem corresponding to nonequilibrium boundary conditions. In Section 7, we present a one-soliton solution and the results of the collision of polarized solitons described by solving the new spectral problem.

2. STATEMENT OF THE PROBLEM AND BASIC EQUATIONS

Suppose that the carrier frequency $\omega = kc$ of an ultrashort pulse with the electric field strength

$$\mathbf{E} = \mathscr{E}(z, t)\exp[i(kz - \omega t)] + \text{c.c.}$$

is resonant with the frequency $\omega_{ba} = (E_b - E_a)/\hbar$ of the allowed optical transition between energy eigenstates $|E_a, j_a, m_a\rangle$ and $|E_b, j_b, m_b\rangle$ of particles in the medium, where E_a and E_b are energies, j_a and j_b are total angular momenta, and m_a and m_b are the magnetic quantum numbers. The eigenstates are magnetically degenerate. The particles are placed in a capacitor, and the Stark shift and splitting are such that the USP is not resonant with any allowed optical transition.

We assume that the USP enters the medium when the static field is switched on and therefore propagates under off-resonance conditions, with a phase velocity c. At $t = 0$, the capacitor electric field is switched off, and the propagating USP is brought into resonance with the medium. Starting from this instant, the particles can be treated as two-level atoms with degenerate states $|E_a, j_a, m_a\rangle$ and $|E_b, j_b, m_b\rangle$ (see Fig. 2). The USP propagation is then described by the equation for the pulse envelope

$$\left(\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right)\mathscr{E} = i2\pi kN\sum_{m_a, m_b} Q^{a*}_{m_a}\mathbf{d}_{m_a m_b}Q^{b}_{m_b}, \qquad (1)$$

and the Schrödinger equation for the wavefunction

$$|\Psi\rangle = \sum_{m_a} Q^a_{m_a}|E_a, j_a, m_a\rangle\exp\left(-\frac{iE_a t}{\hbar}\right)$$

$$+\sum_{m_b} Q^b_{m_b}|E_b, j_b, m_b\rangle\exp(-i\Delta t)\exp\left(-\frac{iE_b t}{\hbar}\right)$$

of resonant particles written for the probability amplitudes as

$$i\hbar\frac{\partial Q^a_{m_a}}{\partial t} = -\sum_{q, m_b} d^q_{m_a m_b}\mathscr{E}^{q*}Q^b_{m_b},$$

$$i\hbar\frac{\partial Q^b_{m_b}}{\partial t} + \hbar\Delta Q^b_{m_b} = -\sum_{q, m_a}(-1)^q\mathscr{E}^q d^{-q}_{m_b m_a}Q^a_{m_a}. \qquad (2)$$

The dynamic Stark effect is neglected here. Using the standard definition of the spherical components of the vector potential $\mathbf{A}$,

$$A^{\pm 1} = \mp 2^{-1/2}(A_x \pm iA_y), \quad A_0 = A_z,$$

we express the matrix elements of spherical components of the dipole moment operator in terms of the reduced dipole moment d as follows:

$$d^q_{m_b m_a} = (-1)^{j_a - m_b}\begin{pmatrix} j_b & 1 & j_a \\ -m_b & q & m_a \end{pmatrix}d,$$

$$d^q_{m_a m_b} = (-1)^{j_b - m_a}\begin{pmatrix} j_a & 1 & j_b \\ -m_a & q & m_b \end{pmatrix}d^*.$$

For $j_b = 1$ and $j_a = 0$, we have

$$d^q_{m_b m_a} = (-1)^{1 - m_b}\begin{pmatrix} 1 & 1 & 0 \\ -m_b & q & 0 \end{pmatrix}d = \frac{d}{\sqrt{3}}\delta_{q m_b},$$

$$d^q_{m_a m_b} = -\begin{pmatrix} 0 & 1 & 1 \\ 0 & q & m_b \end{pmatrix}d^* = -\frac{d^*}{\sqrt{3}}\delta_{q - m_b},$$

$$\left(\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right)\mathscr{E}^q = -i2\pi kNQ^{a*}\frac{d^*}{\sqrt{3}}Q^b_{-q},$$

$$i\hbar\frac{\partial Q^a}{\partial t} = \sum_{m_b}\frac{d^*}{\sqrt{3}}\mathscr{E}^{-q}{}^*Q_q^b,$$

$$i\hbar\frac{\partial Q_{-q}^b}{\partial t} + \hbar\Delta Q_{-q}^b = \mathscr{E}^q\frac{d}{\sqrt{3}}Q^a,$$

for $j_b = 0$ and $j_a = 1$,

$$d_{m_bm_a}^q = \begin{pmatrix} 0 & 1 & 1 \\ 0 & q & m_a \end{pmatrix}d = \frac{d}{\sqrt{3}}\delta_{q-m_a},$$

$$d_{m_am_b}^q = (-1)^{m_a}\begin{pmatrix} 1 & 1 & 0 \\ -m_a & q & 0 \end{pmatrix}d^* = -\frac{d^*}{\sqrt{3}}\delta_{qm_a},$$

$$\left(\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right)\mathscr{E}^q = -i2\pi kNQ_q^{a*}\frac{d^*}{\sqrt{3}}Q^b,$$

$$i\hbar\frac{\partial Q_q^a}{\partial t} = \frac{d^*}{\sqrt{3}}\mathscr{E}^q{}^*Q^b, \quad i\hbar\frac{\partial Q^b}{\partial t} + \hbar\Delta Q^b = \sum_{q,m_a}\mathscr{E}^q\frac{d}{\sqrt{3}}Q_q^a.$$

By introducing the polarization vector $\mathscr{P}$ and the population matrices $\mathcal{N}$ and $\mathcal{M}_{q'q}$ defined as

$$\mathscr{P}^q = \frac{d^*}{\sqrt{3}}Q^{a*}Q_{-q}^b, \quad \mathcal{N} = Q^{a*}Q^a, \quad \mathcal{M}_{q'q} = Q_{-q'}^{b*}Q_{-q}^b$$

for $j_b = 1$ and $j_a = 0$ and

$$\mathscr{P}^q = \frac{d^*}{\sqrt{3}}Q_q^{a*}Q^b, \quad \mathcal{N} = -Q^{b*}Q^b, \quad \mathcal{M}_{q'q} = -Q_{q'}^aQ_q^{a*}$$

for $j_b = 0$ and $j_a = 1$, Eqs. (1) and (2) are rewritten in the form of generalized Maxwell–Bloch equations:

$$\left(\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right)\mathscr{E} = -i2\pi kN\mathscr{P}, \tag{3}$$

$$i\hbar\frac{\partial}{\partial t}\mathscr{P}^q = -\hbar\Delta\mathscr{P}^q - \sum_{q'}\frac{|d|^2}{3}\mathscr{E}^{q'}\mathcal{M}_{q'q} + \mathscr{E}^q\frac{|d|^2}{3}\mathcal{N}, \tag{4}$$

$$\begin{aligned} i\hbar\frac{\partial\mathcal{N}}{\partial t} &= \sum_q(\mathscr{E}^q{}^*\mathscr{P}^q - \mathscr{E}^q\mathscr{P}^q{}^*), \\ i\hbar\frac{\partial\mathcal{M}_{q'q}}{\partial t} &= \mathscr{P}^{q'}{}^*\mathscr{E}^q - \mathscr{E}^{q'}{}^*\mathscr{P}^q. \end{aligned} \tag{5}$$

For simplicity, we assume here that the inhomogeneous broadening of the atomic transition line is negligible.

The initial conditions for Eqs. (3)–(5) for the Stark pulse switched on to start resonant USP propagation at $t = 0$ are

$$\begin{aligned} &\mathscr{E}(z,0) = \mathscr{E}_0(z), \quad \mathscr{P}(z,0) = \mathscr{P}_0(z), \\ &\mathcal{N}(z,0) = \mathcal{N}_0(z), \quad \mathcal{M}_{qq'}(z,0) = \mathcal{M}_{0qq'}(z). \end{aligned} \tag{6}$$

We consider the following boundary conditions:

$$\begin{aligned} &\mathscr{E}(z,t) \longrightarrow 0, \quad \mathscr{P}(z,t) \longrightarrow 0, \quad \mathcal{N}(z,t) \longrightarrow \mathcal{N}^0, \\ &\mathcal{M}_{qq'}(z,t) \longrightarrow \mathcal{M}_{qq'}^0 \\ &\text{as } z \longrightarrow \pm\infty \text{ or } t \longrightarrow \infty. \end{aligned} \tag{7}$$

We assume that the state of the resonant atoms interacting with the Stark pulse has not changed under off-resonance conditions. To ignore the reflected wave after the Stark field is switched off, we also assume that the atoms behave as a gas:

$$\frac{LN\omega|d_{ba}|}{c} \ll 1, \tag{8}$$

where L is the USP length along the z axis and N is the density of two-level atoms with transition frequency ω_{ba}.

We write the generalized Maxwell–Bloch equations (3)–(5) in dimensionless form as

$$\left(\frac{\partial}{\partial\zeta} + \frac{\partial}{\partial l}\right)\boldsymbol{\varepsilon} = -i\mathbf{p}, \tag{9}$$

$$i\frac{\partial}{\partial l}p^q = -\delta p^q - \sum_{q'}\varepsilon^{q'}\mathcal{M}_{q'q} + \varepsilon^q\mathcal{N}, \tag{10}$$

$$\begin{aligned} i\frac{\partial\mathcal{N}}{\partial l} &= \sum_q(\varepsilon^{q*}p^q - \varepsilon^qp^{q*}), \\ i\frac{\partial\mathcal{M}_{q'q}}{\partial l} &= p^{q'*}\varepsilon^q - \varepsilon^{q'*}p^q, \\ \boldsymbol{\varepsilon} = t_0\frac{d}{\sqrt{3}}\mathscr{E}\hbar^{-1}, \quad \mathbf{p} &= \frac{\mathscr{P}}{d}, \quad t_0^2 = \frac{3\hbar}{2\pi kN|d|^2c}, \\ \delta = (\omega - \omega_{ba})t_0, \quad \zeta &= \frac{z}{ct_0}, \quad l = \frac{t}{t_0}. \end{aligned} \tag{11}$$

To be specific, we consider the case of $j_b = 0$ and $j_a = 1$ and recall that the atomic density matrix and the matrix describing the atom–field interaction in the resonance approximation are proportional, respectively, to

$$\mathscr{R} = \begin{pmatrix} \mathcal{N} & p^{-1} & p^1 \\ p^{-1*} & \mathcal{M}_{-1-1} & \mathcal{M}_{-11} \\ p^{1*} & \mathcal{M}_{1-1} & \mathcal{M}_{11} \end{pmatrix},$$

$$\mathscr{V} = \begin{pmatrix} 0 & \varepsilon^{-1} & \varepsilon^1 \\ \varepsilon^{-1*} & 0 & 0 \\ \varepsilon^{1*} & 0 & 0 \end{pmatrix}. \tag{12}$$

Accordingly, we can rewrite Eqs. (9)–(11) as

$$i\left(\frac{\partial}{\partial\zeta}+\frac{\partial}{\partial l}\right)\mathcal{V} = \frac{1}{2}[\mathcal{R}, J_3],$$
$$i\frac{\partial}{\partial l}\mathcal{R} = [\mathcal{R}, \mathcal{V}] + \frac{\delta}{2}[\mathcal{R}, J_3], \tag{13}$$

where $J_3 = \mathrm{diag}(1, -1, -1)$.

Equations (9)–(11) can also be rewritten as a system analogous to (1) and (2),

$$\left(\frac{\partial}{\partial\zeta}+\frac{\partial}{\partial l}\right)\varepsilon^q = -iQ_qQ^*, \tag{14}$$

$$i\frac{\partial Q_q}{\partial l} + \delta Q_q = -\varepsilon^q Q, \quad i\frac{\partial Q}{\partial l} = -\varepsilon^{q*}Q_q, \tag{15}$$

where the new variables are related to those in (9)–(11) as follows:

$$p^q = Q_qQ^*, \quad \mathcal{N} = -QQ^*, \quad \mathcal{M}_{q'q} = -Q_qQ^*_{q'}.$$

Recall that the index $q = \pm1$ denotes the spherical components of both $\boldsymbol{\varepsilon}$ and $\mathbf{p}$. In certain cases, it is convenient to use the index $j = 0, 1, -1$, so that

$$Q_j\big|_{j=0} = Q, \quad Q_j\big|_{j=\pm1} = Q_\pm.$$

Then,

$$R_{jj'} = (-1)^{j+j'}Q^*_jQ_{j'},$$

and

$$\sum_j Q_jQ^*_j = \text{const}.$$

Equations (13) can readily be represented as compatibility conditions for solutions to auxiliary systems of linear equations:

$$\partial_l q = Lq, \tag{16a}$$

$$(\partial_\zeta + \partial_l)q = \hat{A}q, \tag{16b}$$

where

$$\hat{L} = L^{(0)} + \lambda L^{(1)}, \quad \hat{A} = \lambda^{-1}A^{(-1)}. \tag{17}$$

Compatibility condition (13) has the form

$$(\partial_\zeta + \partial_l)\hat{L} = \partial_l\hat{A} + [\hat{A}, \hat{L}]$$

or

$$(\partial_\zeta + \partial_l)(L^{(0)} + \lambda L^{(1)}) = \partial_l\lambda^{-1}A^{(-1)} + [\lambda^{-1}A^{(-1)}, (L^{(0)} + \lambda L^{(1)})],$$

by virtue of (17), and splits into the following equations:

$$(\partial_\zeta + \partial_l)L^{(0)} = [A^{(-1)}, L^{(1)}],$$
$$(\partial_\zeta + \partial_l)L^{(1)} = 0, \quad 0 = \partial_lA^{(-1)} + [A^{(-1)}, L^{(0)}].$$

These equations are equivalent to (13) if

$$\hat{L} = i\left(\lambda + \frac{\delta}{2}\right)J_3 + i\mathcal{V}, \quad \hat{A} = -i(2\lambda)^{-1}\mathcal{R}. \tag{18}$$

We then rewrite Eqs. (16) as

$$\partial_l q = \hat{L}q, \tag{19a}$$

$$\partial_\zeta q = \hat{A}'q, \quad \hat{A}' = \hat{A} - \hat{L}. \tag{19b}$$

The initial conditions at $l = 0$ are reformulated correspondingly,

$$\boldsymbol{\varepsilon}(\zeta, 0) = \boldsymbol{\varepsilon}_0(\zeta), \quad Q_j(\zeta, 0) = Q_{j0}(\zeta),$$

in order to analyze the time evolution of the distributions of the field and atomic probability amplitude (or density matrix). In the limit case of this formulation, the atomic probability amplitudes distribution in space (or density matrix) is prescribed at zero field strength, and the evolution of the field is analyzed. The conventional formulation of this problem (so-called superfluorescence problem) was given in [10].

3. HAMILTONIAN FORM OF GENERALIZED MAXWELL–BLOCH EQUATIONS

Equations (14) and (15) are represented in the Hamiltonian form:

$$\frac{\partial\varepsilon^q}{\partial l} = \{\varepsilon^q(\zeta), H\}, \quad \frac{\partial Q_j}{\partial l} = \{Q_j(\zeta), H\},$$
$$\frac{\partial\varepsilon^{q*}}{\partial l} = \{\varepsilon^{q*}(\zeta), H\}, \quad \frac{\partial Q^*_j}{\partial l} = \{Q^*_j(\zeta), H\}, \tag{20}$$

where $\{,\}$ and H denote the generalized Poisson (Dirac) bracket and the Hamiltonian function, respectively. The Poisson brackets are anticommutative,

$$\{A, B\} = -\{B, A\},$$

satisfy the Jacobi identity

$$\{A, \{B, C\}\} + \{B, \{C, A\}\} + \{C, \{A, B\}\} = 0,$$

and are consistent with the product rule for differentia-

tion with respect to each argument,

$$\{AB, C\} = A\{B, C\} + \{A, C\}B.$$

The Hamiltonian function should be sought in the form

$$H = \int d\zeta \left(i\boldsymbol{\varepsilon}^* \frac{\partial \boldsymbol{\varepsilon}}{\partial \zeta} - \delta \sum_{q = \pm 1} Q_q Q_q^* - \boldsymbol{\varepsilon}^* \cdot \mathbf{p} - \boldsymbol{\varepsilon} \cdot \mathbf{p}^* \right) \quad (21)$$

and the following relations must hold simultaneously:

$$\begin{aligned} \{\varepsilon^q(\zeta), \varepsilon^{q'*}(\zeta')\} &= i\delta_{qq'}\delta(\zeta - \zeta'), \\ \{Q(\zeta), Q^*(\zeta')\} &= -i\delta(\zeta - \zeta'), \\ \{Q_q(\zeta), Q_{q'}^*(\zeta')\} &= -i\delta_{qq'}\delta(\zeta - \zeta'). \end{aligned} \quad (22)$$

The remaining Poisson brackets must vanish. It can be shown by direct calculation that Eqs. (20) combined with (21) and (22) are equivalent to Eqs. (14) and (15). In particular, we have

$$\frac{\partial \varepsilon^q}{\partial l} = \{\varepsilon^q(\zeta), H\} = \left\{ \varepsilon^q(\zeta), \int d\zeta \left(i\boldsymbol{\varepsilon}^* \cdot \frac{\partial \boldsymbol{\varepsilon}}{\partial \zeta} - \delta Q_q Q_q^* - \boldsymbol{\varepsilon}^* \cdot \mathbf{p} - \boldsymbol{\varepsilon} \cdot \mathbf{p}^* \right) \right\}$$

$$= \left\{ \varepsilon^q(\zeta), \int d\zeta \left(i\varepsilon^{q'*} \frac{\partial \varepsilon^{q'*}}{\partial \zeta} - \varepsilon^{q'*} p^{q'} \right) \right\} = -\frac{\partial \varepsilon^q}{\partial \zeta} - ip^q,$$

$$\frac{\partial Q}{\partial l} = \{Q(\zeta), H\}$$

$$= \left\{ Q(\zeta), \int d\zeta \left(i\boldsymbol{\varepsilon}^* \cdot \frac{\partial \boldsymbol{\varepsilon}}{\partial \zeta} - \delta Q_q Q_q^* - \boldsymbol{\varepsilon}^* \cdot \mathbf{p} - \boldsymbol{\varepsilon} \cdot \mathbf{p}^* \right) \right\}$$

$$= \{Q(\zeta), \int d\zeta (-\varepsilon^{q*} Q_q Q^* - \varepsilon^q p^{q*})\} = i\varepsilon^{q*} Q_q, \ldots .$$

4. FUNDAMENTAL POISSON BRACKETS

The Poisson bracket $\{A'_{ij}(\zeta, \lambda), A'_{km}(\zeta', \lambda')\}$ between elements of the matrix A', with λ and ζ corresponding to Eq. (19b), plays a fundamental role [19]. It are conveniently written and calculated in tensor notation. By definition, $(A \otimes B)^{ij}_{km} = A_{ij}B_{km}$ for square matrices A and B of equal size. Accordingly, we express the Poisson bracket as $\{A\otimes, B\}^{ij}_{km} = \{A_{ij}, B_{km}\}$. Then,

$$\begin{aligned} &\{A'(\zeta, \lambda)\otimes, A'(\zeta'', \lambda'')\} \\ &= -(2\lambda)^{-1}(2\lambda'')^{-1}\{\mathcal{R}(\zeta)\otimes, \mathcal{R}(\zeta'')\} \\ &\quad -\{\mathcal{V}(\zeta, \lambda)\otimes, \mathcal{V}(\zeta'', \lambda'')\}, \\ &\{\mathcal{R}(\zeta)\otimes, \mathcal{R}(\zeta'')\} \\ &= [\Pi, I\otimes\mathcal{R}]i\delta(\zeta - \zeta'') = [\mathcal{R}\otimes I, \Pi]i\delta(\zeta - \zeta''), \\ &\{\hat{A}(\zeta, \lambda)\otimes, \hat{A}(\zeta'', \lambda'')\} \\ &= -(-2\lambda)^{-1}(2\lambda'')^{-1}[\Pi, I\otimes\mathcal{R}]i\delta(\zeta - \zeta'') \\ &= -(2\lambda)^{-1}(2\lambda'')^{-1}[\mathcal{R}\otimes I, \Pi]i\delta(\zeta - \zeta''), \\ &\{\hat{A}(\zeta, \lambda)\otimes, \hat{A}(\zeta'', \lambda'')\} = \frac{1}{2(\lambda - \lambda'')} \\ &\times [\Pi, (\hat{A}(\zeta, \lambda)\otimes I + I\otimes\hat{A}(\zeta, \lambda''))]\delta(\zeta - \zeta''). \end{aligned}$$

Finally, we obtain

$$\begin{aligned} \{A'(\zeta, \lambda)\otimes, A'(\zeta', \lambda')\} &= \frac{1}{2(\lambda - \lambda')} \\ \times [\Pi, (A'(\zeta, \lambda)\otimes I + I\otimes A'(\zeta, \lambda'))]\delta(\zeta - \zeta'). \end{aligned} \quad (23)$$

Here, the square bracket denotes a commutator, and Π is a permutation matrix having the following properties:

$$\Pi^2 = I, \quad \Pi(A\otimes B)\Pi = B\otimes A.$$

Thus, the classical r-matrix $r(\lambda) = \Pi/2\lambda$ that determines the Poisson bracket between elements of transfer matrices for spectral problem (19b) (see (25) below and [18]) is similar to the r-matrix for spectral problem (19a) in the theory of self-induced transparency extended to describe interaction of a polarized electromagnetic wave with a system with degenerate resonant levels. This may be interpreted as an algebraic similarity between the Hamiltonian structure of the Maxwell–Bloch equations written in the semi-light-cone gauge and in the original coordinates, at least, with regard to the simplest problem considered in the next section. An analogous similarity has been established for the sine-Gordon equation written in the original and light-cone coordinates [19].

5. SPECTRAL PROBLEM

Consider Eq. (19b) with the operator A' modified by adding a multiple of the identity matrix:

$$\begin{aligned} \partial_\zeta q &= \hat{A}'q, \quad \hat{A}' = -i(2\lambda)^{-1}\hat{\mathcal{R}}' - i\left(\lambda + \frac{\delta}{2}\right)J_3 - i\mathcal{V}, \\ \hat{\mathcal{R}}'_{jj'} &= (-1)^{j+j'}Q_j^* Q_{j'} + N\delta_{jj'}. \end{aligned} \quad (24)$$

Here, the value of N depends on the boundary conditions and is such that

$$U(\lambda) = \lim_{\zeta \to \pm\infty} \hat{A}'(\zeta, \lambda) = -i\left(\lambda + \frac{\delta}{2} - \frac{n_0}{4\lambda}\right)J_3 \quad (25)$$

for $\mathcal{M}^0_{qq'} \sim \delta_{qq'}$. If $j_b = 1$ and $j_a = 0$, then $\mathcal{M}^0_{11} = \mathcal{M}^0_{-1-1} = N_b/3$, $\mathcal{N}^0 = N_a$, and

$$n_0 = N_a - \frac{N_b}{3}, \quad N = \frac{N_a}{2} + \frac{N_b}{6}.$$

If $j_b = 0$ and $j_a = 1$, then $\mathcal{M}^0_{11} = \mathcal{M}^0_{-1-1} = -N_a/3$, $\mathcal{N}^0 = -N_b$, and

$$n_0 = \frac{N_a}{3} - N_b, \quad N = -\frac{N_a}{6} - \frac{N_b}{2}.$$

Define the transfer matrix over the interval (ζ_2, ζ_1),

$$T^{\zeta_2}_{\zeta_1}(\lambda) = \overleftarrow{\exp}\int_{\zeta_1}^{\zeta_2} \hat{A}'d\zeta \quad (26)$$

as the solution to the equation

$$\partial_{\zeta_2} T^{\zeta_2}_{\zeta_1}(\lambda) = \hat{A}'T^{\zeta_2}_{\zeta_1}(\lambda) \quad (27)$$

subject to the condition $T^{\zeta}_{\zeta}(\lambda) = I$. It is obvious that

$$(T^{\zeta_2}_{\zeta_1}(\lambda))^{-1} = T^{\zeta_1}_{\zeta_2}(\lambda), \quad T^{\zeta_3}_{\zeta_2}(\lambda)T^{\zeta_2}_{\zeta_1}(\lambda) = T^{\zeta_3}_{\zeta_1}(\lambda),$$
$$\det T^{\zeta_2}_{\zeta_1}(\lambda) = 1, \quad \partial_{\zeta_2} T^{\zeta_2}_{\zeta_1}(\lambda) = \hat{A}'(\zeta_2)T^{\zeta_2}_{\zeta_1}(\lambda),$$
$$\partial_{\zeta_1} T^{\zeta_2}_{\zeta_1}(\lambda) = -T^{\zeta_2}_{\zeta_1}(\lambda)\hat{A}'(\zeta_1).$$

For rapidly decreasing ε^q and p^q and $\mathcal{M}^0_{qq'} \sim \delta_{qq'}$ as $\zeta \longrightarrow \pm\infty$, we define $e(\zeta, \lambda)$ as the solution to the equation

$$\partial_\zeta e(\zeta, \lambda) = U(\lambda)e(\zeta, \lambda)$$

subject to the condition $e(0, \lambda) = I$:

$$e(\zeta, \lambda) = \exp\left[-i\left(\lambda + \frac{\delta}{2} - \frac{n_0}{4\lambda}\right)J_3\zeta\right].$$

The Jost matrices defined (for real λ) as

$$T_\pm(\zeta, \lambda) = \lim_{\zeta_1 \to \pm\infty} T^{\zeta_2}_{\zeta_1}(\lambda)e(\zeta_1, \lambda)$$

obviously have the asymptotic form

$$T_\pm(\zeta, \lambda) = e(\zeta, \lambda) + o(1) \text{ as } \zeta \longrightarrow \pm\infty.$$

The Jost matrices satisfy Eq. (27). Their triangular representations are

$$T_+(\zeta, \lambda) = e(\zeta, \lambda) + \int_\zeta^\infty \left(\Gamma_{1+}(\zeta, \zeta') + \frac{1}{\lambda}\Gamma_{2+}(\zeta, \zeta')\right)e(\zeta', \lambda)d\zeta',$$
$$T_-(\zeta, \lambda) = e(\zeta, \lambda) + \int_{-\infty}^\zeta \left(\Gamma_{1-}(\zeta, \zeta') + \frac{1}{\lambda}\Gamma_{2-}(\zeta, \zeta')\right)e(\zeta', \lambda)d\zeta', \quad (28)$$

$$e(\zeta, \lambda) = \exp\left(-i\left(\lambda + \frac{\delta}{2} - \frac{n_0}{4\lambda}\right)\sigma_3\zeta\right)$$
$$= \begin{pmatrix} \exp\left[-i\left(\lambda + \frac{\delta}{2} - \frac{n_0}{4\lambda}\right)\zeta\right] & 0 & 0 \\ 0 & \exp\left[\left(\lambda + \frac{\delta}{2} - \frac{n_0}{4\lambda}\right)\zeta\right] & 0 \\ 0 & 0 & \exp\left[i\left(\lambda + \frac{\delta}{2} - \frac{n_0}{4\lambda}\right)\zeta\right] \end{pmatrix},$$

where

$$J_3\Gamma_{1+}(\zeta, \zeta) - \Gamma_{1+}(\zeta, \zeta)J_3 = iJ_3\mathcal{V},$$
$$\Gamma_{2+}(\zeta, \zeta)J_3 - J_3\Gamma_{2+}(\zeta, \zeta) = -iJ_3\frac{2R' + n_0J_3}{4} + J_3\frac{2R' + n_0J_3}{n_0}J_3\Gamma_{2+}(\zeta, \zeta). \quad (29)$$

The triangular representations in (28) can be proved by transforming the equations for their respective kernels into a Goursat-type problem.

The triangular representations entail the analyticity of the Jost matrices. Denote the first, second, and third columns of the Jost matrices by $T^{(1)}_\pm(\zeta, \lambda)$, $T^{(2)}_\pm(\zeta, \lambda)$,

and $T_{\pm}^{(3)}(\zeta, \lambda)$, respectively. If $n_0 > 0$, then $T_{+}^{(2)}(\zeta, \lambda)$, $T_{+}^{(3)}(\zeta, \lambda)$, and $T_{-}^{(1)}(\zeta, \lambda)$ admit analytic continuation into the upper half-plane of complex λ, while $T_{+}^{(1)}(\zeta, \lambda)$, $T_{-}^{(2)}(\zeta, \lambda)$, and $T_{-}^{(3)}(\zeta, \lambda)$ admit analytic continuation into the lower λ half-plane.

The monodromy matrix

$$T(\lambda) = \lim_{\substack{\zeta_1 \to -\infty \\ \zeta_2 \to \infty}} e(-\zeta_2, \lambda) T_{\zeta_1}^{\zeta_2}(\lambda) e(\zeta_1, \lambda) \tag{30}$$

relates the Jost matrices:

$$\begin{aligned} T_{-}(\zeta, \lambda) &= T_{+}(\zeta, \lambda) T(\lambda), \\ T_{-}^{(1)}(\zeta, \lambda)_{11} &= T_{+}^{(1)}(\zeta, \lambda) T_{11}(\lambda) \\ &+ T_{+}^{(2)}(\zeta, \lambda) T_{21}(\lambda) + T_{+}^{(3)}(\zeta, \lambda) T_{31}(\lambda). \end{aligned} \tag{31}$$

Here, indices 1, 2, and 3 correspond to $j = 0, 1, -1$, respectively.

The zeros of $T_{11}(\lambda)$ lying in the upper λ half-plane determine the bounded solutions to the spectral problem. Denoting them by $\lambda_1, \lambda_2, \ldots$, we have

$$T_{-}^{(1)}(\zeta, \lambda_s) = \gamma_s^{(2)} T_{+}^{(2)}(\zeta, \lambda_s) + \gamma_s^{(3)} T_{+}^{(3)}(\zeta, \lambda_s).$$

Rewriting (31) as

$$\frac{1}{T_{11}(\lambda)} T_{-}^{(1)}(\zeta, \lambda) = T_{+}^{(1)}(\zeta, \lambda) + T_{+}^{(2)}(\zeta, \lambda) \frac{T_{21}(\lambda)}{T_{11}(\lambda)} + T_{+}^{(3)}(\zeta, \lambda) \frac{T_{31}(\lambda)}{T_{11}(\lambda)}$$

and making use the analytic continuability of the left-hand side of this relation into the upper λ half-plane, we can use the transform

$$\frac{1}{2\pi} \int \exp\left[i\left(\lambda + \frac{\delta}{2} - \frac{n_0}{4\lambda} \right) \eta \right] d\lambda,$$

$$\frac{1}{2\pi} \int \frac{1}{\lambda} \exp\left[i\left(\lambda + \frac{\delta}{2} - \frac{n_0}{4\lambda} \right) \eta \right] d\lambda.$$

Using the completeness conditions

$$\int_{-\infty}^{\infty} \frac{d\lambda}{\lambda^2} \exp\left[i\left(\lambda - \frac{1}{\lambda} \right) x \right] = 2\pi \delta(x),$$

$$\int_{-\infty}^{\infty} \frac{d\lambda}{\lambda} \exp\left[i\left(\lambda - \frac{1}{\lambda} \right) x \right] = 0,$$

$$\int_{-\infty}^{\infty} d\lambda \exp\left[i\left(\lambda - \frac{1}{\lambda} \right) x \right] = 2\pi \delta(x),$$

we obtain

$$-\Gamma_{1+}^{(1)}(\zeta, \eta) = \sum_{s=2,3} F_{0,s}(\zeta + \eta) g^{(s)} + \sum_{s=2,3} \int_{\zeta}^{\infty} (\Gamma_{1+}^{(s)}(\zeta, \zeta') F_{0,s}(\zeta' + \eta) + \Gamma_{2+}^{(s)}(\zeta, \zeta') F_{1,s}(\zeta' + \eta)) d\zeta',$$

$$-\Gamma_{2+}^{(1)}(\zeta, \eta) = \sum_{s=2,3} F_{1,s}(\zeta + \eta) g^{(s)} + \sum_{s=2,3} \int_{\zeta}^{\infty} (\Gamma_{1+}^{(s)}(\zeta, \zeta') F_{1,s}(\zeta' + \eta) + \Gamma_{2+}^{(s)}(\zeta, \zeta') F_{2,s}(\zeta' + \eta)) d\zeta', \tag{32}$$

$$F_{0,s}(Y) = -\sum_j i \frac{T_{s1}(\lambda_j)}{T'_{11}(\lambda_j)} \exp\left[iY\left(\lambda_j + \frac{\delta}{2} - \frac{n_0}{4\lambda_j} \right) \right] + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{T_{s1}(\lambda)}{T_{11}(\lambda)} \exp\left[iY\left(\lambda + \frac{\delta}{2} - \frac{n_0}{4\lambda} \right) \right] d\lambda,$$

$$F_{1,s}(Y) = -\sum_j i \frac{T_{s1}(\lambda_j)}{\lambda_j T'_{11}(\lambda_j)} \exp\left[iY\left(\lambda_j + \frac{\delta}{2} - \frac{n_0}{4\lambda_j} \right) \right] + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{T_{s1}(\lambda)}{\lambda T_{11}(\lambda)} \exp\left[iY\left(\lambda + \frac{\delta}{2} - \frac{n_0}{4\lambda} \right) \right] d\lambda, \tag{33}$$

$$F_{2,s}(Y) = -\sum_j i \frac{T_{s1}(\lambda_j)}{\lambda_j^2 T'_{11}(\lambda_j)} \exp\left[iY\left(\lambda_j + \frac{\delta}{2} - \frac{n_0}{4\lambda_j} \right) \right] + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{T_{s1}(\lambda)}{\lambda^2 T_{11}(\lambda)} \exp\left[iY\left(\lambda + \frac{\delta}{2} - \frac{n_0}{4\lambda} \right) \right] d\lambda,$$

$$g^{(2)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad G^{(3)} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

$$\Gamma_{1+,1}^{(2)} = -\Gamma_{1+,2}^{(1)*}, \quad \Gamma_{1+,1}^{(3)} = -\Gamma_{1+,3}^{(1)*}, \quad \Gamma_{2+,1}^{(2)} = -\Gamma_{1+,2}^{(1)*},$$

$$\Gamma_{2+,1}^{(3)} = -\Gamma_{1+,3}^{(1)*}, \quad s = 2, 3.$$

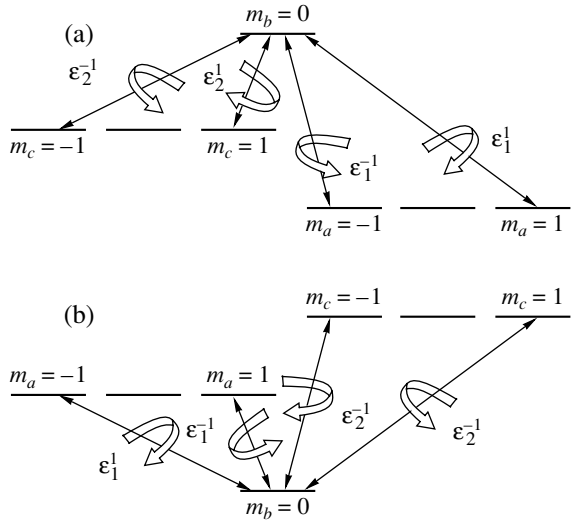


Fig. 3. Transitions between Zeeman sublevels of resonant levels in a double-resonance experiment for a three-level system with angular momentum values 0 and 1.

Finally, we write out the evolution equations for scattering data:

$$\frac{\partial}{\partial l}T(\lambda) = i\left(\lambda + \frac{\delta}{2}\right)[J_3, T(\lambda)]. \quad (34)$$

6. GENERALIZED SPECTRAL PROBLEMS

Gelfand–Levitan–Marchenko-type equations (32), (33) can readily be extended to the $(K + 1) \times (K + 1)$ generalization of spectral problem (24), for which

$$U(\lambda) = \lim_{\zeta \to \pm\infty} \hat{A}'(\zeta, \lambda) = -i\left(\lambda + \frac{\delta}{2} - \frac{n_0}{4\lambda}\right)J_3, \quad (35)$$

with

$$J_3 = \mathrm{diag}(\underbrace{1, -1, \ldots, -1}_{K}).$$

Then, the summation in s in (32) runs from 2 to $K + 1$, $\Gamma^{(1)*}_{1+,1} = -\Gamma^{(1)*}_{1+,s}$, $\Gamma^{(s)*}_{2+,1} = -\Gamma^{(1)*}_{1+,s}$, and $g_i^{(s)} = d_{is}$ ($s = 2, 3, \ldots, K + 1$). The spectral problem with $K = 4$ corresponds to the a two-frequency pulse of polarized light (simulton) interacting with an ensemble of three-level V- or Λ-type atoms with zero angular momentum of the common level b and unit angular momenta of the levels a and c (see Fig. 3). It is also assumed that the detuning from resonance is δ for both $b \longleftrightarrow a$ and $b \longleftrightarrow c$ transitions, and the oscillator strengths of these transitions are equal. The corresponding matrices (18) in the zero-curvature equation are determined by the following matrices (see [4, 20]):

$$\mathcal{R} = \begin{pmatrix} \mathcal{N} & p_1^{-1} & p_1^1 & p_2^{-1} & p_2^1 \\ p_1^{-1*} & \mathcal{M}_{-1-1} & \mathcal{M}_{-11} & \mathcal{T}_{-1-1} & \mathcal{T}_{-11} \\ p_1^{1*} & \mathcal{M}_{1-1} & \mathcal{M}_{11} & \mathcal{T}_{1-1} & \mathcal{T}_{11} \\ p_2^{-1*} & \mathcal{T}^*_{-1-1} & \mathcal{T}^*_{1-1} & \mathcal{H}_{-1-1} & \mathcal{H}_{-11} \\ p_2^{1*} & \mathcal{T}^*_{-11} & \mathcal{T}^*_{11} & \mathcal{H}_{1-1} & \mathcal{H}_{11} \end{pmatrix},$$

$$\mathcal{V} = \begin{pmatrix} 0 & \varepsilon_1^{-1} & \varepsilon_1^1 & \varepsilon_2^{-1} & \varepsilon_2^1 \\ \varepsilon_1^{-1*} & 0 & 0 & 0 & 0 \\ \varepsilon_1^{1*} & 0 & 0 & 0 & 0 \\ \varepsilon_2^{-1*} & 0 & 0 & 0 & 0 \\ \varepsilon_2^{1*} & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (36)$$

Here, the electric field amplitude vectors $\boldsymbol{\varepsilon}_i = (\varepsilon_i^{-1}, \varepsilon_i^1)$ with $i = 1$ and 2 correspond to modes resonant with the $b \longleftrightarrow a$ and $b \longleftrightarrow c$ transitions, respectively; $\mathbf{p}_i = (p_i^{-1}, p_i^1)$ are the polarization amplitude vectors for these transitions; $\mathcal{N}$ is the population of the level b; and the matrices $\mathcal{M}$ and $\mathcal{H}$ describe the distributions over the Zeeman sublevels of a and c, respectively. The matrix $\mathcal{T}$ characterizes the coherence of the two-photon process $a \longleftrightarrow c$.

The spectral problem considered in the preceding section is consistent with the boundary conditions, $\mathcal{M}^0_{qq'} \sim \delta_{qq'}$, and both ε^q and p^q rapidly decreasing as $\zeta \longrightarrow \pm\infty$. These conditions correspond to an equilibrium state of the system before and after its interaction with a propagating USP. Assuming that the system is in a nonequilibrium steady state before and after the pulse propagation, e.g.,

$$\mathcal{M}^0_{qq'} = \begin{pmatrix} m & ir \\ -ir & m \end{pmatrix}, \quad (37)$$

as $\zeta \longrightarrow \pm\infty$, we have

$$U(\lambda) = \lim_{\zeta \to \pm\infty} \hat{A}'(\zeta, \lambda) = -\frac{i}{2\lambda}\begin{pmatrix} \mathcal{N}^0 & 0 & 0 \\ 0 & m^0 & ir \\ 0 & -ir & m \end{pmatrix}$$

$$-i\left(\lambda + \frac{\delta}{2}\right)\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

instead of (25). Here, the positive constant m characterizes the population of the coupled Zeeman sublevels, and the coherence induced between them is described by the parameter r ($|r| \le m$). A nonequilibrium system of this kind can be prepared by using additional cw laser beams that are not resonant with the transition in question. Performing the canonical transformation

$$q = S^{-1}\tilde{q}, \quad \tilde{A}(\zeta, \lambda) = S\hat{A}'(\zeta, \lambda)S^{-1},$$

$$S = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & -i \\ 0 & i & -1 \end{pmatrix},$$

we obtain the spectral problem

$$\partial_\zeta \tilde{q} = \tilde{A}\tilde{q} \tag{38}$$

with operator having the asymptotic form

$$\tilde{U}(\lambda) = \lim_{\zeta \to \pm\infty} \tilde{A}(\zeta, \Lambda)$$

$$= -\frac{i}{4\lambda}\begin{pmatrix} \mathcal{N}^0 & 0 & 0 \\ 0 & 2m-2r & 0 \\ 0 & 0 & 2m+2r \end{pmatrix} \tag{39}$$

$$-i\left(\lambda + \frac{\delta}{2}\right)\begin{pmatrix} 1/2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

This problem cannot be reduced to that considered above and must be analyzed separately. An analogous spectral problem arises when there is no coherence between the Zeeman sublevels and their steady-state populations are different.

7. ONE-SOLITON SOLUTION AND COLLISION OF TWO SOLITONS

Consider again a system in equilibrium. The one-soliton solution is found by retaining one spectral component in each function F, i.e., setting $\lambda_1 = i\sigma + \xi$ and $T_{21}(\lambda) = T_{31}(\lambda) = 0$, and performing standard calculations:

$$\boldsymbol{\varepsilon}(\zeta, l) = \frac{2\sigma\mathbf{l}}{\cosh\left\{2\sigma\left(l - \zeta\left(1 + \frac{1}{4\sigma^2}\right) - l_0\right)\right\}}. \tag{40}$$

Here, $\mathbf{l}$ is the unit polarization vector and l_0 is a constant "phase." Thus, the one-soliton solution to spectral problem (24) has the form of a classical hyperbolic secant pulse. Since the velocity of soliton (40),

$$\frac{4|\sigma|^2}{1 + 4|\sigma|^2}$$

is determined by σ, two propagating solitons of this form can collide and interact. It should be noted here that soliton (40) has an arbitrary polarization $\mathbf{l}$ by virtue of initial and boundary conditions corresponding to a state of equilibrium. When the system deviates from these conditions, the soliton cannot be arbitrarily polarized.

Consider two propagating such pulses. At the point where the resonant interaction is "switched on," the field amplitude is

$$\boldsymbol{\varepsilon}(\zeta, l = 0) = \frac{2\sigma_1\mathbf{l}_1}{\cosh\left(\zeta\left(2\sigma_1 + \frac{1}{2\sigma_1}\right) + l_{01}\right)} + \frac{2\sigma_2\mathbf{l}_2}{\cosh\left(\zeta\left(2\sigma_2 + \frac{1}{2\sigma_2}\right) + l_{02}\right)}. \tag{41}$$

If $\sigma_2 > \sigma_1$ and $l_{02} > l_{01}$, then the second soliton overtakes the first one, their polarization vectors change, and the total field amplitude becomes

$$\boldsymbol{\varepsilon}(\zeta, l) = \frac{2\sigma_1\mathbf{l}_1'}{\cosh\left[2\sigma_1\left(l - \zeta\left(1 + \frac{1}{4\sigma_1^2}\right) - l_{01}'\right)\right]} + \frac{2\sigma_2\mathbf{l}_2'}{\cosh\left[2\sigma_2\left(l - \zeta\left(1 + \frac{1}{4\sigma_2^2}\right) - l_{02}'\right)\right]}, \tag{42}$$

where

$$\mathbf{l}_1' = f\left[-\mathbf{l}_1 + \frac{2\sigma_2}{\sigma_2 - \sigma_1}\mathbf{l}_2(\mathbf{l}_1 \cdot \mathbf{l}_2^*)\right],$$

$$\mathbf{l}_2' = f\left[-\mathbf{l}_2 + \frac{2\sigma_1}{\sigma_2 - \sigma_1}\mathbf{l}_1(\mathbf{l}_2 \cdot \mathbf{l}_1^*)\right],$$

$$f = \left[1 + \frac{4\sigma_1\sigma_2}{(\sigma_2 - \sigma_1)^2}|\mathbf{l}_1 \cdot \mathbf{l}_2^*|^2\right]^{-1/2}, \tag{43}$$

$$(\mathbf{l}_1' \cdot \mathbf{l}_2'^*) = (\mathbf{l}_1 \cdot \mathbf{l}_2^*),$$

$$l_{01}' = l_{01} + \frac{1}{2\sigma_1}\ln\frac{\sigma_1 + \sigma_2}{\sigma_2 - \sigma_1}f,$$

$$l_{02}' = l_{02} - \frac{1}{2\sigma_2}\ln\frac{\sigma_1 + \sigma_2}{\sigma_2 - \sigma_1}f.$$

According to (43), if the solitons are linearly polarized before their collision, then their polarizations will

remain linear after the collisions. However, their respective polarization vectors will rotate to different angles:

$$\begin{aligned}\mathbf{l}_1' \cdot \mathbf{l}_1 &= \cos\theta_1 = -(1 + B_{12}\cos^2\theta)\\ &\times (1 - B_{12}B_{21}\cos^2\theta)^{-1/2},\\ \mathbf{l}_2' \cdot \mathbf{l}_2 &= \cos\theta_2 = -(1 + B_{21}\cos^2\theta)\\ &\times (1 - B_{12}B_{21}\cos^2\theta)^{-1/2},\end{aligned} \qquad (44)$$

$$B_{12} = \frac{2\sigma_2}{\sigma_1 - \sigma_2}, \quad B_{21} = \frac{2\sigma_1}{\sigma_2 - \sigma_1}, \quad \cos\theta = \mathbf{l}_1 \cdot \mathbf{l}_2.$$

As could be expected, these relations are similar to those for the polarization vectors of the solitons in the Manakov spectral problem [8].

Note that the generalized problem corresponding to double resonance in the system schematized by Fig. 3a also has a soliton solution, which describes a two-frequency pulse characterized by a single polarization vector (simulton) [4]. Collisions of simultons result in interrelated rotations of their respective polarization vectors and energy redistribution between the single-mode components of the simulton, which are similar to those described in [20].

8. CONCLUSIONS

Current areas of topical research interest include optical phenomena in so-called artificial media, such as photonic crystals, "left-handed" media, quantum dots and wires, and atoms in magneto-optical traps or optical lattices. The artificial media may also exhibit optical properties induced by various electromagnetic and acoustic fields. It should be expected that technological progress in development of artificial media will make it possible to effectively manipulate resonance conditions and thus control nonlinear propagation of optical pulses. The Stark-pulse technique is the simplest among those in the experimentalist's toolkit. However, the analysis presented in this paper shows that even the simplest modification of resonance conditions leads to new mathematical problems.

The generalized Maxwell–Bloch equations are special in that they admit zero-curvature representation and can be solved by the inverse scattering method, which relies on analysis of the spectral problem for an appropriate linear operator [4]. The conventional USP propagation problem and the problem considered here are solved by inverse scattering methods associated with spectral problems for different operators. Since these spectral problems are not gauge equivalent, the Gelfand–Levitan–Marchenko-type Eqs. (32) and (33) required to solve the new spectral problem are derived in this study. Combined with evolution equation (34), these equations describe resonant interaction between a USP with a prescribed initial spatial profile and a two-level system with magnetically degenerate levels. In this paper, only the simplest soliton solution (40) is presented, and its transformation in a soliton collision is described. The self-induced transparency pulse described by the new model is a secant soliton similar to most soliton solutions. This result is easy to explain when the self-induced transparency pulse interacts with a system in equilibrium: the pulse energy is absorbed at the leading edge and emitted by the excited system at the trailing edge. The pulse shape remains invariant, and the initial conditions of its interaction with the medium determine only the amplitude of the self-induced transparency pulse by the area theorem. These conditions are independent of whether temporal or spatial pulse profile is prescribed. Moreover, the operator in the spectral problem defines a class of isospectral deformations [21]. This may explain the relation between soliton solutions to gauge-inequivalent spectral problems discussed above.

Solutions to different problems should be substantially different only when the region where resonant interaction with a USP is "switched on" deviates from equilibrium conditions. Solution of spectral problem (38) for a nonequilibrium system, with an operator having an asymptotic form (39) different from (25), is beyond the scope of this study.

The approach developed here provides a basis for canonical quantization of the Maxwell–Bloch equations without redefining the spatial and temporal variables. Fundamental Poisson bracket (23) defines commutation relations for the elements of the quantum monodromy matrix $T(\lambda)$, and the quantization is performed in the framework of the canonical approach [22].

ACKNOWLEDGMENTS

I thank I.R. Gabitov, A.I. Maimistov, E.A. Manykin, and A.V. Rybin for helpful discussions.

REFERENCES

1. S. L. McCall and E. L. Hahn, Phys. Rev. **183**, 457 (1969).
2. M. J. Ablowitz, D. J. Kaup, and A. C. Newell, J. Math. Phys. **15**, 1852 (1974).
3. R. K. Bullough, P. M. Jack, P. W. Kitchenside, and R. Saunders, Phys. Scr. **20**, 364 (1979).
4. A. I. Maimistov and A. M. Basharov, *Nonlinear Optical Waves* (Kluwer Academic, Dordrecht, 1999).
5. T. Iizuka, M. Wadati, and T. Yajima, J. Phys. Soc. Jpn. **60**, 2862 (1991).
6. R. G. Brewer and R. L. Shoemaker, Phys. Rev. Lett. **27**, 631 (1971); R. L. Shoemaker and R. G. Brewer, Phys. Rev. A **6**, 2001 (1972).
7. A. M. Basharov and A. I. Maĭmistov, Zh. Éksp. Teor. Fiz. **87**, 1594 (1984) [Sov. Phys. JETP **60**, 913 (1984)].
8. S. V. Manakov, Zh. Éksp. Teor. Fiz. **65**, 505 (1973) [Sov. Phys. JETP **38**, 248 (1974)].

9. A. M. Basharov and V. A. Kashurnikova, Opt. Spektrosk. **99**, 820 (2005).
10. I. R. Gabitov, V. E. Zakharov, and A. V. Mikhaĭlov, Teor. Mat. Fiz. **63**, 11 (1985).
11. V. I. Rupasov, Zh. Éksp. Teor. Fiz. **83**, 1711 (1982) [Sov. Phys. JETP **56**, 989 (1982)].
12. A. I. Alekseev and A. M. Basharov, Zh. Éksp. Teor. Fiz. **77**, 537 (1979) [Sov. Phys. JETP **50**, 272 (1979)]; J. Phys. B **15**, 4269 (1982).
13. A. Schenzle, N. C. Wong, and R. G. Brewer, Phys. Rev. A **21**, 887 (1980).
14. W. L. Chan and K. S. Li, J. Phys. A **27**, 883 (1994).
15. Tong-ke Ning, Deng-yuan Chen, and j. Da-jun Zhang, Physica A (Amsterdam) **339**, 248 (2004).
16. A. A. Zabolotskiĭ, Zh. Éksp. Teor. Fiz. **120**, 749 (2001) [JETP **93**, 657 (2001)].
17. A. A. Zabolotskiĭ, Zh. Éksp. Teor. Fiz. **121**, 1012 (2002) [JETP **94**, 869 (2002)].
18. S. M. Zakharov and É. A. Manykin, Zh. Éksp. Teor. Fiz. **82**, 397 (1982) [Sov. Phys. JETP **55**, 227 (1982)].
19. L. D. Faddeev and L. A. Takhtajan, *Hamiltonian Methods in the Theory of Solitons* (Nauka, Moscow, 1986; Springer, Berlin, 1987).
20. A. M. Basharov and A. I. Maĭmistov, Opt. Spektrosk. **68**, 1112 (1990) [Opt. Spectrosc. **68**, 649 (1990)].
21. I. M. Gel'fand and L. A. Dikiĭ, Funkts. Anal. **11** (2), 11 (1977).
22. E. K. Sklyanin, Zap. Nauchn. Semin. LOMI **95**, 55 (1980).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 868–880.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 1000–1012.
Original Russian Text Copyright © 2005 by Gekht, Bondarenko.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

A Quantum Spin Liquid in a Two-Layer Triangular Antiferromagnet

R. S. Gekht† and I. N. Bondarenko*

Kirenskiĭ Institute of Physics, Siberian Division, Russian Academy of Sciences, Akademgorodok, Krasnoyarsk, 660036 Russia

**e-mail: bondhome@mail.ru*

Received August 13, 2004

Abstract—The possibility of implementing a quantum-spin-liquid-type state in a two-layer triangular spin-1/2 antiferromagnet at $T = 0$ is investigated. The ratio of intra- to interlayer exchange constants (j) is found under which a transition from the classical state with 120° triangular order to a quantum state with zero magnetization per site occurs; in this case, the spins of adjacent layers form singlets that are separated from triplet excitations by an energy gap. Compared with an analogous system with the square lattice, the range of j in which the classical ordered state is realized turns out to be an order of magnitude smaller due to the effects of frustration; in this case, the behavior of thermodynamic quantities is analogous, on the whole, to that in two-layer square lattices; a difference manifests itself in the behavior of the gap in the spectrum of quasiparticles in an external magnetic field h. For small fields h, a j–h phase diagram is constructed that determines the domains in which the 120° and the singlet phases exist. It is established that, in the neighborhood of the second-order phase transition, the contribution, to the thermodynamic quantities, of longitudinal spin fluctuations, which are disregarded in the spin-wave description, is comparable to the contribution of transverse fluctuations. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

It is well known that the formation of a spin gap between a lower singlet state and excited magnetic levels in magnets leads to the situation when there is now magnetic ordering in these magnets at low temperatures. Moreover, the phenomenon of high-temperature superconductivity is likely to have a spin-gap nature. Therefore, a large number of model objects and real compounds possessing the above properties have recently been studied [1–19]. Quantum phenomena most clearly manifest themselves in low-dimensional systems, systems with small coordination number, low-spin systems, and frustrated systems [20–26]. Frustrations, in particular, create or enhance the degeneracy of the ground state [27], thus giving rise to new types of ordering as a result of competition between quantum phenomena and weak perturbations of different nature; on the other hand, they lead to the effective damping of coupling, thus significantly changing the domain where a quantum-disordered phase can be realized.

In the present paper, we consider conditions under which singlet and magnetic phases arise, as well as the thermodynamic properties of a system consisting of two layers of a triangular ferromagnet. Theoretical interest in bilayer systems has been stimulated in part by experimental factors. Experiments show that certain high-temperature semiconductors contain pairs of CuO_2 layers separated from other layers by nonmagnetic interlayers [28, 29]. Layers with a triangular lattice formed by ferromagnetic He^3 have also been deposited in experiments [30].

Investigations in two-dimensional triangular Heisenberg antiferromagnets have shown [31–34] that, at $T = 0$, long-range ordering exists even in spin-1/2 systems; in this case, the magnetization per site is half the classical magnetization and has virtually the same value as that in square lattices [35, 36]. At the same time, it is well known that, under certain relations between intra- and interlayer exchange constants, the interaction between layers in two-layer square antiferromagnets may lead to the transition to a singlet state with complete quantum reduction of the spin [37–39].

The possibility of quantum behavior in bilayer antiferromagnets is clear from the following considerations. In a bilayer system consisting of Heisenberg 1/2 spins with the intralayer exchange constant J_1 and the interlayer constant J_2, in the limiting case of $J_1 = 0$, we have a system of noninteracting dimers in which one of the following four states is realized at each site: a singlet or one of three triplet states, whose energies differ from the ground-state energy by the value of the exchange constant J_2. Obviously, the inclusion of the weak intralayer exchange J_1 does not substantially change the situation: the width of the gap is not equal to J_2 as in the case of noninteracting dimers but is on the

† Deceased.

order of J_2 (a confirmation of this fact is given in the text; see formula (14) below). Therefore, for a weak interlayer exchange ($J_2 \ll J_1$), the gap is small, the triplet states at every site are populated, and the mean magnetization is different from zero and, in a bilayer triangular antiferromagnet, should correspond to the classical 120° ordering. In the limit of large J_2 ($J_2 \gg J_1$), the triplet magnetic levels are separated from the ground state by a too large gap, and the system should live in a singlet state with zero magnetization per site.

The Hamiltonian of the model ($J_1, J_2 \geq 0$) is given by

$$H = J_1 \sum_{\langle i,j \rangle} \mathbf{S}_{1i}\mathbf{S}_{1j} + J_1 \sum_{\langle i,j \rangle} \mathbf{S}_{2i}\mathbf{S}_{2j} + J_2 \sum_i \mathbf{S}_{1i}\mathbf{S}_{2i} + J_2 \sum_i \mathbf{S}_{2i}\mathbf{S}_{1i}, \quad (1)$$

where $\langle i, j \rangle$ is a pair of nearest neighbors in each layer and 1 and 2 are the layer numbers. The apparent redundancy in the last terms is due to the fact that, under a systematic enumeration in $\sum_{\langle i,j \rangle}$, each interacting pair of spins is counted twice, while, in $\sum_i$, it is counted once. A pair of nearest-neighbor spins of adjacent layers is called a dimer.

2. SPIN-WAVE CALCULATIONS

We carried out standard spin-wave calculations in the ordered triangular phase with the use of the Holstein–Primakov transformation from spin operators to the operators of creation and annihilation of magnon deviations from the 120° triangular order. We established that the excitation spectrum consists of two branches each of which contains a Goldstone boson: the first branch contains this boson at a wavevector of $\mathbf{k} = (0, 0)$, which corresponds to fermionic ordering, and the second, at $\mathbf{k} = \mathbf{q} = (4\pi/3, 0)$, which corresponds to the 120° triangular order:

$$E_{k1}^{\mathrm{HP}} = 3J_1 S\sqrt{(1-\nu_k)(1+2\nu_k+2j)},$$
$$E_{k2}^{\mathrm{HP}} = 3J_1 S\sqrt{(1+2\nu_k)(1-\nu_k+2j)} > E_{k1}^{\mathrm{HP}}, \quad (2)$$

$$\nu_k \equiv \frac{1}{3}\left(\cos k_x + 2\cos\frac{k_x}{2}\cos\frac{\sqrt{3}}{2}k_y\right), \quad j \equiv \frac{J_2}{3J_1}. \quad (3)$$

In the first order in $1/S$, we obtained the magnetization per site and the velocity of spin waves in the neighborhood of a symmetry wavevector $\mathbf{k} = \mathbf{q}$ (here, N is the number of dimers).

$$N_0 = S + \frac{1}{2} - \frac{S}{2N}\sum_{\alpha,k}\frac{3J_1 + J_2 + 3J_1\nu_k/2}{2E_{k\alpha}^{\mathrm{HP}}},$$
$$c = \frac{3\sqrt{3}}{2\sqrt{2}}J_1 S\sqrt{1+\frac{4}{9}\frac{J_2}{J_1}}. \quad (4)$$

3. BOND-OPERATOR REPRESENTATION

In another limiting case, a spin-wave description is inapplicable to a singlet phase with zero magnetization per site; therefore, we used the bond-operator representation, which was first introduced in [35] and was later applied to Heisenberg models with a competing interaction [2, 5, 36] and to two-layer antiferromagnets with a square lattice [38, 39]. We introduce a system of dimer states

$$|t_a\rangle = \uparrow\uparrow = |1, 1\rangle,$$
$$|t_b\rangle = -\downarrow\downarrow = |1, -1\rangle,$$
$$|t_c\rangle = \frac{1}{\sqrt{2}}(\uparrow\downarrow + \downarrow\uparrow) = |1, 0\rangle, \quad (5)$$
$$|0\rangle = \frac{-1}{\sqrt{2}}(\uparrow\downarrow - \downarrow\uparrow) = |0, 0\rangle$$

and three bosons a, b, and c that describe a transition from the singlet state $|0\rangle$ to one of the three triplet states:

$$a^+|0\rangle = |t_a\rangle, \quad b^+|0\rangle = |t_b\rangle, \quad c^+|0\rangle = |t_c\rangle. \quad (6)$$

The creation and annihilation operators of the singlet state are defined as

$$s^+|0\rangle = |0\rangle, \quad s|0\rangle = |0\rangle. \quad (7)$$

The operators s^+ and s thus defined are equal to each other and to a constant:

$$s^+ = s \equiv u,$$

which indicates that the singlet state is a condensed state. The operator u defined by the normalization condition allows us to rule out the existence of several triplet excitations at a single site. At a given moment, one of the four states is realized at a site and the operator of the total number of particles at the site is 1:

$$s^+s + a^+a + b^+b + c^+c = 1.$$

Therefore,

$$s^+s = u^2 = 1 - a^+a - b^+b - c^+c \Rightarrow u$$
$$= \sqrt{1-(a^+a + b^+b + c^+c)}. \quad (8)$$

In terms of the new operators, the components of the

ferro- and antiferromagnetism vectors

$$\mathbf{M} = \mathbf{S}_1 + \mathbf{S}_2, \quad \mathbf{L} = \mathbf{S}_1 - \mathbf{S}_2$$

of the dimer are expressed as

$$\begin{aligned} M^z &= a^+a - b^+b, \quad M^+ = \sqrt{2}(a^+c - c^+b), \\ M^- &= \sqrt{2}(c^+a - b^+c), \quad L^z = -(c^+u + uc), \\ L^+ &= \sqrt{2}(a^+u + ub), \quad L^- = \sqrt{2}(b^+u + ua). \end{aligned} \tag{9}$$

Following [38], we introduce a parameter λ in the root of the operator u,

$$u = \sqrt{1 - \lambda(a^+a + b^+b + c^+c)},$$

which allows us to expand u in the approximation of $\lambda \ll 1$. Then, similar to the $1/S$ expansion in the usual spin-wave theory, we set $\lambda = 1$ in the final results. In order that the spin commutation relations

$$\begin{aligned} [M^\alpha, M^\beta] = i\varepsilon_{\alpha\beta\gamma}M^\gamma, \quad [L^\alpha, L^\beta] = i\varepsilon_{\alpha\beta\gamma}M^\gamma, \\ [M^\alpha, L^\beta] = i\varepsilon_{\alpha\beta\gamma}L^\gamma \end{aligned} \tag{10}$$

remain unchanged, we introduce the factor $1/\sqrt{\lambda}$ into the three components of the vector $\mathbf{L}$:

$$\begin{aligned} L^z &= -(c^+u + uc)/\sqrt{\lambda}, \\ L^+ &= \sqrt{2}(a^+u + ub)/\sqrt{\lambda}, \\ L^- &= \sqrt{2}(b^+u + ua)/\sqrt{\lambda}. \end{aligned} \tag{11}$$

Substituting (9) and (11) into the initial Hamiltonian (1) and taking into account that the relation

$$\mathbf{S}_1\mathbf{S}_2 = -\frac{3}{4} + a^+a + b^+b + c^+c$$

holds for a dimer with $S = 1/2$, we obtain ($J^*_{ij} \equiv J_{ij}/\lambda$)

$$\begin{aligned} H = &-\frac{3}{2}J_2N + 2J_2\sum_i (a_i^+a_i + b_i^+b_i + c_i^+c_i) \\ &+ \sum_{ij}\Big\{ J_{ij}(c_i^+a_i - b_i^+c_i)(a_j^+c_j - c_j^+b_j) \\ &+ J^*_{ij}(b_i^+u_i + u_ia_i)(a_j^+u_j + u_jb_j) \\ &+ \frac{1}{2}J^*_{ij}(c_i^+u_i + u_ic_i)(c_j^+u_j + u_jc_j) \\ &+ \frac{1}{2}J_{ij}(a_i^+a_i - b_i^+b_i)(a_j^+a_j - b_j^+b_j)\Big\}, \end{aligned} \tag{12}$$

where the operators a, b, and c satisfy the boson permutation relations

$$\begin{aligned} [a_i, a_j^+] = \delta_{ij}, \quad [a_i^+, a_j^+] = 0, \\ [a_i, a_j] = 0, \quad [a_i, b_j] = 0, \text{ etc.} \end{aligned}$$

4. THE SPECTRUM OF A DISORDERED PHASE

In a disordered state, the bosons a, b, and c are equivalent and a quadratic form of Hamiltonian (12) is expressed as ($u \approx 1$)

$$\begin{aligned} H = &-\frac{9}{2}J_2N + \sum_k A_k(a_k^+a_k + b_k^+b_k + c_k^+c_k \\ &+ a_{-k}a_{-k}^+ + b_{-k}b_{-k}^+ + c_{-k}c_{-k}^+) \\ &+ \sum_k B_k(a_k^+b_{-k}^+ + b_k^+a_{-k}^+ + b_{-k}a_k \\ &+ a_{-k}b_k + c_k^+c_{-k}^+ + c_{-k}c_k), \\ A_k = 3J_1^*(j + \nu_k), \quad B_k = 3J_1^*\nu_k, \quad j \equiv &\frac{J_2}{3J_1^*}. \end{aligned} \tag{13}$$

The stability domain of the singlet phase can be determined by analyzing the excitation spectrum of quasiparticles. The excitation spectrum of a disordered state is determined without taking quantum corrections into account by the diagonalization of the quadratic form. In view of the equivalence of the bosons a, b, and c, the spectrum is triply degenerate and has a gap at the wavevector $\mathbf{q}$ of the 120° triangular order:

$$\begin{aligned} E_k &= \sqrt{A_k^2 - B_k^2} = J_2\sqrt{1 + \frac{1}{j}2\nu_k}, \\ \Delta_{abc} &= J_2\sqrt{1 - 1/j} = E_k(\mathbf{k} = \mathbf{q}). \end{aligned} \tag{14}$$

As explained in the Introduction, the gap in the spectrum of elementary excitations of a disordered phase (including the domain $J_2 \gg J_1$) is on the order of J_2 and is equal to the exact value of J_2 in the case of noninteracting dimers ($J_1 = 0$).

When $j > 1$, the spectrum is everywhere real; when $j < 1$, the spectrum becomes partially imaginary: the system should pass to a new state. At the point of the phase transition, $j = 1$, the gap in the spectrum vanishes; therefore, the energy of excitations associated with the formation of the 120° triangular order vanishes. A Goldstone boson $E_k(\mathbf{k} = \mathbf{q}) = 0$ arises that points to the symmetry reduction associated with the condensation of the new state—the 120° triangular order—for $j < 1$. Thus, in this approximation, the 120° triangular order is stable in the domain of $j < 1$, while the singlet phase is

stable for $j > 1$. The velocity of spin waves at the transition point is $c = (3/2)J_1^*$.

5. MODIFICATION OF OPERATORS IN THE 120° PHASE

In the ordered phase, the operators a, b, and c are modified so as to guarantee that the mean value of the spin at a site corresponds to the 120° triangular order. This can be done by separating the mean value of operators of kind c at the wavevector $\mathbf{k} = \mathbf{q}$. If we set

$$\langle c_q \rangle = \sqrt{N}\alpha \Leftrightarrow c_k\sqrt{N}\alpha\delta_{kq} + \varepsilon_k, \tag{15}$$

then, for the mean value of the spin at a site in the second layer, we obtain

$$\langle S_{2i}^z \rangle = \frac{\sqrt{\beta(1-\beta)}}{\lambda}\cos qR_i \equiv N_0^{\text{mid}}\cos qR_i,$$

where $\beta \equiv \lambda\alpha^2$. One can see that $\langle S_{2i}^z \rangle$ behaves as a projection of the modulus N_0^{mid} onto the z axis at an angle $\alpha_i = qR_i$, where α_i is changed by $4\pi/3 \cdot 1 = 240° \Leftrightarrow -120°$ (rotation of the spin) under the transition from a certain site to the neighboring one ($R_i = 1$). At each site, the spins of the first layer are opposite, as it must be, to the spins of the second. Thus,

$$N_0^{\text{mid}} \equiv \frac{\sqrt{\beta(1-\beta)}}{\lambda}$$

is the mean value of the spin at a site in the zero approximation, and representation (15) guarantees the 120° triangular order.

The equilibrium value of β is determined from the minimum of the ground-state energy. In the mean-field approximation, the energy of the ground state and β ($\partial E_0/\partial\beta = 0$) are given by

$$\begin{aligned} E_0 &= -\frac{3}{2}J_2 N + 2J_2 N\frac{\beta}{\lambda} - 6J_1^* N\frac{\beta}{\lambda}(1-\beta), \\ \beta_0 &= \frac{1}{2}(1-j). \end{aligned} \tag{16}$$

Thus, the mean value of the operators c ($\sim\alpha$) and the mean value of the spin at a site make sense for $j < 1$, i.e., in the ordered phase; at the point $j = 1$ of the phase transition in the mean-field approximation, all the means vanish.

6. THE EXCITATION SPECTRUM OF THE ORDERED PHASE

To determine the excitation spectrum in the ordered phase, one should find a quadratic form of Hamiltonian (12) with regard to relations (15).

The Hamiltonian of the ordered phase can be represented as

$$H = E_0 + H_\perp + H_\parallel,$$

where $H_\perp$ is a part that is quadratic with respect to the operators a and b and $H_\parallel$ is a part that is quadratic with respect to the operators ε. $H_\parallel$ yields the spectrum of longitudinal fluctuations of the spin (the operators c and ε determine the mean value of the spin at a site), and $H_\perp$ determines the spectrum of transverse oscillations.

6.1. The Spectrum of Transverse Oscillations

Let us explain how we determine the quadratic form by the operators a and b ($H_\perp$). To this end, in Hamiltonian (12), it suffices to use, as u, the expression

$$u = \langle u \rangle = \sqrt{1-\beta}$$

in the terms that explicitly contain a, b, and u, and the approximation

$$u = 1 - \frac{\lambda}{2}(a^+a + b^+b)$$

in the terms containing c and u; for c in $H_\perp$, we everywhere use the mean value

$$c = \langle c \rangle.$$

As a result, after the transition to the k space, $H_\perp$ takes the form

$$\begin{aligned} H_\perp &= \sum_k A_k^\perp(a_k^+a_k + b_k^+b_k + a_{-k}a_{-k}^+ + b_{-k}b_{-k}^+) \\ &\quad + B_k^\perp(a_k^+b_{-k}^+ + b_k^+a_{-k}^+ + a_kb_{-k} + b_ka_{-k}), \\ \frac{A_k^\perp}{3J_1^*} &\equiv a_\perp = j + \beta + \nu_k\left(1 - \frac{3}{2}\beta\right), \\ \frac{B_k^\perp}{3J_1^*} &\equiv b_\perp = \nu_k\left(1 - \frac{\beta}{2}\right). \end{aligned} \tag{17}$$

The spectrum of transverse modes is doubly degenerate, is gapless, and contains a Goldstone mode with $\mathbf{k} = \mathbf{q}$ (for any j):

$$\begin{aligned} E_k^\perp &= \sqrt{A_k^{\perp^2} - B_k^{\perp^2}} = J_2\left(1 + \frac{\beta}{j}\right) \\ &\quad \times \sqrt{\left(1 - \frac{\beta}{j+\beta}\nu_k\right)\left(1 + \frac{1-\beta}{j+\beta}2\nu_k\right)}, \end{aligned} \tag{18}$$

$$E_k^\perp(\beta_0) = \frac{3}{2}J_1^*(1+j)\sqrt{(1+2\nu_k)\left(1 - \frac{1-j}{1+j}\nu_k\right)}, \tag{19}$$

where $E_k^{\perp}(\beta_0)$ is the mean-field approximation. The presence of a Goldstone boson in the spectrum of oscillations in the plane of the layer is obviously associated with symmetry breaking due to the 120° Néel ordering in the plane of the layer for $j < 1$. The velocity of spin waves near $\mathbf{k} = \mathbf{q}$ is equal to

$$c = \frac{3}{2}J_1^*\sqrt{(1-\beta_0)\left(1-\frac{\beta_0}{2}\right)}.$$

6.2. The Spectrum of Longitudinal Oscillations

The terms that are quadratic in ε and form the Hamiltonian $H_{\parallel}$ are contained in the following components of Hamiltonian (12):

$$H_{\parallel} = 2J_2\sum_i \varepsilon_i^+\varepsilon_i + \frac{1}{2\lambda}\sum_{ij} J_{ij}\{(c_i^+u_i + u_ic_i)(c_j^+u_j + u_jc_j) - (c_{i0}^+u_{i0} + u_{i0}c_{i0})(c_{j0}^+u_{j0} + u_{j0}c_{j0})\},$$

where it suffices to apply the relations

$$u = \sqrt{1-\lambda c^+c}, \quad c = c_0 + \varepsilon,$$

and expand u in powers of λ. The condensate state that is included in the sum with minus sign is already taken into account in the ground-state energy (16). Upon separating the part quadratic in ε and passing to the k space, we obtain

$$H_{\parallel} = e_{\parallel}^0 + \sum_k \{A_k^{\parallel}(\varepsilon_k^+\varepsilon_k + \varepsilon_{-k}\varepsilon_{-k}^+) + B_k^{\parallel}(\varepsilon_k^+\varepsilon_{-k}^+ + \varepsilon_k\varepsilon_{-k})\}, \quad (20)$$

$$e_{\parallel}^0 = \frac{3}{2}J_1^*N\frac{\beta^2}{1-\beta},$$

$$\frac{A_k^{\parallel}}{3J_1^*} \equiv a_k^{\parallel} = j + 2\beta + \frac{\beta^2}{2(1-\beta)}(1+2\nu_k) + \nu_k(1-3\beta),$$

$$\frac{B_k^{\parallel}}{3J_1^*} \equiv b_k^{\parallel} = \beta + \frac{\beta^2}{2(1-\beta)}(1+2\nu_k) + \nu_k(1-3\beta).$$

The spectrum of longitudinal oscillations has a gap ($E_k^{\parallel}(\beta_0)$ is the mean-field approximation) and is represented as

$$E_k^{\parallel} = \sqrt{A_k^{\parallel 2} - B_k^{\parallel 2}} = 3J_1^*\sqrt{(j+\beta)\frac{j(1-\beta)-\beta(2\beta-3)}{1-\beta}\left[1+\frac{(2\beta-1)^2}{j(1-\beta)-\beta(2\beta-3)}2\nu_k\right]}, \quad (21)$$

$$E_k^{\parallel}(\beta_0) = 3J_1^*\sqrt{1+2\nu_k j^2},$$
$$\Delta_{\parallel}(\beta_0) = 6J_1^*\sqrt{\beta_0(1-\beta_0)} = E_k^{\parallel}(\mathbf{k}=\mathbf{q}). \quad (22)$$

The gap is closed at the phase-transition point $\beta_0 = 0$. In the neighborhood of the critical point ($\beta_0 \longrightarrow 0$), the value of the gap is small ($\Delta_{\parallel}(\beta_0) \sim \sqrt{\beta_0}$); therefore, in calculating various physical quantities, one may expect that the contribution of longitudinal fluctuations will be comparable to the contribution of transverse ones.

7. CORRELATION FUNCTIONS

We investigated the behavior of correlation functions between nearest-neighbor spins in a layer, $\langle \mathbf{S}_{ni}\mathbf{S}_{nj}\rangle$, and between the layers, $\langle \mathbf{S}_{1i}\mathbf{S}_{2i}\rangle$, in both phases:

$$\langle \mathbf{S}_{1i}\mathbf{S}_{1j}\rangle = \left\langle \frac{\mathbf{M}_i+\mathbf{L}_i}{2}\frac{\mathbf{M}_j+\mathbf{L}_j}{2}\right\rangle,$$
$$\langle \mathbf{S}_{2i}\mathbf{S}_{2j}\rangle = \left\langle \frac{\mathbf{M}_i-\mathbf{L}_i}{2}\frac{\mathbf{M}_j-\mathbf{L}_j}{2}\right\rangle, \quad (23)$$

where we used relations (9) and (11) and restricted the analysis to a quadratic approximation with respect to operators. For example, in a disordered phase, we obtained the following expression for $\langle \mathbf{S}_i\mathbf{S}_j\rangle^{1,2}$:

$$\langle \mathbf{S}_i\mathbf{S}_j\rangle^{1,2} = \frac{1}{4\lambda N}\sum_k 2\Big(\langle a_k^+b_{-k}^+\rangle + \langle a_kb_{-k}\rangle + \langle a_k^+a_k\rangle + \langle b_k^+b_k\rangle + \langle c_k^+c_k\rangle + \frac{1}{2}[\langle c_k^+c_{-k}^+\rangle + \langle c_kc_{-k}\rangle]\Big)\cos\mathbf{k}\Delta.$$

The mean values in this expression are obtained by a Bogolyubov transformation to new operators in terms of which the original Hamiltonian (the Hamiltonian of a disordered state in this case) is diagonal. As a result, these means contain constants and the operators of the number of particles of certain kinds with a definite value of $\mathbf{k}$; according to the Bose distribution, these operators are equal to zero in the case, which is considered here, of a magnon gas with a chemical potential of $\mu = 0$ at $T = 0$. The remaining correlation functions are

determined analogously. The results for a disordered phase look as follows ($j > 1$):

$$\langle \mathbf{S}_i \mathbf{S}_j \rangle^{1,2} = \frac{3}{4\lambda N} \sum_k \frac{\cos \mathbf{k}\Delta}{\sqrt{1 + \frac{2\nu_k}{j}}},$$
$$\langle \mathbf{S}_{1i} \mathbf{S}_{2i} \rangle = -\frac{9}{4} + \frac{3}{2N} \sum_k \frac{1 + \frac{\nu_k}{j}}{\sqrt{1 + \frac{2\nu_k}{j}}}. \quad (24)$$

We do not present analytical results for the ordered phase ($j < 1$) in view of their awkwardness. A detailed account of all the results discussed can be found in [40]. The behavior of correlation functions is shown in Fig. 1. As it should be, in the limit of $j \longrightarrow \infty$, the correlation between spins of the same dimer $\langle \mathbf{S}_{1i} \mathbf{S}_{2i} \rangle$ have the asymptotics –3/4, while intralayer correlations $\langle \mathbf{S}_i \mathbf{S}_j \rangle$ tend to zero. At the transition point, the correlation functions are continuous. In the limit of $j = 0$, the terms in $\langle \mathbf{S}_i \mathbf{S}_j \rangle$ that are attributed to the longitudinal fluctuations total zero; i.e., they make zero contribution. In contrast to modified spin-wave methods, the correlations between spins in a layer in the singlet phase have a finite value and increase as approaching the phase-transition point ($j = 1$). The correlations between spins in adjacent layers decrease with j and reach a value of –0.47, which is less than that in the modified spin-wave theory.

8. GROUND-STATE ENERGY WITH REGARD TO FLUCTUATIONS

After the diagonalization, the components $H_\perp$ and $H_\parallel$ of the Hamiltonian of the ordered phase have a standard form that allows us to calculate the ground-state energy E with regard to fluctuation corrections,

$$H_\perp = \sum_k E_k^\perp (\alpha_k^+ \alpha_k + \beta_k^+ \beta_k + \alpha_{-k} \alpha_{-k}^+ + \beta_{-k} \beta_{-k}^+) + 2 \sum_k (E_k^\perp - A_k^\perp),$$
$$H_\parallel = e_\parallel^0 + \sum_k E_k^\parallel (\gamma_k^+ \gamma_k + \gamma_{-k} \gamma_{-k}^+) + \sum_k (E_k^\parallel - A_k^\parallel), \quad (25)$$
$$E = E_0 + e_\parallel^0 + 2 \sum_k (E_k^\perp - A_k^\perp) + \sum_k (E_k^\parallel - A_k^\parallel),$$

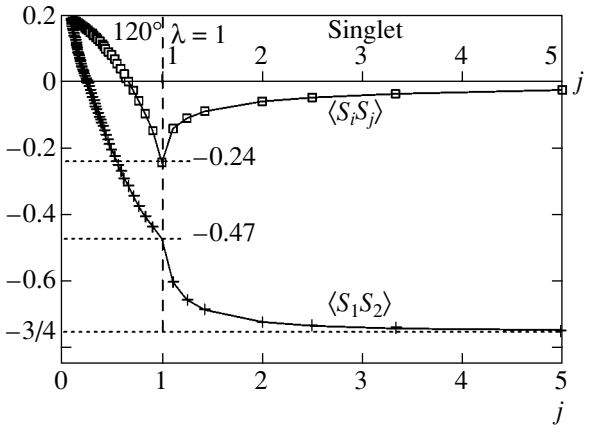


Fig. 1. Correlation functions between nearest-neighbor spins (the mean-field approximation).

and to refine the equilibrium value of the parameter β ($\partial E/\partial \beta = 0$). An equation for the equilibrium β has a self-consistent form:

$$\beta = \beta_0 - \lambda Z_b(\beta),$$
$$Z_b(\beta) = Z_1(\beta) + Z_2(\beta) + Z_3(\beta) + Z_4(\beta) + Z_5(\beta), \quad (26)$$
$$Z_1(\beta) = \frac{\beta(2-\beta)}{8(1-\beta)^2},$$
$$Z_2(\beta) = \frac{1}{N6J_1^*} \sum_k \frac{\partial A_k^\perp}{\partial \beta} \left(\frac{A_k^\perp}{E_k^\perp} - 1 \right),$$
$$Z_3(\beta) = -\frac{1}{N6J_1^*} \sum_k \frac{\partial B_k^\perp}{\partial \beta} \frac{B_k^\perp}{E_k^\perp},$$
$$Z_4(\beta) = \frac{1}{N12J_1^*} \sum_k \frac{\partial A_k^\parallel}{\partial \beta} \left(\frac{A_k^\parallel}{E_k^\parallel} - 1 \right),$$
$$Z_5(\beta) = -\frac{1}{N12J_1^*} \sum_k \frac{\partial B_k^\parallel}{\partial \beta} \frac{B_k^\parallel}{E_k^\parallel}.$$

We calculated $Z_b(\beta)$ to a first approximation by transforming the exact values of β and then used the mean-field approximation β_0 for β. The roots of the equation are determined by the method of interval bisection with an accuracy of 0.01. A family of functions $\beta(j)$ that correspond to different λ is shown in Fig. 2. One can see that there exists an asymptote

$$\beta = \beta_0 = \frac{1}{2}(1 - j)$$

as $\lambda \longrightarrow 0$. A characteristic feature of these functions for large values of λ is the two-valuedness of $\beta(j)$ in the

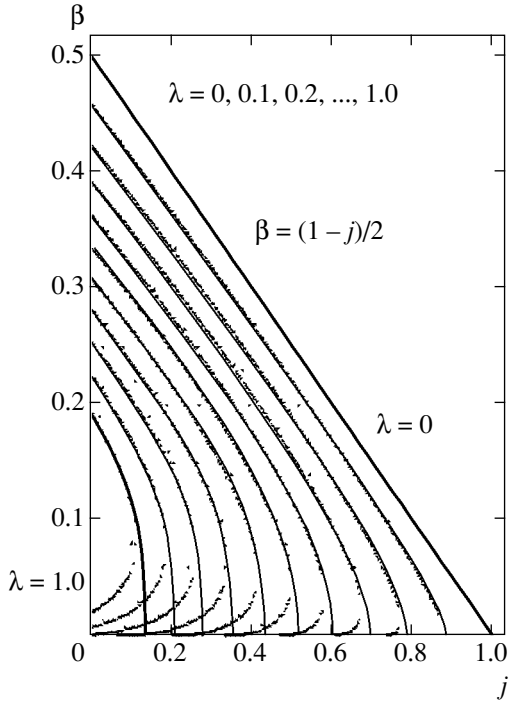


Fig. 2. Equilibrium β with regard to quantum corrections. The dots represent a solution to the self-consistent equation (26) for λ = 0.1, …, 1.0; the solid curves represent an approximation that eliminates the unphysical two-valuedness of the function, which is associated with the divergence of Gaussian fluctuations near the phase transition (β ⟶ 0). The point at which β vanishes for the physical case of λ = 1 is $j = 0.132$.

region of small β, which is physically meaningless. Such results are attributed to Gaussian fluctuations that indefinitely increase as the phase transition (β ⟶ 0) is approached; this fact was discussed in detail in [38]. Therefore, using the values of β that were calculated away from the phase transition, we approximate the functions in the neighborhood of the phase transition (β ≈ 0) so as to avoid the two-valuedness (Fig. 2). According to this approximation, the point at which β vanishes for λ = 1 is given by $j = J_2/3J_1 = 0.132$; i.e., $J_2/J_1 = 0.4$. In a similar system with a square lattice, the point of transition to a disordered phase corresponds to the values of J_2/J_1 ranging from 1.86 to 4.5 depending on the computation method [38, 39]. As expected, the classical 120° state triangular ordered decays much faster than that in the square system: the domain of values of j in which the ordered state is realized is less by an order of magnitude. Such a significant difference may be attributed in part to the in-plane frustratedness of bonds. For equal J_1 and J_2 in square and triangular systems, the effective interaction of two spins in the plane of the triangular lattice $\tilde{J}_1$ proves to be weaker by a factor of two and the ratio J_2/J_1 at the transition point is effectively greater (approaching that in the square lattice), because the spin is not oriented according to the local energy minimum. It is qualitatively clear that an effective increase in J_2/J_1 leads to an increase in the gap between the singlet and the triplet levels, whereby the singlet orbital turns out to be populated earlier (in j) and the 120° triangular order is destroyed.

9. SPONTANEOUS MAGNETIZATION

The mean value of the spin at a site in the 120° phase is equal to

$$N_0 = \left|\langle S_1^z\rangle\right| = \left|\langle S_2^z\rangle\right| = \frac{1}{\sqrt{\lambda}}\langle c^+ u\rangle = \frac{1}{\sqrt{\lambda}} \times \langle(\alpha+\varepsilon^+)\sqrt{1-\lambda[a^+a + b^+b + (\alpha+\varepsilon^+)(\alpha+\varepsilon)]}\rangle. \quad (27)$$

Up to the terms that are quadratic in the operators, we have

$$N_0 = \frac{\alpha\sqrt{1-\beta}}{\sqrt{\lambda}} \times\Bigg[1 - \frac{\lambda}{2(1-\beta)}(\langle a_k^+ a_k\rangle + \langle b_k^+ b_k\rangle + \langle\varepsilon_k^+\varepsilon_k\rangle) - \frac{\lambda}{2(1-\beta)}(\langle\varepsilon_k^+\varepsilon_k\rangle + \langle\varepsilon_k^+\varepsilon_{-k}^+\rangle) - \frac{\lambda\beta}{8(1-\beta)^2} \times(\langle\varepsilon_k\varepsilon_{-k}\rangle + \langle\varepsilon_k^+\varepsilon_{-k}^+\rangle + \langle\varepsilon_k^+\varepsilon_k\rangle + \langle\varepsilon_k\varepsilon_k^+\rangle)\Bigg]. \quad (28)$$

The means are determined by the Bogolyubov transformation, which was obtained in Section 7. As a result, we obtain

$$N_0 = \frac{\sqrt{\beta(1-\beta)}}{\lambda}(1-\lambda Z_a(\beta)),$$

$$Z_a(\beta) = Z_6(\beta) + Z_7(\beta) + Z_8(\beta) + Z_9(\beta),$$

$$Z_6(\beta) = \frac{\beta}{8(1-\beta)^2},$$

$$Z_7(\beta) = -\frac{2-\beta}{8(1-\beta)^2}\frac{1}{N}\sum_k\frac{B_k^\parallel}{E_k^\parallel}, \quad (29)$$

$$Z_8(\beta) = \frac{4-3\beta}{8(1-\beta)^2}\frac{1}{N}\sum_k\left(\frac{A_k^\parallel}{E_k^\parallel} - 1\right),$$

$$Z_9(\beta) = \frac{1}{2(1-\beta)}\frac{1}{N}\sum_k\left(\frac{A_k^\perp}{E_k^\perp} - 1\right).$$

Here,

$$\frac{\sqrt{\beta(1-\beta)}}{\lambda} = \frac{\langle c_0^+ u_0\rangle}{\sqrt{\lambda}}$$

is the approximation of magnetization without fluctuations.

An approximation of spontaneous magnetization, which is independent of the table of values of $\beta(j)$, can be obtained with the use of the first iterative approximation for β:

$$\beta \approx \beta_1 = \beta_0 - \lambda Z_b(\beta_0).$$

Then, setting

$$Z_a \approx Z_a(\beta_0), \quad Z_b \approx Z_b(\beta_0),$$

and taking into account that $\lambda \ll 1$, we obtain the approximation of magnetization that is shown in Fig. 3. This approximation has an significant drawback: the points at which N_0 and β vanish (the phase-transition point) do not coincide. However, it reveals the following characteristic feature of the function: the presence of a small maximum in the region of small j, which is especially manifest for medium values of λ; this behavior is analogous to that in two-layer square lattices [38].

We have established that, in the limit of $j = 0$, longitudinal waves do not contribute to the magnetization because the terms in Z_a that are associated with longitudinal fluctuations total zero. In the other limiting case, in the neighborhood of the phase transition ($\beta \longrightarrow 0$), the magnetization vanishes as $N_0 \propto \sqrt{\beta}$, and all the terms in $Z_a(\beta = 0)$, both transverse and longitudinal, prove to be of the same order of magnitude. Thus, longitudinal spin fluctuations, which are neglected in the spin-wave description, prove to be comparable with transverse fluctuations in the neighborhood of the phase transition.

It may seem that the obtained table of equilibrium values of $\beta(j)$ allows one to determine the exact behavior of magnetization. However, this is a fallacy, because the substitution of the values of $\beta(j)$ determined with regard to fluctuations into the integral functions $Z_a(\beta)$ and $Z_b(\beta)$, which contain the spectrum, and are calculated only in the quadratic approximation with respect to the operators of secondary quantization without taking quantum corrections into account, is an excess of accuracy. In the quadratic approximation, the spectrum is determined in the range of $j \in (0, 1)$, the point of the phase transition is $j = 1$, and the equilibrium value is $\beta = (1/2)(1-j)$. The substitution of the refined values of $\beta(j) < (1/2)(1-j)$, which are determined in the interval

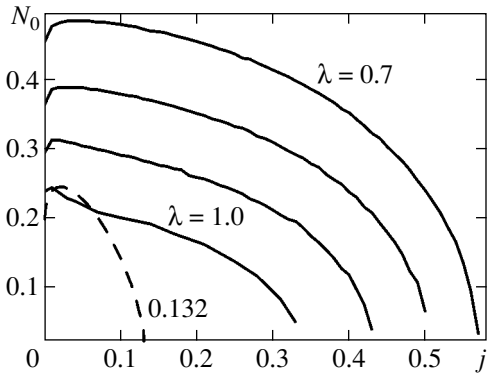


Fig. 3. Spontaneous magnetization with regard to fluctuations. The solid curves are obtained with the use of the first iterative approximation for β, and the dashed curve illustrates the approximate behavior corresponding to the table of equilibrium $\beta(j)$ for $\lambda = 1.0$.

of $j \in (0, 0.132)$, makes the spectra of the ordered phase partially imaginary (for example, the transverse spectrum is positive definite in the entire Brillouin zone only if $\beta \in ((1/2)(1-j), 1)$). Thus, a more accurate determination of the behavior of the magnetization and other quantities requires that one should either determine the excitation spectrum and the equilibrium β in the same consistent approximation, which is often associated with the elimination of unphysical states in higher orders of expansion in the operators, or apply the Monte Carlo method. We restricted ourselves to an approximate determination of the behavior that is based on the knowledge of the phase-transition point and the characteristic form of the function. An approximate behavior of magnetization is shown by a dashed line in Fig. 3. According to this figure, the magnetization vanishes simultaneously with $\beta(j)$ at $j = J_2/3J_1 = 0.132$. The calculations performed allow us to evaluate the mean value of the spin at a site; it ranges from about 1/4 to 0. Thus, the quantum reduction of the spin in the 120° triangular phase ranges from 50 to 100% depending on j.

10. INITIAL SUSCEPTIBILITY

We have calculated the initial susceptibility in a field $H = H_x$ that is perpendicular to the plane of the layers. Instead of the a and b bosons, it is convenient to use operators s and p in the field H_x,

$$s = \frac{a+b}{\sqrt{2}}, \quad p = \frac{a-b}{\sqrt{2}},$$

which satisfy the relation

$$a^+a + b^+b = s^+s + p^+p;$$

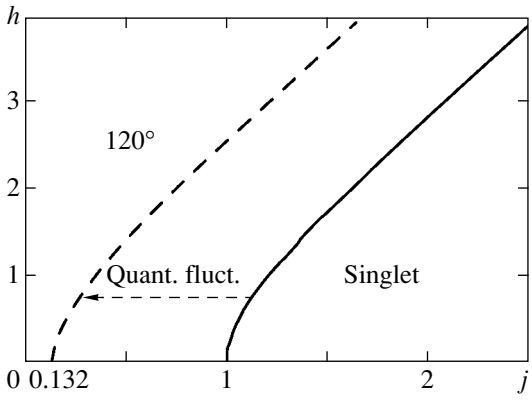


Fig. 4. *j–h* phase diagram for small *h*. The solid curve shows the position of the phase boundary in the mean-field approximation, and the dashed curve shows the same boundary with regard to fluctuations. As the field *h* increases, instead of the 120° structure [41], other seven phases become the ground state in the region of small *j*.

the ferro- and antiferromagnetism vectors are given by

$$M^x = c^+p + p^+c, \quad M^y = i(c^+s - s^+c),$$
$$M^z = s^+p + p^+s, \quad L^x = \frac{1}{\sqrt{\lambda}}(s^+u + us), \tag{30}$$
$$L^y = -\frac{i}{\sqrt{\lambda}}(p^+u - up), \quad L^z = -\frac{1}{\sqrt{\lambda}}(c^+u + uc).$$

The convenience of the new operators is associated with the fact that a field applied along the *x* axis induces the condensation of the *p* field alone, which occurs at the wavevector $\mathbf{k} = \mathbf{q}$. Indeed, the expected value of induced magnetization

$$\langle S_x(i)\rangle = M_\perp(i) = \langle c_i^+ p_i\rangle = \text{const}(i)$$

must not depend on a site; this fact imposes a constraint on the admissible form of the operators *p*. In the general case, we have

$$c_i^+ = \langle c_i^+\rangle + \varepsilon_i^+, \quad p_i = \langle p_i\rangle + \chi_i;$$

therefore,

$$M_\perp = \langle S_x\rangle = \langle c_i^+ p_i\rangle = \langle c_i^+\rangle\langle p_i\rangle + \langle \varepsilon_i^+\chi_i\rangle$$
$$\equiv M_\perp^0(i) + \Delta M_\perp(i) = \text{const}(i),$$
$$M_\perp^0(i) = \langle c_i^+\rangle\langle p_i\rangle = \alpha\exp(-iqR_i)\langle p_i\rangle \tag{31}$$
$$= \text{const}(i) \Rightarrow \langle p_i\rangle = \tilde{\alpha}\exp(iqR_i),$$
$$p_k = \sqrt{N}\tilde{\alpha}\delta_{kq} + \chi_k,$$
$$M_\perp^0 = \alpha\tilde{\alpha} = \frac{\sqrt{\beta\gamma}}{\lambda}, \quad \gamma \equiv \lambda\tilde{\alpha}^2;$$

i.e., bosons *c* and *p* condense at a wavevector $\mathbf{k} = \mathbf{q}$. The induced magnetization and the susceptibility are given by

$$M_\perp = \frac{\sqrt{\beta\gamma}}{\lambda} + \frac{1}{N}\sum_k \langle \varepsilon_k^+\chi_k\rangle \equiv M_\perp^0 + \Delta M_\perp,$$
$$\chi_\perp = \left.\frac{\partial M_\perp}{\partial H_x}\right|_{H_x = 0} \equiv \chi_\perp^0 + \Delta\chi_\perp. \tag{32}$$

The parameters β and γ can be determined from the minimum of the ground-state energy $\tilde{e}_0$ in a magnetic field. In the mean-field approximation ($h \equiv H_x/3J_1^*$), we have

$$\tilde{e}_0 \equiv \frac{\lambda E_0}{N} = -\frac{3}{2}J_2\lambda + 2J_2(\beta + \gamma)$$
$$-2H_x\sqrt{\beta\gamma} + 18J_1^*\beta\gamma - 6J_1^*\beta(1-\beta), \tag{33}$$

$$\gamma_0 = \beta_0\left[\frac{h}{2(j + 3\beta_0)}\right]^2,$$
$$j = 1 - 2\beta_0 + j\left[\frac{h}{2(j + 3\beta_0)}\right]^2. \tag{34}$$

On the curve

$$j_k = \frac{1}{2}[1 + \sqrt{1 + h^2}],$$

the parameters β_0 and γ_0 vanish simultaneously for $h \neq 0$. This function represents a curve of phase transitions in the mean-field approximation because the 120° triangular order vanishes in this case. Figure 4 shows the corresponding *j–h* phase diagram of the model. The dashed curve in this figure illustrates how the phase diagram changes when one takes into account the effect of quantum fluctuations. The magnetic field displaces, as it should, the point of phase transition to a disordered state to a domain of greater values of *j*. However, such behavior of the equilibrium curve is only characteristic of small *h*; as the magnetic field increases, the 120° triangular order is no longer the ground state, and the *j–h* phase plane exhibits an intricate structure in the region of small *j* and *h* and defines the existence domain of seven phases with different types of spin ordering [20, 41, 42]. Note that thermal fluctuations in purely two-dimensional systems lead to complete destruction of ordering at $h = 0$ due to the vanishing of the effective spin length (the Mermin–Wagner theorem). The temperature behavior of quasi-two-dimensional triangular antiferromagnets has recently been investigated in [43] with the example of $RuFe(MnO_4)_2$.

Using the obtained values of β_0 and γ_0, we determine the mean-field approximation to the induced magnetization and the initial susceptibility:

$$M_\perp^0(h \longrightarrow 0) = \frac{\sqrt{\beta_0\gamma_0}}{\lambda}\bigg|_{h\to 0} = \frac{1}{2\lambda}\frac{(1-j)}{(3-j)}h,$$
$$\chi_\perp^0 = \frac{\partial M_\perp^0(h \longrightarrow 0)}{\partial H_x} = \frac{1}{6J_1}\frac{(1-j)}{(3-j)}. \quad (35)$$

In the limiting case of $j = 0$, we have $\chi_\perp^0 = 1/18J_1$, which coincides with the results for single-layer triangular lattices.

To calculate the fluctuation corrections to the magnetization,

$$\Delta M_\perp = \frac{1}{N}\sum_k \langle \varepsilon_k^+ \chi_k \rangle,$$

we should find the eigenfunctions of the Hamiltonian in a magnetic field; in the quadratic approximation, this Hamiltonian can be reduced to

$$H_h = E_0 + H_s + H_{cp},$$

where H_s and H_{cp} are the quadratic forms with respect to the s and $\varepsilon\chi$ operators, respectively. We have

$$H_s = \sum_k 2A_k^s s_k^+ s_k + B_k^s(s_k^+ s_{-k}^+ + s_k s_{-k}),$$
$$\frac{A_k^s}{3J_1^*} \equiv a_s = j + \beta + \nu_k\left(1 - \frac{3}{2}\beta - \frac{3}{2}\gamma\right), \quad (36)$$
$$\frac{B_k^s}{3J_1^*} \equiv b_s = \nu_k\left(1 - \frac{\beta}{2} - \frac{3}{2}\gamma\right).$$

The spectrum of s excitations has a gap,

$$E_k^s = 3J_1^*(j+\beta) \times \sqrt{\left(1 - \frac{\beta}{j+\beta}\nu_k\right)\left(1 + \frac{1-\beta-\frac{3}{2}\gamma}{j+\beta}2\nu_k\right)}, \quad (37)$$

$$\Delta_s = E_k^s(\mathbf{k} = \mathbf{q})|_{h\to 0} \longrightarrow 3J_1^*\frac{2-\beta_0}{4(1+\beta_0)}h. \quad (38)$$

In contrast to square lattices, the gap depends on j; moreover, it is closed in zero field. The spectrum of s excitations in zero field transforms into the spectrum of transverse oscillations of the ordered phase. We have

$$H_{cp} = -J_q^*\frac{\beta(\beta+\gamma)}{2(1-\beta-\gamma)}N + \sum_k \{2A_k^\varepsilon \varepsilon_k^+ \varepsilon_k + B_k^\varepsilon(\varepsilon_k^+ \varepsilon_{-k}^+ + \varepsilon_k \varepsilon_{-k}) + 2A_k^\chi \chi_k^+ \chi_k + B_k^\chi(\chi_k^+ \chi_{-k}^+ + \chi_k \chi_{-k}) + C_k(\varepsilon_k^+ \chi_k + \chi_k^+ \varepsilon_k) + D_k(\varepsilon_k^+ \chi_{-k}^+ + \varepsilon_k \chi_{-k})\}, \quad (39)$$

$$\frac{A_k^\varepsilon}{3J_1^*} \equiv a_\varepsilon = j + 2\beta + \frac{\beta^2}{2(1-\beta-\gamma)}(1+2\nu_k) + \nu_k\left(1 - 3\beta - \frac{3}{2}\gamma\right),$$

$$\frac{B_k^\varepsilon}{3J_1^*} \equiv \beta_\varepsilon = \beta + \frac{\beta^2}{2(1-\beta-\gamma)}(1+2\nu_k) + \nu_k\left(1 - 3\beta - \frac{3}{2}\gamma\right),$$

$$\frac{A_k^\chi}{3J_1^*} \equiv a_\chi = j + \beta + \frac{\beta\gamma}{2(1-\beta-\gamma)}(1+2\nu_k) + \nu_k\left(1 - \gamma - \frac{3}{2}\beta\right),$$

$$\frac{B_k^\chi}{3J_1^*} \equiv b_\chi = \frac{\beta\gamma}{2(1-\beta-\gamma)}(1+2\nu_k) + \nu_k\left(-1 + \gamma + \frac{\beta}{2}\right),$$

$$\frac{C_k}{3J_1^*} \equiv c = -h + \sqrt{\beta\gamma}\left[5 + \frac{\beta}{1-\beta-\gamma}(1+2\nu_k) - 3\nu_k\right],$$

$$\frac{D_k}{3J_1^*} \equiv d = \sqrt{\beta\gamma}\left[1 + \frac{\beta}{1-\beta-\gamma}(1+2\nu_k) - 3\nu_k\right].$$

An analytic expression for the spectrum of *cp* excitations in a magnetic field is given by

$$E_k^{cp} = 3J_1^*\sqrt{m \pm \sqrt{m^2 - \tilde{c}}},$$
$$m = \frac{1}{2}\left[a_\varepsilon^2 - b_\varepsilon^2 + a_\chi^2 - b_\chi^2 + \frac{1}{2}(c^2 - d^2)\right],$$
$$\tilde{c} = c_1 c_2, \quad (40)$$
$$c_1 = \frac{(c-d)^2}{4} - (a_\varepsilon - b_\varepsilon)(a_\chi - b_\chi),$$
$$c_2 = \frac{(c+d)^2}{4} - (a_\varepsilon + b_\varepsilon)(a_\chi + b_\chi).$$

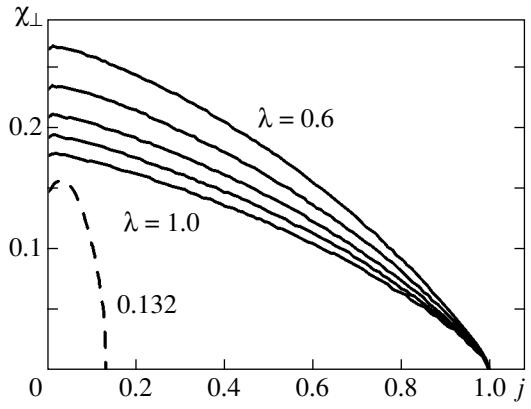


Fig. 5. Initial susceptibility. The solid curves are obtained with the use of the mean-field approximation for β, and the dashed curve represents an approximate behavior that corresponds to β(j) with regard to fluctuations: $\chi_\perp \sim \sqrt{\beta}$; therefore, β and $\chi_\perp$ vanish simultaneously at $j = 0.132$.

Unfortunately, this spectrum cannot be represented as a compact function of the variables (j, h), (β, γ), or others. The spectrum of *cp* excitations contains two branches. The lower branch contains a Goldstone boson for $\mathbf{k} = \mathbf{q}$ and, at $h = 0$, turns into the spectrum of transverse (*ab*) modes of the ordered phase. The upper branch has a gap and turns into the spectrum of longitudinal (*c*) oscillations at $h = 0$; the value of the gap in the mean-field approximation is

$$\Delta_{cp} = E^{cp}_{k(2)}(\mathbf{k} = \mathbf{q}) = 6J_1^*\sqrt{(j+\beta_0)(j-1+3\beta_0)}. \quad (41)$$

The diagonalization of H_s yields

$$H_s = \sum_k (E^s_k - A^s_k) + \sum_k 2E^s_k \gamma^+_k \gamma_k.$$

As $h \longrightarrow 0$, the diagonalization of H_{cp} leads to the following expression (ε_c and ε_p are the spectra for $h \longrightarrow 0$):

$$H_{cp} = E^0_{cp} + \sum_k [(\varepsilon_c - A^\varepsilon_k) + (\varepsilon_p - A^\chi_k) - O(h^2)] + 2\sum_k (\varepsilon_p a^+_k a_k + \varepsilon_c b^+_k b_k),$$
$$E^0_{cp} = 3J_1^* \frac{\beta(\beta+\gamma)}{2(1-\beta-\gamma)} N, \quad \varepsilon_c = \sqrt{A^{\varepsilon 2}_k - B^{\varepsilon 2}_k}, \quad (42)$$
$$\varepsilon_p = \sqrt{A^{\chi 2}_k - B^{\chi 2}_k}.$$

The ground-state energy in a magnetic field for $h \longrightarrow 0$,

$$E_h = E_0 + E^0_{cp} + \sum_k (E^s_k - A^s_k) + \sum_k [(\varepsilon_c - A^\varepsilon_k) + (\varepsilon_p - A^\chi_k) - O(h^2)], \quad (43)$$

allows us to determine the equilibrium values of β and γ (upon setting $\partial E_h/\partial\beta = 0$ and $\partial E_h/\partial\gamma = 0$). It suffices to refine one parameter, γ; the other parameter can be taken in the mean-field approximation because β and γ enter the equation

$$M_\perp = \frac{\sqrt{\beta\gamma}}{\lambda} + \Delta M_\perp$$

as a product. As a result, we obtain

$$\gamma = \beta\left[\frac{h}{2(j+3\beta)+\lambda\varphi}\right]^2,$$

where $\varphi = \varphi(\beta, \gamma, j, h)$ is an integral that contains the coefficients of the Hamiltonians H_s and H_{cp}.

Thus, taking into account a renormalized value of γ and diagonalizing transformations, we obtain the following expressions for the induced magnetization:

$$M_\perp = \frac{\sqrt{\beta_0\gamma}}{\lambda} + \Delta M_\perp$$

and the initial susceptibility,

$$\chi_\perp = \chi^0_\perp + \Delta\chi_\perp,$$

where the mean fluctuation components are given by (here, $h \longrightarrow 0$, $\beta = \beta_0$, and $j \approx 1 - 2\beta_0$)

$$M^0_\perp \approx \frac{\sqrt{\beta_0\gamma}}{\lambda} \approx h\frac{\beta_0}{\lambda[2(1+\beta_0)+\lambda\varphi(h=0)]},$$
$$\chi^0_\perp = \frac{1}{3J_1^*}\frac{\beta_0}{\lambda[2(1+\beta_0)+\lambda\varphi(h=0)]},$$
$$\Delta M_\perp = -h\frac{1}{4(1+\beta_0)}\frac{1}{N}\sum_k u^2_c u^2_p (x_c + x_p) \times \frac{[-2(1-\beta_0)+\beta_0 d'](x_c+x_p)+\beta_0 d'(1+x_c x_p)}{e_c + e_p}, \quad (44)$$

$$\Delta\chi_\perp = -\frac{1}{3J_1^*}\frac{1}{4(1+\beta_0)}\frac{1}{N}$$

$$\times\sum_k \frac{u_\parallel^2 u_\perp^2 (x_\parallel + x_\perp)}{e_\parallel + e_\perp}$$

$$\times\{[-2(1-\beta_0)+\beta_0 d']\,(x_\parallel + x_\perp) + \beta_0 d'(1 + x_\parallel x_\perp)\}.$$

Here,

$$u_i = \sqrt{\frac{1}{2}\left(\frac{a_i}{e_i}+1\right)}, \quad x_i = -\frac{b_i}{a_i + e_i},$$

$$e_i = \sqrt{a_i^2 - b_i^2} = \frac{\varepsilon_i}{3J_1^*}, \quad i = c(\varepsilon), p(\chi),$$

$$\alpha_{c(\varepsilon)}(h=0) = \alpha_k^\parallel, \quad \alpha_{p(\chi)}(h=0) = \alpha_k^\perp,$$

$$\alpha = a, e, u, x, \quad b_\varepsilon(h=0) = b_k^\parallel,$$

$$b_\chi(h=0) = -b_k^\perp, \quad d' \equiv 1 + \frac{\beta_0}{1-\beta_0}(1+2\nu_k) - 3\nu_k.$$

The form of the function $\chi_\perp$ is shown in Fig. 5. This function is nonmonotonic with a small maximum in the neighborhood of small j. The dashed line represents the behavior corresponding to the table of equilibrium values of $\beta(j)$.

In this paper, we did not consider the possibility of the formation of singlet pairs in a layer. In our view, consideration of this possibility should lead to vanishing of the spin at a site of $\chi_\perp$ for $j \approx 0$; thus, the interval of the 120° triangular phase will become even narrower.

ACKNOWLEDGMENTS

This work was supported by the program Quantum Macrophysics of the Presidium of the Russian Academy of Sciences.

REFERENCES

1. M. Hase, I. Terasaki, and K. Uchinokura, Phys. Rev. Lett. **70**, 3651 (1993).
2. Jun Zang, A. R. Bishop, and D. Schmeltzer, Phys. Rev. B **52**, 6723 (1995).
3. A. V. Chubukov, Phys. Rev. B **43**, 3337 (1991).
4. A. V. Chubukov, Pis'ma Zh. Éksp. Teor. Fiz. **49**, 108 (1989) [JETP Lett. **49**, 129 (1989)].
5. A. V. Chubukov and T. Jolicoeur, Phys. Rev. B **44**, 12050 (1991).
6. J. Darriet and J. P. Regnault, Solid State Commun. **86**, 409 (1993).
7. M. Azuma, Z. Hiroi, M. Takano, *et al.*, Phys. Rev. Lett. **73**, 3463 (1994).
8. S. Taniguchi, T. Nishikawa, Y. Yasui, *et al.*, J. Phys. Soc. Jpn. **64**, 2758 (1995).
9. H. Kageyama, K. Yoshimura, R. Stern, *et al.*, Phys. Rev. Lett. **82**, 3168 (1999).
10. B. S. Shastry and B. Sutherland, Physica B (Amsterdam) **108**, 1069 (1981).
11. H. Kageyama, H. Suzuki, M. Nohara, *et al.*, Physica B (Amsterdam) **281**, 667 (2000).
12. H. Kageyama, M. Nishi, N. Aso, *et al.*, Phys. Rev. Lett. **84**, 5876 (2000).
13. R. W. Smith and D. A. Keszler, J. Solid State Chem. **93**, 430 (1991).
14. H. Nojiri, H. Kageyama, K. Onizuka, *et al.*, J. Phys. Soc. Jpn. **68**, 2906 (1999).
15. P. Lemmens, M. Grove, M. Fischer, *et al.*, Phys. Rev. Lett. **85**, 2605 (2000).
16. S. Miyahara and K. Yeda, Phys. Rev. Lett. **82**, 3701 (1999).
17. Y. Sasago, M. Hase, K. Uchinokura, *et al.*, Phys. Rev. B **52**, 3533 (1995).
18. A. V. Syromyatnikov and S. V. Maleyev, Phys. Rev. B **66**, 132408 (2002).
19. H. Kageyama and A. N. Vasil'ev, Priroda (Moscow), No. 2, 21 (2002).
20. H. Shiba and T. Nikuni, in *Recent Advances in Magnetism of Transition Metal Compounds*, Ed. by A. Kotani and N. Suzuki (World Sci., Singapore, 1993), p. 372.
21. R. S. Gekht, Usp. Fiz. Nauk **159**, 261 (1989) [Sov. Phys. Usp. **32**, 871 (1989)].
22. A. S. Borovik-Romanov, B. S. Dumesh, and A. M. Tikhonov, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 724 (1997) [JETP Lett. **66**, 759 (1997)].
23. A. S. Borovik-Romanov, B. S. Dumesh, S. V. Petrov, *et al.*, Zh. Éksp. Teor. Fiz. **113**, 352 (1998) [JETP **86**, 197 (1998)].
24. B. S. Dumesh, S. V. Petrov, and A. M. Tikhonov, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 988 (1998) [JETP Lett. **67**, 1046 (1998)].
25. B. S. Dumesh, M. I. Kurkin, S. V. Petrov, *et al.*, Zh. Éksp. Teor. Fiz. **115**, 2228 (1999) [JETP **88**, 1221 (1999)].
26. B. S. Dumesh, Usp. Fiz. Nauk **170**, 403 (2000) [Phys. Usp. **43**, 365 (2000)].
27. M. F. Collins and O. A. Petrenko, Can. J. Phys. **75**, 605 (1997).
28. T. Siegrist, S. Sunshine, D. W. Murphy, *et al.*, Phys. Rev. B **35**, 7137 (1987).
29. J. M. Tranquada, G. Shirane, B. Keimer, *et al.*, Phys. Rev. B **40**, 4503 (1989).
30. H. Godfrin, R. R. Ruel, and D. D. Osheroff, Phys. Rev. Lett. **60**, 305 (1988).

31. B. Bernu, C. Lhuillier, and L. Pierre, Phys. Rev. Lett. **69**, 2590 (1992).
32. P. Asaria, B. Delamotte, and D. Mouhanna, Phys. Rev. Lett. **70**, 2483 (1993).
33. A. V. Chubukov, S. Sachdev, and T. Senthil, J. Phys.: Condens. Matter **6**, 8891 (1994).
34. R. R. P. Singh, Phys. Rev. B **39**, 9760 (1989).
35. S. Sachdev and R. N. Bhatt, Phys. Rev. B **41**, 9323 (1990).
36. A. V. Chubukov, E. Gagliano, and C. Balseiro, Phys. Rev. B **45**, 7889 (1992).
37. R. Chitva, S. Rao, D. Sen, *et al.*, Phys. Rev. B **52**, 1061 (1995).
38. A. V. Chubukov and D. K. Morr, Phys. Rev. B **52**, 3521 (1995).
39. Guo-Zhu Wei and An Du, J. Phys.: Condens. Matter **7**, 8813 (1995).
40. I. N. Bondarenko, Candidate's Dissertation in Physics and Mathematics (Inst. of Physics, Siberian Division, Russian Academy of Sciences, Krasnoyarsk, 2003).
41. R. S. Gekht and I. N. Bondarenko, Zh. Éksp. Teor. Fiz. **111**, 627 (1997) [JETP **84**, 345 (1997)].
42. E. Rastelli and A. Tassi, J. Phys.: Condens. Matter **8**, 1811 (1996).
43. L. E. Svistov, A. I. Smirnov, L. A. Prozorova, *et al.*, Phys. Rev. B **67**, 094434 (2003).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 881–891.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 1013–1024.
Original Russian Text Copyright © 2005 by Mambetov, Salikhov.

ORDER, DISORDER, AND PHASE TRANSITIONS IN CONDENSED SYSTEMS

Calculation of the Rate Constants of Spin Exchange between Paramagnetic Particles with Arbitrary Spins for Their Diffusion Motion in Nonviscous Liquids

A. E. Mambetov and K. M. Salikhov

Zavoiskiĭ Physicotechnical Institute, Kazan Scientific Center, Russian Academy of Sciences, Sibirskiĭ trakt 10/7, Kazan 29, 420029 Tatarstan, Russia

e-mail: mambetov@kfti.knc.ru

Received June 16, 2005

Abstract—The rate constant of spin exchange between two paramagnetic particles with arbitrary spins is calculated within a model of diffusion passage through a region of exchange interaction that exponentially depends on the interparticle distance. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Studying a bimolecular change in the spin state of paramagnetic particles caused by exchange interaction is important from various standpoints. Spin exchange is one of the model elementary processes, and its investigation can be instructive for analysis of other bimolecular processes, such as energy transfer, electron transfer, the recombination of electron–hole pairs, the recombination of radical pairs, and so on. Indeed, data on the spin-exchange rate can be used to determine the binary-collision frequency, the motion kinematics of colliding partners, and the orbital overlap of unpaired electrons for two meeting paramagnetic particles [1]. Moreover, studying bimolecular spin exchange is interesting in itself, since many processes are related to spin exchange. One of them is the quenching of positronium atoms by paramagnetic additives. The positronium atom can have a total spin equal to unity (*ortho*-positronium, long-lived state, $\tau \propto 10^{-7}$ s) or zero (*para*-positronium, short-lived state, $\tau \propto 10^{-10}$ s) [2]. Spin exchange between the positronium atom and a paramagnetic particle can transform the positronium atom from the long-lived to the short-lived state, thus quenching positronium [2, 3]. Similarly, spin exchange between radical pairs and paramagnetic particles can result in spin catalysis (e.g., see [4]). Spin exchange manifests itself in the shape of electron paramagnetic resonance (EPR) spectra, causing the exchange-induced broadening of spectral lines and the exchange narrowing of a spectrum [1]. The EPR sensitivity can be increased by creating nonequilibrium polarization of electron spins. One of such methods is based on the fact that the nonradiative intersystem crossing from a singlet into a triplet state in electronically excited molecules is spin-selective; as a result, the triplet sublevels are populated with different probabilities. Then, the triplet polarization can be transferred to other paramagnetic particles, e.g., to stable free radicals, due to spin exchange (e.g., see [5]).

Considerable attention has been given to the calculation of the rate constant of spin exchange between spin-1/2 paramagnetic species, e.g., between free radicals [1]. Two collision models have been analyzed. In most cases, the so-called sudden collision model is considered. In this model, exchange interaction between paramagnetic particles with an exchange integral J is turned on suddenly at the instant of contact, at the minimum distance between the centers of the particles, and it is also suddenly turned off after a collision time τ_c, when the particles cease to be in contact. An alternative model takes into account re-encounters between colliding particles in condensed matter, the spin dynamics of the particles in the time intervals between re-encounters, and the extended character of exchange interaction. The interdiffusion of collision partners randomly modulates the exchange interaction. By assuming that the exchange integral decreases exponentially with increasing interparticle distance and that the interdiffusion is described by a continuous diffusion model, the spin-exchange rate constant was he calculated for spin-1/2 paramagnetic species [6]. For paramagnetic particles with a spin $S > 1/2$, the spin-exchange rate constant has not been calculated with allowance for the extended character of the exchange integral and for diffusion passage through an interaction region. We only know of [7], in which the rate constant of spin exchange between triplets was calculated, and [8], in which the rate constant of spin exchange between spin-1/2 particles and paramagnetic ions with an arbitrary spin was calculated. The authors of [8] used the so-called thin reaction layer approximation: an exchange integral was assumed to be constant and equal to J inside a thin layer and to be zero outside this layer. However, it is necessary to calculate the rate constants of spin exchange

between paramagnetic particles having arbitrary spins. For example, the spin exchange between electronically excited metalloporphyrin molecules and free radicals plays a key role in the redistribution of nonequilibrium spin polarization in mixed solutions [5]. The spin of an electronically excited metalloporphyrin can acquire different values depending on the central ion. Therefore, it is necessary to know the rate constant of spin exchange between a pair of paramagnetic particles with an arbitrary spin S (metalloporphyrin) and with a spin of 1/2 (free radical).

In this work, we calculate the rate constant of spin exchange between paramagnetic particles with arbitrary spins upon their diffusion passage through an exchange-interaction region in the case of an exponential decrease in the exchange integral with increasing interparticle distance.

2. MODEL AND KINETIC EQUATIONS

We consider a solution containing two types of paramagnetic particles (particles A and B). Their concentrations, spins, and Larmor frequencies are denoted by C_A and C_B, S^A and S^B, and ω_A and ω_B, respectively. We restrict ourselves to the study of dilute solutions in order to neglect three-particle, four-particle, etc., collisions and to only analyze the bimolecular collisions of paramagnetic particles. Let R_{eff} be the effective radius of spin exchange between colliding particles (it will be defined below). The condition of applicability of the binary collision approximation is

$$\frac{4\pi}{3}R_{\text{eff}}^3(C_A + C_B) \ll 1. \tag{1}$$

To calculate the spin-exchange efficiency in a pair of colliding paramagnetic particles A and B, we have to choose a spin Hamiltonian for this pair. In the general case, it contains the Zeeman interaction of spins with a constant applied magnetic field, the splitting energy in a zero magnetic field, and the exchange and dipole–dipole spin–spin interactions between A and B. We neglect the anisotropic components of the spin Hamiltonian, which substantially simplifies computations but limits the field of applicability of the results obtained. This approximation is reasonable as applied to spin exchange in nonviscous liquids, since the rotational diffusion of paramagnetic particles effectively averages the anisotropic components of the spin Hamiltonian to zero in this case. Thus, the spin Hamiltonian for the pair is chosen to be

$$H = \hbar\omega_0^A S_Z^A + \hbar\omega_0^B S_Z^B + \hbar J(\mathbf{r}) S^A S,^B \tag{2}$$

where $J(\mathbf{r})$ is the exchange integral for the pair of interacting particles:

$$J(\mathbf{r}) = J(r) = J_0 \exp\{-\kappa(r - r_0)\}. \tag{3}$$

The constant κ in Eq. (3) characterizes the rate of decrease of the exchange integral with the distance; r is the distance between the colliding particles; and r_0 is the distance of closest approach. The relative diffusion motion of the partners in the pair randomly modulates the interaction of the pair spins. The spin–spin interaction energy is assumed to be much lower than the kinetic energy of the relative motion of the colliding partners; therefore, we neglect the effect of their spin states on the collision kinematics. As a consequence, the distance r between the colliding particles is taken to be an external (classic) parameter. The relative motion of the colliding particles is described by a continuous diffusion model.

The collision time τ_c in nonviscous liquids is 0.1–1 ns; hence, the inequality $(\omega_A - \omega_B)\tau_c < 1$ can hold true in moderate constant magnetic fields. Then, to calculate the spin-exchange efficiency, we may neglect the difference in the Larmor frequencies during collision and can use the spin Hamiltonian

$$H' = \hbar\omega_0(S_Z^A + S_Z^B) + \hbar J(r) S^A S^B \tag{4}$$

instead of Eq. (2) to calculate the collision efficiency at $t \le \tau_c$. (The collision time τ_c and the effective spin-exchange radius R_{eff} will be determined below by solving a kinetic equation.) The model spin Hamiltonian (4) has the following important property: it retains the square of the total spin $(S^A + S^B)^2$ and its projection on the direction of the applied magnetic field. This property allows us to solve the problem analytically and to find compact analytic expressions for the spin-exchange rate constants, the effective spin-exchange radius, and the effective collision time (or, in other words, the time of meeting of two particles in condensed matter, since a pair of particles that move diffusionally can undergo re-encounters in the course of their meeting.)

The spin-exchange efficiency is not only a function of the spin Hamiltonian. The kinematics of the colliding particles plays a key role. Therefore, to calculate the spin-exchange efficiency, we have to consider a kinetic equation for the spin matrix of a certain pair of particles A and B. To this end, we divide an ensemble of pairs into subensembles of pairs with a given distance r between pair partners A and B. In terms of the model chosen, a kinetic equation for the partial density matrix $\rho(r, t)$ of the pair subensemble with a given distance r between pair partners has the form

$$\frac{\partial\rho(r,t)}{\partial t} = -\frac{i}{\hbar}[H', \rho(r,t)] + D\Delta_r\rho(r,t), \tag{5}$$

where D is the interdiffusion coefficient of A and B. We use kinetic equation (5) for the spin density matrix of the pair of colliding particles A and B to find a collision term in kinetic equations for the one-particle spin density matrices of paramagnetic particles.

Kinetic equations for the one-particle spin density matrices of the subensembles of particles A and B with allowance for bimolecular collisions were proposed in [9, 10]:

$$\begin{aligned}\frac{d\sigma^A(t)}{dt} &= -i[\omega_0^A S_Z^A, \sigma^A(t)] \\ &- C_B \mathrm{Tr}_B[P\sigma^A(t)\sigma^B(t)], \\ \frac{d\sigma^B(t)}{dt} &= -i[\omega_0^B S_Z^B, \sigma^A(t)] \\ &- C_A \mathrm{Tr}_A[P\sigma^A(t)\sigma^B(t)].\end{aligned} \tag{6}$$

The collision term in Eqs. (6) is written from the following considerations. As is well-known from statistical mechanics, a change in the one-particle spin density matrix of, e.g., particle A caused by its pairwise interaction with particles B (V_{AB}) is expressed in terms of the two-particle density matrix ρ_2 of a pair of particles A and B as follows:

$$\begin{aligned}\frac{d\sigma^A(t)}{dt} &= -i[\omega_0^A S_Z^A, \sigma^A(t)] \\ -\frac{i}{\hbar V}\mathrm{Tr}_B \sum_K \int [V_{AB}(\mathbf{r}_K), \rho_2(\mathbf{r}_K, t)] d\mathbf{r}_K.\end{aligned} \tag{7}$$

The sum in the right-hand side of Eq. (7) is taken over all particles B; Tr_B means the convolution over the spin variables of particles B; and the integral is taken over the entire volume V. When passing to the thermodynamic limit for V, we have $N_B \longrightarrow \infty$, so that $N_B/V \longrightarrow C_B$, where N_B is the number of particles B and C_B is their concentration. Taking into account that, statistically, all particles B give the same contribution to collisions with particles A, we can rewrite Eq. (7) in the form

$$\begin{aligned}\frac{d\sigma^A(t)}{dt} &= -i[\omega_0^A S_Z^A, \sigma^A(t)] \\ -\frac{i}{\hbar V} C_B \mathrm{Tr}_B \int [V_{AB}(\mathbf{r}), \rho_2(\mathbf{r}, t)] d\mathbf{r}.\end{aligned} \tag{8}$$

It is well-known from statistical mechanics that an equation for a pair density matrix is linked to a three-particle density matrix. If condition (1) is satisfied, the probability of the fact that at least one particle is near the two chosen partners within the times comparable to their meeting time τ_c is negligibly small. Therefore, to derive a kinetic equation for a one-particle density matrix from Eq. (8) by making allowance for only bimolecular collisions, we can use the density matrix $\rho(\mathbf{r}, t)$ of an averaged isolated pair of particles found by solving Eq. (5) instead of the two-particle density matrix $\rho_2(\mathbf{r}, t)$ in the right-hand side of Eq. (8). To obtain the kinetic equation, the pair density matrix in Eq. (8) should be expressed in terms of one-particle density matrices. To this end, the solution to Eq. (5) for the pair density matrix is written as (see [11])

$$\rho(r, t) \approx G(r, t)\sigma^A(t)\sigma^B(t). \tag{9}$$

We substitute this expression in Eq. (5). In this equation, the interaction of the chosen pair is taken into account in an explicit form. Therefore, since we neglect triple collisions, we do not take into account collision-induced changes in the one-particle matrices (see Eq. (6)) in the expression

$$\frac{d\rho}{dt} = \frac{dG}{dt}\sigma^A\sigma^B + G\frac{d\sigma^A}{dt}\sigma^B + G\sigma^A\frac{d\sigma^B}{dt}$$

in the left-hand side of Eq. (5), and we use the approximation

$$\begin{aligned}\frac{d\sigma^A(t)}{dt} &= -i[\omega_0^A S_Z^A, \sigma^A(t)], \\ \frac{d\sigma^B(t)}{dt} &= -i[\omega_0^B S_Z^B, \sigma^B(t)].\end{aligned}$$

During particle collisions, we neglect the differences in the Zeeman frequencies, that is, $\omega_0^A = \omega_0^B = \omega_0$. Then, we obtain the following equation for $G(r, t)$ written in the Liouville representation [9]:

$$\begin{aligned}\frac{\partial G(r, t)}{\partial t} &= iW(r)G(r, t) \\ &+ i[Q_0, G(r, t)] + D\Delta_r G(r, t),\end{aligned} \tag{10}$$

where

$$\begin{aligned}W(r)_{IK, LM} \\ = J(r)[(S^A S^B)_{MK}\delta_{LI} - (S^A S^B)_{IL}\delta_{KM}],\end{aligned} \tag{11}$$

and

$$\begin{aligned}Q_{0IK, LM} \\ = \omega_0[(S_Z^A + S_Z^B)_{MK}\delta_{LI} - (S_Z^A + S_Z^B)_{IL}\delta_{KM}].\end{aligned} \tag{12}$$

In the right-hand side of Eq. (10), the term with $W(r)$ describes a change in the pair density matrix induced by spin–spin interaction. The term with the commutator having Q_0 is related to a change in the state of a pair of spins caused by their motion in an applied magnetic

field, and the last term takes into account the fluxes of the pair density matrix between the subensembles of pairs with different values of r that are caused by the interdiffusion of pair partners.

As noted above, spin Hamiltonian (4) commutes with the operator of the total spin squared $(S^A + S^B)^2$. Hence, it is convenient to perform computations using the basis of the eigenfunctions of this operator. Here, the nonzero elements of the W and Q_0 matrices are only their diagonal elements $W_{IK,IK}$ and $Q_{0,IK,IK}$, respectively, where $I \neq K$. If $I = K$, then $W_{KK,KK} = 0$ and $Q_{0,KK,KK} = 0$. Therefore, the equations for the diagonal $G_{IK,IK}$ elements become substantially simpler in the basis of the eigenfunctions of the total spin, since the diagonal elements of the commutator $[Q_0, G(r, t)]_{IK,IK}$ in Eq. (10) become zero. Equation (10) for the $G_{IK,IK}$ elements takes the form

$$\frac{\partial G_{IK,IK}(r,t)}{\partial t} = iW_{IK,IK}(r)G_{IK,IK}(r,t) + D\Delta_r G_{IK,IK}(r,t). \quad (13)$$

As a result of the degeneracy of transitions between different levels, $[Q_0, G(r, t)]_{IK,MN}$ can also vanish for some off-diagonal elements. In this case, an equation similar to Eq. (13) is also valid for the corresponding off-diagonal $G_{IK,MN}(r, t)$ element.

Under steady-state conditions, $G(r, t)$ satisfies the equation

$$D\Delta_r G(r) + i[Q_0, G(r)] + iW(r)G(r) = 0 \quad (14)$$

at times $t > \tau_c$. The diagonal $G(r)$ elements satisfy the equation

$$D\Delta_r G(r)_{IK,IK} + iW(r)_{IK,IK}G(r)_{IK,IK} = 0. \quad (15)$$

Let us formulate boundary conditions for the $G(r)$ correlator. When the partners of a pair are separated, the correlation between their states should weaken [11]. Therefore, we suppose that $G(r)$ tends to a unit matrix E with increasing r:

$$G(r) \longrightarrow E \text{ at } r \longrightarrow \infty. \quad (16)$$

At the distance of closest approach, the particles are completely reflected; that is, the spin states of the pair partners have no time to change within the times when the partners are at the minimum distance. Whence, we have the second boundary condition

$$\nabla_r G(r = r_0) = 0. \quad (17)$$

A collision superoperator P describes a change in the pair density matrix in an interaction region; according to Eqs. (8) and (9), it is defined as [10]

$$P = -i\int W(r)G(r)d\mathbf{r}. \quad (18)$$

For the diagonal elements of the collision superoperator, Eq. (18) can be represented in a equivalent form, taking into account that

$$-iW(r)_{IK,IK}G(r)_{IK,IK} = D\Delta_r G(r)_{IK,IK}$$

according to Eq. (15). Therefore, for the diagonal elements of P, Eq. (18) can be rewritten as [9]

$$P_{IK,IK} = D\int \Delta G_{IK,IK}(r)d\mathbf{r}. \quad (19)$$

Using the divergence theorem and boundary condition (17), we can represent Eq. (19) as the diffusion flux of the pair-correlation gradient $D\nabla_r G_{IK,IK}(r)$ through a sphere of radius $r \longrightarrow \infty$:

$$P_{IK,IK} = 4\pi D \lim_{r\to\infty} r^2 \frac{\partial G_{IK,IK}(r)}{\partial r}. \quad (20)$$

For applied computations, the form of Eq. (20) is more convenient. However, to calculate the off-diagonal elements of P, it is better to use Eq. (18).

Using the considerations given above and the model developed (see Eqs. (14), (15)), we rewrite Eq. (14) for an arbitrary element $G_{IK,MN}(r)$ in terms of the basis of the eigenfunctions of the total spin squared:

$$\frac{1}{r^2}\frac{\partial}{\partial t}\left(r^2\frac{\partial G_{IK,MN}(r)}{\partial r}\right) + i(a_{IK,MN}e^{-\kappa r} + b_{IK,MN})G_{IK,MN}(r) = 0, \quad (21)$$

where

$$a_{IK,MN} = \delta_{IM}\delta_{KN}\frac{J_0 e^{\kappa r_0}}{2D}(S_K(S_K + 1) - S_I(S_I + 1))$$

and

$$b_{IK,MN} = \frac{\omega_0}{D}(-S_{ZI} + S_{ZK} + S_{ZM} - S_{ZN}).$$

In the expression for $a_{IK,MN}$, S_K and S_I are the total spins of the pair characterizing the two-particle K and I states, respectively; in the expression for $b_{IK,MN}$, S_{ZI}, S_{ZK}, S_{ZM}, and S_{ZN} are the projections of the total spin on the direction of the applied magnetic field in the two-particle I, K, M, and N states, respectively.

The terms containing a are caused by the spin–spin interaction of pair partners, and the terms containing b describe the spin dynamics in an applied magnetic field. At different values of the a and b coefficients, we

obtain different types of solutions to Eq. (21). If $a_{IK,MN} \neq 0$ and $b_{IK,MN} \neq 0$, Eq. (21) only has a trivial zero solution under boundary conditions (16) and (17). If $a_{IK,MN} = 0$ and $b_{IK,MN} = 0$, the general solution to Eq. (21) has the form $G_{IK,MN}(r) = C_{IK,MN}/r + F_{IK,MN}$. Using these boundary conditions, we have $C_{IK,MN} = 0$ and $F_{IK,MN} = \delta_{IM}\delta_{KN}$. However, according to Eq. (18), the elements

$$G_{IK,MN}(r) = \delta_{IM}\delta_{KN}$$

do not contribute to the spin-exchange efficiency. This is rather obvious, since the situation $a_{IK,MN} = 0$ means the absence of spin–spin interaction between the partners of the colliding pair of particles. The $G(r)$ elements for which $b_{IK,MN} \neq 0$ and $a_{IK,MN} = 0$ also should not contribute to the spin-exchange efficiency, since spin exchange cannot occur in the absence of spin–spin interaction between the partners. Thus, in the situation under study, spin exchange is only contributed by the G elements for which $a_{IK,MN} \neq 0$ and $b_{IK,MN} = 0$; that is, only the diagonal $G_{IK,IK}(r)$ elements have nontrivial solutions (see Eq. (15)). The general solution to Eq. (15) is expressed in terms of the zeroth-order Bessel and Neumann functions:

$$G(r)_{IK,IK} = \frac{C_1}{r}J_0\left(\sqrt{\frac{4a_{IK,IK}}{\kappa^2}i}e^{-\kappa r/2}\right) + \frac{C_2}{r}N_0\left(\sqrt{\frac{4a_{IK,IK}}{\kappa^2}i}e^{-\kappa r/2}\right). \quad (22)$$

Using boundary conditions (16) and (17), we have

$$C_{1IK,IK} = \frac{\pi}{\kappa}\frac{\kappa r_0 e^{-\kappa r_0/2}\sqrt{\frac{4a_{IK,IK}}{\kappa^2}i}N_1\left(e^{-\kappa r_0/2}\sqrt{\frac{4a_{IK,IK}}{\kappa^2}i}\right) - 2N_0\left(e^{-\kappa r_0/2}\sqrt{\frac{4a_{IK,IK}}{\kappa^2}i}\right)}{\kappa r_0 e^{-\kappa r_0/2}\sqrt{\frac{4a_{IK,IK}}{\kappa^2}i}J_1\left(e^{-\kappa r_0/2}\sqrt{\frac{4a_{IK,IK}}{\kappa^2}i}\right) - 2J_0\left(e^{-\kappa r_0/2}\sqrt{\frac{4a_{IK,IK}}{\kappa^2}i}\right)}, \quad (23)$$

$$C_2 = -\frac{\pi}{\kappa}.$$

Substituting Eqs. (22) and (23) into Eq. (20), we obtain an expression for the diagonal elements of the collision superoperator:

$$P_{IK,IK} = 4\pi D\left\{r_0 + \frac{1}{\kappa}\left[\ln\left(\frac{|J_0(S_K(S_K+1) - S_I(S_I+1))|}{2D\kappa^2}\right) + 2C \pm i\frac{\pi}{2}\right] - C_{1IK,IK}\right\}, \quad (24)$$

where $C \approx 0.57721566\ldots$ is the Euler–Mascheroni constant;

$$C_{1IK,IK} = \frac{\pi}{\kappa}\frac{N_1(z_0)z_0\kappa r_0 - 2N_0(z_0)}{J_1(z_0)z_0\kappa r_0 - 2J_0(z_0)},$$

$$z_0 = 2\sqrt{\frac{i|J_0(S_I(S_I+1) - S_K(S_K+1))|}{2D\kappa^2}},$$

and $J_k(x)$ and $N_k(x)$ are the kth-order Bessel and Neumann functions, respectively. The sign of imaginary unit in Eq. (24) is equal to the sign of the expression $J_0(S_K(S_K + 1) - S_I(S_I + 1))/2$. As follows from Eq. (24), the elements of the collision superoperator depend only on the difference in the squares of the total spins of the two-particle states and do not depend on the total-spin projection on any preferred direction. Moreover, the exchange integral and the total spin of the pair appear only in the combination $J_0(S_K(S_K + 1) - S_I(S_I + 1))/2$. This combination gives the difference in the energies of the pair states with the total spins S_K and S_I. The off-diagonal elements of the collision superoperator are equal to zero. The diagonal elements of P having I and K indices denoting states with the same total spin are also equal to zero. Therefore, the indices of the nonzero elements of the collision superoperator are taken to be the values of the total spin characterizing a certain state. Substituting the results obtained for the collision superoperator into Eq. (6), we can obtain kinetic equations for one-particle density matrices.

We are planning to use these equations to analyze the manifestation of spin exchange in EPR spectra. In this case, observable values are the transverse components of spin magnetization; therefore, we have to know kinetic equations for the one-particle matrix elements

$\sigma^A_{m_A+1,\,m_A}$ and $\sigma^B_{m_B+1,\,m_B}$. From Eqs. (6), for the matrix elements $\sigma^A_{m_A+1,\,m_A}$ and $\sigma^B_{m_B+1,\,m_B}$ we obtain the following equations:

$$\begin{aligned}\frac{d\sigma^A_{m_A+1,\,m_A}}{dt} &= i\omega_0^A\sigma^A_{m_A+1,\,m_A} - C_B \\ &\times \sum_{\substack{m_B,\,S,\,S',\,m'_A \\ m''_A,\,m'_B,\,m''_B}} C^{S,\,M+1}_{S^A,\,m_A+1,\,S^B,\,m_B} C^{S',\,M}_{S^A,\,m_A,\,S^B,\,m_B} \\ &\times C^{S,\,M+1}_{S^A,\,m'_A,\,S^B,\,m'_B} C^{S',\,M}_{S^A,\,m''_A,\,S^B,\,m''_B} P_{S,\,S',\,S,\,S'}\sigma^A_{m'_A,\,m''_A}\sigma^B_{m'_B,\,m''_B}, \\ \frac{d\sigma_{m_B+1,\,m_B}}{dt} &= i\omega_0^B\sigma^B_{m_B+1,\,m_B} - C_A \\ &\times \sum_{\substack{m_A,\,S,\,S',\,m'_A \\ m''_A,\,m'_B,\,m''_B}} C^{S,\,M+1}_{S^A,\,m_A,\,S^B,\,m_B+1} C^{S',\,M}_{S^A,\,m_A,\,S^B,\,m_B} \\ &\times C^{S,\,M+1}_{S^A,\,m'_A,\,S^B,\,m'_B} C^{S',\,M}_{S^A,\,m''_A,\,S^B,\,m''_B} P_{S,\,S',\,S,\,S'}\sigma^A_{m'_A,\,m''_A}\sigma^B_{m'_B,\,m''_B},\end{aligned} \tag{25}$$

where $C^{S,\,M}_{S^A,\,m_A,\,S^B,\,m_B}$ are the Clebsch–Gordan coefficients [12].

3. KINETIC EQUATIONS FOR THE TRANSVERSE MAGNETIZATIONS OF *A* AND *B* PARTICLES

In EPR spectroscopic experiments, observable values are the transverse magnetization components

$$\begin{aligned}M^A_- &= \sum_{m_A} C_A\sqrt{(S^A - m_A)(S^A + m_A + 1)}\,\sigma^A_{m_A+1,\,m_A}, \\ M^B_- &= \sum_{m_B} C_B\sqrt{(S^B - m_B)(S^B + m_B + 1)}\,\sigma^B_{m_B+1,\,m_B}.\end{aligned} \tag{26}$$

Assuming that the spins under study are in a near-equilibrium state and linearizing Eq. (25) in small deviations of the one-particle matrices from the diagonal matrices, we obtain kinetic equations for the transverse magnetization components of spin subensembles (see Appendix):

$$\begin{aligned}\frac{\partial M^A_-}{\partial t} &= i\omega_0^A M^A_- - K_1 C_B M^A_- + K_2 C_A M^B_-, \\ \frac{\partial M^B_-}{\partial t} &= i\omega_0^B M^B_- - K_2 C_A M^B_- + K_1 C_B M^A_-.\end{aligned} \tag{27}$$

The spin-exchange rate constants are

$$\begin{aligned}K_1 &= 4\pi D \sum_{S=|S^A-S^B|+1}^{S^A+S^B} \frac{N(S)}{S^A(S^A+1)} l_{\text{eff}}(S), \\ K_2 &= 4\pi D \sum_{S=|S^A-S^B|+1}^{S^A+S^B} \frac{N(S)}{S^B(S^B+1)} l_{\text{eff}}(S),\end{aligned} \tag{28}$$

where

$$N(S) = \frac{(S^2 - (S^A - S^B)^2)((S^A + S^B + 1)^2 - S^2)}{S(2S^A+1)(2S^B+1)},$$

$$l_{\text{eff}}(S) = \left\{\frac{r_0}{2} + \frac{1}{2\kappa}\left[\ln\left(\frac{|J_0|S}{D\kappa^2}\right) + 2C - C_1'(S)\right]\right\},$$

$$C_1'(S) = \pi\text{Re}\left[\frac{N_1(2x_0)x_0\kappa r_0 - N_0(2x_0)}{J_1(2x_0)x_0\kappa r_0 - J_0(2x_0)}\right],$$

$$x_0 = \sqrt{\frac{|J_0|S}{D\kappa^2}}\exp\left\{\frac{i\pi}{4}\right\}, \quad C \approx 0.57721566\ldots.$$

In the expressions for K_1 and K_2, the sums are taken over the transitions in the two-particle system that correspond to a change in S by unity. Each transition is characterized by a specific S parameter due to the nonequidistance of a spectrum. The dependences of the spin-exchange rate constants on the molecular-kinetic parameters are involved in the corresponding $l_{\text{eff}}(S)$ sets. A set of weight multipliers $N(S)/(S^{A/B}(S^{A/B}+1))$ has a purely statistical nature, since it is obtained via the convolution of the products of the Clebsch–Gordan coefficients in the kinetic equations for the transverse magnetizations over all indices (see Appendix).

Equations (28) demonstrate that the K_1 and K_2 coefficients are related as the partner spins squared, i.e., as the magnetic moments of the colliding partners. Therefore, the spin exchange caused by the collisions of paramagnetic particles *A* and *B* broadens the EPR lines of these particles differently. According to Eqs. (28), the exchange-induced broadening of the EPR lines of particles *A* and *B* satisfies the relation

$$\frac{\Delta\omega_{1/2}(S^A)}{\Delta\omega_{1/2}(S^B)} = \frac{C_B S^B(S^B+1)}{C_A S^A(S^A+1)}. \tag{29}$$

Note that a relation similar to Eq. (29) was also obtained in [1] for the exchange-induced broadening of EPR lines in the framework of the sudden collision model applied for the collisions of spin-1/2 paramagnetic particles with particles having arbitrary spins. For the case of collisions of particles with the same spins

(1/2 or 1), Eqs. (28) and (29) reproduce the results obtained in [6] and [7].

We now give explicit expressions for spin-exchange rate constants in certain particular cases. For spin exchange between radicals (S_A) and triplets (S_B), we have

$$K_1 = \frac{128}{27}\pi D l_{\text{eff}}\left(\frac{3}{2}\right), \quad K_2 = \frac{16}{9}\pi D l_{\text{eff}}\left(\frac{3}{2}\right). \tag{30}$$

For spin exchange between two different triplet states, we have

$$K_1 = K_2 = \frac{4}{9}\pi D(4l_{\text{eff}}(1) + 5l_{\text{eff}}(2)). \tag{31}$$

For spin exchange between spin-1/2 particles and those with spin S^B, we obtain

$$\begin{aligned} K_1 &= \frac{32}{3}\pi D\frac{S^B(S^B+1)}{(S^B+1/2)(2S^B+1)}l_{\text{eff}}\left(S^B+\frac{1}{2}\right), \\ K_2 &= \frac{8\pi D}{(S^B+1/2)(2S^B+1)}l_{\text{eff}}\left(S^B+\frac{1}{2}\right). \end{aligned} \tag{32}$$

The spin-exchange rate constants (28) can be rewritten as

$$K_1 = 4\pi D R^A_{\text{eff}}, \quad K_2 = 4\pi D R^B_{\text{eff}}, \tag{33}$$

where R_{eff} is the effective spin-exchange radius.

The effective spin-exchange radius depends on the interaction parameters J_0 and κ, the interdiffusion coefficient D of colliding particles, and the distance r_0 of closest approach. Depending on these parameters, the effective radius can vary in a wide range. For comparison, we note that, in the sudden collision model, these radii are expressed by the formulas

$$\begin{aligned} R^A_{\text{eff}} &= \frac{2}{3}\frac{r_0 S^B(S^B+1)J_0^2\tau_c^2}{1+(1/2+S^B)^2J_0^2\tau_c^2}, \\ R^B_{\text{eff}} &= \frac{r_0}{2}\frac{J_0^2\tau_c^2}{1+(1/2+S^B)^2J_0^2\tau_c^2} \end{aligned} \tag{34}$$

for the spin exchange between paramagnetic particles with a spin $S_A = 1/2$ and particles with an arbitrary spin S_B [1]. According to these formulas, the effective radius cannot exceed $r_0/2$. This result is caused by the fact that the sudden collision model does not take into account the extended character of spin–spin interaction. In the situation of the diffusion passage of the particles through an exchange-interaction region, which is considered in this work, the following expressions for effective radii can be derived from Eqs. (28):

$$\begin{aligned} R^A_{\text{eff}} &= \sum_{S=|S^A-S^B|+1}^{S^A+S^B}\frac{N(S)}{S^A(S^A+1)} \\ &\times\left\{\frac{r_0}{2}+\frac{1}{2\kappa}\left[\ln\left(\frac{|J_0|S}{D\kappa^2}\right)+2C-C_1'(S)\right]\right\}, \\ R^B_{\text{eff}} &= \sum_{S=|S^a-S^B|+1}^{S^A+S^B}\frac{N(S)}{S^B(S^B+1)} \\ &\times\left\{\frac{r_0}{2}+\frac{1}{2\kappa}\left[\ln\left(\frac{|J_0|S}{(D\kappa^2)}+2C-C_1'(S)\right)\right]\right\}. \end{aligned} \tag{35}$$

These expressions for effective radii can be substantially simplified in the limiting cases. At $\sqrt{|J_0|S/D\kappa^2} < 1$, we obtain

$$l_{\text{eff}}(S) \approx \frac{J_0^2S^2}{8D^2\kappa^5}(5+10\kappa r_0+10(\kappa r_0)^2+4(\kappa r_0)^3), \tag{36}$$

$$\begin{aligned} R^{A/B}_{\text{eff}} &\approx \frac{J_0^2S^{B/A}(S^{B/A}+1)}{6D^2\kappa^5} \\ &\times(5+10\kappa r_0+10(\kappa r_0)^2+4(\kappa r_0)^3). \end{aligned} \tag{37}$$

Two limiting cases are possible here, depending on the relation between r_0 and $1/\kappa$. If the exchange integral decreases slowly with increasing interparticle distance, i.e., if $r_0 < 1/\kappa$, we have

$$\begin{aligned} l_{\text{eff}}(S) &\approx \frac{1}{2\kappa}\frac{5J_0^2S^2}{4D^2\kappa^4}, \\ R^{A/B}_{\text{eff}} &\approx \frac{5J_0^2S^{B/A}(S^{B/A}+1)}{6D^2\kappa^5}. \end{aligned} \tag{38}$$

In this case, as could be expected, the effective spin-exchange radius is virtually independent of the minimum distance r_0 between the centers of the colliding particles, and the characteristic meeting time is

$$\tau_c \approx \frac{1}{D\kappa^2}. \tag{39}$$

In other words, this is the time of diffusion passage through a region of width $1/\kappa$.

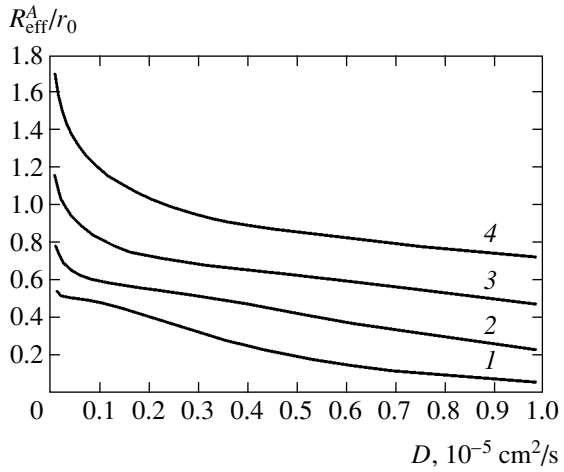


Fig. 1. Dependence of R^A_{eff}/r_0 (see Eqs. (35)) on the interdiffusion coefficient D. Different curves are plotted at different rates of decrease κ of the exchange integral: κ = (*1*) 2×10^8, (*2*) 1×10^8, (*3*) 0.6×10^8, and (*4*) 0.4×10^8 cm^{-1}. The calculations were performed at $S^A = 1/2$, $S^B = 1$, $r_0 = 5\times10^{-8}$ cm, $D = 10^{-7}$–10^{-5} cm^2/s, and $J_0 = 10^{10}$ rad/s.

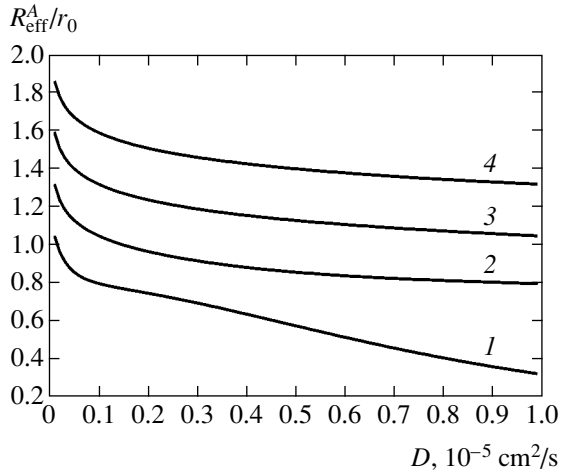


Fig. 2. Dependence of R^A_{eff}/r_0 (see Eqs. (35)) on the interdiffusion coefficient D. Different curves are plotted at different values of the exchange integral J_0: J_0 = (*1*) 10^{10}, (*2*) 10^{11}, (*3*) 10^{12}, and (*4*) 10^{13} rad/s. The calculations were performed at $S^A = 1/2$, $S^B = 1$, $r_0 = 5\times10^{-8}$ cm, $D = 10^{-7}$–10^{-5} cm^2/s, and $\kappa = 10^8$ cm^{-1}.

If the exchange integral decreases rapidly, i.e., if $r_0 > 1/\kappa$, we have

$$l_{\text{eff}}(S) \approx \frac{J_0^2 S^2 r_0^3}{2D^2\kappa^2},$$
$$R^{A/B}_{\text{eff}} \approx \frac{2J_0^2 r_0^3 S^{B/A}(S^{B/A}+1)}{3D^2\kappa^2}. \tag{40}$$

In this case, spin exchange occurs in a sufficiently thin layer (of thickness $1/\kappa$), and the characteristic meeting time is

$$\tau_c \approx \frac{r_0}{D\kappa}. \tag{41}$$

In other words, this is the time of diffusion passage through a thin layer located between spheres with radii r_0 and $r_0 + 1/\kappa$. As would be expected, at small $\sqrt{|J_0 S/D\kappa^2|}$, the spin-exchange rate constants are proportional to $(J_0)^2$ and the $l_{\text{eff}}(S)$ components are proportional to $(J_0 S)^2$.

At $\sqrt{|J_0 S/D\kappa^2|} \gg 1$, we have $C_1'(S) \longrightarrow 0$, and the effective spin-exchange radius depends logarithmically on the exchange integral:

$$l_{\text{eff}}(S) \approx \left\{\frac{r_0}{2} + \frac{1}{2\kappa}\left[\ln\left(\frac{|J_0|S}{D\kappa^2}\right) + 2C\right]\right\},$$
$$R^{A/B}_{\text{eff}} \approx \frac{1}{S^{A/B}(S^{A/B}+1)}$$
$$\times\left\{\left(\frac{r_0}{2} + \frac{1}{2\kappa}\left[\ln\left(\frac{|J_0|}{D\kappa^2}\right) + 2C\right]\right)\right.$$
$$\left.\times L_1(S^A, S^B) + L_2(S^A, S^B)\right\}, \tag{42}$$
$$L_1(S^A, S^B) = \sum_{S=|S^A-S^B|+1}^{S^A+S^B} N(S),$$
$$L_2(S^A, S^B) = \sum_{S=|S^A-S^B|+1}^{S^A+S^B} \frac{N(S)\ln S}{2\kappa}.$$

In this case, the characteristic meeting time is specified by the time of diffusion passage through a region of radius $1/\kappa$ (see Eq. (39)). Since the exchange integral changes exponentially, we should expect a logarithmic dependence of the effective spin-exchange radius on the interaction intensity and the diffusion coefficient at high values of J_0, and a sufficiently small rate of its decrease with increasing distance.

To illustrate the results obtained, we show some examples of the dependences of the effective radius on the molecular-kinetic and magnetic-resonance parameters (Figs. 1, 2). The parameters for these calculations were chosen as follows. At room temperature, the diffusion coefficient of molecules in water is about 10^{-5} cm^2/s. The viscosity inside micellae is two orders

of magnitude higher than the water viscosity; therefore, the diffusion coefficient of paramagnetic particles in micellae should be on the order of 10^{-7} cm^2/s. For interatomic interaction, the exchange integral decreases e times within a distance of about 0.3×10^{-8} cm, which corresponds to the slope of $\kappa = 3.3 \times 10^8$ 1/cm. In molecules, unpaired electrons are delocalized and the exchange interaction can be affected by ligands or chains of chemical bonds; therefore, the rate of decrease of the exchange integral can be smaller than the rate given for interatomic interaction. For atoms, the exchange integral at the distance of their van der Waals radii is about 10^{12} rad/s [1].

An analysis of Figs. 1 and 2 reveals the following features. The effective spin-exchange radius decreases with increasing rate of decrease of the exchange integral with increasing distance between colliding particles. The rate of decrease of the exchange integral affects the volume of the effective spin–spin interaction region. The higher the rate of decrease of the exchange integral, the smaller this volume and the shorter the interaction time of colliding particles. As a result, the effective spin-exchange radius should decrease as the rate of decrease of the exchange integral increases. It is this behavior that is observed in Fig. 1.

The effective spin-exchange radius can substantially exceed the minimum distance r_0 between the centers of the colliding particles, which can be explained by the fact that spin–spin interaction also changes the spin states at distances longer than r_0. The effective spin-exchange radius decreases with increasing diffusion coefficient. This behavior should be expected, since the time within which pair partners are in an interaction region decreases with their increasing mobility. The effective spin-exchange radius increases with the exchange integral J_0 (Fig. 2).

4. CONCLUSIONS

The spin-exchange rate constants K_1 and K_2 can be determined by EPR spectroscopy from the broadening of EPR lines or from the exchange-induced narrowing of an EPR spectrum [1]. To interpret experimental data, one has to calculate the effective spin-exchange radii. In this work, we calculated the effective radii of bimolecular spin exchange for particles with arbitrary spins, as applied to nonviscous liquids. The calculations were performed with allowance for the extended character of exchange interaction on the assumption of diffusion particle motion. The calculation algorithm used here can also be applied to viscous liquids. In this case, the spin Hamiltonian of a system must also take into account anisotropic terms. As a result, linked parabolic equations for the elements of the $G_{IK,MN}(r)$ operator are obtained, and they can be solved only numerically.

Thus, the spin-exchange rate constants have the form

$$K = 4\pi D R_{\text{eff}}(S_A, S_B, J_0, \kappa, D)C \qquad (43)$$

and can be determined from EPR experiments. At least one of the unknown parameters can be found from a comparison of experimental data with theoretical calculations. For example, the concentrations of paramagnetic particles can be determined from a comparison of experiment with theory if the magnetic-resonance parameters, the spins of colliding particles, the exchange integral at the collision radius, the rate of decrease of the exchange integral, the collision radius of the particles, and their interdiffusion coefficient are known from independent sources. This possibility can be very important in studying molecule collisions in complex systems, such as biological objects (e.g., see [1]). The study of the dependence of the spin-exchange rate constant on the diffusion coefficient can be rather fruitful (see Figs. 1, 2). The diffusion coefficient can vary when the temperature or solvent is changed. By comparing the theoretical results obtained with experimental data on spin exchange, we can determine the exchange integral at the maximum approach of colliding particles and the rate of decrease of the exchange integral.

ACKNOWLEDGMENTS

We thank V.N. Lisin for many fruitful discussions.

This work was supported by the Russian Foundation for Basic Research (project no. 03-03-32602), AN RT (project no 06-6.3.1/2004) (FP), CRDF (project no. BRHE REC-007), NIOKR RT (project no. 06-6.3-349), and the Program for Supporting of Leading Scientific Schools (grant no. NSh-1905.2003.2).

APPENDIX

We derive kinetic equations for the transverse magnetizations of the subensembles of spins A and B for one-particle density matrices.

Equations (6) written for the matrix elements of one-particle density matrices that enter into magnetization have the form of Eqs. (25). We consider systems that are close to equilibrium at high temperatures, so that the quadratic terms in Eqs. (25) can be linearized as follows. The diagonal elements of the one-particle density matrices are taken to be equal to $1/(2S^A + 1)$ plus first-order infinitesimals in the σ^A matrix and equal to $1/(2S^B + 1)$ plus first-order infinitesimals in the σ^B matrix. The off-diagonal elements are assumed to be first-order infinitesimals. If we retain only terms up to

the first order of smallness inclusively in Eqs. (25), then the first-order infinitesimals

$$C^{S, M+1}_{S^A, m_A+1; S^B, m_B} C^{S', M}_{S^A, m_A; S^B, m_B} C^{S, M+1}_{S^A, m'_A+1; S^B, m'_B} \times C^{S', M}_{S^A, m'_A; S^B, m'_B} P_{S, S'; S, S'} \frac{\sigma^A_{m'_A+1, m'_A}}{2S^B+1},$$

$$C^{S, M+1}_{S^A, m_A+1; S^B, m_B} C^{S', M}_{S^A, m_A; S^B, m_B} C^{S, M+1}_{S^A, m'_A; S^B, m'_B+1} \times C^{S', M}_{S^A, m'_A; S^B, m'_B} P_{S, S'; S, S'} \frac{\sigma^B_{m'_B+1, m'_B}}{2S^A+1}$$

would correspond to these terms for the first equation in Eqs. (25), and the first-order infinitesimals for the second equation in Eqs. (25) would have the form

$$C^{S, M+1}_{S^A, m_A; S^B, m_B+1} C^{S', M}_{S^A, m_A; S^B, m_B} C^{S, M+1}_{S^A, m'_A+1; S^B, m'_B} \times C^{S', M}_{S^A, m'_A; S^B, m'_B} P_{S, S'; S, S'} \frac{\sigma^A_{m'_A+1, m'_A}}{2S^B+1},$$

$$C^{S, M+1}_{S^A, m_A; S^B, m_B+1} C^{S', M}_{S^A, m_A; S^B, m_B} C^{S, M+1}_{S^A, m'_A; S^B, m'_B+1} \times C^{S', M}_{S^A, m'_A; S^B, m'_B} P_{S, S'; S, S'} \frac{\sigma^B_{m'_B+1, m'_B}}{2S^A+1}.$$

The expressions for the zeroth-order terms contain the products of the Clebsch–Gordan coefficients that have the same sums of the projections of spins A and B and the projections of the total spin differing by unity. According to the definition of the Clebsch–Gordan coefficients, this means that these terms are identically zero. Thus, the linearized equations form a closed system for the matrix elements entering into the transverse magnetizations of subensembles A and B.

We now write kinetic equations for the transverse magnetizations of subensembles A and B (convolution over all indices is implied in the right-hand sides of these equations):

$$\frac{\partial M^A_-}{\partial t} = C_A \sqrt{(S^A - m_A)(S^A + m_A + 1)} \frac{d\sigma^A_{m_A+1, m_A}}{dt} = i\omega^A_0 M^A_- - C_A C_B \sqrt{(S^A - m_A)(S^A + m_A + 1)} \times \ldots$$
$$\times \left(C^{S, M+1}_{S^A, m_A+1; S^B, m_B} C^{S', M}_{S^A, m_A; S^B, m_B} C^{S, M+1}_{S^A, m_A+1; S^B, m'_B} C^{S', M}_{S^A, m_A; S^B, m'_B} P_{S, S'; S, S'} \frac{\sigma^A_{m'_A+1, m'_A}}{2S^B+1} + \ldots \right.$$
$$\left. + C^{S, M+1}_{S^A, m_A+1; S^B, m_B} C^{S', M}_{S^A, m_A; S^B, m_B} C^{S, M+1}_{S^A, m_A; S^B, m'_B+1} C^{S', M}_{S^A, m_A; S^B, m'_B} P_{S, S'; S, S'} \frac{\sigma^B_{m'_B+1, m'_B}}{2S^A+1} \right),$$

$$\frac{\partial M^B_-}{\partial t} = C_B \sqrt{(S^B - m_B)(S^B + m_B + 1)} \frac{d\sigma^B_{m_B+1, m_B}}{dt} = i\omega^B_0 M^B_- - C_A C_B \sqrt{(S^B - m_B)(S^B + m_B + 1)} \times \ldots$$
$$\times \left(C^{S, M+1}_{S^A, m_A; S^B, m_B+1} C^{S', M}_{S^A, m_A; S^B, m_B} C^{S, M+1}_{S^A, m_A+1; S^B, m'_B+1} C^{S', M}_{S^A, m_A; S^B, m'_B} P_{S, S'; S, S'} \frac{\sigma^A_{m'_B+1, m'_B}}{2S^A+1} + \ldots \right.$$
$$\left. + C^{S, M+1}_{S^A, m_A; S^B, m_B+1} C^{S', M}_{S^A, m_A; S^B, m_B} C^{S, M+1}_{S^A, m_A+1; S^B, m'_B} C^{S', M}_{S^A, m_A; S^B, m'_B} P_{S, S'; S, S'} \frac{\sigma^A_{m'_A+1, m'_A}}{2S^B+1} \right).$$

Using the well-known properties of the Clebsch–Gordan coefficients [12], we find that, after convolution over m_A and m_B, the terms with the S and S' indices that differ by less than unity are retained in the equations obtained. The terms with the same S and S' give no contribution, since the corresponding $P_{S, S; S, S}$ multipliers are equal to zero. Note that the convolution over these projections does not require the linearization of the pair density matrix; this requirement appears for convolution over the other indices. Thus, the transverse magnetization transfer rate constants are only contributed by the diagonal $P_{S1, S2; S1, S2}$ elements with S_1 and S_2 differing by unity, irrespective of whether the pair density matrix was linearized or not. This circumstance is related to the fact that the transverse magnetizations are irreducible first-order tensor operators; in the basis of the total spin of a two-spin system, they have nonzero matrix elements between the states whose total spins differ by not more than unity (the triangle rule).

By making allowance for this circumstance and using the well-known relations for the vector summation coefficients [12], we convolute the right-hand sides

of these equations over all projections of the angular momenta of the subsystems. As a result, we have

$$\frac{\partial M_-^A}{\partial t} = i\omega_0^A M_-^A - \left[\frac{C_B M_-^A}{S^A(S^A+1)} - \frac{C_A M_-^B}{S^B(S^B+1)}\right]$$
$$\times \sum_{S=|S^A-S^B|+1}^{S^A+S^B} \frac{(S^2-(S^A-S^B)^2)((S^A+S^B+1)^2-S^2)}{4(2S^A+1)(2S^B+1)S}$$
$$\times (P_{S,S-1;S,S-1} + P_{S-1,S;S-1,S}),$$

$$\frac{\partial M_-^B}{\partial t} = i\omega_0^B M_-^B - \left[\frac{C_A M_-^B}{S^B(S^B+1)} - \frac{C_B M_-^A}{S^A(S^A+1)}\right]$$
$$\times \sum_{S=|S^A-S^B|+1}^{S^A+S^B} \frac{(S^2-(S^A-S^B)^2)((S^A+S^B+1)^2-S^2)}{4(2S^A+1)(2S^B+1)S}$$
$$\times (P_{S,S-1;S,S-1} + P_{S-1,S;S-1,S}).$$

We introduce the designations $l_{\text{eff}}(S) = (P_{S,S-1;S,S-1} + P_{S-1,S;S-1,S})/4$ and

$$N(S) = \frac{(S^2-(S^A-S^B)^2)((S^A+S^B+1)^2-S^2)}{S(2S^A+1)(2S^B+1)},$$

and obtain Eqs. (27) and (28).

REFERENCES

1. K. I. Zamaraev, Yu. N. Molin, and K. M. Salikhov, *Spin Exchange* (Nauka, Novosibirsk, 1977; Springer-Verlag, Heidelberg, 1980), p. 243.
2. V. I. Gol'danskiĭ, *Physical Chemistry of Positron and Positronium* (Nauka, Moscow, 1968) [in Russian].
3. A. B. Doktorov, K. M. Salikhov, and Yu. N. Molin, Dokl. Akad. Nauk SSSR **205**, 1385 (1972).
4. A. L. Buchachenko, Usp. Khim. **62**, 1139 (1993).
5. O. I. Gnezdilov, A. E. Mambetov, A. A. Obynochny, *et al.*, Appl. Magn. Reson. **25**, 157 (2003).
6. K. M. Salikhov, Teor. Éksp. Khim. **10**, 310 (1974).
7. R. M. Bayazitov, Thesis (Kazan State Univ., Kazan, 1993).
8. A. B. Doktorov and N. N. Lukzen, Khim. Fiz. **2**, 764 (1983).
9. K. M. Salikhov, Doctoral Dissertation in Physics and Mathematics (Kazan, 1974).
10. A. B. Doktorov, Physica A (Amsterdam) **90**, 109 (1978).
11. N. N. Bogolyubov, *Selected Works on Statistical Physics* (Mosk. Gos. Univ., Moscow, 1979), p. 343 [in Russian].
12. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Moscow, 1975; World Sci., Singapore, 1988).

Translated by K. Shakhlevich

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 892–906.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 1025–1040.
Original Russian Text Copyright © 2005 by Dickmann, Zhilin, Kulakovskiĭ.

ELECTRONIC PROPERTIES OF SOLIDS

Excitonic Representation: Collective Excitation Spectra in the Quantized Hall Regime and Spin Biexciton

S. M. Dickmann, V. M. Zhilin, and D. V. Kulakovskiĭ

Institute of Solid-State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia
e-mail: dickmann@issp.ac.ru

Received December 25, 2004

Abstract—The excitonic representation method for describing collective excitations in the quantized Hall regime makes it possible to simplify analysis of the spectra and to obtain new results in the strong magnetic field limit, when $E_C \ll \hbar\omega_c$ (ω_c is the cyclotron frequency and E_C is the characteristic Coulomb energy). For an integer odd filling factor ν greater than unity (i.e., for $\nu = 3, 5, 7, \ldots$), the spectra of one-cyclotron magnetoplasma excitations are calculated. For unit filling factor, the existence of a spin biexciton (bound state of two spin waves) corresponding to excitation with a spin change ($\delta S = \delta S_z = -2$) is proved. The exact equation determining the ground state of the biexciton is derived in the thermodynamic limit $N_\phi \longrightarrow \infty$ (N_ϕ is the system degeneracy). The exchange energy of this state is lower than for a single spin wave (with $\delta S = \delta S_z = -1$) for the same value of the 2D wavevector **q**. In the limit $q \longrightarrow \infty$ corresponding to the decay of a biexciton into a pair of quasiparticles one of which is a trion with a spin of –3/2, the energy is found to be lower than the energy $(e^2/\varepsilon l_B)\sqrt{\pi/2}$ required for exciting an electron–hole pair in the strictly 2D case (l_B is the magnetic length and ε is the dielectric constant), although this energy is higher than another "classical" result $(e^2/2\varepsilon l_B)\sqrt{\pi/2}$, corresponding to the excitation of a skyrmion–antiskyrmion pair ($|\delta S| = |\delta S_z| \gg 1$). The solution of the exact equation gives the trion binding energy and the activation gap for quasiparticles whose excitation corresponds to a change in the total spin by $\delta S = \delta S_z = -3$. The energy of a spin biexciton is calculated for values of the wavevector such that $ql_B \gg 1$. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The Coulomb interaction is one of the main factors determining the behavior of the electron system in the regime of the quantum Hall effect [1]. If a 2D electron gas is an insulator, this interaction determines or substantially renormalizes the activation gap in the excitation of dissipative current. A typical example is the fractional quantum Hall effect, in which the gap for filling factors $\nu = 1/3, 1/5, \ldots$ is completely determined by the interaction [1, 2]. At the same time, the Zeeman gap ($\nu = 1, 3, 5, \ldots$; the total spin is equal to the number of electrons at the upper Landau level, $S = N_\phi/2$, where $N_\phi = L^2/2\pi l_B^2$ and $L \times L$ is the area of the system) and the cyclotron gap ($\nu = 2. 4. \ldots$; the system is not polarized, $S = 0$), which already exist in the absence of correlations, are found to be much larger in actual practice due to the interaction even for an integer filling factor.

In all quantum-Hall dielectrics of this type, low-lying excitations are zero-charge (neutral) collective excitations of the type of electron density waves [3], excitons [2, 4–10], or exciton complexes [6, 10–14]. The emergence of Fermi quasiparticles in the system [15] is the result of the limiting transition to infinitely large values of the wavevector in the spectra of these excitations [4–7]. Thus, generation of charge carriers (and, hence, the existence of an activation gap) is closely associated with the lower part of the spectrum of a 2D electron gas.

It should be noted that this problem remains topical. The analytic solution turns out to be possible only for three simplest excitations even for an integer filling factor and even in the first approximation in the small parameter $r_C = E_C/\hbar\omega_c$ ($E_C = \alpha e^2/\varepsilon l_B$, where α is the form factor determined by the finite electron gas width; $\alpha < 1$ in all cases). These excitations are a cyclotron magnetoplasma wave corresponding to an upward electron transition to a Landau level ($\delta n = 1$) without a change in spin ($\delta S = \delta S_z = 0$ [5–7], a spin wave with an odd filling factor ($\delta S = \delta S_z = -1$ and $\delta n = 0$) [4–7] and a spin-cyclotron exciton with even filling factor ($\delta S = \delta S_z = -1$ and $\delta n = 1$) [7]. In the remaining cases, Coulomb corrections to the exciton spectra were calculated not exactly, but only in the framework of the reduced Coulomb interaction model, which is a version of the mean field approximation as applied to exciton excitations [7–9].

Here, we will use the excitonic representation for calculating Coulomb corrections in the simple and familiar cases, as well as in more complex cases. We consider a 2D electron gas with an integer filling factor ν and confine our analysis to a first-order approxima-

tion in parameter r_C. Such a system is an insulator or a quantum-Hall ferromagnet for an odd ν. (In the latter case, all electron spins in the ground state at the half-filled upper Landau level are aligned along the magnetic field in view of the fact that the Lande factor in gallium arsenide structures is negative.) The simplest excitations in this system are excitons formed by an electron placed on an unfilled or half-filled Landau level with or without spin flip and an effective hole emerging thereby on the previous level.

The exciton creation operator is defined by the formula [10–13, 16–21]

$$\hat{\mathfrak{D}}^{\dagger}_{ab\mathbf{q}} = \frac{1}{\sqrt{N_\phi}}\sum_p \exp(-iq_x p l_B) b^{\dagger}_{p+q_y l_B/2} a_{p-q_y l_B/2}, \quad \hat{\mathfrak{D}}_{ab\mathbf{q}} \equiv \hat{\mathfrak{D}}^{\dagger}_{ba-\mathbf{q}}, \tag{1.1}$$

where a_p ($a^{\dagger}_p$) and b_p ($b^{\dagger}_p$) are the annihilation (creation) operators for an electron in the states corresponding to the Landau gauge conditions; subscripts a and b denote one-electron states with different numbers n of Landau levels and with spin sublevels σ;

$$a = (n_a, \sigma_a), \tag{1.2}$$

and p is the internal quantum number of the Landau level. The one-electron energy $E_{a(b)}$ corresponding to subscripts a (or b) is given by

$$E_a = \hbar\omega_c(n_a + 1/2) - |g\mu_B B|\sigma_a, \tag{1.3}$$

where g is the Lande factor. In addition to notation (1.2), we will use another notation for sublevels: subscripts $a = n$ and $a = \bar{n}$ correspond to $n = (n, \uparrow)$ and $\bar{n} = (n, \downarrow)$.

The basis of one-exciton states is determined with the help of operators (1.1):

$$|ab;\ \mathbf{q}\rangle = \hat{\mathfrak{D}}^{\dagger}_{ab\mathbf{q}}|0\rangle, \tag{1.4}$$

where the ground state $|0\rangle$ is set in the zeroth approximation. For an integer filling factor, this means that states (1.4) exist if sublevel a in the ground state is completely filled, while sublevel b is completely empty. Then states (1.4) are mutually orthogonal and normalized,

$$\langle \mathbf{q}_2;\ b_2 a_2 | a_1 b_1;\ \mathbf{q}_1\rangle = \delta_{a_1, a_2}\delta_{b_1, b_2}\delta_{\mathbf{q}_1, \mathbf{q}_2};$$

here, $\delta_{a,b} \equiv \delta_{n_a, n_b}\delta_{\sigma_a, \sigma_b}$. To estimate the energies $\epsilon(\mathbf{q})$ of one-exciton excitations, we can consider the secular equation

$$\det\left|\{\langle 0|\hat{\mathfrak{D}}_{a'b'\mathbf{q}}[\hat{H}, \hat{\mathfrak{D}}^{\dagger}_{ab\mathbf{q}}]|0\rangle - \epsilon(\mathbf{q})\delta_{a,a'}\delta_{b,b'}\}\right| = 0. \tag{1.5}$$

Generally speaking, this approach does not lead to a correct result since system of equations (1.5) corresponds to the above-mentioned mean-field approximation for excitons. However, for some important cases associated with low-lying excitations (i.e., for specific sets of ab pairs), the total Hamiltonian including the Coulomb interaction can be diagonalized to a first approximation in r_C in the basis of one-exciton states (1.4). In this case, the commutator in Eqs. (1.5) is determined only by one-electron terms in the Hamiltonian and by the above-mentioned reduced interaction, i.e., by the $\hat{H}^{\mathrm{ED}}_{\mathrm{int}}$ part of Coulomb Hamiltonian $\hat{H}_{\mathrm{int}}$, which can be diagonalized in the basis of all possible one-exciton states for arbitrary ab pairs. In this case, we can carry out the substitution $\hat{H}_{\mathrm{int}} \longrightarrow \hat{H}^{\mathrm{ED}}_{\mathrm{int}}$ in Eqs. (1.5). It was mentioned above that this equation leads to a correct result for the spectrum $\epsilon = \epsilon_{ab}(\mathbf{q})$ to a first approximation in r_C for some simplest excitations. It should be noted that, after the substitution of $\hat{H}_{\mathrm{int}}$ for $\hat{H}^{\mathrm{ED}}_{\mathrm{int}}$, not only the Zeeman energy $g\mu_B B\hat{S}_z$ is conserved (since $[\hat{S}_z,\ \hat{H}_{\mathrm{int}}] = [\hat{S}_z,\ \hat{H}^{\mathrm{ED}}_{\mathrm{int}}] \equiv 0$ in all cases), but also the one-electron orbital energy corresponding to the operator

$$\hat{H}_1 = \frac{\hbar\omega_c}{2}\sum_{a,p} a^{\dagger}a_p.$$

In other words, the effect of "Landau level mixing" is ignored in this model since $[\hat{H}_1,\ \hat{H}^{\mathrm{ED}}_{\mathrm{int}}] \equiv 0$ (although $[\hat{H}_1,\ \hat{H}_{\mathrm{int}}] \neq 0$ in the general case). The fact that the result is exact for some sets of ab pairs in the first order in r_C only leads to second- and higher-order corrections in the interaction. Changes in the orbital and Zeeman energies are determined by the values of $\hbar\omega_c\delta n$ and $g\mu_B B\delta S_z$, where quantum numbers δn and δS_z are given by

$$\delta n = n_b - n_a = n'_b - n'_a = \ldots, \quad \delta S_z = \sigma_b - \sigma_a = \sigma'_b - \sigma'_a = \ldots. \tag{1.6}$$

These conditions must hold for all $a'b'$ pairs taken into account in Eqs. (1.5). Other $a'b'$ pairs either do not appear in these equations, or they correspond to the case $n'_b - n'_a \neq \delta n$ and, hence, lead to higher-order corrections in r_C. As a result, we can always write

$$\epsilon(\mathbf{q}) = \hbar\omega_c\delta n - |g\mu_B B|\delta S_z + E_{\delta n, -\delta S_z}(\mathbf{q}), \tag{1.7}$$

where $E_{\delta n, -\delta S_z}(\mathbf{q})$ is the energy associated with the Coulomb interaction.

2. EXCITONIC REPRESENTATION

Henceforth, we will use variables with dimensions of length and wavevector in units of l_B and $1/l_B$, respectively. The total Hamiltonian has the form

$$\hat{H} = g\mu_B B\hat{S}_z + \hat{H}_1 + \hat{H}_{\text{int}}, \tag{2.1}$$

where the Coulomb part can be written as follows:

$$\hat{H}_{\text{int}} = \frac{1}{2N_\phi}\sum_{\substack{p,\,p'\mathbf{q}\\ a,\,b,\,c,\,d}} V_{bdca}(\mathbf{q}) \times \exp[iq_x(p-p')]b^\dagger_{p-q_y}d^\dagger_{p'}c_{p'-q_y}a_p. \tag{2.2}$$

Here, the annihilation operators a_p, b_p, c_p, and d_p correspond to their binary indices (1.2). The factors

$$V_{bdca}(\mathbf{q}) = \frac{e^2V(q)}{2\pi\varepsilon l_B}h_{n_an_b}(-\mathbf{q})h_{n_cn_d}(\mathbf{q})\delta_{\sigma_a,\sigma_b}\delta_{\sigma_c,\sigma_d} \tag{2.3}$$

can be expressed in terms of the dimensionless Fourier component $V(q)$ of the Coulomb potential ($V(q) = 2\pi/q$ in the ideal 2D case) and the functions (see, for example, [6, 7, 9, 13, 17]),

$$h_{kn}(\mathbf{q}) = \sqrt{\frac{k!}{n!}}\exp\left(-\frac{q^2}{4}\right)(q_-)^{n-k}L_k^{n-k}\left(\frac{q^2}{2}\right), \tag{2.4}$$

where $q_- = i(q_x - iq_y)/\sqrt{2}$ and L_n^j are generalized Laguerre polynomials. It should be noted that functions (2.4) are also encountered in the description of the interaction between an exciton and the external field [12, 14, 16–18].

Before considering the excitonic representation of Hamiltonian (2.2), we write the commutation relations for operators (1.1):

$$[\hat{\mathfrak{D}}^\dagger_{cd\mathbf{q}_1}, \hat{\mathfrak{D}}^\dagger_{ab\mathbf{q}_2}] \equiv \frac{1}{\sqrt{N_\phi}} \times\left\{\exp\left[-\frac{i}{2}(\mathbf{q}_1\times\mathbf{q}_2)_z\right]\delta_{b,c}\hat{\mathfrak{D}}^\dagger_{ad\mathbf{q}_1+\mathbf{q}_2} - \exp\left[\frac{i}{2}(\mathbf{q}_1\times\mathbf{q}_2)_z\right]\delta_{a,d}\hat{\mathfrak{D}}^\dagger_{cb\mathbf{q}_1+\mathbf{q}_2}\right\}. \tag{2.5}$$

A part of these relations was given in [19] in a different formulation and for a single Landau level, when a, b, c, and d are only $(0, \uparrow)$ or $(0, \downarrow)$ (see also [12–14, 16–18]). Commutation relations (2.5) thus form a Lie algebra. If sublevel a is filled and b is completely empty in state $|0\rangle$, we have $a_p^\dagger|0\rangle = b_p|0\rangle \equiv 0$. In terms of exciton operators, these equalities correspond to the equalities

$$\hat{A}^\dagger_{\mathbf{q}}|0\rangle = \delta_{0,\mathbf{q}}|0\rangle, \quad \hat{B}^\dagger_{\mathbf{q}}|0\rangle = 0, \tag{2.6}$$

where we introduced the operators of "displacement" within sublevels:

$$\hat{A}^\dagger_{\mathbf{q}} = \frac{1}{\sqrt{N_\phi}}\hat{\mathfrak{D}}^\dagger_{aa\mathbf{q}}, \quad \hat{B}^\dagger_{\mathbf{q}} = \frac{1}{\sqrt{N_\phi}}\hat{\mathfrak{D}}^\dagger_{bb\mathbf{q}}.$$

Using now definition (1.1) and commutation relations (2.5), we can write Hamiltonian (2.2) in the excitonic representation:

$$\hat{H}_{\text{int}} = \frac{1}{2}\sum_{\mathbf{q},\,a,\,b,\,c,\,d} V_{bdca}(\mathbf{q})\hat{\mathfrak{D}}^\dagger_{ab\mathbf{q}}\hat{\mathfrak{D}}^\dagger_{cd-\mathbf{q}} - \sum_{\substack{\mathbf{q},\,a,\,b\\ (n_a\le n_b)}}\left(1-\frac{1}{2}\delta_{a,b}\right)V_{baba}(\mathbf{q})\hat{B}^\dagger_0. \tag{2.7}$$

It should be noted that the one-electron part of Hamiltonian (2.1) in this representation has the form

$$g\mu_B B\hat{S}_z + \hat{H}_1 = \frac{1}{2}|g\mu_B B|\sum_n(\hat{N}_{\bar{n}} - \hat{N}_n) + \hbar\omega_c\sum_n\left(n+\frac{1}{2}\right)(\hat{N}_n + \hat{N}_{\bar{n}}), \tag{2.8}$$

where $\hat{N}_n = \sqrt{N_\phi}\hat{\mathfrak{D}}^\dagger_{nn\mathbf{0}}$ and $\hat{N}_{\bar{n}} = \sqrt{N_\phi}\hat{\mathfrak{D}}^\dagger_{\bar{n}\bar{n}\mathbf{0}}$.

The next step is extraction of the exciton-diagonalizable part $\hat{H}^{\text{ED}}_{\text{int}}$ from expression (2.7). This operator at least should not change the one-electron energy (i.e., it must commute with Hamiltonian (2.8)); consequently, we retain in Eq. (2.7) only the terms for which $n_a + n_c = n_b + n_d$. A part of these terms form an operator for which states (1.4) are eigenstates. This diagonal part is given by

$$\hat{H}^{\text{di}}_{\text{ED}} = \sum_a\hat{H}_a + \sum_{\substack{a,\,b\\ (a\ne b,\,n_a\le n_b)}}\hat{H}_{ab}, \tag{2.9}$$

where

$$\hat{H}_a = \frac{1}{2}\sum_{\mathbf{q}}V_{aaaa}(q)(N_\phi\hat{A}^\dagger_{\mathbf{q}}\hat{A}_{\mathbf{q}} - \hat{A}^\dagger_0), \tag{2.10}$$

$$\hat{H}_{ab} = \sum_{\mathbf{q}}[V_{abba}(q)N_\phi\hat{A}^\dagger_{\mathbf{q}}\hat{B}_{\mathbf{q}} + V_{baba}(q)(\hat{\mathfrak{D}}^\dagger_{ab\mathbf{q}}\hat{\mathfrak{D}}_{ab\mathbf{q}} - \hat{B}^\dagger_0)]. \tag{2.11}$$

To find the contribution to the excitation energy, we must calculate the action of commutator $[\hat{H}_{\rm ED}^{\rm di}, \hat{\mathfrak{D}}_{ab\mathbf{q}}^{\dagger}]$ on the ground state $|0\rangle$ for the chosen ab pair. In doing so, we should not retain in $\hat{H}_{\rm ED}^{\rm di}$ the terms which obviously commute with $\hat{\mathfrak{D}}_{ab\mathbf{q}}^{\dagger}$ in accordance with relation (2.5). Using rules (2.5) and (2.6), we can also easily verify that, although commutator $[\hat{H}_b, \hat{\mathfrak{D}}_{ab\mathbf{q}}^{\dagger}]$ differs from zero, its action on state $|0\rangle$ is always equal to zero. Thus, we can easily prove that

$$\begin{aligned}&[\hat{H}_{\rm ED}^{\rm di}, \hat{\mathfrak{D}}_{ab\mathbf{q}}^{\dagger}]|0\rangle\\ &\equiv \left[\hat{H}_a + \hat{H}_{ab} + \sum_{c\neq a,b}(\hat{H}_{ac} + \hat{H}_{bc}), \hat{\mathfrak{D}}_{ab\mathbf{q}}^{\dagger}\right]|0\rangle \qquad (2.12)\\ &= \mathscr{E}_{ab}(q)\hat{\mathfrak{D}}_{ab\mathbf{q}}^{\dagger}|0\rangle\end{aligned}$$

in all cases. At the same time, if $\nu > 1$ and we consider excitations with $\delta n = n_b - n_a > 0$, the off-diagonal part is also significant in the expression for $\hat{H}_{\rm int}^{\rm ED}$. Acting on state (1.4), this off-diagonal part leads to a linear combination of other one-exciton states $|a'b';\ \mathbf{q}\rangle$, but with the same spin $\delta S_z = \sigma_b - \sigma_a$ and with the same value of δn. Thus, we can write

$$\hat{H}_{\rm ED} = \hat{H}_{\rm ED}^{\rm di} + \sum_{a,b}\hat{H}_{ab}^{\rm off\text{–}di}. \qquad (2.13)$$

In contrast to definition (2.9), summation in this formula is carried out only over ab pairs, for which sublevel a is filled and sublevel b is empty in state $|0\rangle$. The terms constituting the sum (2.13) are as follows:

$$\begin{aligned}\hat{H}_{ab}^{\rm off\text{–}di} = \sum_{\substack{c,d\neq a,b\\ n_a+n_d=n_b+n_c}}\sum_{\mathbf{q}}&[V_{adcb}(q)\hat{\mathfrak{D}}_{cd\mathbf{q}}\hat{\mathfrak{D}}_{ab\mathbf{q}}\\ &+ V_{adbc}(q)\hat{\mathfrak{D}}_{ca\mathbf{q}}^{\dagger}\hat{\mathfrak{D}}_{db\mathbf{q}}].\end{aligned} \qquad (2.14)$$

It can easily be verified that

$$[\hat{H}_{ab}^{\rm off\text{–}di}, \hat{\mathfrak{D}}_{ab\mathbf{q}}^{\dagger}]|0\rangle = \sum_{a',b'\neq a,b}\mathscr{E}_{a'b'}^{(ab)}(q)\hat{\mathfrak{D}}_{a'b'\mathbf{q}}^{\dagger}|0\rangle. \qquad (2.15)$$

Here, summation is carried out over the $a'b'$ pairs, such that $\delta n = n_b - n_a = n_{b'} - n_{a'} = \ldots$, $\delta S_z = \sigma_b - \sigma_a = \sigma_{b'} - \sigma_{a'} = \ldots$. Thus, Eqs. (2.12) and (2.15) form a system of a finite number of equations that determine the eigenstates of Hamiltonian $\hat{H}_{\rm ED}$ and, accordingly, the energies of these states for a given set of quantum numbers δn, δS_z, and $\mathbf{q}$.

The remaining terms of the Coulomb Hamiltonian $\hat{\mathscr{H}}_{\rm int} = \hat{H}_{\rm int} - \hat{H}_{\rm ED}$, with which operator $\hat{\mathfrak{D}}_{ab\mathbf{q}}^{\dagger}$ does not commute, have the form

$$\begin{aligned}\hat{H}'_{ab} = &\sum_{\substack{c,d\neq a,b\\ g\neq a}}\sum_{\mathbf{q}}V_{gacd}(q)\hat{\mathfrak{D}}_{dg\mathbf{q}}^{\dagger}\hat{\mathfrak{D}}_{ac\mathbf{q}}\\ &+ \sum_{\substack{c,d\neq a,b\\ g\neq b}}\sum_{\mathbf{q}}V_{cdbg}(q)\hat{\mathfrak{D}}_{dc\mathbf{q}}^{\dagger}\hat{\mathfrak{D}}_{db\mathbf{q}} + \text{H.c.}\end{aligned} \qquad (2.16)$$

Acting on exciton (1.4), operators (2.16) generally generate "additional" two-exciton states that cannot be reduced to one-exciton states. Nevertheless, a part of terms in sums (2.16) commute with the one-electron Hamiltonian (2.8); in other words, these terms lead to two-exciton states corresponding to the same quantum numbers δn, δS_z, and $\mathbf{q}$.

Even in the framework of the first approximation in r_C, these terms must be taken into account in calculating the excitation energy. Additional two-exciton states emerge in all cases when $\delta n \geq 2$ or even for $\delta n = 1$ if $\delta S_z = -1$.

Let us consider a specific example. In the case of unit filling factor, for the spin-cyclotron excitation ($\delta n = -\delta S_z = 1$; i.e., $a = 0$, $b = \bar{1}$), the terms

$$\hat{H}'_{0\bar{1}} = \sum_{\mathbf{q}}V_{1010}(q)\hat{\mathfrak{D}}_{01\mathbf{q}}^{\dagger}\hat{\mathfrak{D}}_{\bar{0}\bar{1}\mathbf{q}} + \text{H.c.} \qquad (2.17)$$

must be taken into account along with $\hat{H}_{\rm ED} = \hat{H}_a + \hat{H}_{ab}$ (it should be noted that the contributions from other operators appearing in Eq. (2.12), as well as off-diagonal components of $\hat{H}_{ab}^{\rm off\text{–}di}$ in Eqs. (2.15) are reduced to zero).

The excitonic representation is suitable in the case when a 2D electron gas is an insulator (i.e., in the absence of free electrons and holes). This representation indicates a change from the Fermi creation operators, which generate the eigenstates of the ideal gas and are multiply degenerate in parameter p, to the exciton operators, which act on the vacuum state $|0\rangle$ and form the system of basis states diagonalizing the Hamiltonian

$$g\mu_B B\hat{S}_z + \hat{H}_1 + \hat{H}_{\rm int}^{\rm ED}.$$

The latter includes a considerable part of the Coulomb interaction; consequently, for states of the new basis (1.7), which are classified by the natural quantum number q, degeneracy is removed. In some cases, we can ignore the difference between operators $\hat{H}_{\rm int}^{\rm ED}$ and

$\hat{H}_{\text{int}}$ and obtain correct results for the excitation energy in the first order in r_C. In a number of other cases, we can treat the terms $\hat{\mathcal{H}}_{\text{int}}$ as a perturbation [13] and use exciton states as basis vectors.

The excitonic representation has a number of other obvious advantages. First, it is independent of the specific features of the gauge used for single-electron wavefunctions. In a different gauge, the definition of operator (1.1) changes, but commutation relations (2.5) and the representation of the total Hamiltonian in the form (2.7) and (2.8) remain unchanged. Second, the Hamiltonian of the Coulomb interaction from the four-operator expression (2.2) is transformed into the two-operator expression (2.7). Finally, the excitonic representation gives the explicit form of the eigenstates of a many-electron system and makes it possible to easily calculate (using commutation algebra (2.5)) the transition matrix elements that determine the probability of various kinetic processes in an electron gas in accordance with the Fermi golden rule. These transitions can be associated, for example, with electron–phonon, electron–impurity, and other interactions (in this case, the interactions are renormalized into exciton–phonon [12, 17, 18], exciton–impurity [14, 20], exciton–exciton [16], and other interactions).

In special cases, operators (1.1) were used in [10, 11] for studying an abstract two-component Fermi system with a symmetric model of interaction. It subsequently turned out that in actual practice this model corresponds to "intervalley waves" in a 2D semiconductor for $\nu = 1$ in the strong magnetic field approximation [19, 22] or to spin waves in a quantum-Hall ferromagnet [5–7, 12, 14, 16, 21]. Specific many-exciton states associated with the presence of the so-called zero-exciton condensate in the system were also considered in [10–12, 14].

3. SOME RESULTS FOR LOW-LYING EXCITATION SPECTRA

Thus, in the first order in r_C, excitations are characterized by quantum numbers δn, δS_z, and $\mathbf{q}$ and their energies are defined by formula (1.7). Since filling factor ν is also a parameter of the system, we can classify eigenstates using the notation $|n; \delta n, -\delta S_z, \mathbf{q}\rangle$. For an integral filling factor in the framework of the exciton-diagonalizable part of Hamiltonian $\hat{H}_{\text{int}}^{\text{ED}}$, solutions are found with the help of Eqs. (2.12) and (2.15). To calculate energy $E_{\delta n, -\delta S_z}$, we must project these equations onto all possible states $|a'b'; \mathbf{q}\rangle$ corresponding to chosen values of δn and δS_z (1.6) and then find the quantities $\mathscr{E}_{a'b'}^{(ab)}(q)$ and $\mathscr{E}_{ab}(q)$. After this, the problem reduces to solving the secular equation

$$\begin{aligned}\det\big|&(\mathscr{E}_{ab} - E_{\delta n, -\delta S_z})\delta_{a,a'}\delta_{b,b'} \\ &+ \mathscr{E}_{a'b'}^{(ab)}(1 - \delta_{a,a'}\delta_{b,b'})\big| = 0\end{aligned} \tag{3.1}$$

(here, the indices of matrix elements are ab and $a'b'$ pairs, in which sublevels a and a' are occupied, while b and b' are empty). Such an algorithm is conventional and is much simpler in actual practice than the Green function method used in the well-known paper [7]. In fact, for each specific type of states $|\nu; \delta n, -\delta S_z, \mathbf{q}\rangle$, we can easily choose specific operators $\hat{H}_a$, $\hat{H}_b$, $\hat{H}_{ab}$, $\hat{H}_{ac}$, $\hat{H}_{bc}$, and $\hat{H}_{ab}^{\text{off–di}}$, which make contributions to the commutators of Eqs. (2.12) and (2.15).

However, as a matter of fact, this method (as well as the one used in [7]) gives a correct result only for excitations

$$|\nu = 2m + 1; \delta n = 0, -\delta S_z = 1, \mathbf{q}\rangle,$$
$$|2m + 1; 1, 0, \mathbf{q}\rangle,$$
$$|2m; 1, 0, \mathbf{q}\rangle, \quad |2m; 1, 1, \mathbf{q}\rangle$$

(here m is an integer). In this case, the order of Eq. (3.1) is not higher than the second. We will not give here the results for the states

$$|2m + 1; 0, 1, \mathbf{q}\rangle, \quad |1; 1, 0, \mathbf{q}\rangle,$$
$$|2m; 1, 0, \mathbf{q}\rangle, \quad |2m; 1, 1, \mathbf{q}\rangle,$$

since these results were obtained in [5–7].

However, we will demonstrate how to calculate the energy of state $|2m + 1; 1, 0, \mathbf{q}\rangle$ for $m \geq 1$ since this quantity is given incorrectly in [7] (see also [23]) and the correct solution has not been reported to our knowledge. The basis vectors of excitation $|2m + 1; 1, 0, \mathbf{q}\rangle$ are excitons $|\overline{m-1}\,\overline{m}; \mathbf{q}\rangle$ and $|m\,m+1; \mathbf{q}\rangle$. Details of transformation of Eqs. (2.12) and (2.15) are given in Appendix 1. It should be noted here that, if $m < 2$ (i.e., $\nu < 5$), the sum over c of the operators on the left-hand side of Eq. (2.12) makes zero contribution to $\mathscr{E}_{ab}$; however, for all $m \geq 1$, we obtain

$$\begin{aligned}\mathscr{E}_{\overline{m-1}\,\overline{m}}(q) &= E_m \equiv u_m(q) + v_m(q), \\ \mathscr{E}_{m\,m+1}(q) &= E_{m+1},\end{aligned} \tag{3.2}$$

where

$$u_m(q) = \frac{1}{N_\phi}\sum_p V_{m-1mmm-1}(p)\{1-\exp(i\mathbf{p}\times\mathbf{q})_z\}$$
$$= \frac{e^2}{\varepsilon l_B}\int_0^\infty \frac{pdp}{2\pi}V(p)\exp\left(-\frac{p^2}{2}\right)L_{m-1}\left(\frac{p^2}{2}\right) \quad (3.3)$$
$$\times L_m\left(\frac{p^2}{2}\right)[1-J_0(pq)],$$

$$v_m(q) = V_{m-1mm-1m}(q)$$
$$= \frac{e^2q^2V(q)}{4\pi\varepsilon l_B m}\exp\left(-\frac{q^2}{2}\right)\left[L^1_{m-1}\left(\frac{q^2}{2}\right)\right]^2, \quad (3.4)$$

J_0 is the Bessel function, while off-diagonal matrix elements determined from Eq. (2.15) are given by

$$\mathscr{C}^{(\overline{m-1}\bar{m})}_{mm+1} = \mathscr{C}^{(mm+1)}_{\overline{m-1}\bar{m}} = W_m(q),$$

where

$$W_m(q) = V_{mmm-1m+1}(q)$$
$$= \frac{e^2q^2V(q)}{4\pi\varepsilon l_B\sqrt{(m+1)m}}\exp\left(-\frac{q^2}{2}\right) \quad (3.5)$$
$$\times L^1_{m-1}\left(\frac{q^2}{2}\right)L^1_m\left(\frac{q^2}{2}\right)$$

(these off-diagonal terms were obtained in [7, 23]). Consequently, the energy $\epsilon = \hbar\omega_c + E^\pm_{1,0}(q)$ of both modes is determined by the following Coulomb corrections:

$$E^\pm_{1,0}(q) = \frac{E_m(q)+E_{m+1}(q)}{2}$$
$$\pm\sqrt{\left[\frac{E_m(q)-E_{m+1}(q)}{2}\right]^2+[W_m(q)]^2}. \quad (3.6)$$

Figure 1 shows the results of calculations. It is noteworthy that the Coulomb energy of both modes is positive for all values of q. Exactly the same property is inherent in the corrections calculated in the first order in the interaction for other spinless excitations [5–7]. It should be noted that, conversely, the second-order correction in r_C is negative in most cases. In particular, for $q = 0$, it is this correction that determines the negative exchange shift for the spin–cyclotron mode $|2m; 1, 1, \mathbf{q} = 0\rangle$ [24].

More complex excitations cannot be calculated in the exciton basis (1.7). In this case, we must extend the basis and go beyond the scope of the reduced model

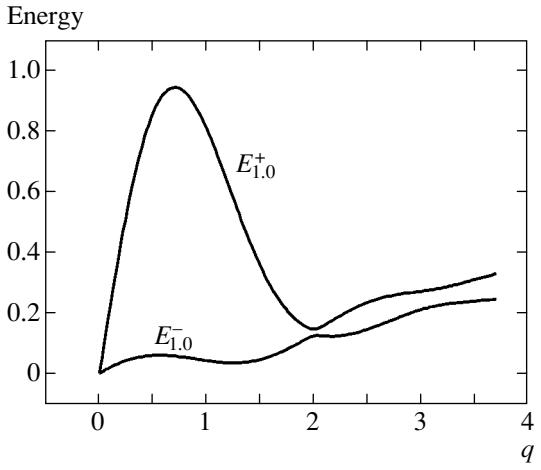


Fig. 1. Dispersion of magnetoplasma modes $E^+_{1,0}(q)$ and $E^-_{1,0}(q)$ vs. the wavevector plotted in units of $1/l_B$. Calculations are made using formula (3.6) for filling factor $\nu = 3$ ($m = 1$). The energy in units of $e^2/\varepsilon l_B$ is plotted along the vertical axis.

based on Hamiltonian $\hat{H}^{\mathrm{ED}}_{\mathrm{int}}$ to obtain the solution. For example, for an excitation of the type $|\nu = 1; 2, 0, \mathbf{q}\rangle$, exciton $|02; \mathbf{q}\rangle$ exhibits quantum fluctuation to the two-exciton state $\hat{\mathfrak{D}}^\dagger_{01\mathbf{q}-\mathbf{p}}\hat{\mathfrak{D}}^\dagger_{01\mathbf{p}}|0\rangle$ (orthogonal to $|02; \mathbf{q}\rangle$) due to the presence of $\hat{H}'_{02} = \sum_{\mathbf{p}} V_{1120}(\mathbf{p})\hat{\mathfrak{D}}^\dagger_{01\mathbf{p}}\hat{\mathfrak{D}}_{12\mathbf{p}}$ in expression (2.16). In this case, we must seek the solution in an extended basis; i.e., we must represent the state in the form

$$|1; 2, 0, \mathbf{q}\rangle = \hat{\mathfrak{D}}^\dagger_{02\mathbf{q}}|0\rangle$$
$$+\frac{1}{\sqrt{N_\phi}}\sum_{\mathbf{s}}\varphi_{20}(\mathbf{s})\hat{\mathfrak{D}}^\dagger_{01\mathbf{q}/2-\mathbf{s}}\hat{\mathfrak{D}}^\dagger_{01\mathbf{q}/2+\mathbf{s}}|0\rangle, \quad (3.7)$$

after which the solution (in the first order in r_C) can be reduced to the following specific mathematical problem: substitute expression (3.7) into the Schrödinger equation,

$$\hat{H}_{\mathrm{int}}|1; 2, 0, \mathbf{q}\rangle = E_{2,0}(q)|1; 2, 0, \mathbf{q}\rangle,$$

where

$$\hat{H}_{\mathrm{int}} = \hat{H}_0 + \hat{H}_1 + \hat{H}_{01} + \hat{H}_{02} + \hat{H}'_{02},$$

and project this equation onto $\hat{\mathfrak{D}}^\dagger_{02\mathbf{q}}|0\rangle$ and $\hat{\mathfrak{D}}^\dagger_{01\mathbf{q}-\mathbf{p}}\hat{\mathfrak{D}}^\dagger_{01\mathbf{p}}|0\rangle$. This gives a system of two integral equations for the function φ_{20}, which contain the sought quantity $E_{2,0}(q)$ as a parameter.

Analogously, the spin–cyclotron excitation for $\nu = 1$ can be presented in the form

$$|1; 1, 1, \mathbf{q}\rangle = \left[\hat{\mathcal{Q}}^{\dagger}_{0\bar{1}\mathbf{q}}|0\rangle + \frac{1}{\sqrt{N_\phi}}\sum_{\mathbf{s}}\varphi_{11}(\mathbf{s})\hat{\mathcal{Q}}^{\dagger}_{0\bar{0}\mathbf{q}/2-\mathbf{s}}\hat{\mathcal{Q}}^{\dagger}_{01\mathbf{q}/2+\mathbf{s}}|0\rangle\right] \tag{3.8}$$

(exciton $|0\bar{1}; \mathbf{q}\rangle$ is mixed with the combination of single-cyclotron magnetoplasmon and a spin wave due to the presence of terms (2.17) in the Hamiltonian). After this, the problem for φ_{11} and $E_{1,1}(q)$ is reduced to solving two integral equations.

We will not give here the cumbersome numerical calculation of the energies of states (3.7) and (3.8). We will consider instead a simple two-exciton excitation $|\nu = 2m + 1; \delta n = 0, -\delta S_z = 2, \mathbf{q}\rangle$, which is not mixed with any one-exciton excitation.

4. TWO-EXCITON STATES

First, we will consider some general properties of two-exciton states. We will discuss only one type of such excitations, namely, the case when the sublevel indices in both exciton operators are the same,

$$|ab; \mathbf{q}, \mathbf{s}\rangle = \frac{1}{2}\hat{\mathcal{Q}}^{\dagger}_{ab\mathbf{q}/2-\mathbf{s}}\hat{\mathcal{Q}}^{\dagger}_{ab\mathbf{q}/2+\mathbf{s}}|0\rangle \tag{4.1}$$

(naturally, $|ab; \mathbf{q}, \mathbf{s}\rangle \equiv |ab; \mathbf{q}, -\mathbf{s}\rangle$). This state is orthogonal to any other two-exciton state $|a'b'; \mathbf{q}', \mathbf{s}'\rangle$ only if $(a, b) \neq (a', b')$ or $\mathbf{q}' \neq \mathbf{q}$. At the same time, orthogonality for $\mathbf{s} \neq \mathbf{s}'$ is determined in a special way. Let us consider the following combination:

$$\sum_{\mathbf{s}} f_{\mathbf{s}}|ab; \mathbf{q}, \mathbf{s}\rangle. \tag{4.2}$$

In fact, physical meaning can be attached only to the projection of this state on some other state. Using commutation algebra (2.5), we can find that

$$\sum_{\mathbf{s}'} f_{\mathbf{s}'}\langle \mathbf{s}, \mathbf{q}; ba|ab; \mathbf{q}', \mathbf{s}'\rangle \equiv \delta_{\mathbf{q},\mathbf{q}'} f^{(a)}_{\mathbf{s}}. \tag{4.3}$$

Here, we define the "antisymmetrization" transformation in the following way:

$$f^{(a)}_{\mathbf{s}} = \frac{1}{2}\left(f_{\mathbf{s}} - \frac{1}{N_\phi}\sum_{\mathbf{s}'} f_{\mathbf{s}'}\exp[i(\mathbf{s}\times\mathbf{s}')_z]\right). \tag{4.4}$$

Applying this transformation twice is equivalent to the single transformation $\{f^{(a)}_{\mathbf{s}}\}^{(a)} = f^{(a)}_{\mathbf{s}}$. Consequently, if we carry out the substitution $f_{\mathbf{s}} \longrightarrow f_{\mathbf{s}} + g_{\mathbf{s}} - g^{(a)}_{\mathbf{s}}$ in the sum (4.3) ($g_{\mathbf{s}}$ is an arbitrary function), this will not affect the result. Thus, function $f_{\mathbf{s}}$ is determined ambiguously and contribution comes only from its "antisymmetric" component $f^{(a)}_{\mathbf{s}}$. Analysis shows that such a property of the envelope function $f_{\mathbf{s}}$ in representation (4.2) is the result of commutation relations for the Fermi operators in definition (1.1) (i.e., a consequence of permutation antisymmetry of the electron gas wavefunction; see [6]). If $f_{\mathbf{s}} = \delta_{\mathbf{s},\mathbf{s}_0}$, we can use relation (4.3) to derive the orthogonality condition for the basis states in the form

$$\langle \mathbf{s}, \mathbf{q}; ba|ab; \mathbf{q}_0, \mathbf{s}_0\rangle = \delta_{\mathbf{q},\mathbf{q}_0}\bar{\delta}^{(a)}_{\mathbf{s},\mathbf{s}_0},$$

where $\bar{\delta}_{\mathbf{s},\mathbf{s}_0} = (\delta_{\mathbf{s}_0,\mathbf{s}} + \delta_{-\mathbf{s}_0,\mathbf{s}})/2$ and, accordingly,

$$\bar{\delta}^{(a)}_{\mathbf{s},\mathbf{s}_0} = \frac{1}{4}\left[\delta_{\mathbf{s}_0,\mathbf{s}} + \delta_{\mathbf{s}_0,-\mathbf{s}} - \frac{1}{N_\phi}(\exp[i(\mathbf{s}\times\mathbf{s}_0)_z] + \exp[i(\mathbf{s}_0\times\mathbf{s})_z])\right].$$

Normalizability of state (4.2) indicates the finiteness of sum $\sum_{\mathbf{s}} f_{\mathbf{s}} f^{(a)}_{\mathbf{s}}$. If we pass from summation to integration with respect to $\mathbf{s}$ and, accordingly, redesignate $f_{\mathbf{s}} = N_\phi^{-1/2} f(\mathbf{s})$, the normalizability is ensured by the finiteness of the integral,

$$\int d\mathbf{s} f^*(\mathbf{s}) f^{(a)}(\mathbf{s}) < \infty. \tag{4.5}$$

We assume that exciton $|ab; \mathbf{q}\rangle$ belongs to simple excitations for which energy $\epsilon(\mathbf{q})$ (1.7) can be calculated in a first order in r_C using the exciton-diagonalizable Hamiltonian $\hat{H}_{\mathrm{ED}}$. Applying the total Hamiltonian (see formulas (2.1) and (2.7)) to state (4.1), we find that the energy of this state is given by

$$\epsilon\left(\frac{\mathbf{q}}{2} + \mathbf{s}\right) + \epsilon\left(\frac{\mathbf{q}}{2} - \mathbf{s}\right) + O\left(\frac{E_C}{N_\phi}\right).$$

Consequently, this state corresponds (to within E_C/N_ϕ) to two noninteracting excitons. In particular, if $\nu = 2m + 1$ and $a = m$, while $b = \bar{m}$, two-exciton state (4.1) is formed by two noninteracting spin excitons (two spin waves) with the total energy $|g\mu_B B| + E_{0,1}(\mathbf{q}/2 - \mathbf{s}) + E_{0,1}(\mathbf{q}/2 + \mathbf{s})$. Here, $E_{0,1}$ is the exchange energy defined as (see [4–7])

$$E_{0,1}(\mathbf{q}) = \frac{1}{N_\phi}\sum_{\mathbf{p}} V_{mmmm}(p)\{1 - \exp i(\mathbf{p}\times\mathbf{q})_z\} = \frac{e^2}{\varepsilon l_B}\int_0^\infty \frac{pdp}{2\pi} V(p)\exp\left(-\frac{p^2}{2}\right)\left[L_m\left(\frac{p^2}{2}\right)\right]^2[1 - J_0(pq)], \tag{4.6}$$

which can easily be verified using expression (2.12).

Let us now calculate the energy of two-exciton state (4.2) in the case when $a = m$ and $b = \bar{m}$. To the first order in interaction, we can confine our analysis to the reduced Hamiltonian (as in the case of a single spin exciton); in this case, the Schrödinger equation can be reduced to

$$\sum_{\mathbf{s}} \varphi_{02,\mathbf{s}}(\mathbf{s})[\hat{H}_m + \hat{H}_{\bar{m}} + \hat{H}_{m\bar{m}}, \hat{\mathfrak{I}}^{\dagger}_{m\bar{m}\mathbf{q}/2-\mathbf{s}}\hat{\mathfrak{I}}^{\dagger}_{m\bar{m}\mathbf{q}/2+\mathbf{s}}]|0\rangle$$

$$= E\sum_{\mathbf{s}} \varphi_{02,\mathbf{s}}(\mathbf{s})|m\bar{m};\ \mathbf{q}, \mathbf{s}\rangle$$

(we denoted $f(\mathbf{s}) = \varphi_{02,\mathbf{s}}$). Let us now project this equation onto state (4.1) with $ab = m\bar{m}$. Using formulas (2.10) and (2.11), commutation relations (2.5), and identities (2.6), (4.3), and (4.4), we arrive at the integral equation for function φ_{02},

$$E\varphi_{02}^{(a)}(\mathbf{s}) = \{\mathcal{L}_{\mathbf{q}}[\varphi_{02}(\mathbf{s})]\}^{(a)},$$

where operator $\mathcal{L}_{\mathbf{q}}[\varphi_{02}(\mathbf{s})]$ is given by

$$\mathcal{L}_{\mathbf{q}}[\varphi_{02}(\mathbf{s})] = \left[E_{0,1}\left(\left|\frac{\mathbf{q}}{2} + \mathbf{s}\right|\right) + E_{0,1}\left(\left|\frac{\mathbf{q}}{2} - \mathbf{s}\right|\right)\right]$$

$$\times \varphi_{02}(\mathbf{s}) + \frac{1}{\pi}\int d\mathbf{p}\varphi_{02}(\mathbf{p})V_{nnnn}(|\mathbf{p} + \mathbf{s}|) \qquad (4.7)$$

$$\times\left\{\cos[(\mathbf{p}\times\mathbf{s})_z] - \cos\left[\frac{(\mathbf{p}\times\mathbf{q})_z}{2} + \frac{(\mathbf{s}\times\mathbf{q})_z}{2}\right]\right\}.$$

Here, $E_{0,1}$ is defined by formula (4.6). In our case, a distinguishing feature of operator (4.7) is that

$$\{\mathcal{L}_{\mathbf{q}}[f]\}^{(a)} = \mathcal{L}_{\mathbf{q}}[f^{(a)}],$$

consequently, the equations for $\varphi_{02}^{(a)}$ and φ_{02} (prior to antisymmetrization) appear identical. Formally, we could simply solve the equation for φ_{02},

$$E\varphi_{02}(\mathbf{s}) = \mathcal{L}_{\mathbf{q}}[\varphi_{02}(\mathbf{s})], \qquad (4.8)$$

without imposing any constraints on the symmetry of the solutions.

For unit filling factor, Eq. (4.8) was derived in [25] using a different approach. This equation describes an excitation corresponding to $\delta n = 0$ and $\delta S = \delta S_z = -2$. Obviously, the solutions to this equation include singular normalized functions (see condition (4.5)), such that

$$|\varphi_{02}(\mathbf{s})|^2 = C[\delta(\mathbf{s} - \mathbf{s}_0) + \delta(\mathbf{s} + \mathbf{s}_0)]. \qquad (4.9)$$

These are functions from the continuum, which correspond to a pair of free spin waves. The exciton–exciton interaction in this case is negligibly weak. Indeed, substituting expression (4.9) into (4.8) nullifies the integral term in the thermodynamic limit $N_\phi \longrightarrow \infty$ so that the energy becomes

$$E = E_{0,1}(|\mathbf{q}/2 + \mathbf{s}_0|) + E_{0,1}(|\mathbf{q}/2 - \mathbf{s}_0|).$$

It should be noted that the lower edge of such a spin-exciton band is determined (for a given value of total momentum $\mathbf{q}$ of the pair) by

$$E_m(q) = \min_{\mathbf{s}}[E_{0,1}(|\mathbf{q}/2 + \mathbf{s}|) + E_{0,1}(|\mathbf{q}/2 - \mathbf{s}|)]. \qquad (4.10)$$

Obviously, we always have $E_m(q) < E_{0,1}(q)$; i.e., if we disregard the Zeeman energy, the formation of such a pair is thermodynamically more advantageous than the formation of a single spin wave with the same value of q. The upper edge of the band is independent of q and is attained for $s_0 \longrightarrow \infty$.

The following question is of fundamental importance: does a regular solution to Eq. (4.8), which is integrable in quadratures and for which the integral term in Eq. (4.7) differs from zero, also exist in addition to solutions (4.9)? The existence of such a solution would indicate the existence of a spin biexciton. We can give an affirmative answer to this question (at least, for unit filling factor (for $m = 0$)): a spin biexciton, viz., the bound state of two spin waves, must exist for all finite values of q except the point $q = 0$. This conclusion follows from analysis of Eq. (4.8) in the limit $q \longrightarrow \infty$.

5. DECAY OF A BIEXCITON: CHARGED SPIN EXCITON (TRION)

We assume that the sought excitation (a spin biexciton) exists. As in the case of a magnetic exciton [4, 5, 7], transition $q \longrightarrow \infty$ indicates its decay. Indeed, for large values of q, at least one exciton in the two-exciton state (4.1), (4.2) is characterized by a large mean distance between the electron and effective hole constituting it (this quantity for exciton is $\mathbf{R} = \mathbf{q} \times \mathbf{e}_z$, where $\mathbf{e}_z$ is a unit vector along the z axis; see [4, 7]). At the same time, if the function $f(\mathbf{s})$ is localized within an interval $|\mathbf{s}| \lesssim 1$, state (4.2) obviously transforms for $q \longrightarrow \infty$ into the four-particle state corresponding to two noninteracting electrons and two holes. The energy of such a state is not minimal, but the maximum possible energy of the four-particle excitation. To determine the state with the minimal energy (i.e., the ground state in the case of decay with $q \longrightarrow \infty$), we must seek a function $f(\mathbf{s})$, which is obviously localized within an interval $|\mathbf{s} - \mathbf{q}/2| \lesssim 1$. One of the excitons decays into an electron and a hole, while the other exciton forms a bound state with one of decay quasiparticles (i.e., it forms a cooperative Fermi excitation—a charged spin exciton (trion)). These states are well known in the regime of the quantum Hall effect (see [15, 26] and the literature cited therein).

Let us now turn to Eq. (4.8). It is difficult to find a correct limit as $q \longrightarrow \infty$ (namely, with φ_{02} localized in the neighborhood of $\mathbf{s} \approx \mathbf{q}/2$) in expression (4.7). We transform this equation as follows. First, we choose the direction of $\mathbf{q}$ parallel to the y axis: $q_x = 0$, $q_y = q$. We retain only the antisymmetric part in Eqs. (4.7), (4.8) (i.e., we assume that $\varphi_{02} = \varphi_{02}^{(a)}$) and apply to it the Fourier transformation in variable s_x. Preserving the same notation s_x for the new conjugate variable, we obtain instead of Eq. (4.8)

$$\begin{aligned} E'\Phi(s_x, s_y) &= \frac{1}{2\pi} \\ &\times \int_{-\infty}^{\infty} dw \Bigg\{ U\left(w + \frac{s_x - s_y}{2}, w + \frac{s_y - s_x}{2}\right) \\ &\times \Phi\left(w + \frac{s_x + s_y}{2}, w - \frac{s_x + s_y}{2}\right) \\ &+ U\left(w + \frac{s_x + s_y}{2}, w - \frac{s_x + s_y}{2}\right) \\ &\times \Phi\left(w + \frac{s_x - s_y}{2}, w + \frac{s_y - s_x}{2}\right) \\ &- [U(q/2 + s_y, w + s_x) + U(q/2 - s_y, w + s_x)]\Phi(w, s_y) \\ &- [U(q/2 + s_x, w + s_y) \\ &+ U(q/2 - s_x, w + s_y)]\Phi(s_x, w) \Bigg\}, \end{aligned} \tag{5.1}$$

where

$$\begin{aligned} \Phi(s_x, s_y) &= \int_{-\infty}^{\infty} ds_x' \exp(-is_x s_x')\varphi_{02}(s_x', s_y), \\ U(s_x, s_y) &= \int_{-\infty}^{\infty} ds_x \exp(-is_x s_x')V_{mmmm}(s_x', s_y), \\ E' &= E - \frac{1}{\pi}\int d^2p\, V_{mmmm}(p) \equiv E - 2E_{0,1}(\infty). \end{aligned} \tag{5.2}$$

It follows from the property of antisymmetry ($\varphi_{02} = \varphi_{02}^{(a)}$) and parity of φ_{02} in each variable that

$$\begin{aligned} \Phi(s_x, s_y) &= \Phi(-s_x, s_y) \\ &= \Phi(s_x, -s_y) = -\Phi(s_y, s_x); \end{aligned} \tag{5.3}$$

i.e., function Φ is even and possesses the Fermi permutation symmetry. Equation (5.1) is special in that the unknown function is a function of two variables, whereas Eq. (5.1) contains a single integral. This simplifies numerical solution of the problem.

As in the case of function φ_{02}, the function Φ in variable s_y is localized in the neighborhood of $s_y \approx q/2$. Thus, in view of properties (5.3), the neighborhoods of all four points on the (s_x, s_y) plane, i.e., $(q/2, \pm q/2)$, $(\pm q/2, q/2)$ must be important for function Φ. We carry out the substitution

$$\Phi(q/2 + x, q/2 + y) = F(x, y),$$

where function F is on the order of unity in both variables in the vicinity of point $x = y = 0$. It follows from properties (5.3) that

$$\begin{aligned} \Phi(-q/2 + x, q/2 + y) &= F(-x, y), \\ \Phi(q/2 + x, -q/2 + y) &= F(x, -y) \end{aligned}$$

and, in addition, $F(x, y) = -F(y, x)$. At the same time, function F apparently does not exhibit parity in variables x and y. If we now perform the substitution of variables

$$s_x = q/2 + x, \quad s_y = q/2 + y,$$

it can be seen in view of the above arguments concerning the localization of F that some terms in the integrand become negligibly small for $q \longrightarrow \infty$. As a result, we obtain

$$\begin{aligned} E'F(x, y) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dw \Bigg\{ U\left(w + \frac{x - y}{2}, w + \frac{y - x}{2}\right) \\ &\times F\left(\frac{x + y}{2} + w, \frac{x + y}{2} - w\right) - U(y, w + x) \\ &\times F(-w, y) - U(x, w + y)F(x, -w) \Bigg\}. \end{aligned} \tag{5.4}$$

Thus, we have derived the equation for function $F(x, y)$, which plays the role of the wavefunction of a 2D trion, viz., a new quasiparticle with spin $\delta S_z = -3/2$. (The other quasiparticle emerging as a result of decay has a spin of $-1/2$.) The following two circumstances should be emphasized. First, Eq. (5.4) is a limiting case for the biexciton equation (5.1), which is equivalent to Eq. (4.8); consequently we can state that this equation describes a biexciton for $q \gg 1$. Second, it is important that Eq. (5.4), as well as Eqs. (4.8) and (5.1), is asymptotically exact in the thermodynamic limit; in other words, in the given approximation, this equation corresponds to an infinitely large number of particles ($N_\phi = \infty$). In this respect, our solution differs considerably from the approaches used in [15, 26].

In the special case of unit filling factor (provided that the gas is perfectly two-dimensional), i.e., for

$$\nu = 1 \quad (m = 0),$$
$$V_{mmmm} = \frac{e^2}{\varepsilon l_B q}\exp\left(-\frac{q^2}{2}\right),$$

the minimal energy (4.10) of two "free" spin excitons with a total momentum $q = \infty$ is $E_m(\infty) = E_{0,1}(\infty) = \sqrt{\pi/2}$ (here and below, we give numerical values of energy in units of $e^2/\varepsilon l_B$). In this case, the momentum of one spin exciton is infinitely large, while the momentum of the other exciton is equal to zero, Energy $E_{3/2} = E' + 2E_{0,1}(\infty)$ is, as expected, lower than this quantity by the trion binding energy $\delta E_{3/2} = E_m(\infty) - E_{3/2} > 0$. In spite of its apparently simple form, numerical solution of Eq. (5.4) is a complicated problem. Indeed, it is binding energy $\delta E_{3/2}$, which is much smaller than energy $E' \approx E_{0,1}(\infty)$, that is a physically significant quantity. In actual practice, this means that the smallest eigenvalue E' should be determined with an error not exceeding 10^{-4}. It should be noted that kernel U can be represented in an integral form (see Appendix 2) convenient for computation and tabulation, which can be subsequently used for substituting into this equation.

Let us solve Eq. (5.4) with the help of expansion in an orthogonal basis, for which we can take Hermite polynomials H_i ($i = 0, 1, 2, \ldots$) with the corresponding weight function

$$F(x, y) = \sum_{i=k=0}^{i_{\max}, k_{\max}} \frac{1}{2^{(i+k)/2}\sqrt{\pi i!k!}} A_{ik} H_i(x) H_k(y) \times \exp\left(-\frac{x^2 + y^2}{2}\right), \quad (5.5)$$

where $A_{ik} = -A_{ki}$. Substituting expansion (5.5) into Eq. (5.4) and projecting it subsequently onto basis functions

$$\frac{1}{\sqrt{2^{i'+k'}\pi i'!k'!}} H_{i'}(x) H_{k'}(y)\exp\left(-\frac{x^2 + y^2}{2}\right),$$

we arrive at a secular equation in E', in which matrix elements $I_{i'k',ik}$ are triple integrals (if we disregard integration in the representation for kernel U). One integration can be carried out analytically; however, the high accuracy necessary for determining E' requires high accuracy in evaluating integrals and a large number of terms in expansion (5.5). The total number of integrals $I_{i'k',ik}$ to be evaluated is $K(K + 1)/2$, where K is the number of different (with allowance for substitution $x \longleftrightarrow y$) factors A_{ik} in expansion (5.5). Consequently, $K = M(M + 1)/2$, where $M = i_{\max} = k_{\max}$. Calculations are performed up to $M = 27$. Details are given in Appendix 2. We give here the result for the binding energy:

$$\delta E_{3/2} = 0.05444. \quad (5.6)$$

Our solution contains no information on the charge of quasiparticles emerging for $q \longrightarrow \infty$. It follows from symmetry considerations [27] that a quasiparticle with a spin of –1/2 may turn out to be an electron or a hole with the same probability; analogously, an excitation with $\delta S_z = -3/2$ may have a charge of $+e$ or $-e$. This feature of the solution does not at all mean that the ground state of a biexciton determined for $q \longrightarrow \infty$ is degenerate. Conditionally speaking, the representation of states (4.1) and (4.2) determines only the mutual orientation of the momenta of interacting excitons. We can in principle pass from state (4.2) in the conjugate space to a four-particle wavefunction in a real 2D space. The latter function must be such that its amplitude for $q \longrightarrow \infty$ is the same for both configurations: a hole plus a negative trion ($h + X^-$) and an electron plus a positive trion ($e + X^+$).

Trions X^+ and X^- possess energy $E_{0,1}(\infty) - \delta E_{3/2}$ and $-\delta E_{3/2}$, respectively. Both excitations (X^+ and X^-) correspond to the triplet state of a trion. Indeed, the change in the total spin $\mathbf{S}$ of the system is $\delta S = \delta S_z = -1$; this is also valid when a spin exciton with a nonzero momentum is added both to the electron (transition from the $b_p^\dagger|0\rangle$ to $\hat{\mathfrak{D}}_{0\bar{0}\mathbf{q}}^\dagger b_p^\dagger|0\rangle$ state) and to the effective hole (transition $a_p|0\rangle \longrightarrow \hat{\mathfrak{D}}_{0\bar{0}\mathbf{q}}^\dagger a_p|0\rangle$). Such a charged triplet exciton corresponds to the bound state determined by us. A singlet exciton does not form a bound state.[1]

Concluding the section, we note that other methods also exist for calculating the binding energy of the trion. First, we can mention the method of exact diagonalization for a finite number of electrons $N_e = N_\phi \pm 1$ (the case when $N_\phi = 80$ is considered, for example, in [15]); the error of this method is determined by the finiteness of the value of N_ϕ. Second, it is an approach based on the fact that the given many-particle problem can be reduced to the quantum-mechanical three-body problem (for $h + h + e$ or $h + e + e$) in the thermodynamic

[1] In terms of the excitonic representation, a charged singlet exciton would correspond to the limiting case $\mathbf{s} = \mathbf{q}/2 \longrightarrow \infty$ of the two-exciton state (4.1), in which one exciton is a "zero" exciton (i.e., its momentum is exactly equal to zero). The zero spin exciton corresponds to a change in the spin numbers $\delta S_z = -1$, but $\delta S = 0$. It is just the rotation of the spin system as a whole about the direction of the z axis. The zero exciton has zero dipole moment and does not form a bound state with an electron or a hole.

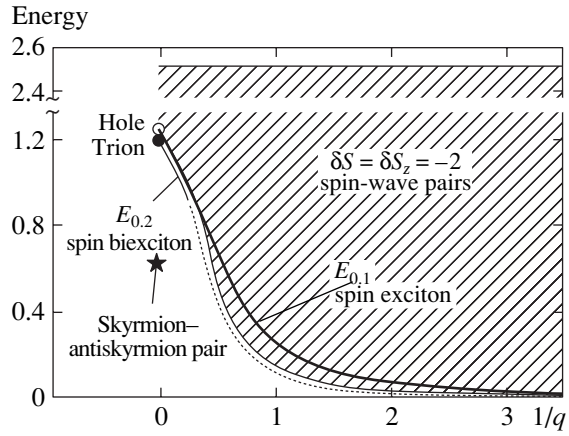


Fig. 2. Dependence of excitation energies on the wavevector. The hatched region of spin-wave pair energy $E_{0,1}(\mathbf{q}/2 + \mathbf{s}) + E_{0,1}(\mathbf{q}/2 - \mathbf{s})$ corresponds to solutions (4.9), for which the integral term in Eq. (4.7) vanishes. The lower boundary of this region is determined by formula (4.10). The solid curve in this region describes the spin exciton energy $E_{0,1}(q)$ (for $q \longrightarrow \infty$, this energy is equal to the energy of the free hole formed at the lower spin sublevel since the electron energy is zero). The solid curve outside this region is the spin biexciton energy $E_{0,2}(q)$ (see formula (6.1)), while the dashed curve describes the predicted behavior of this energy for $q \lesssim 1$. The energies of an effective hole (○) and a positive trion (●), as well as the energy of the formation of a skyrmion–antiskyrmion pair (★), are indicated. Calculations are performed for the ideal 2D case. The energy in units of $e^2/\varepsilon l_B$ is plotted along the vertical axis and the reciprocal wavevector in units of l_B is plotted along the horizontal axis.

limit if we disregard the Landau level mixing. However, the three-body Schrödinger equation with the Coulomb interaction for each pair of quasiparticles cannot be written exactly since only its projection onto one level makes sense. In this case, the limited number of basis functions of the Landau level determines the methodical error of the calculation performed in [26]. Nevertheless, the results obtained in [26] are very close to our result (namely, $\delta E_{3/2} = 0.0544$ [15] and 0.0545 [26].

6. DISCUSSION

The physical meaning of the formation of a trion can be grasped even from the following simple considerations. If the spin exciton momentum is small ($q \ll 1$), its energy is a quadratic function of the momentum ($E_{0,1} \approx q^2/2M_{s^-X}$, where M_{s^-X} is the spin exciton mass). At the same time, the field produced by the exciton due to the dipole moment $e\mathbf{q} \times \mathbf{e}_z$ is linear in q; consequently, for an appropriate effective arrangement of the charges, the total energy might be lower than for a quasiparticle with the same charge, but with a spin of –1/2.

Let us discuss again the existence of a biexciton. It should be emphasized that we are speaking of the state of the system corresponding to the lowest energy for preset quantum numbers $\delta S = \delta S_z = -2$, and $\mathbf{q}$. We have determined this state for $q \longrightarrow \infty$. Obviously, it also exists for finite but large values of q (and, hence, for the large distance between a trion and a quasiparticle in the real space). The energy is the sum of the trion energy and the energy of the remote quasiparticle minus its binding energy with the trion.

For definiteness, we consider a biexciton formed by a positively charged exciton X^+ and electron e for large values of q. The energy of trion X^+ is $E_{0,1}(\infty) - \delta E_{3/2}$. The electron energy at infinity is zero. Let us now estimate the binding energy of X^+ and e. We know the binding energy of hole h and electron e ($\delta E_{0,1}(q) = E_{0,1}(\infty) - E_{0,1}(q)$). At large distances, the power expansion in $R^{-1} = 1/q$ holds. Naturally, in addition to charge, the trion also possesses a quadrupole moment, but the field produced by the positive charge plays the major role for large values of R; i.e., the trion does not differ from a hole in this respect. Consequently, in the main approximation in $1/q$, the binding energy of X^+ and e is the same as the binding energy of h and e. In this approximation, we obtain the following expression for the biexciton energy:

$$E_{0,2}(q) \approx E_{0,1}(\infty) - \delta E_{3/2} - \delta E_{0,1}(q) = E_{0,1}(q) - \delta E_{3/2} \quad (q \gg 1). \tag{6.1}$$

In Fig. 2, the exciton energy $E_{0,1}$ is plotted as a function of the reciprocal wavevector (solid curve); the biexciton energy $E_{0,2}(q)$ is also presented (the solid curve passing to the dashed curve). For $1/q = 0$, this quantity is equal to the trion energy $E_{0,1}(\infty) - \delta E_{3/2}$. The hatched region in the figure corresponds to the band of free spin exciton pairs. The lower edge $E_m(q)$ of this band is determined by Eq. (4.10), while the upper edge is bounded by the straight line $2E_{0,1}(\infty)$. All states of this band naturally correspond to the same quantum numbers $\delta n = 0$ and $\delta S = \delta S_z = -2$; consequently, the $E_{0,2}(q)$ curve must pass below $E_m(q)$. Otherwise (i.e., in the case of equality $E_{0,2}(q_c) = E_m(q_c)$ for a finite value of q_c), it would turn out that the lowest state of the system for fixed quantum numbers δS, δS_z, and $\mathbf{q}$ is degenerate, which cannot be explained by symmetry considerations or certain other special factors. At the same time, for $q = 0$, both energy E_m and the biexciton energy $E_{0,2}$ obviously vanish. In other words, the biexciton does not exist at point $q = 0$, and function φ_{02} describing it is formally reduced to singular form (4.9) with $s_0 = 0$.[2] It should be noted that point $q \equiv 0$ is "pricked out" for the states corresponding to the edge of the spin-

[2] Coulomb energy $E_{0,2}$ cannot become negative. Otherwise, it would turn out that the ferromagnetic ground state $|0\rangle$ with filling factor $\nu = 1$ decays since the total spin turns out to be smaller than its maximal value $S_0 = N_\phi/2$.

exciton band also. This is due to the fact that excitation $|0\bar{0};\mathbf{0},\mathbf{0}\rangle = (1/2)\hat{\mathfrak{D}}^{\dagger}_{0\bar{0}\mathbf{0}}\hat{\mathfrak{D}}^{\dagger}_{0\bar{0}\mathbf{0}}|0\rangle$ corresponding to the Coulomb energy $E_m(0) = 0$ has spin numbers $\delta S_z = -2$ and $\delta S = 0$, not $\delta S_z = \delta S = -2$ (see [12, 14] and Footnote 1 at the end of Section 5).

We found that the gap that is associated with the creation of fermions X^+ and e (or X^- and h) upon the transition $q \longrightarrow \infty$ leading to biexciton dissociation is smaller than the value of $E_{0,1}(\infty)$ determined by the decay of a spin exciton. The gain in energy (5.6) in field $B = 10$ T corresponds approximately to 9 K in the ultra-two-dimensional limit. We have disregarded the difference in the Zeeman energies for excitations with $\delta S_z = -1$ and $\delta S_z = -2$. At the same time, this energy ($|g\mu_B B| \approx 2.9$ K for $B = 10$ T) apparently does not compensate the substantial difference in the Coulomb energies. We can also consider larger quantum numbers $|\delta S_z|$ analyzing, for example, a 3-exciton, etc. The numerical calculations made in [15] show that the Fermi gap for the conventional Lande factor for GaAs ($g = -0.44$) is determined by small values of $|\delta S_z| \sim 1$, and a fermion excitation is realized for $\nu = 1$ in the form of the so-called skyrmion only as $g \longrightarrow 0$ (in this case, $\delta S_z = -\infty$ and the gap is equal to $(1/2)\sqrt{\pi/2}$). Our result for the gap with $\delta S_z = -2$ is therefore intermediate between the case of the decay of a spin exciton into an electron and an effective hole and the case of formation of a skyrmion–antiskyrmion pair.

Let us now consider the activation energy for an excitation corresponding to the change $\delta S = \delta S_z = -3$. Obviously, for $q \longrightarrow \infty$, such a 3-exciton must also decay into two Fermi excitations. If these excitations have spins $\delta S_z = -3/2$, the energy of the gap has already been calculated by us. Taking into account the Zeeman energy, we can write the gap energy as

$$3|g\mu_B B| + E_{0,1}(\infty) - 2\delta E_{3/2}, \qquad (6.2)$$

i.e., the decrease in the gap due to the exchange interaction amounts to $2\delta E_{3/2} \approx 0.109$. The "asymmetric" version of dissociation into fermions with $\delta S_z = -1/2$ and $\delta S_z = -5/2$ leads to a smaller gain in energy. Indeed, such a decay would be more advantageous in the case when a quasiparticle with a spin of $-5/2$ ensures an energy gain $\delta E_{5/2}$ larger than $2\delta E_{3/2}$. However, in accordance with the physical meaning (which is also confirmed by calculations [15]), the inequality

$$\delta E_{K+\frac{3}{2}} - \delta E_{K+1/2} < \delta E_{K+1/2} - \delta E_{K-1/2}$$

must hold for any K ($K \geq 1$, $\delta E_{1/2} = 0$). For $K = \infty$, the binding energy is $\delta E_{\infty} = 0.25\sqrt{\pi/2} \approx 0.31$ [28].

Naturally, the above arguments pertain only to the approximation linear in parameter r_C. Even a second-order correction (taking into account the fact that this parameter is not small in actual practice) may introduce a substantial correction to the calculated values of the gaps, which reduce their values. In addition, it should be borne in mind that the excitation spectrum is in fact truncated for a finite q due to inevitable presence of a smooth random potential in the 2D channel (because of impurities located behind the spacer). Indeed, in the case of a spin exciton, its existence is due to the fact that the force $|dE_{0,1}(\mathbf{q})/d\mathbf{q}|$ of interaction between an electron and an effective hole exceeds the force $|\nabla\phi|$ determined by random potential $\phi(\mathbf{r})$ (the characteristic values of $|\nabla\phi| \sim 0.05$–0.1 K/nm). This condition is violated even for $q \gtrsim 4$; i.e., the gap turns out to be much smaller than the value formally determined for $q \longrightarrow \infty$. One more correction appears if we take into account the thickness of the 2D gas in the direction of the z axis. This correction reduces the excitation energy by approximately 30%. Thus, the experimental data on the fermion gap (which is determined from the activation energy in the regime of the quantum Hall effect [29]) generally do not contradict the theoretical concepts. If corrections for disorder, the finiteness of r_C, and the finiteness of the gas thickness are taken into account, the gap turns out to be quite close to the value given by formula (6.2).

ACKNOWLEDGMENTS

The authors are grateful to S.V. Iordanskiĭ and I.V. Kukushkin for fruitful discussions.

This study was supported by the Russian Foundation for Basic Research. One of the authors (S. M. D.) is obliged to the MINERVA foundation for support and to the Weizmann Institute of Science, Rehovot, Israel for the hospitality offered during the preparation of this paper. S. M. D. and V. M. Zh. also thank Max-Planck-Institut für Physik komplexer Systeme, Dresden, Germany for hospitality and the opportunity to work in the computer network of the institute while performing numerical calculations.

APPENDIX 1

Let us consider by way of example Eqs. (2.12) and (2.15) in the specific case of filling factor $\nu = 2m + 1$ and pairs $(ab) = (\overline{m-1}\,\overline{m})$ and $(a'b') = (mm+1)$. Using the definitions of operators $\hat{H}_a$ and $\hat{H}_{ab}$ (see Eqs. (2.10) and (2.11) as well as the definition of "displacement" operators $\hat{A}^{\dagger}_{\mathbf{q}}$, $\hat{B}^{\dagger}_{\mathbf{q}}$, $\hat{C}^{\dagger}_{\mathbf{q}}$, … in terms of operators (1.1)), commutation relations (2.5), and rules (2.6)

of action of the operators, we can directly find

$$[\hat{H}_a + \hat{H}_{ab}, \hat{\mathfrak{Q}}^{\dagger}_{ab\mathbf{q}}]|0\rangle = \left\{ V_{m-1mm-1m}(\mathbf{q}) \right.$$

$$+ \frac{1}{N_\phi}\sum_{\mathbf{p}}[V_{m-1m-1m-1m-1}(\mathbf{p}) - V_{m-1mm-1m}(\mathbf{p}) \quad \text{(A.1.1)}$$

$$\left. - V_{m-1mmm-1}(\mathbf{p})\exp[i(\mathbf{p}\times\mathbf{q})_z]\right\}\hat{\mathfrak{Q}}^{\dagger}_{\overline{m-1}\bar{m}\mathbf{q}}|0\rangle$$

(we do not mark by the bar the spin index in quantities $V_{\ldots}$ since we always have $V_{klmn} = V_{\bar{k}\bar{l}\overline{mn}} = V_{\bar{k}lm\bar{n}} = V_{kl\bar{m}n}$).

In the same way, we obtain

$$\sum_{\mathbf{c}}[(\hat{H}_{ac} + \hat{H}_{bc}), \hat{\mathfrak{Q}}^{\dagger}_{ab\mathbf{q}}]|0\rangle$$

$$= \left\{\frac{1}{N_\phi}\sum_{n_c=0}^{m-2}\sum_{\mathbf{p}}[V_{m-1n_cm-1,n_c}(\mathbf{p}) \right. \quad \text{(A.1.2)}$$

$$\left. - V_{mn_cm,n_c}(\mathbf{p})]\right\}\hat{\mathfrak{Q}}^{\dagger}_{\overline{m-1}\bar{m}\mathbf{q}}|0\rangle$$

(this contribution emerges when $m \geq 2$). We substitute expressions (2.3) and (2.4) for $V_{\ldots}$ into these formulas and carry out summation over n_c in (A.1.2). For this purpose, we use the following functional identity for the generalized Laguerre polynomials:

$$\sum_{k=0}^{n}\left\{\frac{k!}{n!}z^{n-k}[L_k^{n-k}(z)]^2 \right. \quad \text{(A.1.3)}$$

$$\left. - \frac{k!}{(n+1)!}z^{n+1-k}[L_k^{n+1-k}(z)]^2\right\} = L_n(z)L_{n+1}(z).$$

Summing expressions (A.1.1) and (A.1.2) and changing from summation to integration with respect to $\mathbf{p}$, we obtain quantity $\mathscr{E}_{\overline{m-1}\bar{m}}(q)$ in accordance with Eq. (2.12) (see Eq. (3.2) and (3.4)). The other matrix element $\mathscr{E}_{mm+1}(q)$ can be calculated analogously.

As regards Eq. (2.15), the contribution to its left-hand side for the pair $(ab) = (\overline{m-1}\bar{m})$ comes only from the following terms appearing in sum (2.14):

$$\sum_{\mathbf{q}}[V_{m-1m+1mm}(q)\hat{\mathfrak{Q}}^{+}_{mm+1\mathbf{q}}\hat{\mathfrak{Q}}_{\overline{m-1}\bar{m}\mathbf{q}} \quad \text{(A.1.4)}$$

$$+ V_{m-1m+1m,m}(q)\hat{\mathfrak{Q}}^{+}_{m\overline{m-1}\mathbf{q}}\hat{\mathfrak{Q}}_{m+1\bar{m}\mathbf{q}}].$$

Accordingly, the contribution for the pair $(a'b') = (mm+1)$ is determined by the terms

$$\sum_{\mathbf{q}}[V_{mmm-1m+1}(q)\hat{\mathfrak{Q}}^{+}_{\overline{m-1}\bar{m}\mathbf{q}}\hat{\mathfrak{Q}}_{mm+1\mathbf{q}} \quad \text{(A.1.5)}$$

$$+ V_{mmm+1,m-1}(q)\hat{\mathfrak{Q}}^{+}_{\overline{m-1}m\mathbf{q}}\hat{\mathfrak{Q}}_{\bar{m}m+1\mathbf{q}}].$$

Using commutation relations (2.5), we find from Eqs. (2.15) that the off-diagonal elements $\mathscr{E}^{(\overline{m-1}\bar{m})}_{mm+1} = \mathscr{E}^{(mm+1)}_{\overline{m-1}\bar{m}}$ are determined by quantity (3.5).

It should be noted that identity (A.1.3) is also useful in calculating excitations $|2m; 1, 0, \mathbf{q}\rangle$ and $|2m; 1, 1, \mathbf{q}\rangle$ to simplify the sum over c in Eq. (2.12) only if $m \geq 2$ (i.e., $\nu \geq 4$).

APPENDIX 2

In Eq. (5.4), we must specify the function of kernel U. Definition (5.2) of this function implies that

$$U(x, y) \equiv \exp\left(-\frac{y^2}{2}\right) \quad \text{(A.2.1)}$$

$$\times \int_{-\infty}^{\infty} du\,\frac{\exp(-u^2/2 - i|x|u)}{\sqrt{u^2 + y^2}}$$

in units of $e^3/\varepsilon l_B$ (we consider the case when $m = 0$ and $V(q) = 2\pi/q$). However, this formula is inconvenient for tabulating the values of $U(x, y)$ since the integral in this formula converges slowly and the accuracy of calculations is insufficient. We transform expression (A.2.1) as follows. Taking into account the presence of branching point $u = -i|y|$, we modify the integration contour. For $|x| < |y|$, it consists of segments $\mathscr{G}_1 + \mathscr{G}_4$, where

$$\mathscr{G}_1 = (e^{3i\pi/2}|x| - \infty, e^{3i\pi/2}|x| - 0),$$
$$\mathscr{G}_4 = (e^{-i\pi/2}|x| + 0, e^{-i\pi/2}|x| + \infty).$$

If, however, $|x| > |y|$, integration is carried out over contour $\mathscr{G}_1 + \mathscr{G}_2 + \mathscr{G}_3 + \mathscr{G}_4$, where

$$\mathscr{G}_2 = [e^{3i\pi/2}(|x| - |y|) + e^{-i\pi/2}|y| - 0,$$
$$0\cdot e^{3i\pi/2} + e^{-i\pi/2}|y| - 0],$$

$$\mathscr{G}_3 = (e^{-i\pi/2}|y| + 0, e^{-i\pi/2}|x| + 0)$$

(integration over an infinitely small loop in the vicinity of $u = -i|y|$ makes zero contribution).

First, we perform the substitution $u = -i|x| + z$ in Eq. (A.2.1) and integrate over segment $\mathcal{G}_1 + \mathcal{G}_4$. The corresponding contribution is

$$U_1(x, y) \equiv \exp\left(-\frac{y^2}{2}\right) \times \int\limits_{\mathcal{G}_1+\mathcal{G}_4} du \frac{\exp(-u^2/2 - i|x|u)}{\sqrt{u^2 + y^2}}$$
$$\equiv \exp\left(-\frac{x^2+y^2}{2}\right) \int\limits_{-\infty}^{\infty} \frac{\exp(-z^2/2)}{\sqrt{(-i|x| + z)^2 + y^2}} \qquad \text{(A.2.2)}$$
$$\equiv \sqrt{2}\exp\left(-\frac{y^2+x^2}{2}\right) \int\limits_0^{\infty} dz \exp\left(-\frac{z^2}{2}\right) \times \frac{\sqrt{R_{x,y}(z) + z^2 + y^2 - x^2}}{R_{x,y}(z)},$$

where

$$R_{x,y}(z) = \sqrt{(z^2 + y^2 - x^2)^2 + 4x^2z^2}. \qquad \text{(A.2.3)}$$

To integrate over contour $\mathcal{G}_2 + \mathcal{G}_3$, we carry out the substitution

$$u = |y|e^{-i\pi/2} + \begin{cases} w^2 e^{3i\pi/2} & \text{for } \mathcal{G}_2, \\ w^2 e^{-i\pi/2} & \text{for } \mathcal{G}_3 \end{cases}$$

in Eq. (A.2.1) and obtain the result, which can be written in the following form convenient for numerical calculations:

$$U_2(x, y) \equiv \exp\left(-\frac{y^2}{2}\right) \int\limits_{\mathcal{G}_2+\mathcal{G}_3} du \frac{\exp(-u^2/2 - i|x|u)}{\sqrt{u^2 + y^2}}$$
$$\equiv 2\sqrt{2}\exp(-|xy|) \qquad \text{(A.2.4)}$$
$$\times \int\limits_0^{\sqrt{|x|-|y|}} dw \frac{\exp[w^4/2 + w^2(|y| - |x|)]}{\sqrt{|y| + w^2/2}}.$$

The final result is

$$U(x, y) = \begin{cases} U_1(x, y), & |x| < |y|, \\ U_1(x, y) + U_2(x, y), & |x| > |y|. \end{cases} \qquad \text{(A.2.5)}$$

Formulas (A.2.2)–(A.2.5) make it possible to create an array of values of $U(x, y)$ for use in the numerical solution of Eq. (5.4). All calculations were made with an accuracy to four decimal places. With increasing maximal index M of the Hermite polynomials used in expansion (5.5), the sought minimal eigenvalue of matrix $I_{i'k', ik}$ decreases (i.e., its absolute value increases since $E' < 0$). Figure 3 demonstrates this dependence. With increasing M, the dependence naturally becomes weaker. The value of jump $D(M) = E'(M) - E'(M + 1)$ is shown in the inset. It can be seen that $D(M)$ exhibits a characteristic quasi-periodic behavior, which makes it possible to extrapolate the $D(M)$ curve analytically to the range of values of $M > 27$ and in this way to calculate the value of $E'(\infty)$. Extrapolation can be carried out in various ways, which leads to a certain spread in the results for $E'(\infty)$. The spread is found to be within an error of 10^{-5}; consequently, we arrive at the value of $E'(\infty) = -\sqrt{\pi/2} - \delta E_{3/2}$, where the value of $\delta E_{3/2}$ is given in Section 5.

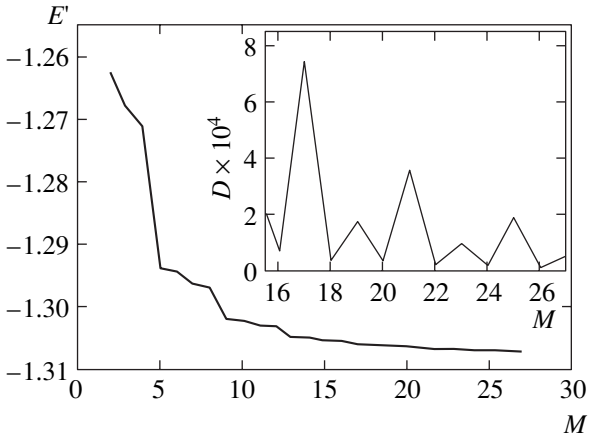


Fig. 3. Calculated quantity E' (in units of $e^2/\varepsilon l_B$) as a function of number M, viz., the main index of Hermite polynomials in representation (5.5). The inset shows the quantity $D(M) = E'(M) - E'(M + 1)$.

REFERENCES

1. *The Quantum Hall Effect*, Ed. by R. R. Prange and S. M. Girvin, 2nd ed. (Springer, New York, 1990; Mir, Moscow, 1989).
2. R. B. Laughlin, Physica B (Amsterdam) **126**, 254 (1984).
3. S. M. Girvin, P. H. MacDonald, and P. M. Platzman, Phys. Rev. B **33**, 2481 (1986).
4. I. V. Lerner and Yu. E. Lozovik, Zh. Éksp. Teor. Fiz. **78**, 1167 (1980) [Sov. Phys. JETP **51**, 588 (1980)].
5. Yu. A. Bychkov, S. V. Iordanskiĭ, and G. M. Eliashberg, Pis'ma Zh. Éksp. Teor. Fiz. **33**, 152 (1981) [JETP Lett. **33**, 143 (1981)].
6. Yu. A. Bychkov and E. I. Rashba, Zh. Éksp. Teor. Fiz. **85**, 1826 (1983) [Sov. Phys. JETP **58**, 1062 (1983)].
7. C. Kallin and B. I. Halperin, Phys. Rev. B **30**, 5655 (1984).
8. A. Pinczuk, B. S. Dennis, D. Heiman, *et al.*, Phys. Rev. Lett. **68**, 3623 (1992).

9. J. P. Longo and C. Kallin, Phys. Rev. B **47**, 4429 (1993).
10. A. B. Dzyubenko and Yu. E. Lozovik, Fiz. Tverd. Tela (Leningrad) **25**, 1442 (1983) [Sov. Phys. Solid State **25**, 874 (1983)]; **26**, 1540 (1984) [**26**, 938 (1984)].
11. A. B. Dzyubenko and Yu. E. Lozovik, J. Phys. A **24**, 415 (1991).
12. S. M. Dikman and S. V. Iordanskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 43 (1996) [JETP Lett. **63**, 50 (1996)]; Zh. Éksp. Teor. Fiz. **110**, 238 (1996) [JETP **83**, 128 (1996)].
13. S. Dickmann and Y. Levinson, Phys. Rev. B **60**, 7760 (1999).
14. S. Dickmann, Phys. Rev. Lett. **93**, 206804 (2004).
15. J. J. Palacios, D. Yoshioka, and A. H. MacDonald, Phys. Rev. B **54**, R2296 (1996).
16. S. M. Dikman and S. V. Iordanskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 543 (1999) [JETP Lett. **70**, 543 (1999)].
17. S. Dickmann, Phys. Rev. B **61**, 5461 (2000).
18. S. Dickmann and Y. Levinson, Physica E (Amsterdam) **5**, 153 (2000).
19. Yu. A. Bychkov and S. V. Iordanskiĭ, Fiz. Tverd. Tela (Leningrad) **29**, 2442 (1987) [Sov. Phys. Solid State **29**, 1405 (1987)].
20. S. Dickmann, Pis'ma Zh. Éksp. Teor. Fiz. **81**, 144 (2005) [JETP Lett. **81**, 112 (2005)].
21. Yu. A. Bychkov, T. Maniv, and I. D. Vagner, Phys. Rev. B **53**, 10148 (1996).
22. M. Rasolt, B. I. Halperin, and D. Vanderbilt, Phys. Rev. Lett. **57**, 126 (1986).
23. C. Kallin and B. I. Halperin, Phys. Rev. B **31**, 3635 (1985).
24. S. Dickmann and I. V. Kukushkin, Phys. Rev. B **71**, 241310(R) (2005).
25. Yu. A. Bychkov, S. V. Iordanskiĭ, and G. M. Eliashberg, Poverkhnost, No. 10, 33 (1982).
26. D. M. Whittaker and A. J. Shields, Phys. Rev. B **56**, 15185 (1997).
27. H. A. Fertig, L. Brey, R. Côté, *et al.*, Phys. Rev. B **55**, 10671 (1997).
28. H. A. Fertig, L. Brey, R. Côté, and A. H. MacDonald, Phys. Rev. B **50**, 11018 (1994).
29. A. Usher, R. J. Nicholas, J. J. Harris, and C. T. Foxon, Phys. Rev. B **41**, 1129 (1990).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 907–912.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 1041–1046.
Original Russian Text Copyright © 2005 by Margulis, Shorokhov.

ELECTRONIC PROPERTIES OF SOLIDS

Quasi-Ballistic Electron Transport in Quantum Wires

V. A. Margulis* and A. V. Shorokhov
Ogarev Mordovian State University, Saransk, 430000 Russia
**e-mail: theorphysics@mrsu.ru*
Received April 13, 2005

Abstract—Electron transport in a three-dimensional quantum wire is analyzed by taking into account electron scattering by a single point impurity. It is shown that the magnetoconductance plotted versus chemical potential μ has narrow peaks and closely located peaks separated by a dip when the scattering length is positive and negative, respectively. The peaks lie near the conductance steps. The thermopower plotted versus μ has narrow peaks and closely located peaks separated by a dip when the scattering length is positive and negative, respectively. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Since the discovery of conductance quantization in quantum wires, scattering by impurities in these nanostructures has been a subject of considerable interest [1–14]. It was shown in [1–3, 12, 13] that even a single impurity can cause conductance quantization breakdown in quantum wires and constrictions (quasi-ballistic transport regime). The breakdown is most significant in the neighborhoods of conductance steps. Electron scattering in quantum channels, wires, and constrictions has been studied both for extended [15, 16] and point impurities [1–14, 17–20]. Analyses of resonance peaks of Breit–Wigner and Fano types in the conductance of certain nanostructures were presented in [15, 16].

Magnetic field applied to a nanostructure induces band hybridization in the electron spectrum and enhances the lateral confinement of electrons. Moreover, conductance may depend not only on magnetic field strength, but also on its direction [21]. Note also that magnetic field may change the characteristics of conductance steps.

In a number of papers, it was shown that the conductance curve may contain both peaks and dips near conductance steps; i.e., the presence of an impurity may cause both resonant reflection and resonant transmission.

Thermopower in quantum wires was examined in [22, 23]. A variety of model potentials have been proposed to describe geometrical confinement in a wire. In this paper, we used the symmetric harmonic potential

$$U(x, y) = m^*\omega_0^2(x^2 + y^2)/2,$$

where m^* is the effective mass and ω_0 is the characteristic frequency of the confinement potential ($l_0 = \sqrt{\hbar/m^*\omega_0}$, where l_0 is the effective wire radius).

The wire Hamiltonian describing the unperturbed one-electron states in the absence of impurities is

$$H_0 = \frac{1}{2m^*}\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right)^2 + \frac{m^*\omega_0^2}{2}(x^2 + y^2). \qquad (1)$$

In this paper, we analyze quasi-ballistic transport in a quantum wire with symmetric cross section. We examine scattering by impurities as manifested in such transport characteristics as magnetoconductance and magnetothermopower. In this study, we consider only the case of a magnetic field **B** parallel to the wire, because an exact analytical expression for the magnetoconductance of a quantum wire can be obtained by using a Dirac-delta model potential in this particular case.

We use symmetric gauge

$$\mathbf{A} = \left(-\frac{yB}{2}, \frac{xB}{2}, 0\right)$$

for the vector potential corresponding to **B**. Hamiltonian (1) has a well-known spectrum:

$$E_{mnp} = \frac{\hbar\omega_c}{2}m + \frac{\hbar\Omega}{2}(2n + |m| + 1) + \frac{p^2}{2m^*}, \qquad (2)$$

where ω_c is the cyclotron frequency and $\Omega = \sqrt{\omega_c^2 + 4\omega_0^2}$ ($n = 0, 1, \ldots, m \in \mathbf{Z}$).

The solution to the Schrödinger equation with Hamiltonian (1) is conveniently written in cylindrical coordinates:

$$\psi_{mnp}^0 = \exp\frac{ipz}{\hbar}\frac{\exp(im\varphi)}{\sqrt{2\pi}}R_{mn}(\rho), \qquad (3)$$

where

$$R_{mn}(\rho) = c_{mn}\rho^{|m|}\exp\left(-\frac{\rho^2}{4l^2}\right)L_n^{|m|}\left(\frac{\rho^2}{2l^2}\right). \quad (4)$$

Here, $L_n^{|m|}$ denotes generalized Laguerre polynomials, $l = \sqrt{\hbar/m^*\Omega}$, and

$$c_{mn} = \frac{1}{l^{|m|+1}}\sqrt{\frac{n!}{2^{|m|}(n+|m|)!}}. \quad (5)$$

According to (3) and (4), the electron density reaches a maximum on the symmetry axis of the wire and exponentially toward its periphery. Therefore, only impurities located near the wire axis can effectively scatter the electron modes propagating in the wire. Comparing the values of l and l_0, we see that the magnetic field squeezes the electron wavefunction in the plane perpendicular to the magnetic field vector. In what follows, we analyze the particularly interesting case when the impurity located on the axis of the wire and the effects due to scattering are most pronounced. We consider short-range impurities, which are adequately modeled by Dirac-delta potentials.

2. QUASI-BALLISTIC TRANSPORT DUE TO SCATTERING BY IMPURITIES

The ballistic regime of conductance in a parallel magnetic field was studied in [12]. The contribution to conductance due to impurity scattering is expressed as

$$\frac{G^i}{G_0} = \gamma^2\{-2(\mathrm{Im}\zeta)^2 + (N+1)[\psi(N+\delta) - \psi(\delta) + (N+\delta)^{-1}]\}\{(1+\gamma\mathrm{Re}\zeta)^2 + (\gamma\mathrm{Im}\zeta)^2\}^{-1}, \quad (6)$$

where $\gamma = a/\sqrt{2}\,l$, G_0 is the conductance quantum, a is the scattering length (which may be either positive or negative [24]), $\psi(x)$ is the logarithmic derivative of the Euler gamma function,

$$\zeta = \zeta\left(\frac{1}{2}, \frac{1}{2} - \frac{\mu}{\hbar\Omega}\right)$$

is the generalized Riemann zeta function, N is an integer, and $0 < \delta < 1$. In (6), we use the chemical potential

$$\mu = \hbar\Omega(N + \delta + 1/2).$$

Expression (6) was derived in [12], where it was shown that only s waves are scattered (for which the magnetic quantum number is $m = 0$).

It is important for further analysis that $\gamma \sim 0.1$ for realistic values of a and l; i.e., the contribution to conductance due to scattering (proportional to γ^2) is small.

However, the value of G^i is not small in a certain range of μ.

It follows from the Hermite and shift formulas for the zeta function [25] that

$$\mathrm{Re}\zeta\left(\frac{1}{2}, \frac{1}{2} - \frac{\mu}{\hbar\Omega}\right) = \zeta\left(\frac{1}{2}, 1-\delta\right) \approx \frac{1}{\sqrt{1-\delta}} + \frac{1}{2\sqrt{2-\delta}} - 2\sqrt{2-\delta} + \frac{1}{24\sqrt{(2-\delta)^3}}. \quad (7)$$

Using the shift formula for the imaginary part of the zeta function, we obtain

$$\mathrm{Im}\zeta\left(\frac{1}{2}, \frac{1}{2} - \frac{\mu}{\hbar\Omega}\right) = -\sum_{n=0}^{N}(n+\delta)^{-1/2}. \quad (8)$$

It follows from (7) and (8) that $\mathrm{Re}\zeta(1/2, 1-\delta)$ remains finite as $\delta \longrightarrow 0$, while

$$\mathrm{Im}\zeta\left(\frac{1}{2}, \frac{1}{2} - \frac{\mu}{\hbar\Omega}\right) \longrightarrow \infty.$$

Combining expressions (6)–(8), we find that

$$\lim_{\delta\to 0} G^i/G_0 = N - 1.$$

When $\gamma > 0$ ($a > 0$), the curve of $G^i(\mu)$ has the shape of a very narrow peak. Since conductance steps are observed only at very low temperatures, the thermal smearing of the peak can be neglected. This is demonstrated by the following estimates. When $\delta \ll 1$, expressions (6)–(8) yield

$$\frac{G^i}{G_0} \approx \frac{N-1}{1+\gamma^{-2}\delta}$$

in a small neighborhood of the peak. Using the standard approach, we estimate the thermal smearing of the Fermi–Dirac distribution as

$$\frac{G^i(\mu, T)}{G^i(\mu, 0)} \approx 1 + \frac{\pi^2}{3}\left(\frac{k_B T}{\hbar\Omega}\right)^2 \frac{1+\delta_0\gamma^{-2}}{(\delta_0+\gamma^2)^2}, \quad (9)$$

with

$$\mu = \hbar\Omega(N + \delta_0 + 1/2).$$

For $T \approx 1$ K, $\gamma \sim 0.1$, and $\delta_0 \sim 0.01$, the estimated correction is on the order of 10^{-2}. For $1 - \delta \ll 1$, expression (6) yields

$$\frac{G^{\mathrm{i}}}{G_0} \approx \frac{A(N)(1-\delta)}{(1+\gamma^{-1}\sqrt{1-\delta})^2}.$$

When $\gamma > 0$, the correction to conductance due to impurity is small everywhere outside this neighborhood. When $\gamma < 0$, G^{i} is not small if $1 + \gamma \mathrm{Re}\,\zeta = 0$, because the factors γ^2 in the numerator and denominator of (6) cancel out. The resonance condition is also satisfied for small γ when δ is close to unity.

An analogous estimate yields

$$\frac{G^{\mathrm{i}}(\mu, T)}{G^{\mathrm{i}}(\mu, 0)} \approx 1 - \frac{\pi^2}{4}\left(\frac{k_{\mathrm{B}}T}{\hbar\Omega}\right)^2 \times \frac{\gamma^{-1}}{(1+\gamma^{-1}\sqrt{1-\delta_0})^2(1-\delta_0)^{3/2}},$$

where δ_0 evaluated from the condition $1 + \gamma \mathrm{Re}\,\zeta = 0$ is 0.01. The numerical estimate for the correction is on the order of 0.01. In Fig. 1b, the two conductance peaks are very close to one another. However, estimates show that they are separated by an interval $\Delta\mu \approx 7k_{\mathrm{B}}T$ at $T \approx 1$ K, which is much greater than $k_{\mathrm{B}}T$; i.e., the thermal smearing of the peaks is small.

When $N > 1$ and $\gamma < 0$, there is a narrow antiresonance (dip in the curve) on the left of a conductance step, where the correction is negative since

$$2(\mathrm{Im}\,\zeta)^2 > (N+1)\left[\psi(N+\delta) + \psi(\delta) + \frac{1}{N+1}\right].$$

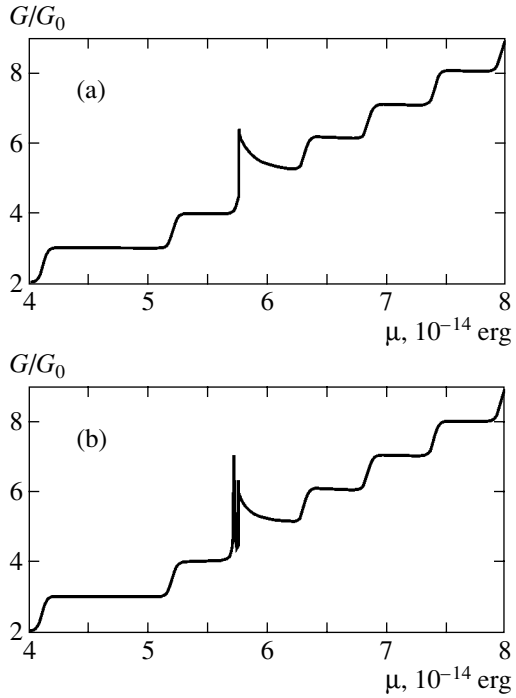


Fig. 1. Conductance vs. chemical potential for positive (a) and negative (b) scattering lengths: $\Omega = 1.8 \times 10^{13}$ s^{-1}, $B = 2$ T, $\gamma = 0.1$, and $T = 1.5$ K.

The ballistic contribution G^{b} to conductance can be derived from the results obtained in [21]. For convenience, we henceforth represent it as $G^{\mathrm{b}} = G^{\mathrm{mon}} + G^{\mathrm{osc}}$. The monotonic part of conductance is expressed as

$$\frac{G^{\mathrm{mon}}}{G_0} = \frac{1}{2\hbar^2\omega_0^2}\left[\mu^2 + \frac{\pi^2 k_{\mathrm{B}}^2 T^2}{3} - \frac{1}{12}\hbar^2(\omega_{\mathrm{c}}^2 + \omega_0^2)\right]. \quad (10)$$

The oscillating part of conductance is

$$\frac{G^{\mathrm{osc}}}{G_0} = 2\pi k_{\mathrm{B}}T\sum_{n=1}^{\infty}(-1)^{n+1}\left[\frac{1}{\hbar(\Omega+\omega_{\mathrm{c}})}\frac{\cos[2\pi n\mu/\hbar(\Omega+\omega_{\mathrm{c}})]}{\sinh[2\pi^2 n k_{\mathrm{B}}T/\hbar(\Omega+\omega_{\mathrm{c}})]\sin[\pi n(\Omega-\omega_{\mathrm{c}})/(\Omega+\omega_{\mathrm{c}})]}\right. + \left.\frac{1}{\hbar(\Omega-\omega_{\mathrm{c}})}\frac{\cos[2\pi n\mu/\hbar(\Omega-\omega_{\mathrm{c}})]}{\sinh[2\pi^2 n k_{\mathrm{B}}T/\hbar(\Omega-\omega_{\mathrm{c}})]\sin[\pi n(\Omega+\omega_{\mathrm{c}})/(\Omega-\omega_{\mathrm{c}})]}\right]. \quad (11)$$

The ratio $G^{\mathrm{osc}}/G^{\mathrm{mon}}$ is estimated as

$$\frac{G^{\mathrm{osc}}}{G^{\mathrm{mon}}} \sim \frac{\pi\hbar(\Omega-\omega_{\mathrm{c}})}{\mu}\frac{k_{\mathrm{B}}T}{\mu}. \quad (12)$$

This result has a consequence important for further analysis: when $\Omega < \mu$, $G^{\mathrm{osc}}/G^{\mathrm{mon}} \ll 1$ at temperatures not higher than 1 K. Figure 1 shows curves of magnetoconductance in the quasi-ballistic regime, with plateaus corresponding to propagating electron modes with $m \neq 0$ and peaks at the conductance steps. When $\gamma > 0$ and $m = 0$, there is a single peak on the right of a step; when $\gamma < 0$, there are two peaks (at a step and on the left of it) separated by a narrow dip. The curves correspond to the analytical results presented above. Figure 1a demonstrates that the plateau width may vary. This result was explained in [26].

Now, let us consider the thermopower in a quantum wire. Thermopower S is expressed in terms of conductance by the well-known formula [22]

$$S = \frac{\pi^2 k_{\mathrm{B}}^2 T}{3e}\frac{G'(\mu)}{G(\mu)}. \quad (13)$$

Since $G^{mon} \gg G^{osc}$ and $G^{i} \gg G^{osc}$, we have

$$S \approx \frac{\pi^2 k_B^2 T}{3e} \times \left[\frac{(G^{i})'}{G^{i}+G^{mon}} + \frac{(G^{mon})'}{G^{i}+G^{mon}} + \frac{(G^{osc})'}{G^{i}+G^{mon}}\right]. \quad (14)$$

Using (6), we obtain

$$\frac{(G^{i})'}{G_0} = -2\gamma\frac{\gamma \mathrm{Im}\zeta(\mathrm{Im}\zeta)' + (1+\gamma\mathrm{Re}\zeta)(\mathrm{Re}\zeta)'}{(1+\gamma\mathrm{Re}\zeta)^2 + (\gamma\mathrm{Im}\zeta)^2}\frac{G^{i}}{G_0} + \frac{\gamma^2}{(1+\gamma\mathrm{Re}\zeta)^2 + (\gamma\mathrm{Im}\zeta)^2}\left\{4\mathrm{Im}\zeta(\mathrm{Im}\zeta)' + \left[(N+1)\left[\psi(N+\delta) - \psi(\delta) + \frac{1}{N+\delta}\right]\right]\right\}. \quad (15)$$

To calculate the derivatives in (15), we use the fact that both G^{i} and $(G^{i})'$ do not vanish only in small right neighborhoods of conductance steps (when $\gamma > 0$) or only in small left and right neighborhoods of steps (when $\gamma < 0$). Therefore, we can set N = const in these neighborhoods and use derivatives with respect to δ instead of derivatives with respect to μ. Using the formula [25]

$$2\pi\zeta(s, v) = -\Gamma(1-s)\int_{-\infty}^{(0+)}(-t)^{s-1}\frac{e^{-vt}}{1-e^{-t}}dt, \quad (16)$$

we obtain

$$(\mathrm{Re}\zeta)' = \zeta'\left(\frac{1}{2}, 1-\delta\right) = \frac{1}{2}\zeta\left(\frac{3}{2}, 1-\delta\right), \quad (\mathrm{Im}\zeta)' = -\frac{1}{2}\mathrm{Im}\zeta\left(\frac{3}{2}, -N-\delta\right) \quad (17)$$

for N = const.

Since it holds that

$$\psi'(N+\delta) - \psi'(\delta) = -\sum_{k=0}^{N-1}\frac{1}{(k+\delta)^2} \quad (18)$$

for derivatives with respect to δ, we can substitute

$$(N+1)\left[\psi(N+\delta) - \psi(\delta) + \frac{1}{N+\delta}\right]' = -\sum_{k=0}^{N}\frac{1}{(k+\delta)^2} \quad (19)$$

into (15).

Using the formulas for derivatives obtained above, we find

$$(G^{i})' = (G_1^{i})' + (G_2^{i})',$$

where

$$\frac{(G_1^{i})'}{G_0} = \gamma\frac{G^{i}}{G_0} \times \left\{\frac{\gamma\mathrm{Im}\zeta(1/2, -N-\delta)\mathrm{Im}\zeta(3/2, -N-\delta)}{[1+\gamma\mathrm{Re}\zeta(1/2, 1-\delta)]^2 + [\gamma\mathrm{Im}\zeta(1/2, 1-\delta)]^2} - \frac{[1+\gamma\zeta(1/2, 1-\delta)]\zeta(3/2, 1-\delta)}{[1+\gamma\mathrm{Re}\zeta(1/2, 1-\delta)]^2 + [\gamma\mathrm{Im}\zeta(1/2, 1-\delta)]^2}\right\}, \quad (20)$$

$$\frac{(G_2^{i})'}{G_0} = -\gamma^2\frac{2\mathrm{Im}\zeta(1/2, -N-\delta)\mathrm{Im}\zeta(3/2, -N-\delta) - (N+1)\sum_{k=0}^{N}(k+\delta)^{-2}}{[1+\gamma\mathrm{Re}\zeta(1/2, 1-\delta)]^2 + [\gamma\mathrm{Im}\zeta(1/2, -N-\delta)]^2}. \quad (21)$$

Here, use is made of the identities

$$\zeta(3/2, 1-\delta) = \sum_{n=1}^{\infty}(n-\delta)^{-3/2},$$

$$\zeta(3/2, -N-\delta) = \sum_{n=0}^{\infty}(N+\delta-n)^{-3/2}.$$

Substituting (20), (21), (10), and (11) into (13), we obtain an expression for the thermopower (which is too cumbersome to be written out here). Figure 2 shows the thermopower as a function of chemical potential. The curves demonstrate that the thermopower drops to a negative value near a conductance peak when $\gamma = 0.1$ (Fig. 2a). Outside the neighborhoods of the peaks, the thermopower exhibits a nearly sinusoidal behavior. However, the thermopower curve obtained for $\gamma = -0.1$ (Fig. 2b) has two closely located peaks and two dips near those conductance steps where the contribution of scattering is substantial. Outside the neighborhoods of

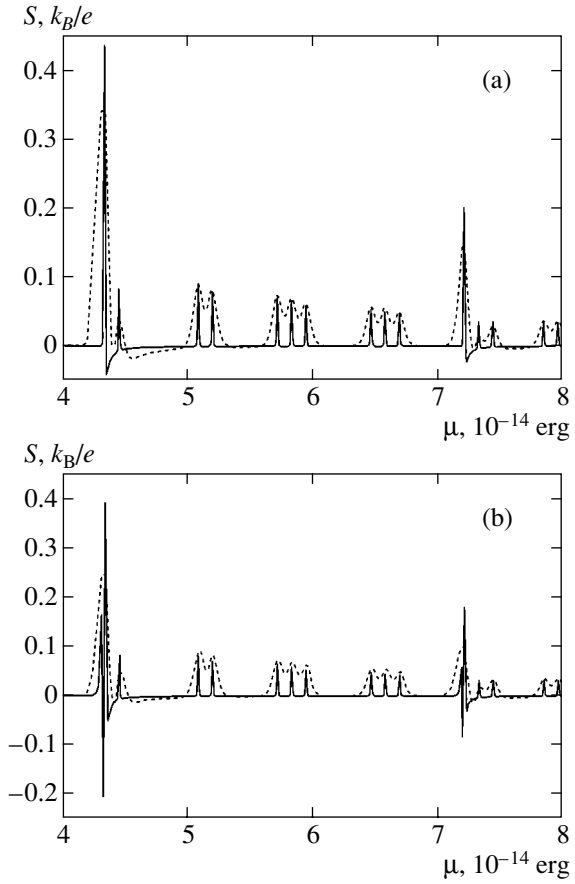


Fig. 2. Thermopower vs. chemical potential for positive (a) and negative (b) scattering lengths: $\Omega = 1.2 \times 10^{13}$ s^{-1}, $B = 5$ T, $\gamma = 0.1$, $T = 0.2$ K (solid curves) and $T = 1.5$ K (dotted curves).

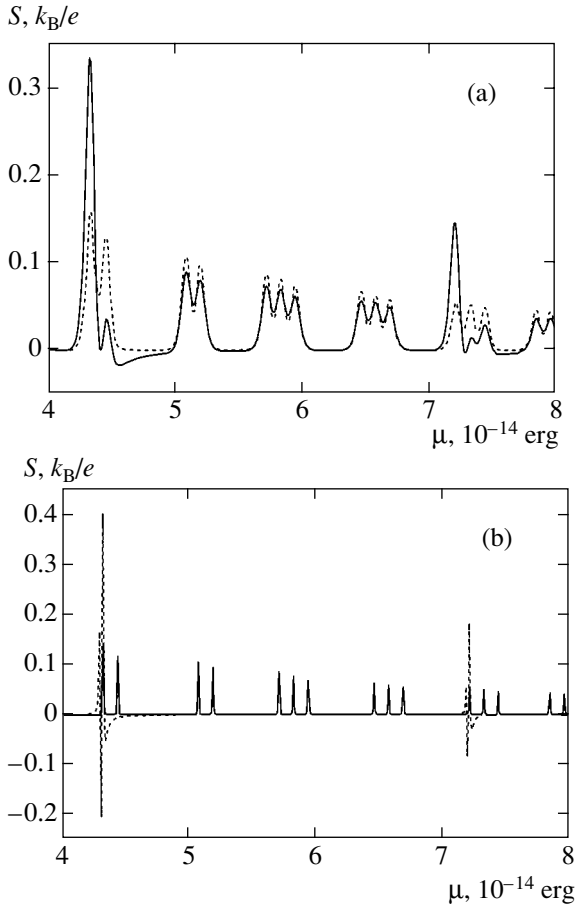


Fig. 3. Curves of $S(\mu)$ for ballistic (dotted curves) and quasi-ballistic (solid curves) transport regimes: $\Omega = 1.2 \times 10^{13}$ s^{-1}, $B = 5$ T, $\gamma = 0.1$, $T = 1.5$ K (a) and $T = 0.2$ K (b).

the peaks, the thermopower also exhibits a nearly sinusoidal behavior.

3. CONCLUSIONS

Electron magnetotransport in three-dimensional quantum wires in parallel magnetic fields is analyzed. The contribution of the Dirac-delta potential to the conductance and thermopower as functions of chemical potential is examined.

It is shown that the shape of the conductance curve depends on the sign of the scattering length a. When a is positive, the curve of $G(\mu)$ has a peak at a conductance step. The conductance curve obtained for $a < 0$ has an analogous peak, but it also has an additional peak and a dip on the left of the step. Figure 3 compares the curves of $S(\mu)$ obtained for ballistic and quasi-ballistic transport regimes corresponding to different signs of the scattering length.

In the limit of $\omega_c, \omega_0 \longrightarrow 0$, the cases of $a > 0$ and $a < 0$ correspond to the existence and absence of a bound state, respectively [24]; i.e., the quantum well represented by a Dirac-delta potential has a larger effective depth when $a > 0$.

It is shown that the thermopower as a function of chemical potential exhibits a nearly sinusoidal behavior outside the neighborhoods of the conductance peaks. Near magnetoconductance peaks, the thermopower curve has a sharp peak and a dip when the impurity scattering length is positive and two closely located peaks and two dips when the scattering length is negative. The analytical results of this study are illustrated by curves of $G(\mu)$ and $S(\mu)$.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 05-02-16145.

REFERENCES

1. Y. B. Levinson, M. J. Lubin, and E. V. Sukhorukov, Phys. Rev. B **45**, 11936 (1992).
2. I. B. Levinson, M. I. Lyubin, and E. V. Sukhorukov, Pis'ma Zh. Éksp. Teor. Fiz. **54**, 405 (1991) [JETP Lett. **54**, 401 (1991)].

3. M. I. Lyubin, Pis'ma Zh. Éksp. Teor. Fiz. **57**, 346 (1993) [JETP Lett. **57**, 361 (1993)].
4. D. H. Gobden, N. K. Patel, M. Pepper, *et al.*, Phys. Rev. B **44**, 1938 (1991).
5. A. B. Fowler, G. L. Timp, J. J. Wainer, and R. A. Webb, Phys. Rev. Lett. **57**, 138 (1986).
6. T. E. Kopley, P. L. McEuen, and R. G. Wheller, Phys. Rev. Lett. **61**, 1654 (1988).
7. S. J. Bending and M. R. Beasley, Phys. Rev. Lett. **55**, 324 (1985).
8. Y. Xu, A. Matsuda, and M. R. Beasley, Phys. Rev. B **42**, 1492 (1990).
9. N. W. Dellow, P. H. Beton, C. J. Landerak, *et al.*, Phys. Rev. Lett. **68**, 1754 (1992).
10. A. K. Geim, P. C. Main, N. La Scala, *et al.*, Phys. Rev. Lett. **72**, 2061 (1994).
11. A. K. Geim, T. J. Foster, A. Nogaret, *et al.*, Phys. Rev. B **50**, 8074 (1994).
12. V. A. Geĭler and V. A. Margulis, Zh. Éksp. Teor. Fiz. **111**, 2215 (1997) [JETP **84**, 1209 (1997)].
13. V. A. Geĭler, V. A. Margulis, and L. I. Filina, Zh. Éksp. Teor. Fiz. **113**, 1376 (1998) [JETP **86**, 751 (1998)].
14. V. Vargiamidis and H. M. Polatoglon, Phys. Rev. B **67**, 245303 (2003).
15. C. S. Kim and A. M. Satanin, Physica E (Amsterdam) **4**, 211 (1999).
16. Ch. S. Kim, A. M. Satanin, O. N. Roznova, and V. B. Shtenberg, Zh. Éksp. Teor. Fiz. **121**, 1157 (2002) [JETP **94**, 992 (2002)].
17. D. Boese, M. Lischka, and L. E. Reichl, Phys. Rev. B **61**, 5632 (2000).
18. S. A. Gurvitz and Y. B. Levinson, Phys. Rev. B **47**, 10578 (1993).
19. P. F. Bagwell, Phys. Rev. B **41**, 10354 (1990).
20. E. S. Avotina and Yu. A. Kolesnichenko, Fiz. Nizk. Temp. **30**, 209 (2004) [Low Temp. Phys. **30**, 153 (2004)].
21. V. A. Geyler and V. A. Margulis, Phys. Rev. B **61**, 1716 (2000).
22. C. R. Proetto, Phys. Rev. B **44**, 9096 (1991).
23. I. A. Kokurin, V. A. Margulis, and A. V. Shorokhov, J. Phys.: Condens. Matter **16**, 8015 (2004).
24. Yu. N. Demkov and V. N. Ostrovskiĭ, *Zero-Range Potentials and Their Applications in Atomic Physics* (Leningr. Gos. Univ., Leningrad, 1975; Plenum, New York, 1988).
25. *Higher Transcendental Functions (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1953; Nauka, Moscow, 1973), Vol. 1.
26. V. A. Geĭler and V. A. Margulis, Fiz. Tekh. Poluprovodn. (St. Petersburg) **33**, 1141 (1999) [Semiconductors **33**, 1040 (1999)].

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 913–918.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 1047–1053.
Original Russian Text Copyright © 2005 by Gorshunov, Prokhorov, Spektor, Volkov, Dressel, Dumm, Matsumura.

ELECTRONIC PROPERTIES OF SOLIDS

Kondo Scattering and Properties of TmSe in the Infrared Region

B. P. Gorshunov[a], A. S. Prokhorov[a], I. E. Spektor[a], A. A. Volkov[a], M. Dressel[b], M. Dumm[b], and T. Matsumura[c]

[a]*Prokhorov General Physics Institute, Russian Academy of Sciences, ul. Vavilova 38, Moscow, 117942 Russia*
e-mail: gorshunov@ran.gpi.ru
[b]*Physikalisches Institut 1, Universität Stuttgart, Pfaffenwaldring 57, Stuttgart, D-70550 Germany*
[c]*Department of Physics, Tohoku University, Japan*
Received May 31, 2005

Abstract—The conductivity and permittivity spectra of the intermediate-valence semiconductor TmSe have been measured by terahertz and infrared spectroscopy in a frequency range of 10–10^4 cm^{-1} and a temperature range of 5–300 K. At low temperatures (5 K < T < 100 K), the spectra contain a gap $\Delta \approx 2.5$ meV, whose appearance is considered to be related to conduction-electron localization at local magnetic moments. At high temperatures (100 K < T < 300 K), the dielectric response is specified by two electronic components: "light" conduction electrons and "heavy" hybridized *f–d* states. The microscopic parameters of both components, such as the concentration, mobility, effective mass, relaxation frequency, and the plasma frequency, are determined.
© 2005 Pleiades Publishing, Inc.

1. INTRODUCTION

Despite significant progress in understanding the physical processes that occur in heavy-fermion and intermediate-valence compounds [1, 2], there is no generally accepted viewpoint regarding the nature of their ground state. Some recent transport, optical, magnetic, and thermodynamic experiments indicate a substantially larger variety of their low-temperature properties as compared to the predictions of the periodic Anderson model. Therefore, it is necessary to comprehensively study the interaction mechanisms of conduction electrons with localized magnetic moments in intermetallic compounds. One of the most effective methods for studying such mechanisms is optical spectroscopy, which can directly probe the dynamic properties of charge carriers [1]. The terahertz–subterahertz frequency range is of particular importance, since it corresponds to the energies (of about several millielectronvolts) characteristic of mass renormalization and the fermion relaxation frequency. Measurements in this range, which had not been achieved earlier, have become possible only recently due to the creation of BWO spectrometers (BWO is a backward wave oscillator, a coherent terahertz–subterahertz radiation oscillator) [3]. As a result, researchers can now systematically study correlation effects in heavy-fermion and intermediate-valence systems by terahertz BWO spectroscopy. Even the first such measurements performed for typical representatives of these families (SmB_6, UPd_2Al_3, UPt_3) revealed new information on low-temperature excitations and specific features in the density of states at energies E < 1 meV, which are likely to be the characteristic features of the ground state for the materials of these classes [4, 5]. To search for general dependences in the low-temperature properties of intermediate-valence compounds, we continued studying their spectroscopic characteristics. In this work, we study thulium selenide TmSe.

Unlike other members of the family of intermediate-valence semiconductors, both valence sates of the thulium ion in TmSe (Tm^{2+}, Tm^{3+}) are magnetic; the average thulium valence is +2.75 [2]. At a Néel temperature T_N = 3.5 K, TmSe transforms into an antiferromagnetic phase [6]. Because of these two features, TmSe is considered to exhibit a number of properties that are atypical of this class of materials: its transport properties and neutron absorption depend on the magnetic field, the pressure, and the temperature, and its optical absorption is complex at $E \sim 1$ meV. Eventually, these specific features should be related to the microscopic mechanisms of ionic-valence fluctuations and to charge dynamics, which are being extensively studied. It should be noted that researchers mainly investigate the magnetically ordered phase existing at $T < T_N$. However, even at higher temperatures (up to room temperature), some properties of TmSe cannot be explained. These are the absence of the activation behavior of the electrical resistivity and the Hall constant, as in the related YbB_{12} and SmB_6 compounds; complex behavior of the electrical resistivity ρ as temperature decreases to 35–40 K, below which its behavior is characteristic of Kondo scattering ($\rho \propto -\log T$); and the behavior of the dynamic conductivity in the infrared region, which cannot be explained in terms of the conductivity mecha-

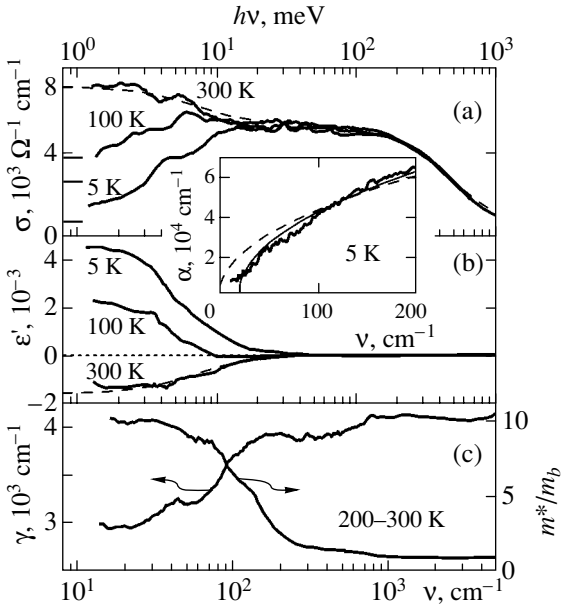


Fig. 1. Frequency dependences of (a) the dynamic conductivity σ, (b) the permittivity ε', (c) the relaxation frequency γ, and (d) the effective mass m^*/m_b for TmSe at various temperatures. The dashed lines in (a, b) for $T = 300$ K illustrate the ε'(ν) and σ(ν) dependences processed by the least squares method using Eqs. (1) and (2) of the Drude model of conduction. The heavy horizontal segments demonstrate the static conductivities measured at 300, 100, and 5 K (from top to bottom). The frequency dependences of γ and m^*/m_b were obtained within the framework of the generalized Drude model (see text). The inset shows the frequency dependence of the absorptivity α at $T = 5$ K. The solid and dashed lines illustrate the results of fitting by the least squares method using the expression $\alpha \propto \sqrt{\nu - \Delta/h}$ (see text).

nism that is typical of conductors and is described by the Drude model.

2. EXPERIMENTAL

We studied high-quality TmSe single crystals grown by the method described in [7]. The crystals were dark orange and 3×4 mm^2 in size. We measured the frequency dependences of the coefficient $R(\nu)$ of reflection from a specially prepared flat polished crystal face with two spectrometers, namely, Bruker IFS113V ($20 \text{ cm}^{-1} < \nu < 10^4 \text{ cm}^{-1}$) and a terahertz quasi-optical BWO spectrometer [3] ($10 \text{ cm}^{-1} < \nu < 30 \text{ cm}^{-1}$). The spectra recorded were joined to yield a general picture, which was analyzed with the Kramers–Kronig relations to obtain the frequency dependences of the conductivity (σ(ν)) and the permittivity (ε'(ν)). The reflectance spectra were approximated to a zero frequency using the Hagen–Rubens relations [8]

$$R = 1 - \sqrt{4\nu/\sigma_0},$$

where σ_0 is the measured static conductivity. The high-frequency extrapolations were based on the reflectance spectra recorded in [9]. The resistivity ρ was measured by the standard four-probe method on a sample in the form of a parallelepiped 0.15×0.04 mm^2 in cross section and several millimeters in length. The contacts were made of gold conductors fixed to the sample by a special technique with a silver paste; the contact resistance was less than 10 Ω. For comparison with the frequency dependences of the dynamic conductivity, the resistivity was calculated to give the static conductivity 1/ρ. All measurements were carried out in the paramagnetic phase at temperatures of 5–300 K in a zero magnetic field.

Figure 1 shows the frequency dependences of the conductivity and permittivity of TmSe. On the whole, they are similar to those measured in [2, 9] (note that the conductivity in [9] is given in terms of s^{-1}). At frequencies $\nu > 100 \text{ cm}^{-1}$, ε'(ν) and σ(ν) are virtually temperature-independent and have a shape typical of conductors: at $100 \text{ cm}^{-1} < \nu < 1000 \text{ cm}^{-1}$, ε'(ν) and σ(ν) have no dispersion; at $\nu > 1000 \text{ cm}^{-1}$, the conductivity decreases and the permittivity does not increase (on the chosen scale). To obtain the characteristics of charge carriers, we processed these spectra by the least squares method using the corresponding expressions within the framework of the Drude model of conduction [8]

$$\sigma(\nu) = \sigma_0\gamma^2(\gamma^2 + \nu^2)^{-1}, \quad (1)$$

$$\varepsilon'(\nu) = -2\sigma_0\gamma(\gamma^2 + \nu^2)^{-1}, \quad (2)$$

where $\sigma_0 = \nu_p^2/2\gamma$, ν_p is the plasma frequency, and γ is the carrier relaxation frequency. The values of $\nu_p = 39000 \text{ cm}^{-1}$ ($h\nu_p = 4.8$ eV) and $\gamma = 4800 \text{ cm}^{-1}$ thus obtained agree with the data of [9].

At low frequencies ($\nu < 100 \text{ cm}^{-1}$), the ε'(ν) and σ(ν) dependences deviate strongly from the simple Drude curves at both high and low temperatures. Although these deviations were mentioned in [9], they have not been studied. It will be low frequencies that we now direct our attention to when analyzing the dielectric response of TmSe. At $T = 300$ K, the conductivity increases and the permittivity decreases as the frequency decreases below 100 cm^{-1}. This indicates an additional Drude-type dispersion mechanism operating at low frequencies; in other words, this evidences that the dynamic response of TmSe is contributed by an additional component of mobile charge carriers that differs from the component specifying the properties of the compound at $\nu > 100 \text{ cm}^{-1}$. We processed the spectra at $T = 300$ K using a sum of Drude relations (1) and (2) (Fig. 1, dashed line) to determine the parameters of this component: $\gamma = 70 \text{ cm}^{-1}$ and $\nu_p = 2750 \text{ cm}^{-1}$. As the temperature decreases from 300 to 200 K, the ε'(ν) and σ(ν) spectra are almost unchanged. Upon further

cooling, the low-frequency hump in the frequency dependence of the conductivity begins to lose its spectral weight. At $T < 50$–100 K, a trough forms at its place; this trough is similar to the trough in the frequency dependence of the conductivity that appears due to the appearance of an energy gap in the density of states, e.g., in semiconductors or superconductors [10, 11] and other heavy-fermion and intermediate-valence compounds [4, 5]. The nature of this gap feature in the frequency dependence of the conductivity of TmSe will be discussed below. The permittivity behaves according to the dynamic conductivity: the gap feature in the frequency dependence of the conductivity (a decrease in σ toward low frequencies) causes a characteristic increase in ε' to about 4500 at $T = 5$ K at frequencies $\nu = 10$–20 cm^{-1}.

Figure 2 shows the temperature dependence of the static conductivity of TmSe. A specific feature near the magnetic phase transition temperature ($T_N = 3.5$ K) is clearly visible. Beginning from $T = 300$ K, σ_0 decreases; at $T_N < T < 20$–30 K, its behavior is described by the Kondo dependence [12] $\sigma_0 = (a - b\log T)^{-1}$ (where a and b are constants), which can be used to estimate the Kondo temperature for TmSe ($T_K \approx 20$–30 K). Note that, as in [9], at $T = 200$–300 K, the static conductivity is well below the low-frequency dynamic conductivity; at $T < 50$ K, this difference almost disappears.

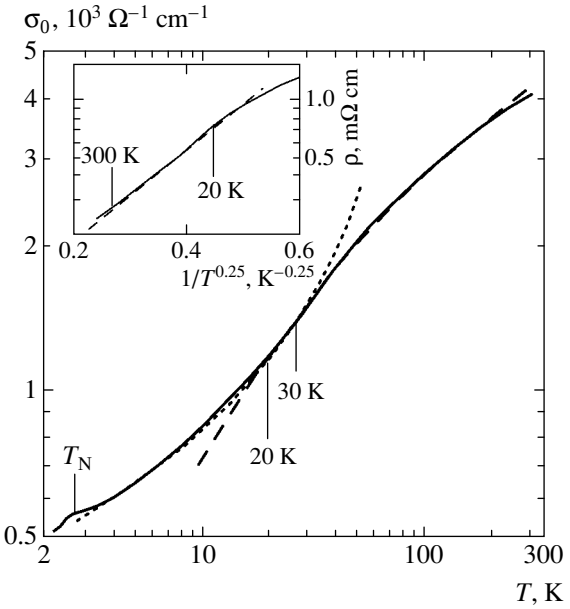


Fig. 2. Temperature dependence of the static conductivity of TmSe. The temperature T_N corresponds to the magnetic phase transition. Short dashes show the dependence processed by the least squares method using the formula $\sigma_0 = (a - b\log T)^{-1}$, which describes incoherent Kondo scattering. Long dashes illustrate the dependence processed by the formula $\sigma_0 \propto \exp(T_0/T)^{-1/4}$, which describes the conductivity of disordered systems. The inset shows the temperature behavior of the resistivity, which exhibits the Mott dependence of ρ in the temperature range 20–300 K (the dashed straight line).

3. DISCUSSION OF THE RESULTS

The detected features of the low-frequency response of TmSe can be explained using the schematic diagram of the density of states shown in Fig. 3 [2, 9]. Figure 3 shows a portion of this diagram, namely, the bottom of the 5*d* conduction band overlapping with the $4f^{13}$ level of the Tm^{2+} ion of the localized electrons. The hybridization of the 5*d* and 4*f* electrons results in a Kondo peak (resonance) of width W at the Fermi level E_F, and this peak includes mixed states of the $4f^{13}$–$4f^{12}5d$ type.

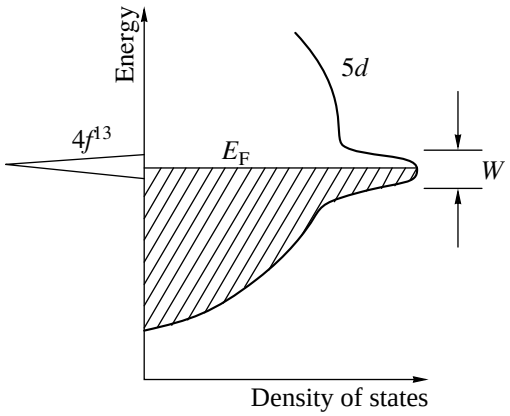


Fig. 3. Portion of the schematic diagram of the density of states for TmSe: the bottom of the conduction band, the 5*d*-electron band, and the $4f^{13}$ electronic level coinciding with the Fermi level. The Kondo resonance width is denoted as *W*.

We believe that the Kondo resonance in the density of states of TmSe exists even at room temperature. (Although the resonance is thought to appear only near T_K [2, 13], its presence at $T \gg T_K$ was experimentally observed for a number of heavy-fermion compounds [2, 14].) This means that, in the vicinity of $\pm W/2$ near the Fermi level, the electronic states of TmSe are hybridized and have a complex *f*–*d* character, whereas, at higher energies, the degree of hybridization decreases and electrons mainly have the *d* character. In this situation, probing electromagnetic radiation with a quantum energy of $h\nu < W$ should be sensitive to the hybridized *f*–*d* electrons, and the response at higher frequencies ($h\nu > W$) is determined by the *d* electrons. This behavior causes two corresponding Drude components in the frequency dependence of the dielectric response of TmSe at temperatures $T = 200$–300 K, namely, the low-frequency ($\nu < 100$ cm^{-1}) and high-frequency ($\nu > 100$ cm^{-1}) components. Using the $\varepsilon'(\nu)$ and $\sigma(\nu)$ dependences, we can estimate the width of the Kondo resonance: $W/h \approx 100$ cm^{-1}, or $W \approx 10$ meV. Then, using the relation $W \approx k_B T_K$ (where k_B is the

Parameters of the mobile *d* and hybridized *f–d* electrons for TmSe at $T = 200$–300 K: the plasma frequency ν_p, the relaxation frequency γ, the effective mass m^*, the concentration $n = \nu_p^2 m^*/4\pi e^2$, and the mobility $\mu = e/2\pi m^*\gamma$

	ν_p, cm^{-1}	γ, cm^{-1}	m^*/m_0	n, cm^{-3}	μ, $cm^2/V\,s$
d electrons	39000	4800	1.6	6.8×10^{20}	1.2
f–d electrons	2750	70	16	3.4×10^{19}	8

Boltzmann constant), which holds true for intermediate-valence systems [2, 15], we estimate the resonance width at a lower temperature: $W(T \approx T_K) \approx 2$–$3$ meV. In other words, an increase in the temperature leads to the broadening of the Kondo resonance, which is consistent with the theoretical calculations of [16].

The effects of hybridization and the corresponding renormalization of the effective mass and the electron relaxation frequency are usually analyzed within the framework of the so-called generalized Drude model [1, 17]. This model introduces a complex relaxation frequency

$$\Gamma(\omega) = \tau^{-1}(\omega) - i\omega\lambda(\omega),$$

which replaces the relaxation frequency in the standard expression for the complex Drude conductivity:

$$\sigma^*(\omega) = \sigma(\omega) + i\sigma_2(\omega) = \frac{\omega_p^2}{4\pi}[\Gamma(\omega) - i\omega]^{-1} = \frac{\omega_p^2}{4\pi}\left[\tau^{-1}(\omega) - i\omega\frac{m^*(\omega)}{m_b}\right]^{-1}. \quad (3)$$

Here, $\sigma_2 = \omega(\varepsilon - \varepsilon')/4\pi$, $\omega_p = 2\pi\nu_p$, τ is the relaxation time, m_b is the carrier mass in the conduction band, $\omega = 2\pi\nu$, and ε is the high-frequency permittivity. The quantity $\lambda(\omega) = m^*(\omega)/m_b - 1$ characterizes the increase in the effective mass and is related to the relaxation time τ by the Kramers–Kronig relations [17]. The relaxation frequency τ^{-1} and the effective mass m^* can be expressed in terms of σ and σ_2:

$$\tau^{-1}(\omega) = \frac{\omega_p^2\sigma(\omega)}{4\pi(\sigma^2 + \sigma_2^2)}, \quad (4)$$

$$\frac{m^*(\omega)}{m_b} = \frac{\omega_p^2\sigma_2(\omega)}{4\pi\sigma(\sigma^2 + \sigma_2^2)}. \quad (5)$$

We used Eqs. (4) and (5) to calculate the frequency dependences of $\gamma = (2\pi\tau)^{-1}$ and m^*/m_b shown in Fig. 1c. As is seen, the renormalization effects are negligible at high frequencies and the relaxation frequency is about 4200 cm^{-1}. This value corresponds to the relaxation of carriers in the *d* state and is close to the value (γ = 4800 cm^{-1}) calculated from the $\varepsilon'(\nu)$ and $\sigma(\nu)$ spectra with Drude relations (1) and (2). At $\nu < 100$ cm^{-1}, the relaxation frequency decreases, which means an increase in the lifetime of the hybridized *f–d* states. The effective mass behaves similarly: at high frequencies ($\nu > 100$ cm^{-1}), the response is determined by carriers in the *d* state, for which $m^* = m_b \approx 1.6m_0$ [9] (where m_0 is the free-electron mass), whereas, at $\nu < 100$ cm^{-1}, an addition of the *f* states results in an increase in the carrier mass to $m^* \approx 10m_b \approx 16m_0$.

Using the data obtained for the effective mass, we can calculate the microscopic parameters of the two components of mobile charges that are responsible for the high- and low-frequency responses of TmSe at T = 200–300 K, namely, the parameters of the uncorrelated conduction electrons and the heavy hybridized *f–d* electrons. The results obtained are given in the table.

Let us consider the frequency dependences of the dielectric spectrum of TmSe at low temperatures. We assume that the gap feature in the low-temperature conductivity spectra cannot be related to the opening of a hybridization gap in the density of states, as in the related intermediate-valence semiconductors (such as YbB_{12} or SmB_6) [2, 18, 19], due to the following reasons. The Hall constant in TmSe do not demonstrate activation behavior, and its magnetic susceptibility does not increase with decreasing temperature. These factors are two typical signs of the presence of a gap in the density of states [2]. Second, the paramagnetic TmSe phase contains an odd number of *f* and *d* electrons per unit cell; therefore, according to the Luttinger theorem [13], the hybridization of the mobile *d* and localized *f* electrons cannot cause the formation of an energy gap in the density of states [20, 21]. A hybridization gap can appear only during a phase transition into the antiferromagnetically ordered phase at $T = T_N$, when the unit cell of TmSe doubles and acquires an even number of *f* and *d* electrons. Indeed, such a gap (Δ_{hybr} = 1–2 meV) is observed in the infrared conductivity spectra [2]. Third, in the case of TmSe, the coherence temperature (below which quasiparticles should behave like the Fermi liquid and a hybridization gap should open) is equal to $T^* = \Delta_{hybr}/5k_B$ = 2–5 K [2, 22], which is well below the temperatures at which the low-frequency gap behavior of the conductivity spectra of TmSe is observed.

We believe that the decrease in the low-frequency ($\nu < 100$ cm^{-1}) conductivity of TmSe is related to the appearance of a mobility gap in the spectrum of electronic states at low temperatures and to the fact that, at the microscopic level, this gap is caused by conduction-electron localization at magnetic moments. A typical sign of incoherent electron scattering by magnetic moments is the characteristic behavior of the static con-

ductivity, $\sigma_0 = (a - b\log T)^{-1}$ [12], at $T < T_K$ (Fig. 2). This decrease in the static conductivity with decreasing temperature should result in the corresponding suppression of the low-frequency spectral weight in the conductivity spectrum up to the frequencies corresponding to the gap energy or the electron binding energy at a magnetic center, $k_B T_K$ = 1.7–2.6 meV (for T_K = 20–30 K). The gap width can also be determined from the measured dependence of the electromagnetic-radiation absorptivity, $\alpha = 4\pi k/\lambda$ (where k is the extinction coefficient and λ is the radiation wavelength), shown in the inset to Fig. 1 for T = 5 K. The smooth solid line demonstrates the spectrum processed by the least squares method using an expression for α at the edge of fundamental absorption in semiconductors caused by an energy gap Δ ($\alpha \propto \sqrt{-\Delta/h}$ [10]). The best fit is obtained at Δ = 2.5 meV, which agrees with the estimate calculated above. The dashed line in the inset shows the calculation result for $\Delta = 0$, and it again evidences the presence of a finite energy gap. Nonzero absorption at frequencies below Δ/h = 20 cm^{-1} is related to nonzero conductivity at $\nu < 100$ cm^{-1}.

Note that the specific features of the dielectric response detected in this work in TmSe can also be typical of other intermediate-valence and heavy-fermion compounds. As the temperature of such materials decreases, first (at $T = T_K$), incoherent electron scattering by magnetic moments appears, which results in the Kondo behavior of the resistivity ($\rho \propto -\log T$). In addition, this scattering should cause a mobility gap in the absorption spectrum (in the frequency dependence of the dynamic conductivity). The gap width is determined by the electron binding energy at a magnetic center. Upon a further decrease in the temperature and a transition into the coherent state at $T < T^*$, spin scattering becomes coherent and the mobility gap becomes a real hybridization gap. According to the reasons given above, such a gap in the case of TmSe appears only in the magnetically ordered phase below T_N = 3.5 K, which is found to be virtually coincident with the coherence temperature (T^* = 2–5 K).

Finally, we consider the temperature behavior of the dynamic and static conductivities and the resistivity of TmSe at temperatures $T > T_K$. As is seen from Fig. 2, the static conductivity and resistivity in this range are well described by Mott's relation $\sigma_0(T) \propto \exp(T_0/T)^{-1/4}$ (or $\rho(T) \propto \exp(T_0/T)^{1/4}$) (Fig. 2, inset), which is characteristic of disordered three-dimensional systems [23]. At a dimensionality n, this relation is written as

$$\rho(T) \propto \exp(T_0/T)^{1/(n+1)},$$

where T_0 is a constant. According to [23–25], the low-frequency conductivity of such systems increases with the frequency as $\sigma(\nu) \propto \nu^s$ ($s \sim 1$), which also agrees with our observations for TmSe: at $T > 50$ K, the dynamic conductivity at $\nu \approx 10$ cm^{-1} is significantly higher than the static conductivity. These two specific features of the temperature–frequency behavior of the conductivity of TmSe indicate that the localization or disordering effects can influence the transport properties of this compound at 50 K < T < 300 K. It should be noted, however, that the realization of the Mott mechanism of conduction at such high temperatures seems to be unlikely and that the nature of electron transport in TmSe at T = 50–300 K requires further investigation.

4. CONCLUSIONS

We measured the frequency dependences of the conductivity and permittivity of the intermediate-valence semiconductor TmSe using quasi-optical BWO and infrared Fourier spectroscopies in the frequency range 10–10^4 cm^{-1} and the temperature range 5–300 K. At T = 200–300 K, the spectra are determined by two subsystems of free carriers: light ($m^* = 1.6m_0$) conduction electrons and heavy ($m^* = 16m_0$) electronic states appearing due to the hybridization of the mobile d and localized f electrons. The microscopic parameters of both components, such as the concentration, mobility, relaxation frequency, and the plasma frequency, were determined. At T = 5 K, a gap was found in the dielectric spectra. We showed that its nature could not be caused by coherence effects in electron scattering by magnetic moments, as in the related intermediate-valence compounds (such as YbB_{12} or SmB_6). This gap is assumed to be associated with conduction-electron localization at magnetic moments, and it is assumed to be a mobility gap.

ACKNOWLEDGMENTS

We are grateful to P. Haas, D. Faltermeier, and G. Untereiner for technical assistance.

This work was supported by the fundamental research program "Radiophysics Problems," Department of Physical Sciences of the Russian Academy of Sciences, and by the German Research Society.

REFERENCES

1. L. Degiorgi, Rev. Mod. Phys. **71**, 687 (1999).
2. P. Wachter, *Intermediate Valence and Heavy Fermions* (North-Holland, Amsterdam, 1993).
3. G. V. Kozlov and A. A. Volkov, in *Millimeter and Submillimeter Wave Spectroscopy of Solids*, Ed. by G. Gruner (Springer, Berlin, 1998).
4. B. Gorshunov, N. Sluchanko, A. Volkov, *et al.*, Phys. Rev. B **59**, 1808 (1999).
5. M. Dressel, N. Kasper, K. Petukhov, *et al.*, Phys. Rev. B **66**, 035110 (2002).
6. H. Bjerrum-Moller, S. M. Shapiro, and R. J. Birgeneau, Phys. Rev. Lett. **39**, 1021 (1977).

7. T. Matsumura, S. Nakamura, T. Goto, *et al.*, J. Phys. Soc. Jpn. **67**, 612 (1998).
8. A. V. Sokolov, *Optical Properties of Metals* (Fizmatgiz, Moscow, 1961; Elsevier, New York, 1967).
9. B. Battlog, Phys. Rev. B **23**, 1827 (1981).
10. T. S. Moss, G. J. Burrell, and B. Ellis, *Semiconductor Opto-Electronics* (Butterworths, London, 1973; Mir, Moscow, 1976).
11. M. Tinkham, *Introduction to Superconductivity* (McGraw-Hill, New York, 1975; Atomizdat, Moscow, 1980).
12. A. A. Abrikosov, *Fundamentals of the Theory of Metals* (Nauka, Moscow, 1987; North-Holland, Amsterdam, 1988).
13. N. B. Brandt and V. V. Moshchalkov, Adv. Phys. **33**, 373 (1984).
14. B. Bucher, Z. Schlesinger, P. C. Canfield, and Z. Fisk, Phys. Rev. Lett. **72**, 522 (1994).
15. R. M. Martin, Phys. Rev. Lett. **48**, 362 (1982).
16. M. J. Rozenberg, G. Kotliar, and H. Kajueter, Phys. Rev. B **54**, 8452 (1996).
17. M. Dressel and G. Gruner, *Electrodynamics of Solids* (Cambridge Univ. Press, Cambridge, 2002).
18. A. Menth, E. Buehler, and T. H. Geballe, Phys. Rev. Lett. **22**, 295 (1969).
19. J. W. Allen, B. Batlogg, and P. Wachter, Phys. Rev. B **20**, 4807 (1979).
20. J. M. Luttinger, Phys. Rev. **119**, 1153 (1960).
21. R. M. Martin and J. W. Allen, J. Appl. Phys. **50**, 7561 (1979).
22. G. Güntherodt, W. A. Thompson, F. Holtzberg, and Z. Fisk, Phys. Rev. Lett. **49**, 1030 (1982).
23. N. F. Mott and E. A. Davis, *Electronic Processes in Non-Crystalline Materials* (Clarendon, Oxford, 1971; Mir, Moscow, 1974).
24. J. C. Dyre and T. B. Schroder, Rev. Mod. Phys. **72**, 873 (2000).
25. H. Bottger and V. V. Bryksin, *Hopping Conduction in Solids* (Akademie, Berlin, 1985).

Translated by K. Shakhlevich

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 919–925.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 1054–1060.
Original English Text Copyright © 2005 by Sergeenkov.

ELECTRONIC PROPERTIES OF SOLIDS

Polarization Effects Induced by a Magnetic Field in Intrinsically Granular Superconductors¶

S. Sergeenkov

Centro de Física das Interacções Fundamentais, Instituto Superior Técnico, 1049-001 Lisboa, Portugal
Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia
e-mail: ssa@thsun1.jinr.ru

Received June 3, 2005

Abstract—Based on the previously suggested model of nanoscale dislocation-induced Josephson junctions and their arrays, we study the magnetic-field-induced electric polarization effects in intrinsically granular superconductors. In addition to the new phenomenon of chemomagnetoelectricity, the model also predicts a few other interesting effects, including charge analogs of Meissner paramagnetism (at low fields) and a "fishtail" anomaly (at high fields). The conditions under which these effects can be experimentally measured in nonstoichiometric high-T_c superconductors are discussed. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Both granular superconductors and artificially prepared arrays of Josephson junctions (JJAs) proved useful in studying the numerous quantum (charging) effects, including blockade of Cooper pair tunneling [1], Bloch oscillations [2], propagation of quantum ballistic vortices [3], spin-tunneling related effects with specially designed SFS-type junctions [4, 5], novel Coulomb effects in SINIS-type nanoscale junctions [6], and recently observed geometric quantization phenomena [7] (see, e.g., [8] for a recent review on charge and spin effects in mesoscopic two-dimensional Josephson junctions).

More recently, it was realized that JJAs can also be used as quantum channels to transfer quantum information between distant sites [9–11] through the implementation of the so-called superconducting qubits, which involve both charge and phase degrees of freedom (see, e.g., [12] for a review on quantum-state engineering with Josephson-junction devices).

At the same time, imaging of the granular structure in underdoped $Bi_2Sr_2CaCu_2O_{8+\delta}$ crystals [13] revealed an apparent charge segregation of its electronic structure into superconducting domains (on the order of a few nanometers) located in an electronically distinct background. In particular, it was found that, at low levels of hole doping ($\delta \le 0.2$), the holes become concentrated at certain hole-rich domains. Tunneling between such domains leads to intrinsic granular superconductivity in high-T_c superconductors (HTS). As was shown earlier [14], granular superconductivity-based phenomena can shed some light on the origin and evolution of the so-called paramagnetic Meissner effect (PME), which manifests itself in both high-T_c and conventional superconductors [15, 16].

In this paper, within a previously suggested [14] model of JJAs created by a regular two-dimensional network of twin-boundary dislocations with strain fields acting as an insulating barrier between hole-rich domains in underdoped crystals, we address another class of interesting phenomena that are actually dual to the chemomagnetic effects described in [14]. Specifically, we discuss the possible existence of a nonzero electric polarization $P(B, \delta)$ (chemomagnetoelectric effect) and the related change of the charge balance in an intrinsically granular nonstoichiometric material under the influence of an applied magnetic field. In particular, we predict an anomalous low-field magnetic behavior of the effective junction charge $Q(B, \delta)$ and concomitant magnetocapacitance $C(B, \delta)$ in the paramagnetic Meissner phase and a charge analog of a "fishtail-like" anomaly at high magnetic fields.

2. MODEL

We recall that the regular two-dimensional dislocation networks of oxygen-depleted regions with the size d_0 of a few Burgers vectors, observed in HTS single crystals [13, 17–20], can provide a quite realistic possibility for the existence of a two-dimensional Josephson network within the CuO plane [21, 22]. In this regard, it is also important to mention the pioneering works by Khaikin and Khlyustikov [23–25] on twinning-induced superconductivity in dislocated crystals.

At the same time, in underdoped crystals, there is a realistic possibility to facilitate oxygen transport via the

¶ The text was submitted by the author in English.

so-called osmotic mechanism [14, 19, 20, 26], which relates the local value of the chemical potential

$$\mu(\mathbf{x}) = \mu(0) + \nabla\mu \cdot \mathbf{x}$$

with the local concentration of point defects as

$$c(\mathbf{x}) = \exp(-\mu(\mathbf{x})/k_B T),$$

and allows explicit incorporation of the oxygen deficiency parameter δ into our model by relating it to the excess oxygen concentration of vacancies $c_v \equiv c(0)$ as

$$\delta = 1 - c_v.$$

Assuming the relation between the variation of mechanical and chemical properties of planar defects,

$$\mu(\mathbf{x}) = K\Omega_0\epsilon(\mathbf{x}),$$

where

$$\epsilon(\mathbf{x}) = \epsilon_0\exp(-|\mathbf{x}|/d_0)$$

is the screened strain field produced by tetragonal regions in a *d*-wave orthorhombic YBCO crystal, Ω_0 is an effective atomic volume of the vacancy, and K is the bulk elastic modulus, we can study the properties of twin-boundary induced JJs under the intrinsic chemical pressure $\nabla\mu$ (created by the variation of the oxygen doping parameter δ). More specifically, a single SIS-type junction (comprising a Josephson network) is formed around the twin-boundary due to a local depression of the superconducting order parameter $\Delta(\mathbf{x}) \propto \epsilon(\mathbf{x})$ over distance d_0, thus producing a weak link with the Josephson coupling

$$J(\delta) = \epsilon(\mathbf{x})J_0 = J_0(\delta)\exp(-|\mathbf{x}|/d_0),$$

where

$$J_0(\delta) = \epsilon_0 J_0 = (\mu_v/K\Omega_0)J_0$$

(here, $J_0 \propto \Delta_0/R_n$, with R_n being the resistance of the junction). We note that, in accordance with observations, for a stoichiometric situation (when $\delta \approx 0$), the Josephson coupling $J(\delta) \approx 0$ and the system loses its explicitly granular signature.

To describe the influence of chemomagnetic effects on the charge balance of an intrinsically granular superconductor, we use the model of two-dimensional overdamped Josephson junction array based on the well-known Hamiltonian

$$\mathcal{H} = \sum_{i,j}^{N} J_{ij}(1 - \cos\phi_{ij}) + \sum_{i,j}^{N}\frac{q_i q_j}{2C_{ij}}. \qquad (1)$$

We introduce a short-range (nearest neighbor) interaction between N junctions (which are formed around oxygen-rich superconducting areas with phases ϕ_i), arranged in a two-dimensional lattice with coordinates $\mathbf{x_i} = (x_i, y_i)$. The areas are separated by oxygen-poor insulating boundaries (created by twin-boundary strain fields $\epsilon(\mathbf{x}_{ij})$), producing a short-range Josephson coupling

$$J_{ij} = J_0(\delta)\exp(-|\mathbf{x}_{ij}|/d).$$

Thus, typically for granular superconductors, the Josephson energy of the array varies exponentially with the distance $\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ between neighboring junctions (with d being the average junction size). As usual, the second term in the right-hand side of Eq. (1) accounts for Coulomb effects, where $q_i = -2en_i$ is the junction charge, with n_i being the pair number operator. Naturally, the same strain fields $\epsilon(\mathbf{x}_{ij})$ are also responsible for dielectric properties of oxygen-depleted regions via the δ-dependent capacitance tensor

$$C_{ij}(\delta) = C[\epsilon(\mathbf{x}_{ij})].$$

If, in addition to the chemical pressure

$$\nabla\mu(\mathbf{x}) = K\Omega_0\nabla\epsilon(\mathbf{x}),$$

the network of superconducting grains is under the influence of an applied frustrating magnetic field $\mathbf{B}$, the total phase difference through the contact is given by

$$\phi_{ij} = \phi_{ij}^0 + \frac{\pi w}{\Phi_0}(\mathbf{x}_{ij} \wedge \mathbf{n}_{ij}) \cdot \mathbf{B} + \frac{\nabla\mu \cdot \mathbf{x}_{ij} t}{\hbar}, \qquad (2)$$

where ϕ_{ij}^0 is the initial phase difference (see below),

$$\mathbf{n}_{ij} = \frac{\mathbf{X}_{ij}}{|\mathbf{X}_{ij}|}, \quad \mathbf{X}_{ij} = \frac{\mathbf{x}_i + \mathbf{x}_j}{2},$$

and

$$w = 2\lambda_L(T) + l,$$

with λ_L being the London penetration depth of the superconducting area and l, the insulator thickness (which, within the scenario discussed here, is simply equal to the twin-boundary thickness [26]).

As usual, to safely neglect the influence of the self-field effects in a real material, the corresponding Josephson penetration length

$$\lambda_J = \sqrt{\frac{\Phi_0}{2\pi\mu_0 j_c w}}$$

must be larger than the junction size d. Here, j_c is the critical current density of the superconducting (hole-

rich) area. As we see below, this condition is rather well satisfied for HTS single crystals.

3. CHEMOMAGNETOELECTRICITY

In what follows, we are interested in the behavior of the magnetic-field-induced electric polarization (chemomagnetoelectricity) in chemically induced granular superconductivity described by a two-dimensional JJA. We recall that a conventional (zero-field) pair polarization operator within the model under discussion is given by [27, 28]

$$\mathbf{p} = \sum_{i=1}^{N} q_i \mathbf{x}_i. \tag{3}$$

In view of Eqs. (1)–(3) and the usual "phase-number" commutation relation

$$[\phi_i, n_j] = i\delta_{ij},$$

it can be shown that the evolution of the pair polarization operator is determined by the equation of motion

$$\frac{d\mathbf{p}}{dt} = \frac{1}{i\hbar}[\mathbf{p}, \mathcal{H}] = \frac{2e}{\hbar}\sum_{i,j}^{N} J_{ij}\sin\phi_{ij}(t)\mathbf{x}_{ij}. \tag{4}$$

Solving this equation, we obtain the net value of the magnetic-field-induced longitudinal electric polarization

$$P(\delta, \mathbf{B}) \equiv \langle p_x(t)\rangle$$

(along the x axis) and the corresponding effective junction charge

$$Q(\delta, \mathbf{B}) = \frac{2eJ_0}{\hbar\tau d}\int_0^\tau dt\int_0^t dt' \times \int\frac{d^2x}{S}\sin\phi(\mathbf{x}, t')x\exp(-|\mathbf{x}|/d), \tag{5}$$

where $S = 2\pi d^2$ is the properly defined normalization area, τ is the characteristic time (see Discussion), and we made the usual substitution

$$\frac{1}{N}\sum_{i,j} A_{ij}(t) \longrightarrow \frac{1}{S}\int d^2x A(\mathbf{x}, t)$$

valid in the long-wavelength approximation [28].

To capture the very essence of the superconducting analog of the chemomagnetoelectric effect, we assume for simplicity in what follows that a stoichiometric sample (with $\delta \approx 0$) does not have any spontaneous polarization at zero magnetic field, that is, $P(0, 0) = 0$.

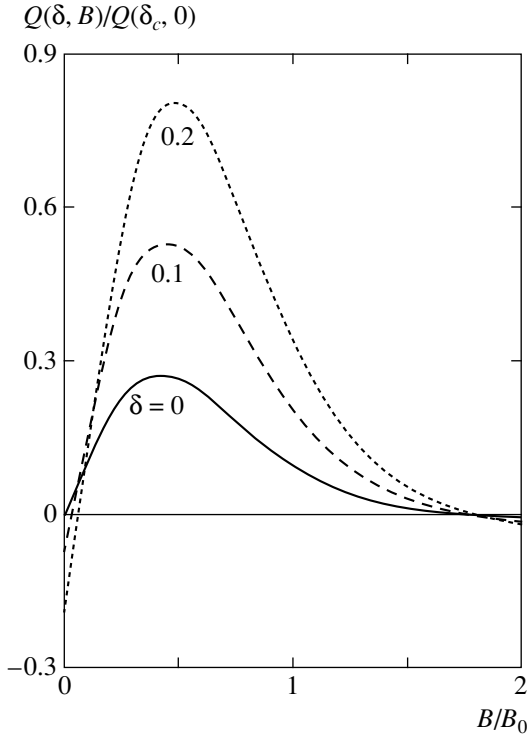


Fig. 1. The effective junction charge $Q(\delta, B)/Q(\delta_c, 0)$ (chemomagnetoelectric effect) as a function of the applied magnetic field B/B_0, according to Eq. (6), for different values of the oxygen deficiency parameter: $\delta \approx 0$ (solid line), $\delta = 0.1$ (dashed line), and $\delta = 0.2$ (dotted line).

According to Eq. (5), this condition implies that $\phi_{ij}^0 = 2\pi m$ for the initial phase difference with $m = 0, \pm1, \pm2, \ldots$.

Choosing the applied magnetic field along the c axis (and normal to the CuO plane), that is, $\mathbf{B} = (0, 0, B)$, we finally obtain

$$Q(\delta, B) = Q_0(\delta)\frac{2\tilde{b} + b(1-\tilde{b}^2)}{(1+b^2)(1+\tilde{b}^2)^2} \tag{6}$$

for the magnetic field behavior of the effective junction charge in chemically induced granular superconductors. Here,

$$Q_0(\delta) = e\tau J_0(\delta)/\hbar$$

with $J_0(\delta)$ defined earlier,

$$b = B/B_0, \quad \tilde{b} = b - b_\mu,$$

and

$$b_\mu = B_\mu/B_0 \approx (k_B T\tau/\hbar)\delta,$$

where

$$B_\mu(\delta) = (\mu_\nu\tau/\hbar)B_0$$

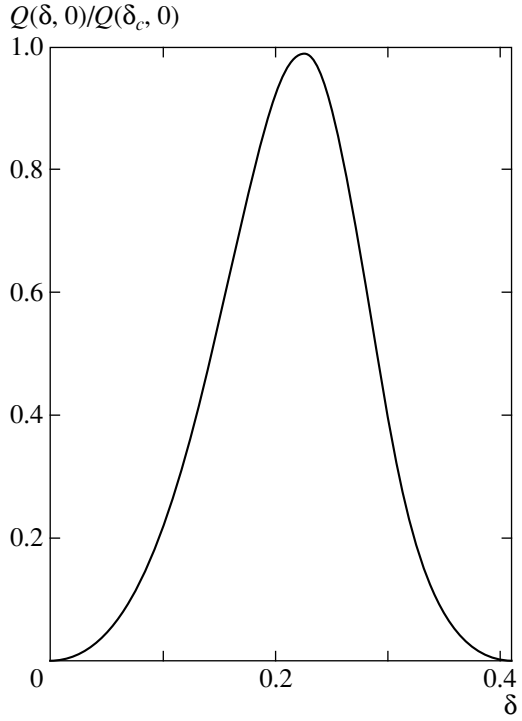


Fig. 2. Chemically induced effective junction charge $Q(\delta, 0)/Q(\delta_c, 0)$ in zero applied magnetic field (true chemoelectric effect).

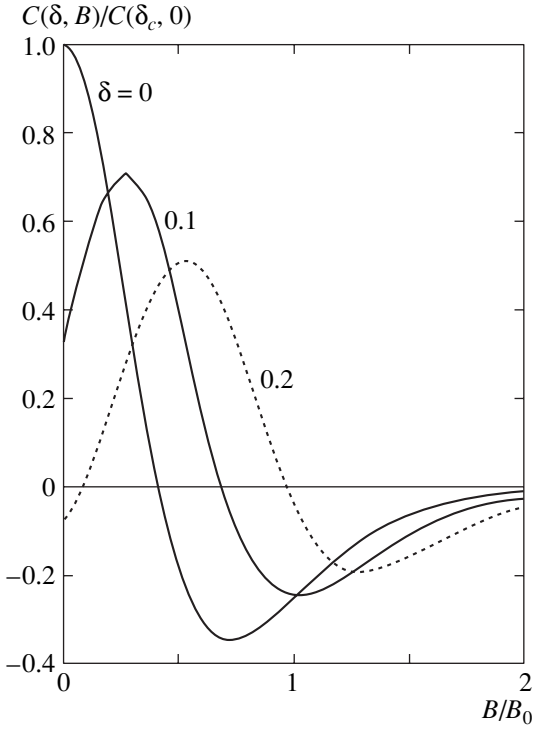


Fig. 3. The effective flux capacitance $C(\delta, B)/C(\delta_c, 0)$ as a function of the applied magnetic field B/B_0, according to Eq. (7), for different values of the oxygen deficiency parameter: $\delta \approx 0$ (solid line), $\delta = 0.1$ (dashed line), and $\delta = 0.2$ (dotted line).

is the chemically induced contribution (which disappears in optimally doped systems with $\delta \approx 0$) and

$$B_0 = \Phi_0/wd$$

is the characteristic Josephson field.

Figure 1 shows changes in the initial (stoichiometric) effective junction charge Q (solid line) with oxygen deficiency δ. We note a sign change of Q (dotted and dashed lines) driven by nonzero values of δ at low magnetic fields (a charge analog of a chemically induced PME). According to Eq. (6), the effective charge changes its sign as soon as the chemomagnetic contribution $B_\mu(\delta)$ exceeds the applied magnetic field B (see Discussion).

At the same time, Fig. 2 presents a true chemoelectric effect with the concentration (deficiency) induced effective junction charge $Q(\delta, 0)$ in zero magnetic field. We note that $Q(\delta, 0)$ exhibits a maximum around $\delta_c \approx 0.2$ (in agreement with the classical percolative behavior observed in nonstoichiometric $YBa_2Cu_3O_{7-\delta}$ samples [17]).

It is also of interest to consider the magnetic field behavior of the concomitant effective flux capacitance

$$C \equiv \frac{\tau dQ(\delta, B)}{d\Phi},$$

which, in view of Eq. (6), is given by

$$C(\delta, B) = C_0(\delta)\frac{1 - 3b\tilde{b} - 3\tilde{b}^2 + b\tilde{b}^3}{(1 + b^2)(1 + \tilde{b}^2)^3}, \qquad (7)$$

where

$$\Phi = SB, \quad C_0(\delta) = \tau Q_0(\delta)/\Phi_0.$$

Figure 3 depicts the behavior of the effective flux capacitance $C(\delta, B)$ in an applied magnetic field for different values of the oxygen deficiency parameter: $\delta \approx 0$ (solid line), $\delta = 0.1$ (dashed line), and $\delta = 0.2$ (dotted line). We note a decrease of the magnetocapacitance amplitude and its peak shifting with increase of δ and a sign change at low magnetic fields, which is another manifestation of the charge analog of a chemically induced PME (cf. Fig. 1).

4. CHARGE ANALOG OF THE "FISHTAIL" ANOMALY

So far, we neglected a possible field dependence of the chemical potential μ_v of oxygen vacancies. We recall, however, that, in sufficiently high applied magnetic fields B, the field-induced change of the chemical potential

$$\Delta\mu_v(B) \equiv \mu_v(B) - \mu_v(0)$$

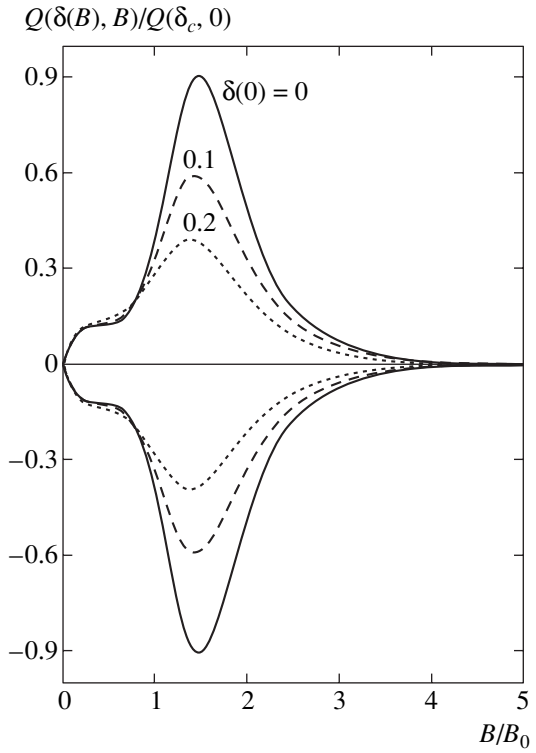


Fig. 4. A fishtail-like behavior of the effective charge $Q(\delta(B), B)/Q(\delta_c, 0)$ in the applied magnetic field B/B_0 in the presence of a magnetoconcentration effect (with field-induced oxygen vacancies $\delta(B)$) for three values of the field-free deficiency parameter: $\delta(0) \approx 0$ (solid line), $\delta(0) = 0.1$ (dashed line), and $\delta(0) = 0.2$ (dotted line).

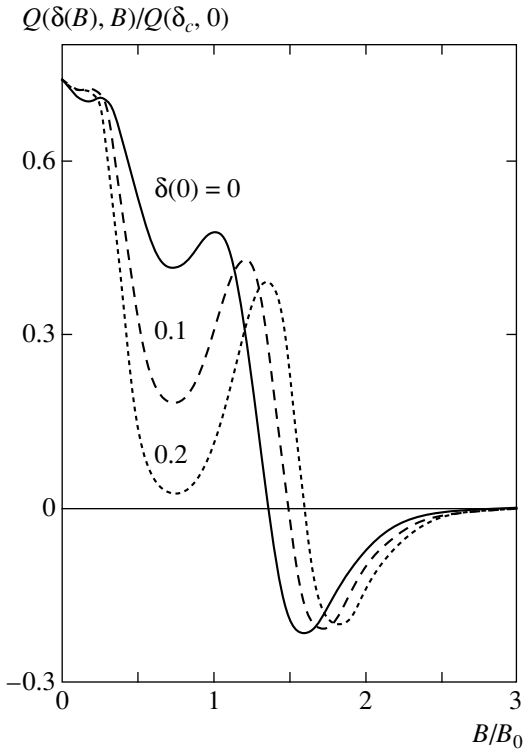


Fig. 5. The behavior of the effective flux capacitance $C(\delta(B), B)/C(\delta_c, 0)$ in the applied magnetic field B/B_0 in the presence of a magnetoconcentration effect for three values of the field-free deficiency parameter: $\delta(0) \approx 0$ (solid line), $\delta(0) = 0.1$ (dashed line), and $\delta(0) = 0.2$ (dotted line).

becomes tangible and should be taken into account [14, 29, 30]. As a result, we obtain a superconducting analog of the so-called magnetoconcentration effect [14] with field-induced creation of oxygen vacancies

$$c_v(B) = c_v(0)\exp(-\Delta\mu_v(B)/k_BT),$$

which, in turn, leads to a fishtail-like behavior of the high-field chemomagnetization (see [14] for more details).

Figure 4 shows the field behavior of the effective junction charge in the presence of the above-mentioned magnetoconcentration effect. As is clearly seen, $Q(\delta(B), B)$ exhibits a fishtail-like anomaly typical of the previously discussed [14] chemomagnetization in underdoped crystals with intragrain granularity (for symmetry and better visual effect, we also plotted $-Q(\delta(B), B)$ in the same figure). This more complex structure of the effective charge appears when the applied magnetic field B matches the intrinsic chemomagnetic field $B_\mu(\delta(B))$ (which now also depends on B via the magnetoconcentration effect). We note that the fishtail structure of $Q(\delta(B), B)$ manifests itself even at zero values of the field-free deficiency parameter $\delta(0)$ (solid line in Fig. 4), thus confirming the field-induced nature of intrinsic granularity [13, 17–20]. Likewise, Fig. 5 depicts the evolution of the effective flux capacitance $C(\delta(B), B)$ in the applied magnetic field B/B_0 in the presence of a magnetoconcentration effect (cf. Fig. 3).

5. DISCUSSION

Thus, the present model predicts the appearance of two interrelated phenomena (dual to the previously discussed behavior of chemomagnetism [14]), a charge analog of Meissner paramagnetism at low fields and a charge analog of the fishtail anomaly at high fields. To see whether these effects can be actually observed in a real material, we estimate the order of magnitude of the main model parameters.

Using the values $\lambda_L(0) \approx 150$ nm, $d \approx 10$ nm, and $j_c \approx 10^{10}$ A/m^2 typical [17, 19] of HTS single crystals, we estimate the characteristic field as $B_0 \approx 0.5$ T and the chemomagnetic field as $B_\mu(\delta) \approx 0.5B_0$. Therefore, the predicted charge analog of PME should be observable for applied magnetic fields $B < 0.25$ T. We note that, for the above set of parameters, the Josephson length is on the order of $\lambda_J \approx 1$ μm, which means that the small-junction approximation assumed in this paper is valid and the "self-field" effects can be safely neglected.

Furthermore, the characteristic frequencies $\omega \approx \tau^{-1}$ needed to probe the effects suggested here are related to the processes governed by tunneling relaxation times $\tau \approx \hbar/J_0(\delta)$. Because the deficiency parameter $\delta = 0.1$ for oxygen, the chemically induced zero-temperature Josephson energy in nonstoichiometric YBCO single crystals is on the order of $J_0(\delta) \approx k_B T_C \delta \approx 1$ meV, we obtain the required frequencies $\omega \approx 10^{13}$ Hz and the estimates of the effective junction charge $Q_0 \approx e = 1.6 \times 10^{-19}$ C and flux capacitance $C_0 \approx 10^{-18}$ F. We note that the above estimates fall into the range of parameters used in typical experiments for studying single-electron tunneling effects both in JJs and JJAs [1, 2, 12, 31], thus suggesting quite an optimistic possibility of observing the predicted field-induced effects experimentally in nonstoichiometric superconductors with pronounced networks of planar defects or in artificially prepared JJAs. (It is worth mentioning that a somewhat similar behavior of the magnetic-field-induced charge and related flux capacitance has been observed in 2D electron systems [32].)

Finally, it can be easily verified that, in view of Eqs. (1)–(5), the field-induced Coulomb energy of the oxygen-depleted region within our model is given by

$$E_C(\delta, B) \equiv \left\langle \sum_{i,j}^{N} \frac{q_i q_j}{2C_{ij}} \right\rangle = \frac{Q^2(\delta, B)}{2C(\delta, B)} \qquad (8)$$

with $Q(\delta, B)$ and $C(\delta, B)$ defined by Eqs. (6) and (7).

A thorough analysis of the above expression reveals that, in the PME state (when $B \ll B_\mu$), the chemically induced granular superconductor is always in the so-called Coulomb blockade regime (with $E_C > J_0$), while, in the fishtail state (for $B \geq B_\mu$), the energy balance tips in favor of tunneling (with $E_C < J_0$). In particular,

$$E_C(\delta, B = 0.1B_\mu) = \frac{\pi}{2} J_0(\delta)$$

and

$$E_C(\delta, B = B_\mu) = \frac{\pi}{8} J_0(\delta).$$

It would be also interesting to verify this phenomenon of field-induced weakening of the Coulomb blockade experimentally.

6. CONCLUSIONS

In conclusion, within a realistic model of two-dimensional Josephson junction arrays created by a two-dimensional network of twin boundary dislocations (with strain fields acting as an insulating barrier between hole-rich domains in underdoped crystals), a few novel electric polarization related effects expected to occur in an intrinsically granular material under applied magnetic fields were predicted, including a phenomenon of chemomagnetoelectricity, an anomalous low-field magnetic behavior of the effective junction charge (and flux capacitance) in the paramagnetic Meissner phase, and a charge analog of a fishtail-like anomaly at high magnetic fields, as well as field-dependent weakening of the chemically induced Coulomb blockade. The experimental conditions needed to observe the effects predicted here in nonstoichiometric high-T_c superconductors were discussed.

ACKNOWLEDGMENTS

This work was done during my stay at the Center for Physics of Fundamental Interactions (Instituto Superior Técnico, Lisboa) and was partially funded by the FCT. I thank P. Sacramento and V. Vieira for hospitality and interesting discussions on the subject. I am also indebted to the referee for drawing my attention to the pioneering works by Khaikin and Khlyustikov on twinning-induced superconductivity in dislocated crystals.

REFERENCES

1. M. Iansity, A. J. Johnson, C. J. Lobb, *et al.*, Phys. Rev. Lett. **60**, 2414 (1988).
2. D. B. Haviland, L. S. Kuzmin, P. Delsing, *et al.*, Z. Phys. B **85**, 339 (1991).
3. H. S. J. van der Zant, Physica B (Amsterdam) **222**, 344 (1996).
4. V. V. Ryazanov, V. A. Oboznov, A. Yu. Rusanov, *et al.*, Phys. Rev. Lett. **86**, 2427 (2001).
5. A. A. Golubov, M. Yu. Kupriyanov, and Ya. V. Fominov, JETP Lett. **75**, 588 (2002).
6. P. M. Ostrovsky and M. V. Feigel'man, JETP Lett. **79**, 489 (2004).
7. S. Sergeenkov and F. M. Araujo-Moreira, JETP Lett. **80**, 580 (2004).
8. I. V. Krive, S. I. Kulinich, R. I. Shekhter, *et al.*, Low Temp. Phys. **30**, 554 (2004).
9. L. B. Ioffe, M. V. Feigel'man, A. Ioselevich, *et al.*, Nature **415**, 503 (2002).
10. D. Born, V. I. Shnyrkov, W. Krechet, *et al.*, Phys. Rev. B **70**, 180501 (2004).
11. A. B. Zorin, Zh. Éksp. Teor. Fiz. **125**, 1423 (2004) [JETP **98**, 1250 (2004)].
12. Yu. Makhlin, G. Schön, and A. Shnirman, Rev. Mod. Phys. **73**, 357 (2001).
13. K. M. Lang, V. Madhavan, J. E. Hoffman, *et al.*, Nature **415**, 412 (2002).
14. S. Sergeenkov, JETP Lett. **77**, 94 (2003).
15. V. Kataev, N. Knauf, W. Braunisch, *et al.*, JETP Lett. **58**, 636 (1993); A. K. Geim, S. V. Dubonos, J. G. S. Lok, *et al.*, Nature **396**, 144 (1998).
16. C. De Leo and G. Rotoli, Phys. Rev. Lett. **89**, 167001 (2002).

17. M. Daeumling, J. M. Seuntjens, D. C. Larbalestier, *et al.*, Nature **346**, 332 (1990); V. F. Gantmakher, A. M. Neminskii, and D. V. Shovkun, JETP Lett. **52**, 630 (1990).
18. G. Yang, P. Shang, S. D. Sutton, *et al.*, Phys. Rev. B **48**, 4054 (1993).
19. A. Gurevich and E. A. Pashitskii, Phys. Rev. B **56**, 6213 (1997).
20. B. H. Moeckley, D. K. Lathrop, and R. A. Buhrman, Phys. Rev. B **47**, 400 (1993).
21. E. Z. Meilikhov, Zh. Éksp. Teor. Fiz. **110**, 1453 (1996) [JETP **83**, 803 (1996)].
22. S. Sergeenkov, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 36 (1999) [JETP Lett. **70**, 36 (1999)].
23. I. N. Khlyustikov and M. S. Khaikin, Zh. Éksp. Teor. Fiz. **75**, 1158 (1978) [Sov. Phys. JETP **48**, 583 (1978)]; M. S. Khaikin and I. N. Khlyustikov, JETP Lett. **33**, 158 (1981).
24. I. N. Khlyustikov and M. S. Khaikin, JETP Lett. **34**, 198 (1981); JETP Lett. **36**, 164 (1982); JETP Lett. **38**, 224 (1983).
25. I. N. Khlyustikov, Zh. Éksp. Teor. Fiz. **94** (3), 314 (1988) [Sov. Phys. JETP **67**, 607 (1988)]; **96**, 2073 (1989) [**69**, 1171 (1989)].
26. S. Sergeenkov, J. Appl. Phys. **78**, 1114 (1995).
27. S. Sergeenkov, J. Phys. I **7**, 1175 (1997).
28. S. Sergeenkov, JETP Lett. **76**, 170 (2002).
29. A. A. Abrikosov, *Fundamentals of the Theory of Metals* (Nauka, Moscow, 1987; Elsevier, Amsterdam, 1988).
30. S. Sergeenkov and M. Ausloos, Zh. Éksp. Teor. Fiz. **116**, 257 (1999) [JETP **89**, 140 (1999)].
31. P. J. M. van Bentum, H. van Kempen, L. E. C. van de Leemput, *et al.*, Phys. Rev. Lett. **60**, 369 (1988).
32. W. Chen, T. P. Smith, M. Buttiker, *et al.*, Phys. Rev. Lett. **73**, 146 (1994).

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 926–941.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 1061–1077.
Original Russian Text Copyright © 2005 by Slunyaev.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

A High-Order Nonlinear Envelope Equation for Gravity Waves in Finite-Depth Water

A. V. Slunyaev

Institute of Applied Physics, Russian Academy of Sciences, Nizhni Novgorod, 603950 Russia
e-mail: Slunyaev@hydro.appl.sci-nnov.ru

Received April 8, 2005

Abstract—A third-order nonlinear envelope equation is derived for surface waves in finite-depth water by assuming small wave steepness, narrow-band spectrum, and small depth as compared to the modulation length. A generalized Dysthe equation is derived for waves in relatively deep water. In the shallow-water limit, one of the nonlinear dispersive terms vanishes. This limit case is compared with the envelope equation for waves described by the Korteweg–de Vries equation. The critical regime of vanishing nonlinearity in the classical nonlinear Schrödinger equation for water waves (when $kh \approx 1.363$) is analyzed. It is shown that the modulational instability threshold shifts toward the shallow-water (long-wavelength) limit with increasing wave intensity. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The classical nonlinear Schrödinger (NLS) equation was originally derived in nonlinear optics [1, 2] and subsequently applied to various problems in physics [3]. For water waves, it was derived in [4–6]. The dynamics described by the equation for water waves can be qualitatively different, depending on the relative values of coefficients, which are functions of water depth. As applied to waves in relatively deep water ($kh > 1.363$, where k is the carrier wavenumber and h is the unperturbed water depth), the equation describes the nonlinear modulational instability discovered in [7] and subsequently observed in experiments. In the opposite limit, the modulated wavetrain is stable. For three-dimensional water waves described by modified NLS equations, the structure of the modulational instability domain in the parameter space is more complicated [8, 9].

The NLS equation is a uniquely versatile integrable nonlinear partial differential equation. In many cases, its analysis can be simplified by invoking the theory of its soliton solutions. However, its applicability is restricted in terms of nonlinearity magnitude and spectral width. These disadvantages are frequently eliminated by adding higher order nonlinear and dispersive terms. Models of this kind are currently used to describe short pulse propagation in optical fiber links [10]. The Dysthe equation originally derived in [11] and its subsequent modifications are most widely employed to describe ocean wave propagation. Related equations (for different physical variables) were presented in different forms in various papers. Frequently, coefficients in these equations were mutually inconsistent. Most of these equations were derived for waves in infinitely deep water. Recently, a closed second-order nonlinear Schrödinger equation was derived for the displacement envelope [12]:

$$i\left(\frac{\partial A}{\partial t} + V\frac{\partial A}{\partial x}\right) + \varepsilon\beta_1\frac{\partial^2 A}{\partial x^2} + \varepsilon\alpha_1|A|^2 A + i\varepsilon^2\beta_2\frac{\partial^3 A}{\partial x^3} + i\varepsilon^2\alpha_{21}|A|^2\frac{\partial A}{\partial x} + i\varepsilon^2\alpha_{22}A^2\frac{\partial A^*}{\partial x} = 0, \quad (1)$$

where ε is a small parameter. Equation (1) is reduced to the classical NLS equation by dropping the terms of order ε^2. In the limit of infinitely deep water, the coefficients in (1) obtained in [12] are equal to those in the equations used by Dysthe and Trulsen with coworkers in their recent publications (e.g., see [13]). Equation (1) arises in other applications as well. A review of papers focused on generalized NLS equations can be found in [12].

Modulational instability is an important property of NLS-type equations. The modulational instability vanishes as the depth decreases below the critical value corresponding to $kh \approx 1.363$, because the nonlinear coefficient α_1 changes sign. Since the cubic nonlinearity is anomalously weak in the critical region, quintic nonlinearity must be taken into account to ensure correct description of nonlinear effects. For equations derived by the conventional asymptotic method, this requirement is dictated by a change in the relative values of the small nonlinearity and dispersion parameters. In particular, this approach was used in [14, 15]. Originally, the nonlinear equation for $kh \approx 1.363$ was derived in [14], where the coefficients corresponding to the critical region were given and the modulational instability domain was analyzed. In [15], an alternative derivation

of the equation (written in somewhat different form) was presented:

$$i\left(\frac{\partial A}{\partial t} + V\frac{\partial A}{\partial x}\right) + \varepsilon\beta_1\frac{\partial^2 A}{\partial x^2} + \sqrt{\varepsilon}\alpha_1|A|^2A + i\varepsilon\alpha_{21}|A|^2\frac{\partial A}{\partial x} + i\varepsilon\alpha_{22}A^2\frac{\partial A^*}{\partial x} + \varepsilon\alpha_{31}|A|^4A = 0 \quad (2)$$

with $\alpha_1 = O(\varepsilon^{1/2})$ and coefficients expressed as functions of depth. In addition to the higher order nonlinear term, Eq. (2) contains nonlinear dispersive terms. It was noted that the coefficients in the new equation were inconsistent with those in [14], and different conclusions were made about the instability domain.

Equations (1) and (2) contain similar nonlinear dispersive terms. However, different coefficients of these terms were obtained in [12, 14, 15]; in particular, they increase indefinitely as $kh \longrightarrow \infty$ [15]. Equation (2) was derived from a generalized Korteweg–de Vries (KdV) equation in [16], where an analysis of modulational instability was also presented. Vanishing nonlinearity can be associated with new effects in dynamics of large-amplitude waves (e.g., see [17, 18]).

In this paper, the conventional asymptotic expansion scheme is used to derive an equation describing weakly nonlinear, weakly modulated finite-depth water waves. The equation is the generalization of (1) taking into account the next-order nonlinear terms. In the shallow-water limit, the generalized NLS equation derived here is compared with the envelope equation for waves based on the Korteweg–de Vries equation. It is shown that the nonlinear dispersion coefficients in the derived equation increase indefinitely (to the asymptotic order taken into account) as $kh \longrightarrow \infty$. A next-order generalization of the Dysthe equation is obtained for deep-water waves, with finite coefficients. In the special case of $kh \longrightarrow 1.363$, the generalized NLS equation has the form of (2). An analysis of modulational instability using the calculated coefficients is presented.

2. STATEMENT OF THE PROBLEM

The classical statement of the problem is considered (e.g., see [12, 19]). Since the calculations required to derive the coefficients in the envelope equation present a formidable task, the Maple symbolic computation package was employed. Here, the procedure is presented in a form suitable for algorithmic implementation. Consider weakly nonlinear waves on the surface of an incompressible inviscid fluid in the presence of gravity. The system of equations consists of two boundary conditions on the fluid surface,

$$\varphi_z = \eta_t + \varphi_x\eta_x, \quad z = \eta, \quad (3)$$

$$\varphi_t + g\eta + \frac{1}{2}(\varphi_x^2 + \varphi_z^2) = 0, \quad z = \eta, \quad (4)$$

and the Laplace equation for the "bulk" water motion supplemented with the bottom impermeability condition,

$$\Delta\varphi = 0, \quad -h \le z \le \eta, \quad (5)$$

$$\frac{\partial\varphi}{\partial z} = 0, \quad z = -h. \quad (6)$$

Here, z and x are the vertical and horizontal coordinates, respectively; $\eta(x, t)$ is the surface displacement; and $\varphi(x, z, t)$ is the velocity potential ($\mathbf{v} = \nabla\varphi$). The assumption of potential flow is frequently used in analyses of surface waves. For weakly modulated waves (with narrow-band spectrum), the solution can be represented as a superposition of harmonics:

$$\begin{pmatrix} \varphi(x, z, t) \\ \eta(x, z, t) \end{pmatrix} = \delta\sum_n \begin{pmatrix} \varphi_n(x, z, t) \\ \eta_n(x, z, t) \end{pmatrix} E^n,$$
$$E^n = \begin{cases} 1, & n = 0, \\ (1/2)\exp(in\omega t - ink x), & n \ne 0, \end{cases} \quad (7)$$

where ω and k denote the frequency and wavevector of the fundamental harmonic (carrier wave). Small-amplitude approximation is introduced by assuming that $\delta \ll 1$ in (7). To ensure that $\varphi(x, z, t)$ and $\eta(x, t)$ are real-valued functions, the following conditions must hold:

$$\varphi_{-n} = \varphi_n^*, \quad \eta_{-n} = \eta_n^*.$$

Since the surface displacement is small, the conditions on the boundary $z = \eta(x, t)$ can be set for $z = 0$ after the functions of depth are represented by Taylor series expansions. Then, system (3), (4) is rewritten as

$$\sum_{j=0}^{\infty}\frac{\eta^j\partial_z^{j+1}\varphi}{j!} = \eta_t + \eta_x\sum_{j=0}^{\infty}\frac{\eta^j\partial_z^j\varphi_x}{j!}, \quad z = 0, \quad (8)$$

$$\sum_{j=0}^{\infty}\frac{\eta^j\partial_z^j\varphi_t}{j!} + g\eta + \frac{1}{2}\left(\sum_{j=0}^{\infty}\frac{\eta^j\partial_z^j\varphi_x}{j!}\right)^2 + \frac{1}{2}\left(\sum_{j=0}^{\infty}\frac{\eta^j\partial_z^{j+1}\varphi}{j!}\right)^2 = 0, \quad z = 0. \quad (9)$$

Weak dispersion, i.e., slow variation of the envelope as compared to the carrier wavelength, is taken into account by introducing a slow coordinate x_1,

$$\frac{\partial}{\partial x} \longrightarrow \frac{\partial}{\partial x_0} + \mu\frac{\partial}{\partial x_1}, \quad \mu \ll 1, \quad (10)$$

and slow times t_i,

$$\frac{\partial}{\partial t} \longrightarrow \frac{\partial}{\partial t_0} + \varepsilon\frac{\partial}{\partial t_1} + \varepsilon^2\frac{\partial}{\partial t_2} + \varepsilon^3\frac{\partial}{\partial t_3} + \varepsilon^4\frac{\partial}{\partial t_4} + \ldots, \quad \varepsilon \ll 1. \tag{11}$$

The solution is represented as a series expansion in powers of ε:

$$\begin{pmatrix} \varphi_n \\ \eta_n \end{pmatrix} = \sum_{m=0}^{\infty} \varepsilon^m \begin{pmatrix} \varphi_{nm} \\ \eta_{nm} \end{pmatrix}. \tag{12}$$

Thus, the potential and displacement are represented as

$$\varphi(x, z, t) = \delta \sum_{n=-\infty}^{\infty} E^n \sum_{m=0}^{\infty} \varepsilon^m \varphi_{nm}(x_0, x_1, z, t_i), \quad i = 1, 2, \ldots, \tag{13}$$

$$\eta(x, t) = \delta \sum_{n=-\infty}^{\infty} E^n \sum_{m=0}^{\infty} \varepsilon^m \eta_{nm}(x_0, x_1, t_i), \quad i = 1, 2, \ldots. \tag{14}$$

The relative values of the small parameters δ and μ may vary depending on the relative strength of dispersion effects and nonlinearity. The small parameter ε is used to represent the desired solution as an asymptotic series. It should be specified by taking into account the relative values of δ and μ. In the classical NLS equation, where the cubic nonlinearity and the first dispersive correction correspond to the same asymptotic order, it is required that $\varepsilon \sim \mu \sim \delta$. In what follows, it is assumed that $\mu = \varepsilon$ and $\delta = \varepsilon$; i.e., a single small parameter ε is used.

Since the dominant contribution to the solution is due to the first harmonic, only the zeroth, first, and "minus first" (conjugate) harmonics are retained: $\varphi_{n0} = 0$ for $|n| > 1$. The terms with φ_{00} represent the mean wave-induced flow, and $\eta_{n0} = 0$ for $n \neq \pm 1$. Since all harmonics other than the zeroth one are excited by nonlinear interaction between components, it can be assumed that $\varphi_{nm} = 0$ and $\eta_{nm} = 0$ for $|n| > m$. These conditions are imposed to reduce the required calculations, even though can be obtained in the course of analysis.

Next, expressions (13) and (14) are substituted into Eqs. (8) and (9), change of variables is performed according to (10) and (11), and the terms corresponding to the same order of harmonic (exponent n of E) and the same asymptotic order (exponent m of ε) are collected alternately. Hereinafter, the resulting equations are labeled by the double index $\{n, m\}$. To derive a closed system of equations, Laplace equation (5) subject to boundary condition (6) is solved. Then, Eq. (9) is used to successively find the displacement components η_{nm}. Finally, these expressions are substituted into Eq. (8) to obtain the desired evolution equation.

3. SOLUTION OF THE LAPLACE EQUATION

Consider Eq. (5) supplemented with boundary condition (6). By using the slow coordinates and series expansions of functions introduced above, it is rewritten as

$$L_n\varphi_n - 2\mu ink\partial_{x_1}\varphi_n + \mu^2\partial_{x_1}^2\varphi_n = 0, \quad L_n = \partial_z^2 - n^2k^2,$$

for all n or

$$L_n\varphi_{nm} - 2ink\partial_{x_1}\varphi_{n,m-1} + \partial_{x_1}^2\varphi_{n,m-2} = 0 \tag{15}$$

for all n and m. Expansion (15) of the Laplace equation can be solved order by order for $m \geq 0$. Its solutions can be written as

$$\{0, 0\}: \varphi_{00} = A_{00}(x_i, t_i), \tag{16}$$

$$\{0, 1\}: \varphi_{01} = A_{01}(x_i, t_i), \tag{17}$$

$$\{0, 2\}: \varphi_{02} = A_{02}(x_i, t_i) - \frac{1}{2}(z + h)^2\partial_{x_1}^2 A_{00}, \tag{18}$$

$$\{0, 3\}: \varphi_{03} = A_{03}(x_i, t_i) - \frac{1}{2}(z + h)^2\partial_{x_1}^2 A_{01}, \tag{19}$$

$$\{0, 4\}: \varphi_{04} = A_{04}(x_i, t_i) - \frac{1}{2}(z + h)^2\partial_{x_1}^2 A_{02} + \frac{1}{24}(z + h)^4\partial_{x_1}^4 A_{00} \tag{20}$$

for $0 \leq m \leq 4$ and $n = 0$ and

$$\{n, 0\}: \varphi_{n0} = A_{n0}(x_i, t_i)\frac{\cosh nk(z + h)}{\cosh nkh}, \tag{21}$$

$$\{n, 1\}: \varphi_{n1} = A_{n1}(x_i, t_i)\frac{\cosh nk(z + h)}{\cosh nkh} + i\partial_{x_1}A_{n0}(z + h)\frac{\sinh nk(z + h)}{\cosh nkh}, \tag{22}$$

$$\{n, 2\}: \varphi_{n2} = \left(A_{n2}(x_i, t_i) - \partial_{x_1}^2 A_{n,0}\frac{(z + h)^2}{2}\right) \times \frac{\cosh nk(z + h)}{\cosh nkh} + i\partial_{x_1}A_{n1}(z + h)\frac{\sinh nk(z + h)}{\cosh nkh}, \tag{23}$$

$$\{n, 3\}: \varphi_{n3} = \left(A_{n3}(x_i, t_i) - \partial_{x_1}^2 A_{n1}\frac{(z + h)^2}{2}\right) \times \frac{\cosh nk(z + h)}{\cosh nkh} + \left(i\partial_{x_1}A_{n2}(z + h) - i\partial_{x_1}^3 A_{n0}\frac{(z + h)^3}{6}\right) \tag{24}$$

$$\times \frac{\sinh nk(z+h)}{\cosh nkh},$$

$$\{n, 4\}: \varphi_{n4} = \left(A_{n4}(x_i, t_i) - \partial_{x_1}^2 A_{n2}\frac{(z+h)^2}{2} + \partial_{x_1}^4 A_{n0}\frac{(z+h)^4}{24}\right)\frac{\cosh nk(z+h)}{\cosh nkh} + \left(i\partial_{x_1}A_{n3}(z+h) - i\partial_{x_1}^3 A_{n1}\frac{(z+h)^3}{6}\right) \times \frac{\sinh nk(z+h)}{\cosh nkh} \quad (25)$$

for $n \neq 0$. The functions $A_{nm}(x_i, t_i)$ in (16)–(25) are yet undetermined.

4. CLASSICAL NLS EQUATION

Order {1, 0} terms are used to find the dispersion relation

$$\omega = \sqrt{kg\sigma}, \quad \sigma = \tanh(kh). \quad (26)$$

Order {1, 1} terms yield

$$\frac{\partial A_{10}}{\partial t_1} + V\frac{\partial A_{10}}{\partial x_1} = 0. \quad (27)$$

This equation describes linear waves propagating in a medium characterized by the dispersion relation (26) with the group velocity

$$V = \frac{g}{2\omega}(\sigma + kh(1-\sigma^2)). \quad (28)$$

Order {1, 2} terms lead to an evolution equation containing the zeroth and second harmonics. The equation for the zeroth harmonic derived by collecting order {0, 2} terms is

$$gh\frac{\partial^2 A_{00}}{\partial x_1^2} - \frac{\partial^2 A_{00}}{\partial t_1^2} = \Gamma_1\frac{\partial}{\partial x_1}|A_{10}|^2. \quad (29)$$

Equation (29) is combined with (27) to obtain

$$\left(\frac{\partial}{\partial t_1} + V\frac{\partial}{\partial x_1}\right)\left(\frac{\partial^2}{\partial t_1^2} - V_{\text{long}}^2\frac{\partial^2}{\partial x_1^2}\right)A_{00} = 0, \quad V_{\text{long}} = \sqrt{gh}, \quad (30)$$

where V_{long} is the linear long-wave velocity corresponding to the free solution for the zeroth harmonic. In the model considered here,

$$\frac{\partial A_{00}}{\partial t_1} + V\frac{\partial A_{00}}{\partial x_1} = 0; \quad (31)$$

i.e., the leading-order zeroth harmonic is assumed to propagate with the velocity of the driving (surface) wave. Then, Eq. (29) becomes

$$\frac{\partial A_{00}}{\partial x_1} = \gamma_1|A_{10}|^2, \quad \gamma_1 = \Gamma_1/V_{\text{d}}^2, \quad V_{\text{d}}^2 = gh - V^2. \quad (32)$$

The expressions for the coefficients obtained below are given in the Appendix.

Order {2, 1} terms yield an expression for the second harmonic having the form

$$A_{21} = i\chi_1 A_{10}^2. \quad (33)$$

Order {1, 2} terms combined with (27), (31), and (33) lead to the equation

$$i\frac{\partial A_{10}}{\partial t_2} + \beta_1\frac{\partial^2 A_{10}}{\partial x_1^2} + \tilde{\rho}_{11}|A_{10}|^2 A_{10} + \tilde{\rho}_{12}A_{10}\frac{\partial A_{00}}{\partial x_1} = 0. \quad (34)$$

A closed evolution equation is derived from (34) and (32):

$$i\frac{\partial A_{10}}{\partial t_2} + \beta_1\frac{\partial^2 A_{10}}{\partial x_1^2} + \tilde{\alpha}_1|A_{10}|^2 A_{10} = 0. \quad (35)$$

Since an equation for the displacement is to be derived, we use the relation between velocity potential and displacement

$$\eta_{10} = i\lambda A_{10}, \quad \lambda = -\frac{\omega}{g}, \quad (36)$$

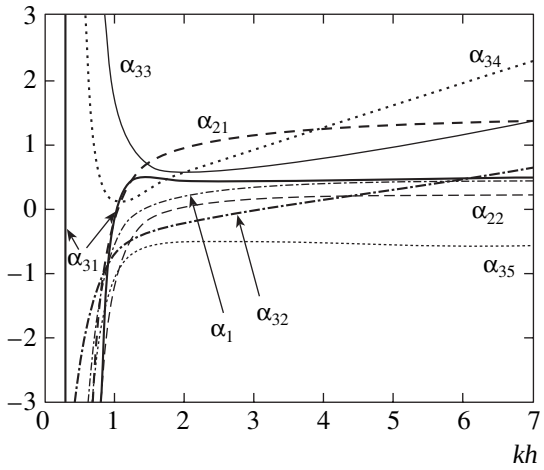
which is obtained by solving Eq. (9). In physical variables, the NLS equation is written as

$$i\left(\frac{\partial \eta_{10}}{\partial t} + V\frac{\partial \eta_{10}}{\partial x}\right) + \varepsilon\beta_1\frac{\partial^2 \eta_{10}}{\partial x^2} + \varepsilon\alpha_1|\eta_{10}|^2\eta_{10} = 0, \quad \alpha_1 = \tilde{\alpha}_1/\lambda^2, \quad (37)$$

and the surface displacement is given by the formula

$$\eta = \text{Re}(\eta_{10}\exp(i\omega t - ikx)). \quad (38)$$

Equation (37) is the classical first-order NLS equation with well-known coefficients. In particular, the coefficient β_1 is strictly positive, whereas the nonlinear coefficient α_1 changes sign as $kh \longrightarrow 1.363$ (see fig-



Dimensionless nonlinear coefficients in Eq. (66) as functions of *kh* (see table).

ure), which implies a change from focusing (for $\alpha_1 > 0$) to defocusing type of Eq. (37).

5. SECOND-ORDER NONLINEAR EVOLUTION EQUATION (SEDLETSKY EQUATION)

The second-order nonlinear NLS equation for finite-depth water waves was derived in [12]. On the whole, the present analysis follows that derivation, except for a different treatment of the derivative of the zeroth-harmonic velocity potential with respect to t_2. It was assumed in [12] that the propagation velocity of the leading-order zeroth harmonic (φ_{00}) is V not only to the first (as in (31)), but also to the next, asymptotic order. However, this assumption is inconsistent with relation (32) between the first and zeroth harmonics. By virtue of (32), the derivative of A_{00} with respect to t_2 is calculated by using (35) and does not vanish. When the corresponding term is retained, corrections to nonlinear dispersion coefficients must be taken into account (however, this does not entail any change in the limit of $kh \longrightarrow \infty$).

To the second asymptotic order, the surface displacement is expressed as

$$\begin{aligned} \eta_1 &= \eta_{10} + \varepsilon\eta_{11} + O(\varepsilon^2), \\ \eta_{11} &= i\lambda\left(A_{11} + ip_1\frac{\partial A_{10}}{\partial x_1}\right). \end{aligned} \tag{39}$$

The equation derived here describes the evolution of the combination $A_{10} + \varepsilon A_{11}$. The function A_{11} (the homogeneous part of solution (22) to the Laplace equation) is yet undetermined. To derive an equation for displacement, functions of the potential must be expressed in terms of the displacement represented by (39). The calculations can be simplified by using the fact that A_{11} is an arbitrary function and setting,

$$A_{11} = -ip_1\frac{\partial A_{10}}{\partial x_1}, \tag{40}$$

in which case

$$\eta_1 = i\lambda A_{10} + O(\varepsilon^2). \tag{41}$$

The second-order nonlinear evolution equation for the first harmonic derived from the asymptotic expansion to order {1, 3} is

$$\begin{aligned} &i\frac{\partial A_{10}}{\partial t_3} + i\beta_2\frac{\partial^3 A_{10}}{\partial x_1^3} + iP_{21}|A_{10}|^2\frac{\partial A_{10}}{\partial x_1} \\ &+ iP_{22}A_{10}^2\frac{\partial A_{10}^*}{\partial x_1} + sA_{10}\frac{\partial A_{00}}{\partial t_2} + \tilde{\rho}_{12}A_{10}\frac{\partial A_{01}}{\partial x_1} = 0. \end{aligned} \tag{42}$$

An expression for the second to last term in (42) can be found by using (32) and (35):

$$\frac{\partial A_{00}}{\partial t_2} = -i\beta_1\gamma_1\left(A_{10}\frac{\partial A_{10}^*}{\partial x_1} - A_{10}^*\frac{\partial A_{00}}{\partial x_1}\right). \tag{43}$$

(The corresponding contribution to nonlinear dispersion coefficients was ignored in [12].) Then, (42) is rewritten as

$$\begin{aligned} &i\frac{\partial A_{10}}{\partial t_3} + i\beta_2\frac{\partial^3 A_{10}}{\partial x_1^3} + i\tilde{\rho}_{21}|A_{10}|^2\frac{\partial A_{10}}{\partial x_1} \\ &+ i\tilde{\rho}_{22}A_{10}^2\frac{\partial A_{10}^*}{\partial x_1} + \tilde{\rho}_{12}A_{10}\frac{\partial A_{01}}{\partial x_1} = 0. \end{aligned} \tag{44}$$

Equation (44) contains a zeroth-harmonic term. The equation for this term is derived from the asymptotic expansion to order {0, 3}:

$$\frac{\partial A_{01}}{\partial x_1} = i\gamma_2\left(A_{10}\frac{\partial A_{10}^*}{\partial x_1} - A_{10}^*\frac{\partial A_{10}}{\partial x_1}\right). \tag{45}$$

In addition to the relations found above, the derivation of Eqs. (44) and (45) makes use of the condition

$$\frac{\partial A_{01}}{\partial t_1} + V\frac{\partial A_{01}}{\partial x_1} = 0 \tag{46}$$

and the expression

$$A_{22} = \chi_2 A_{10}\frac{\partial A_{10}}{\partial x_1} \tag{47}$$

for the second-harmonic component derived from the asymptotic expansion to order {2, 2}. Equations (44)

and (45) are combined to obtain the second-order evolution equation

$$i\frac{\partial A_{10}}{\partial t_3} + i\beta_2\frac{\partial^3 A_{10}}{\partial x_1^3} + i\tilde{\alpha}_{21}|A_{10}|^2\frac{\partial A_{10}}{\partial x_1} + i\tilde{\alpha}_{22}A_{10}^2\frac{\partial A_{10}^*}{\partial x_1} = 0. \quad (48)$$

Combining (48) with (27) and (35) and using (41), we obtain a complete second-order equation for the displacement envelope:

$$i\left(\frac{\partial \eta_{10}}{\partial t} + V\frac{\partial \eta_{10}}{\partial x}\right) + \varepsilon\beta_1\frac{\partial^2\eta_{10}}{\partial x^2} + \varepsilon\alpha_1|\eta_{10}|^2\eta_{10} + i\varepsilon^2\beta_2\frac{\partial^3\eta_{10}}{\partial x^3} + i\varepsilon^2\alpha_{21}|\eta_{10}|^2\frac{\partial\eta_{10}}{\partial x} + i\varepsilon^2\alpha_{22}\eta_{10}^2\frac{\partial\eta_{10}^*}{\partial x} = 0, \quad (49)$$

$$\alpha_{21} = \tilde{\alpha}_{21}/\lambda^2, \quad \alpha_{22} = \tilde{\alpha}_{22}/\lambda^2.$$

The total surface displacement

$$\eta = \varepsilon^2\eta_{01} + \varepsilon\mathrm{Re}(\eta_{10}\exp(i\omega t - ikx)) + \varepsilon^2\mathrm{Re}(\eta_{21}\exp(2i\omega t - 2ikx)) + O(\varepsilon^2) \quad (50)$$

contains contributions of the zeroth, first, and second harmonics. The expressions for the displacement components are found by solving (9):

$$\eta_{01} = \tilde{r}_{01}|A_{10}|^2 = r_{01}|\eta_{10}|^2, \quad r_{01} = \tilde{r}_{01}/\lambda^2, \quad (51)$$

$$\eta_{21} = \tilde{r}_{21}A_{10}^2 = r_{21}\eta_{10}^2, \quad r_{21} = -\tilde{r}_{21}/\lambda^2. \quad (52)$$

An equation similar to (49) was derived in [12], with coefficients equal to those obtained here for $s = 0$. It is a generalization of the classical NLS equation to the next asymptotic order. However, the nonlinear term (with coefficient α_1) vanishes when $kh \approx 1.363$, and the resulting equation (as well as (37)) fails to describe the nonlinear wave dynamics. For this reason, the evolution equation extended to the next asymptotic order is considered below.

6. THIRD-ORDER NONLINEAR EVOLUTION EQUATION

To the third asymptotic order, the surface displacement is expressed as

$$\eta_1 = i\lambda A_{10} + \varepsilon^2\eta_{12} + O(\varepsilon^3), \quad (53)$$

where

$$\eta_{12} = i\lambda\left(A_{12} + p_{21}\frac{\partial^2 A_{10}}{\partial x_1^2} + p_{22}|A_{10}|^2A_{10}\right). \quad (54)$$

Here, freedom in defining A_{12} can be used:

$$A_{12} = -p_{21}\frac{\partial^2 A_{10}}{\partial x_1^2} - p_{22}|A_{10}|^2A_{10}. \quad (55)$$

Then,

$$\eta_1 = i\lambda A_{10} + O(\varepsilon^3). \quad (56)$$

The equation for the first harmonic derived from the asymptotic expansion to order {1, 4} is

$$i\frac{\partial A_{10}}{\partial t_4} + \beta_3\frac{\partial^4 A_{10}}{\partial x_1^4} + P_{31}|A_{10}|^4A_{10} + P_{32}|A_{10}|^2\frac{\partial^2 A_{10}}{\partial x_1^2} + P_{33}A_{10}^2\frac{\partial^2 A_{10}^*}{\partial x_1^2} + P_{34}A_{10}\frac{\partial A_{10}}{\partial x_1}\frac{\partial A_{10}^*}{\partial x_1} + P_{35}A_{10}^*\left(\frac{\partial A_{10}}{\partial x_1}\right)^2 + sA_{10}\frac{\partial A_{01}}{\partial t_2} + sA_{10}\frac{\partial A_{00}}{\partial t_3} + \tilde{\rho}_{12}A_{10}\frac{\partial A_{02}}{\partial x_1} = 0, \quad (57)$$

and the related equation for the zeroth harmonic is

$$\frac{\partial A_{02}}{\partial x_1} = \gamma_{31}|A_{10}|^4 + \gamma_{32}\left(A_{10}\frac{\partial^2 A_{10}^*}{\partial x_1^2} + A_{10}^*\frac{\partial A_{10}}{\partial x_1^2}\right) + \gamma_{33}\frac{\partial A_{10}}{\partial x_1}\frac{\partial A_{10}^*}{\partial x_1}. \quad (58)$$

These equations are derived by using the condition

$$\frac{\partial A_{02}}{\partial t_1} + V\frac{\partial A_{02}}{\partial x_1} = 0 \quad (59)$$

and the following expressions for components of the second and third harmonics:

$$A_{23} = i\chi_{31}|A_{10}|^2A_{10}^2 + i\chi_{32}A_{10}\frac{\partial^2 A_{10}}{\partial x_1^2} + i\chi_{33}\left(\frac{\partial A_{10}}{\partial x_1}\right)^2, \quad (60)$$

$$A_{32} = \nu_1 A_{10}^3. \quad (61)$$

The derivatives of the zeroth-harmonic components multiplied by s in (57) are calculated by using the rela-

tions obtained above:

$$\frac{\partial A_{01}}{\partial t_2} = \gamma_2\tilde{\alpha}_1|A_{10}|^4 + \gamma_2\beta_1\left(A_{10}\frac{\partial^2 A_{10}^*}{\partial x_1^2} + A_{10}^*\frac{\partial^2 A_{10}}{\partial x_1^2} - 2\frac{\partial A_{10}}{\partial x_1}\frac{\partial A_{10}^*}{\partial x_1}\right), \tag{62}$$

$$\frac{\partial A_{00}}{\partial t_3} = -\gamma_1\frac{\tilde{\alpha}_{21}+\tilde{\alpha}_{22}}{2}|A_{10}|^4 - \gamma_1\beta_2\left(A_{10}\frac{\partial^2 A_{10}^*}{\partial x_1^2} + A_{10}^*\frac{\partial^2 A_{10}}{\partial x_1^2} - \frac{\partial A_{10}}{\partial x_1}\frac{\partial A_{10}^*}{\partial x_1}\right). \tag{63}$$

Substituting (62) and (63) into (57) yields

$$i\frac{\partial A_{10}}{\partial t_4} + \beta_3\frac{\partial^4 A_{10}}{\partial x_1^4} + \tilde{\rho}_{31}|A_{10}|^4A_{10} + \tilde{\rho}_{32}|A_{10}|^2\frac{\partial^2 A_{10}}{\partial x_1^2} + \tilde{\rho}_{33}A_{10}^2\frac{\partial^2 A_{10}^*}{\partial x_1^2} + \tilde{\rho}_{34}A_{10}\frac{\partial A_{10}}{\partial x_1}\frac{\partial A_{10}^*}{\partial x_1} + \tilde{\rho}_{35}A_{10}^*\left(\frac{\partial A_{10}}{\partial x_1}\right)^2 + \tilde{\rho}_{12}A_{10}\frac{\partial A_{02}}{\partial x_1} = 0. \tag{64}$$

A closed third-order nonlinear evolution equation for the first harmonic is obtained by combining Eq. (64) with (58):

$$i\frac{\partial A_{10}}{\partial t_4} + \beta_3\frac{\partial^4 A_{10}}{\partial x_1^4} + \tilde{\alpha}_{31}|A_{10}|^4A_{10} + \tilde{\alpha}_{32}|A_{10}|^2\frac{\partial^2 A_{10}}{\partial x_1^2} + \tilde{\alpha}_{33}A_{10}^2\frac{\partial^2 A_{10}^*}{\partial x_1^2} + \tilde{\alpha}_{34}A_{10}\frac{\partial A_{10}}{\partial x_1}\frac{\partial A_{10}^*}{\partial x_1} + \tilde{\alpha}_{35}A_{10}^*\left(\frac{\partial A_{10}}{\partial x_1}\right)^2 = 0, \tag{65}$$

and the corresponding complete equation for displacement is

$$i\left(\frac{\partial \eta_{10}}{\partial t} + V\frac{\partial \eta_{10}}{\partial x}\right) + \varepsilon\beta_1\frac{\partial^2\eta_{10}}{\partial x^2} + \varepsilon\alpha_1|\eta_{10}|^2\eta_{10} + i\varepsilon^2\beta_2\frac{\partial^3\eta_{10}}{\partial x^3} + i\varepsilon^2\alpha_{21}|\eta_{10}|^2\frac{\partial\eta_{10}}{\partial x} + i\varepsilon^2\alpha_{22}\eta_{10}^2\frac{\partial\eta_{10}^*}{\partial x} + \varepsilon^3\beta_3\frac{\partial^4\eta_{10}}{\partial x^4} + \varepsilon^3\alpha_{31}|\eta_{10}|^4\eta_{10} + \varepsilon^3\alpha_{32}|\eta_{10}|^2\frac{\partial^2\eta_{10}}{\partial x^2} + \varepsilon^3\alpha_{33}\eta_{10}^2\frac{\partial^2\eta_{10}^*}{\partial x^2} + \varepsilon^3\alpha_{34}\eta_{10}\frac{\partial\eta_{10}}{\partial x}\frac{\partial\eta_{10}^*}{\partial x} + \varepsilon^3\alpha_{35}\eta_{10}^*\left(\frac{\partial\eta_{10}}{\partial x}\right)^2 = 0, \tag{66}$$

$$\alpha_{31} = \tilde{\alpha}_{31}/\lambda^4, \quad \alpha_{32} = \tilde{\alpha}_{32}/\lambda^2, \quad \alpha_{33} = \tilde{\alpha}_{33}/\lambda^2,$$

$$\alpha_{34} = \tilde{\alpha}_{34}/\lambda^2, \quad \alpha_{35} = \tilde{\alpha}_{35}/\lambda^2.$$

The displacement envelope is expressed as

$$\eta = \varepsilon^2(\eta_{01} + \varepsilon\eta_{02}) + \varepsilon\mathrm{Re}(\eta_{10}\exp(i\omega t - ikx)) + \varepsilon^2\mathrm{Re}((\eta_{21} + \varepsilon\eta_{22})\exp(2i\omega t - 2ikx)) + \varepsilon^3\mathrm{Re}(\eta_{32}\exp(3i\omega t - 3ikx)) + O(\varepsilon^4) \tag{67}$$

with components given by (51), (52), and the formulas

$$\eta_{02} = i\tilde{r}_{02}\left(A_{10}\frac{\partial A_{10}^*}{\partial x} - A_{10}^*\frac{\partial A_{10}}{\partial x}\right) = ir_{02}\left(\eta_{10}\frac{\partial\eta_{10}^*}{\partial x} - \eta_{10}^*\frac{\partial\eta_{10}}{\partial x}\right), \quad r_{02} = \tilde{r}_{02}/\lambda^2, \tag{68}$$

$$\eta_{22} = i\tilde{r}_{22}A_{10}\frac{\partial A_{10}}{\partial x} = ir_{22}\eta_{10}\frac{\partial\eta_{10}}{\partial x}, \quad r_{22} = -\frac{\tilde{r}_{22}}{\lambda^2}, \tag{69}$$

$$\eta_{32} = i\tilde{r}_{32}A_{10}^3 = r_{32}\eta_{10}^3, \quad r_{32} = -\frac{\tilde{r}_{32}}{\lambda^3}. \tag{70}$$

Equation (66) is a generalization of the equation derived in [12] to the next asymptotic order.

Note that the right-hand sides of (32), (45), and Eq. (58) integrated by parts are similar in form to the first three conservation laws for the first-order NLS equation [19]:

$$I_1 = \int_{-\infty}^{\infty}|A_{10}|^2dx_1, \tag{71}$$

$$I_2 = \int_{-\infty}^{\infty}\left(A_{10}\frac{\partial A_{10}^*}{\partial x_1} - A_{10}^*\frac{\partial A_{10}}{\partial x_1}\right)dx_1, \tag{72}$$

$$I_3 = \int_{-\infty}^{\infty}\left(|A_{10}|^4 - \frac{2\beta_1}{\tilde{\alpha}_1}\left|\frac{\partial A_{10}}{\partial x_1}\right|^2\right)dx_1. \tag{73}$$

This fact is directly related to the possibility of representing the zeroth-harmonic components of potential

Coefficients in equations corresponding to special cases

	Special cases of generalized NLS equation (66)			Generalized Dysthe equation (81)
	$kh \longrightarrow 0$	$kh = 1.363$	$kh \longrightarrow \infty$	
Vk/ω	1	0.6793	1/2	
$\beta_1 k^2/\omega$	$(kh)^2/2$	0.2657	1/8	
$\alpha_1/k^2/\omega$	$(-9/16)/(kh)^4$	0.0002	1/2	
$\beta_2 k^3/\omega$	$(kh)^2/6$	−0.1114	−1/16	
$\alpha_{21}/k/\omega$	$(-9/8)/(kh)^4$	0.6833	3/2	
$\alpha_{22}/k/\omega$	$(-27/16)/(kh)^4$	−0.2678	1/4	
$\beta_3 k^4/\omega$	$(19/72)(kh)^4$	−0.0066	−5/128	
$\alpha_{31}/k^4/\omega$	$(81/1024)/(kh)^{10}$	0.3864	1/2	
α_{32}/ω	$(-15/16)/(kh)^2$	−0.4433	$kh/6$	−5/8
α_{33}/ω	$(9/4)/(kh)^4$	0.7798	$kh/6$	3/32
α_{34}/ω	$(9/8)/(kh)^4$	0.2394	$kh/3$	−3/16
α_{35}/ω	$(-9/16)/(kh)^4$	−0.6467	−19/32	
$\gamma_1\omega/k^3$	$(-3/4)/(kh)^2$	−0.4935	$-1/(2kh)$	–
$\gamma_2\omega/k^2$	$(-3/4)/(kh)^2$	−0.1351	$1/(8kh)$	–
$\gamma_{31}\omega^3/k^7$	$(-81/64)/(kh)^4$	−0.6239	$-1/(8kh)$	–
$\gamma_{32}\omega/k$	$(3/4)/(kh)^2$	0.1343	$-kh/12$	–
$\gamma_{33}\omega/k$	$(-3/4)/(kh)^2$	−0.3705	$-kh/6$	–
r_{01}/k	$(-3/4)/(kh)^3$	−0.4479	$-1/(4kh)$	0
r_{02}	$(-3/4)/(kh)^3$	−0.3116	$(-1/32)/(kh)^2$	0
r_{21}/k	$(3/4)/(kh)^3$	0.8265	1/2	
r_{22}	$(-3/2)/(kh)^3$	−0.2669	1/2	
r_{32}/k^2	$(27/64)/(kh)^6$	0.8250	3/8	

by expressions (43), (62), and (63), which do not contain integrals. Indeed, it follows from (32) that

$$\frac{\partial}{\partial t_2}\int_{-\infty}^{\infty}\gamma_1|A_{10}|^2 dx_1 = \frac{\partial}{\partial t_2}A_{00}\Big|_{x_1=-\infty}^{x_1=\infty}. \qquad (74)$$

By substituting (43), the integral in (74) is found to vanish. Therefore, the integral on the left-hand side in (74) is a conserved quantity. The integral I_1 is also conserved in the second-order NLS equation by virtue of (63), and expression (62) implies that I_2 is an integral of motion of the first-order NLS equation. It is well known that the classical NLS equation has an infinite number of conservation laws. However, it can be shown that the integral I_1 is conserved in Eq. (66) only when a certain relation between the nonlinear dispersion coefficients α_{32}, α_{33}, and α_{35} is satisfied. This implies breakdown of conservation laws of high-order NLS equations in the general case. Even though expression (58) is similar in form to I_3, it does not entail (73) in the general case. Since terms analogous to those with coefficients s in (42) and (57) should be expected to appear in higher order asymptotic expansions, higher order equations may contain nonlocal terms (derivatives of zeroth-order harmonic components of the potential expressed in terms of indefinite integrals).

7. LIMIT OF INFINITELY DEEP WATER ($kh \longrightarrow \infty$)

The coefficients α_{32}, α_{33}, and α_{34} in Eq. (66) are divergent in the limit of $kh \longrightarrow \infty$ (see table and figure). Asymptotic expansion can be used only if higher order terms are smaller than lower order ones. This requirement is violated for the zeroth-harmonic expansion

$$\varphi_0 = \varepsilon\varphi_{00} + \varepsilon^2\varphi_{01} + \varepsilon^3\varphi_{02} + \varepsilon^4\varphi_{03} + \varepsilon^5\varphi_{04} + O(\varepsilon^6).$$

Indeed, as $kh \longrightarrow \infty$, expressions (18) and (19) tend to the following limits:

$$\varphi_{02} \xrightarrow[kh\to\infty]{} A_{02} + \frac{k^2}{2\omega} \times\left(z + \frac{h}{2} + \frac{1}{8k}\right)\frac{\partial}{\partial x_1}|A_{10}|^2, \tag{75}$$

$$\varphi_{03} \xrightarrow[kh\to\infty]{} A_{03} - i\frac{k}{8\omega} \times\left(z + \frac{h}{2} - \frac{1}{8k}\right)\frac{\partial}{\partial x_1}\left(A_{10}\frac{\partial A_{10}^*}{\partial x_1} - A_{10}^*\frac{\partial A_{10}}{\partial x_1}\right). \tag{76}$$

Thus, both φ_{02} and φ_{03} increase indefinitely if A_{10} and A_{01} remain finite in the limit of infinitely deep water. The secular terms can be eliminated by using the freedom in defining A_{02} and A_{03}:

$$A_{02} \longrightarrow A_{02} - \frac{k^2 h}{4\omega}\frac{\partial}{\partial x_1}|A_{10}|^2, \tag{77}$$

$$A_{03} \longrightarrow A_{03} + i\frac{kh}{16\omega}\frac{\partial}{\partial x_1}\left(A_{10}\frac{\partial A_{10}^*}{\partial x_1} - A_{10}^*\frac{\partial A_{10}}{\partial x_1}\right), \tag{78}$$

as done in [12]. However, this is not the only possible choice. Substitutions of this kind cannot be used to change the closed equation or the coefficients therein, but can modify coefficients in the system of equations for the first and second harmonics.

As $kh \longrightarrow \infty$, the component φ_{04} behaves as follows:

$$\varphi_{04} \xrightarrow[kh\to\infty]{} A_{04} + \frac{k^6}{8\omega^3}\left(z + \frac{h}{2} + \frac{2}{k}\right)\frac{\partial}{\partial x_1}|A_{10}|^4 + \frac{1}{\omega}P_1\left(k^2h^3, kh^2, k^2h^2z, h, khz, k^2z^2h, \frac{1}{k}, z, kz^2, k^2z^3\right) \times\frac{\partial}{\partial x_1}\left(A_{10}\frac{\partial^2 A_{10}^*}{\partial x_1^2} + A_{10}^*\frac{\partial^2 A_{10}}{\partial x_1^2}\right) + \frac{1}{\omega}P_2\left(k^2h^3, kh^2, h, khz, k^2z^2h, \frac{1}{k}, z, kz^2, k^2z^3\right) \times\frac{\partial}{\partial x_1}\left(\frac{\partial A_{10}}{\partial x_1}\frac{\partial A_{10}^*}{\partial x_1}\right), \tag{79}$$

where P_1 and P_2 are polynomials containing the terms in parentheses. It is obvious that a bounded φ_{04} can always be obtained by appropriate choice of A_{04}, whereas the corresponding derivatives of φ_{04} with respect to z remain indefinitely increasing functions as $kh \longrightarrow \infty$. The substitution of such expressions as (77) and (78) does not lead to any substantial change in the form of (79). Therefore, the evolution equation must contain terms with coefficients that increase indefinitely as $kh \longrightarrow \infty$; i.e., the ensuing asymptotic series is divergent.

Indefinitely increasing terms appear in the series because the depth h is not small as compared to the length scale L of the zeroth harmonic, which makes it impossible to represent the zeroth harmonic in the solution to the Laplace equation as a power series (see (15)). Whereas this does not rule out analysis in the limit of $kh \longrightarrow \infty$ for the classical NLS equation or the second-order generalized equation, Eq. (66) cannot be used to describe waves in infinitely deep water.

An evolution equation in the limit of $kh \longrightarrow \infty$ can be derived as an envelope equation by in the manner of [11]. Suppose that $h \gg L$ ($L/h = O(\mu)$). The mode structure of the wave-induced flow is not prescribed,

$$\varphi_{0m} = A_{0m}(z, x_i, t_i), \quad m \geq 0. \tag{80}$$

By following the derivation of (66), a generalized Dysthe equation is obtained:

$$i\left(\frac{\partial\eta_{10}}{\partial t} + \frac{\omega}{2k}\frac{\partial\eta_{10}}{\partial x}\right) + \varepsilon\frac{\omega}{8k^2}\frac{\partial^2\eta_{10}}{\partial x^2} + \varepsilon\frac{\omega k^2}{2}|\eta_{10}|^2\eta_{10} - i\varepsilon^2\frac{\omega}{16k^3}\frac{\partial^3\eta_{10}}{\partial x^3} + i\varepsilon^2\frac{3\omega k}{2}|\eta_{10}|^2\frac{\partial\eta_{10}}{\partial x} + i\varepsilon^2\frac{\omega k}{4}\eta_{10}^2\frac{\partial\eta_{10}^*}{\partial x} - \varepsilon^3\frac{5\omega}{128k^4}\frac{\partial^4\eta_{10}}{\partial x^4} + \varepsilon^3\frac{\omega k^4}{2}|\eta_{10}|^4\eta_{10} - \varepsilon^3\frac{5\omega}{8}|\eta_{10}|^2\frac{\partial^2\eta_{10}}{\partial x^2} + \varepsilon^3\frac{3\omega}{32}\eta_{10}^2\frac{\partial^2\eta_{10}^*}{\partial x^2} - \varepsilon^3\frac{3\omega}{16}\eta_{10}\frac{\partial\eta_{10}}{\partial x}\frac{\partial\eta_{10}^*}{\partial x} - \varepsilon^3\frac{19}{32}\eta_{10}^*\left(\frac{\partial\eta_{10}}{\partial x}\right)^2 + k\eta_{10}\frac{\partial A_{02}}{\partial x} = 0, \tag{81}$$

$$\frac{\partial A_{02}}{\partial z} = \frac{\omega}{2}\frac{\partial}{\partial x}|\eta_{10}|^2, \quad z = 0, \tag{82}$$

$$\frac{\partial^2 A_{02}}{\partial z^2} + \frac{\partial^2 A_{02}}{\partial x^2} = 0, \quad -h < z < 0, \tag{83}$$

$$\frac{\partial A_{02}}{\partial z} = 0, \quad z = -h. \tag{84}$$

Equation (81) is derived by using condition (59), the

identities $A_{00} \equiv 0$ and $A_{01} \equiv 0$, and the relation

$$\frac{A_{03}}{\partial z} = -i\frac{\omega}{8k}\frac{\partial}{\partial x}\left(\eta_{10}\frac{\partial^2\eta_{10}^*}{\partial x^2} - \eta_{10}^*\frac{\partial^2\eta_{10}}{\partial x^2}\right), \quad z = 0. \tag{85}$$

Note that (82) and (85) are identical to the expressions for these derivatives that follow from (75) and (76), respectively. The last (wave-induced) term in (81) is similar to the analogous term in the classical Dysthe equation. This is due to the assumption that $L/h = O(\mu)$ (more restrictive than $L/h = O(1)$), which is required to simplify the derived equation. Equation (81) is applicable if $kh \geq 1/\varepsilon^2$.

The third-order NLS equation obtained by dropping the last term in (81) is valid in the limit of infinitely deep water, and the coefficients therein are equal to the bounded ones in Eq. (66) (see table). Note that the contribution of the zeroth harmonic (wave-induced flow) is not only represented by the last term in (81), but it is also contained in other nonlinear coefficients.

The total displacement envelope field described by Eq. (81) is represented by expression (67). An analysis of the limit behavior of r_{ij} using the fact that $(kh)^{-1}$ is small shows that the contribution of the zeroth harmonic exceeds the accuracy of the present model, even though this contribution is retained in Dysthe's model [13].

8. SHALLOW-WATER LIMIT ($kh \longrightarrow 0$)

As $kh \longrightarrow 0$, the coefficients tend to infinity under the normalization conditions used here (see table), but the terms in asymptotic series expansions remain small. With increasing water depth, the sign of α_{31} changes from plus to minus and then back to plus (see figure).

The NLS equation for $kh \longrightarrow 0$ can also be derived by starting from the KdV (shallow-water) equation [5]

$$\frac{\partial\eta}{\partial t} + \sqrt{gh}\frac{\partial\eta}{\partial x} + \frac{3\sqrt{gh}}{2h}\eta\frac{\partial\eta}{\partial x} + \frac{h^2\sqrt{gh}}{6}\frac{\partial^3\eta}{\partial x^3} = 0. \tag{86}$$

This method for deriving the envelope equation was used to verify the coefficients obtained above, and complete agreement was demonstrated. The third-order nonlinear equation for weakly modulated waves based on the KdV equation is similar to (66), except for $\alpha_{32} = 0$; the remaining coefficients are given in the table (see the $kh \longrightarrow 0$ column in table). The coefficient α_{32} vanishes because it behaves as a different power of kh (see table).

When terms of the next order are retained in the KdV equation, the values of the coefficients in the envelope equation are corrected. In particular, the corrections to α_1 obtained by using Eq. (86) modified by including a cubic nonlinear term may result in a change from focusing to defocusing type of the equation [20]. However, this change does not occur in water wave dynamics.

In [16, 20], higher order envelope equations were derived for the KdV equation generalized by adding a cubic nonlinear term. The derivation presented in [16] followed [15] in relying on a modified relation between small parameters and made use of the condition for vanishing nonlinearity in the first-order envelope equation. In the present study, the envelope equation is derived by starting from a generalized KdV equation without using this condition (classical relation between small parameters), and agreement with the results of [16] is achieved for near-critical values of α_1. In [20], a solution was sought to the second asymptotic order, whereas the present derivation shows that third-order corrections contribute to α_{31} as well.

9. VANISHING CUBIC NONLINEARITY ($kh \approx 1.363$)

A nonlinear evolution equation for the modulated envelope in the limit of $kh \longrightarrow 1.363$ was originally derived in [14] and later in [15]. In the former paper, numerical values of the coefficients corresponding to $kh = 1.363$ were given; in the latter, expressions for the coefficients as functions of depth were presented (different from those in the former and divergent in the limit of infinitely deep water).

To formulate the evolution equation for $kh \approx 1.363$, one should rewrite Eq. (66), factoring out the small parameters associated with nonlinearity (δ) and dispersion (μ):

$$\begin{aligned} &i\left(\varepsilon\frac{\partial\eta_{10}}{\partial t} + \mu V\frac{\partial\eta_{10}}{\partial x}\right) + \mu^2\beta_1\frac{\partial^2\eta_{10}}{\partial x^2} + \delta^2\alpha_{10}|\eta_{10}|^2\eta_{10} \\ &+ i\mu^3\beta_2\frac{\partial^3\eta_{10}}{\partial x^3} + i\delta^2\mu\alpha_{21}|\eta_{10}|^2\frac{\partial\eta_{10}}{\partial x} + i\delta^2\mu\alpha_{22}\eta_{10}^2\frac{\partial\eta_{10}^*}{\partial x} \\ &+ \mu^4\beta_3\frac{\partial^4\eta_{10}}{\partial x^4} + \delta^4\alpha_{31}|\eta_{10}|^4\eta_{10} \\ &+ \delta^2\mu^2\alpha_{32}|\eta_{10}|^2\frac{\partial^2\eta_{10}}{\partial x^2} \\ &+ \delta^2\mu^2\alpha_{33}\eta_{10}^2\frac{\partial^2\eta_{10}^*}{\partial x^2} + \delta^2\mu^2\alpha_{34}\eta_{10}\frac{\partial\eta_{10}}{\partial x}\frac{\partial\eta_{10}^*}{\partial x} \\ &+ \delta^2\mu^2\alpha_{35}\eta_{10}^*\left(\frac{\partial\eta_{10}}{\partial x}\right)^2 = 0. \end{aligned} \tag{87}$$

When the term with α_1 is dropped and the next-order nonlinear and first-order dispersive terms (with α_{31} and β_1, respectively) are taken into account as terms of similar asymptotic order, a new relation between the small nonlinearity and dispersion parameters is

obtained: $\delta^2 \sim \mu$. Then, retaining only the first-order corrections to the KdV equation, we obtain the modified equation

$$i\left(\frac{\partial\eta_{10}}{\partial t} + V\frac{\partial\eta_{10}}{\partial x}\right) + \varepsilon\beta_1\frac{\partial^2\eta_{10}}{\partial x^2} + \sqrt{\varepsilon}\alpha_1|\eta_{10}|^2\eta_{10}$$
$$+ i\varepsilon\alpha_{21}|\eta_{10}|^2\frac{\partial\eta_{10}}{\partial x} + i\varepsilon\alpha_{22}\eta_{10}^2\frac{\partial\eta_{10}^*}{\partial x} \quad (88)$$
$$+ \varepsilon\alpha_{31}|\eta_{10}|^4\eta_{10} = 0.$$

Here, $\varepsilon = \mu = \delta^2$, and the cubic nonlinear term is retained under the assumption that α_1 is an order $\varepsilon^{1/2}$ or smaller quantity.

Equation (88) describes nonlinear dynamics of surface waves when $kh \approx 1.363$; i.e., it should be used instead of (37) in this parameter region. Even though the terms contained in Eq. (88) are obtained by using expansions to order ε^5, it is essentially a first-order equation. In this study, the coefficients in this equation are well defined for any kh (in contrast to [14]) and are always finite (in contrast to [15]). Their values corresponding to $kh = 1.363$ are listed in the table. The surface displacement is given by formula (38).

The modulational instability of the simplest (plane-wave) solution to Eq. (88) was analyzed in [15, 16]. If the solution is represented as

$$\eta_{10} = A\exp(i((\Omega + \omega)t - (K + k)x)), \quad (89)$$

where K and Ω are the wavenumber and frequency detunings, respectively, then the modulational instability criterion is

$$\beta_1\alpha_1 + D > 0, \quad D = \beta_1 K(\alpha_{22} - \alpha_{21})$$
$$+ A^2\left(2\beta_1\alpha_{31} - \frac{1}{2}\alpha_{22}^2\right). \quad (90)$$

The wavenumber interval corresponding to unstable perturbations is determined by the relation

$$0 < \Delta K < \left|\frac{A}{\beta_1}\right|\sqrt{2(\beta_1\alpha_1 + D)} \quad (91)$$

(the total wavenumber is $k + K + \Delta K$), and the largest growth rate exponent is

$$(\mathrm{Im}\Omega)_{\max} = \frac{A^2}{\beta_1}(\beta_1\alpha_1 + D)$$
$$\text{for } \Delta K_{\max} = \left|\frac{A}{\beta_1}\right|\sqrt{\beta_1\alpha_1 + D}. \quad (92)$$

The type of instability varies, depending on the sign of D. For $kh = 1.363$, the results listed in the table can be used to obtain

$$D \approx \omega^2\left(-0.2527\frac{K}{k} + 0.2311(kA)^2\right). \quad (93)$$

The coefficients in (93) differ from those found in [15]. However, since the sign of D corresponding to the Stokes wave ($K = 0$) is the same, the conclusions about unstable behavior made in that study remain valid in the present context: there always exist unstable perturbations of the Stokes wave when $kh > 1.363$, whereas modulational instability develops for $kh < 1.363$ only if

$$D > -\beta_1\alpha_1 > 0. \quad (94)$$

The latter condition corresponds to an instability threshold for intense waves shifted from $kh = 1.363$ toward the shallow-water limit. The shift can be estimated by setting $K = 0$ and $kA = 0.1$ (typical steepness of ocean waves) or $kA = 0.4$: the critical depth corresponds to $kh \approx 1.3508$ and $kh \approx 1.2520$, respectively.

10. CONCLUSIONS

A third-order NLS equation for waves in finite-depth water that generalizes the equation obtained in [12] is derived without assuming that the wave-induced-flow velocity is equal to the group velocity of the fundamental wave to high orders of accuracy. The derived equation is valid for waves characterized by a small steepness and narrow-band spectrum when the water depth is small as compared to the modulation length. The latter condition is formally violated as $kh \longrightarrow \infty$ (power-series representation of the components of the zeroth harmonic of velocity potential breaks down). In this limit, a third-order generalized Dysthe equation with finite coefficients is obtained that takes into account weak wave-induced flow. The generalized NLS equation obtained in the shallow-water limit is consistent with the envelope equation based on the KdV equation. In the special case of vanishing cubic nonlinearity ($kh \approx 1.363$), a generalization of the first-order NLS equation is derived and the coefficients therein are calculated. The analysis of modulational instability performed in this case shows that the modulational instability threshold shifts toward the shallow-water limit with increasing wave intensity.

ACKNOWLEDGMENTS

I thank Yu.V. Sedletsky for stimulating discussions and E.N. Pelinovsky for attention to this study. This work was supported by INTAS, grant no. 04-83-3032; the Russian Foundation for Basic Research, project no. 05-05-64265; and by the Talanov Leading Science School, grant NSh-1637.2003.2.

APPENDIX

Coefficients of Equations

In what follows, the coefficients of terms corresponding to higher asymptotic orders are expressed in

terms of lower order coefficients. These representations are suitable for verifying calculations and understanding the contributions of various effects. The expression for V is substituted into the coefficients in some equations and is not substituted into others to reduce the size of expressions.

Coefficients of Field Components

$$p_1 = h\sigma + \frac{V}{\omega},$$

$$p_{21} = \frac{\beta_1}{\omega} + k^2h^2\frac{3\sigma^4 - 2\sigma^2 + 1}{4k^2\sigma^2} + \frac{kh}{2k^2\sigma} + \frac{1}{4k^2},$$

$$p_{22} = \frac{\tilde{\alpha}_1}{\omega} + \frac{k^2V\sigma^2 - \omega k}{\omega^2}\gamma_1 + k^2\frac{-\sigma^4 - 2\sigma^2 + 1}{\omega(\sigma^2 + 1)}\chi_1$$

$$+ k^4\frac{\sigma^2(5\sigma^2 - 8)}{8\omega^2}.$$

Coefficients of Nonlinear Interaction Terms

$$\chi_1 = 3k^2\frac{\sigma^4 - 1}{8\omega\sigma^2},$$

$$\chi_2 = \left(kh\frac{-\sigma^2 + 3}{k\sigma} + \frac{1}{k}\right)\chi_1$$

$$+ 3k^2h\frac{(\sigma^4 - 1)(3\sigma^2 + 1)}{16\omega\sigma^3} + 9k\frac{-\sigma^4 + 1}{16\omega\sigma^2},$$

$$\chi_{31} = -2\frac{\sigma^2 + 1}{\omega\sigma^2}\tilde{\alpha}_1\chi_1 + k^2\frac{\sigma^4 + 4\sigma^2 + 3}{8\omega^2\sigma^2}\tilde{\alpha}_1$$

$$+\left(k^2h\frac{(\sigma^2 - 1)(3\sigma^2 - 1)}{2\omega\sigma^3} + k\frac{\sigma^2 + 5}{2\omega\sigma^2}\right)\gamma_1\chi_1$$

$$+\left(k^4h\frac{(\sigma^2 - 1)^2(\sigma^2 + 1)}{4\omega^2\sigma} - k^3\frac{2\sigma^6 + \sigma^4 + 2\sigma^2 + 3}{8\omega^2\sigma^2}\right)\gamma_1$$

$$+ k^4\frac{\sigma^4 - 4\sigma^2 + 19}{4\omega^2}\chi_1 + 3k^2\frac{(1 - \sigma^4)(5\sigma^2 + 3)}{4\omega\sigma^2(3\sigma^2 + 1)}\nu_1$$

$$+ k^6\frac{-21\sigma^6 + 4\sigma^4 - 27\sigma^2 - 52}{96\omega^3},$$

$$\chi_{32} = -2\frac{\sigma^2 + 1}{\omega\sigma^2}\beta_1\chi_1 + k^2\frac{\sigma^4 + 4\sigma^2 + 3}{8\omega^2\sigma^2}\beta_1$$

$$+\left(kh\frac{\sigma^2 - 3}{2k\sigma} - \frac{1}{2k}\right)\chi_2$$

$$+\left(k^2h^2\frac{\sigma^6 - 9\sigma^4 + 15\sigma^2 + 1}{8k^2\sigma^4} + kh\frac{-\sigma^4 + 6\sigma^2 - 1}{4k^2\sigma^3}\right.$$

$$\left.+\frac{\sigma^2 + 1}{8k^2\sigma^2}\right)\chi_1 + k^2h^2\frac{-17\sigma^8 + 4\sigma^6 + 14\sigma^4 - 4\sigma^2 + 3}{32\omega\sigma^4}$$

$$+ kh\frac{(\sigma^4 - 1)(9\sigma^2 + 1)}{16\omega\sigma^3} - \frac{(\sigma^2 + 1)^2}{32\omega\sigma^2},$$

$$\chi_{33} = \left(kh\frac{\sigma^2 - 3}{2k\sigma} - \frac{1}{2k}\right)\chi_2$$

$$+\left(k^2h^2\frac{\sigma^6 - 9\sigma^4 + 15\sigma^2 + 1}{8k^2\sigma^4}\right.$$

$$\left.+ kh\frac{-\sigma^4 + 6\sigma^2 - 1}{4k^2\sigma^3} + \frac{\sigma^2 + 1}{8k^2\sigma^2}\right)\chi_1$$

$$- k^2h^2\frac{(\sigma^2 - 1)^2(1 + \sigma^2)}{8\omega\sigma^2}$$

$$+ kh\frac{(\sigma^4 - 1)(2\sigma^2 + 1)}{8\omega\sigma^3} + \frac{-\sigma^4 + 1}{8\omega\sigma^2},$$

$$\nu_1 = k^2\frac{-3\sigma^6 - 13\sigma^4 + 5\sigma^2 + 3}{8\omega\sigma^2(\sigma^2 + 1)}\chi_1$$

$$+ k^4\frac{(\sigma^2 - 1)(9\sigma^4 - 1)}{64\omega^2\sigma^2}.$$

Linear Dispersion Coefficients

$$\beta_1 = -\frac{1}{2}\frac{\partial^2\omega}{\partial k^2}, \quad \beta_2 = -\frac{1}{6}\frac{\partial^3\omega}{\partial k^3}, \quad \beta_3 = \frac{1}{24}\frac{\partial^4\omega}{\partial k^4}.$$

These coefficients can be calculated by using dispersion relation (26).

Coefficients of Wave-Induced Flow Components

$$\gamma_1 = \frac{k^2V(\sigma^2 - 1) - 2\omega k}{4V_d^2},$$

$$\gamma_2 = 2\frac{V}{V_d^2}\gamma_1\beta_1 + k^2\frac{\sigma^2 - 1}{4V_d^2}\beta_1 + \frac{\omega^2 - k^2V^2(\sigma^2 + 1)}{4\omega V_d^2},$$

$$\gamma_{31} = k^2\frac{\sigma^2 - 1}{8V_d^2}(\tilde{\alpha}_{21} + \tilde{\alpha}_{22}) + \frac{V}{V_d^2}(\tilde{\alpha}_{21} + \tilde{\alpha}_{22})\gamma_1$$

$$-2\frac{V}{V_d^2}\tilde{\alpha}_1\gamma_2-\frac{1}{V_d^2}\tilde{\alpha}_1\gamma_1\beta_1+\frac{k\omega+3k^2V(\sigma^2+1)}{4\omega V_d^2}\tilde{\alpha}_1$$

$$-3\frac{V}{2V_d^2}\gamma_1^2+\left(k^3h\frac{\sigma^4-1}{4\sigma V_d^2}-k^2\frac{\sigma^2+1}{2V_d^2}\right)\gamma_1$$

$$+\left(\omega k^2h\frac{(\sigma^2-1)(-7\sigma^4-2\sigma^2+1)}{2\sigma(\sigma^2+1)^2V_d^2}\right.$$

$$\left.+\omega k\frac{3\sigma^4-6\sigma^2-5}{2(\sigma^2+1)^2V_d^2}\right)\chi_1^2$$

$$+\left(k^4h\frac{(\sigma^2-1)^2(7\sigma^2+1)}{4\sigma(\sigma^2+1)V_d^2}+k^3\frac{-2\sigma^4+5\sigma^2+1}{2(\sigma^2+1)V_d^2}\right)\chi_1$$

$$+k^6h\frac{\sigma(\sigma^2-1)(-3\sigma^2+11)}{32\omega V_d^2}+k^5\frac{\sigma^2(3\sigma^2-23)}{32\omega V_d^2},$$

$$\gamma_{32}=-2\frac{V}{V_d^2}\gamma_2\beta_1+2\frac{V}{V_d^2}\gamma_1\beta_2-\frac{1}{V_d^2}\gamma_1\beta_1^2$$

$$+h^2\frac{V_d^2-2V^2}{6V_d^2}\gamma_1+k^2\frac{\sigma^2-1}{4V_d^2}\beta_2+k^2V\frac{\sigma^2+1}{2\omega V_d^2}\beta_1$$

$$+\omega k^3h^3\frac{(\sigma^2-1)^4}{32k\sigma^3V_d^2}+3\omega k^2h^2\frac{(-\sigma^2+1)^3}{32k\sigma^2V_d^2}$$

$$+\omega kh\frac{(\sigma^2-1)(3\sigma^2+1)}{32k\sigma V_d^2}-\omega\frac{\sigma^2+3}{32kV_d^2},$$

$$\gamma_{33}=4\frac{V}{V_d^2}\gamma_2\beta_1-2\frac{V}{V_d^2}\gamma_1\beta_2+\frac{2}{V_d^2}\gamma_1\beta_1^2$$

$$+h^2\frac{V_d^2-2V^2}{3V_d^2}\gamma_1+k^2\frac{-\sigma^2+1}{4V_d^2}\beta_2+\frac{k\omega-k^2V(\sigma^2+1)}{2\omega V_d^2}\beta_1$$

$$+\omega k^3h^3\frac{(\sigma^2-1)^3(\sigma^2+1)}{32k\sigma^3V_d^2}+\omega k^2h^2\frac{(\sigma^2-1)^2(-3\sigma^2+1)}{32k\sigma^2V_d^2}$$

$$+\omega kh\frac{(\sigma^2-1)(3\sigma^2-1)}{32k\sigma V_d^2}-\omega\frac{\sigma^2+1}{32kV_d^2}.$$

Coefficients in Equations

$$\tilde{\rho}_{11}=k^2\frac{\sigma^4-3}{2(\sigma^2+1)}\chi_1+k^4\frac{-5\sigma^4+16\sigma^2-3}{16\omega},$$

$$\tilde{\rho}_{12}=\frac{2\omega k-k^2V(\sigma^2-1)}{2\omega},$$

$$\tilde{\alpha}_1=\tilde{\rho}_{11}+\tilde{\rho}_{12}\gamma_1,$$

$$P_{21}=\left(k^2h^2\frac{-4\sigma^6+7\sigma^4-2\sigma^2-1}{8\sigma^2}\right.$$

$$\left.+kh\frac{4\sigma^4-9\sigma^2+3}{4\sigma}+\frac{-4\sigma^2+19}{8}\right)\gamma_1+k^2\frac{-\sigma^4+3}{2(\sigma^2+1)}\chi_2$$

$$+\left(k^2h\frac{-3\sigma^6+7\sigma^4-9\sigma^2-3}{4\sigma(\sigma^2+1)}+3k\frac{\sigma^4-5}{4(\sigma^2+1)}\right)\chi_1$$

$$+k^4h\frac{11\sigma^6-23\sigma^4+9\sigma^2+3}{16\omega\sigma}+k^3\frac{-11\sigma^4+40\sigma^2-9}{16\omega},$$

$$P_{22}=\left(-k^2h^2\frac{(\sigma^2-1)^2}{8}+kh\frac{\sigma^4-5\sigma^2+2}{4\sigma}\right.$$

$$\left.+\frac{-\sigma^2+8}{8}\right)\gamma_1$$

$$+\left(k^2h\frac{(\sigma^2-1)(\sigma^4+3)}{4\sigma(\sigma^2+1)}-k\frac{\sigma^4+3}{4(\sigma^2+1)}\right)\chi_1$$

$$+k^4h\frac{-3\sigma^6-5\sigma^4+11\sigma^2-3}{32\omega\sigma}+3k^3\frac{\sigma^4-1}{32\omega},$$

$$\tilde{\rho}_{21}=P_{21}+s\beta_1\gamma_1,\quad\tilde{\rho}_{22}=P_{22}-s\beta_1\gamma_1,$$

$$s=k^2\frac{\sigma^2-1}{2\omega},$$

$$\tilde{\alpha}_{21}=\tilde{\rho}_{21}-\tilde{\rho}_{12}\gamma_2,\quad\tilde{\alpha}_{22}=\tilde{\rho}_{22}+\tilde{\rho}_{12}\gamma_2,$$

$$P_{31}=\frac{1}{2\omega}\tilde{\alpha}_1^2+\frac{k^2V(\sigma^2-1)-2\omega k}{2\omega^2}\tilde{\alpha}_1\gamma_1$$

$$+k^2\frac{3\sigma^2+1}{\omega(\sigma^2+1)}\tilde{\alpha}_1\chi_1+3k^4\frac{\sigma^4-12\sigma^2+3}{16\omega^2}\tilde{\alpha}_1$$

$$+\left(k^4h^2\frac{(\sigma^2-1)^3}{8\omega}-k^3h\frac{(\sigma^2-1)^3}{4\omega\sigma}+k^2\frac{\sigma^4-\sigma^2+4}{8\omega}\right)\gamma_1^2$$

$$+\left(k^4h\frac{(\sigma^2-1)(2\sigma^6-7\sigma^4+8\sigma^2+1)}{4\omega\sigma(\sigma^2+1)}\right.$$

$$\left.+k^3\frac{-2\sigma^6+7\sigma^4-20\sigma^2-5}{4\omega(\sigma^2+1)}\right)\gamma_1\chi_1$$

$$+\left(k^6h\frac{\sigma(\sigma^2-1)^2(-7\sigma^2+13)}{16\omega^2}\right.$$

$$+ k^5\frac{7\sigma^6 - 23\sigma^4 + 49\sigma^2 - 9}{16\omega^2}\Big)\gamma_1$$

$$+ k^2\frac{\sigma^4 - 3}{2(\sigma^2 + 1)}\chi_{31} + k^4\frac{\sigma^2(\sigma^6 - 17\sigma^4 - \sigma^2 + 37)}{2\omega(\sigma^2 + 1)^2}\chi_1^2$$

$$+ 3k^2\frac{-\sigma^4 - 4\sigma^2 + 3}{2(\sigma^2 + 1)}\chi_1\nu_1$$

$$+ k^6\frac{-36\sigma^8 + 318\sigma^6 - 516\sigma^4 - 82\sigma^2 + 27}{48\omega^2(\sigma^2 + 1)}\chi_1$$

$$+ 3k^4\frac{19\sigma^6 + 5\sigma^4 - 51\sigma^2 + 3}{16\omega(3\sigma^2 + 1)}\nu_1$$

$$+ k^8\frac{\sigma^2(117\sigma^6 - 578\sigma^4 + 1175\sigma^2 - 378)}{384\omega^3},$$

$$P_{32} = \frac{2}{\omega}\tilde{\alpha}_1\beta_1 + \frac{V_d^2 - 2\omega h\sigma V}{\omega^2}\tilde{\alpha}_1$$

$$+\Big(-k^2h^2\frac{(\sigma^2 - 1)^2}{8} + kh\frac{\sigma^4 - 5\sigma^2 + 2}{4\sigma} + \frac{-\sigma^2 + 8}{8}\Big)\gamma_2$$

$$+ \frac{k^2V(\sigma^2 - 1) - 2\omega k}{2\omega^2}\gamma_1\beta_1$$

$$+\Big(-k^3h^3\frac{(\sigma^2 - 1)^2(12\sigma^4 + \sigma^2 + 1)}{16k\sigma^3}$$

$$+ k^2h^2\frac{24\sigma^6 - 53\sigma^4 + 16\sigma^2 + 5}{16k\sigma^2}$$

$$+ kh\frac{-12\sigma^4 + 27\sigma^2 - 7}{16k\sigma} + \frac{3}{16k}\Big)\gamma_1 + k^2\frac{-\sigma^2 + 1}{2\omega}\beta_1\chi_1$$

$$+ k^4\frac{5\sigma^4 - 16\sigma^2 + 3}{8\omega^2}\beta_1 + k^2\frac{\sigma^4 - 3}{2(\sigma^2 + 1)}\chi_{32}$$

$$+\Big(k^2h\frac{-3\sigma^6 + 7\sigma^4 - 9\sigma^2 - 3}{8\sigma(\sigma^2 + 1)} + 3k\frac{\sigma^4 - 5}{8(\sigma^2 + 1)}\Big)\chi_2$$

$$+\Big(k^2h^2\frac{-7\sigma^8 + 5\sigma^6 + \sigma^4 + 11\sigma^2 - 2}{8\sigma^2(\sigma^2 + 1)}$$

$$+ kh\frac{3\sigma^6 - 4\sigma^4 + 9\sigma^2 + 4}{4\sigma(\sigma^2 + 1)} + \frac{\sigma^4 - \sigma^2 + 6}{8(\sigma^2 + 1)}\Big)\chi_1$$

$$+ 3k^4h^2\frac{(\sigma^2 - 1)(11\sigma^6 - 17\sigma^4 - \sigma^2 - 1)}{32\omega\sigma^2}$$

$$+ k^3h\frac{(\sigma^2 - 1)(-17\sigma^4 + 28\sigma^2 + 3)}{16\omega\sigma} + k^2\frac{\sigma^4 - 10\sigma^2 + 3}{32\omega},$$

$$P_{33} = \frac{V_d^2 - 2\omega h\sigma V}{2\omega^2}\tilde{\alpha}_1$$

$$+\Big(k^2h^2\frac{(\sigma^2 - 1)^2}{8} + kh\frac{-\sigma^4 + 5\sigma^2 - 2}{4\sigma} + \frac{\sigma^2 - 8}{8}\Big)\gamma_2$$

$$+\Big(-3k^3h^3\frac{(\sigma^2 - 1)^2(\sigma^2 + 1)}{16k\sigma}$$

$$+ k^2h^2\frac{5\sigma^6 - 18\sigma^4 + 3\sigma^2 + 2}{16k\sigma^2}$$

$$+ kh\frac{-\sigma^4 + 13\sigma^2 - 4}{16k\sigma} + \frac{-\sigma^2 + 2}{16k}\Big)\gamma_1$$

$$+ k^2\frac{\sigma^4 + 6\sigma^2 + 1}{2\omega(\sigma^2 + 1)}\beta_1\chi_1 + k^4\frac{-7\sigma^4 - 4\sigma^2 + 3}{16\omega^2}\beta_1$$

$$+\Big(k^2h^2\frac{(\sigma^2 - 1)(-5\sigma^6 - 4\sigma^4 + 3\sigma^2 - 2)}{8\sigma^2(\sigma^2 + 1)}$$

$$+ kh\frac{(\sigma^2 - 1)^2(\sigma^2 + 2)}{4\sigma(\sigma^2 + 1)} + \frac{3\sigma^4 - 5\sigma^2 - 6}{8(\sigma^2 + 1)}\Big)\chi_1$$

$$+ k^4h^2\frac{(\sigma^2 - 1)(21\sigma^6 - 43\sigma^4 + \sigma^2 - 3)}{64\omega\sigma^2}$$

$$+ k^3h\frac{(\sigma^2 - 1)(-5\sigma^4 + 10\sigma^2 - 3)}{32\omega\sigma}$$

$$+ k^2\frac{-11\sigma^4 + 46\sigma^2 - 9}{64\omega},$$

$$P_{34} = \frac{2}{\omega}\tilde{\alpha}_1\beta_1 + 2\frac{V_d^2 - 2\omega h\sigma V}{\omega^2}\tilde{\alpha}_1$$

$$+\Big(k^2h^2\frac{(\sigma^2 - 1)^2(3\sigma^2 + 1)}{8\sigma^2}$$

$$+ kh\frac{(\sigma^2 - 1)(-3\sigma^2 + 1)}{4\sigma} + \frac{3\sigma^2 - 11}{8}\Big)\gamma_2$$

$$+\left(-k^2h\frac{(\sigma^2-1)(3\sigma^2+1)}{4\omega\sigma}+3k\frac{\sigma^2-1}{4\omega}\right)\gamma_1\beta_1$$

$$+\left(-k^3h^3\frac{(\sigma^2-1)^2(7\sigma^2+3)}{8k\sigma}\right.$$

$$+k^2h^2\frac{13\sigma^6-38\sigma^4+11\sigma^2+6}{8k\sigma^2}$$

$$\left.+kh\frac{-5\sigma^4+25\sigma^2-8}{8k\sigma}+\frac{-\sigma^2+2}{8k}\right)\gamma_1$$

$$+\left(k^2h\frac{(\sigma^2-1)(\sigma^4+3)}{4\sigma(\sigma^2+1)}-k\frac{\sigma^4+3}{4(\sigma^2+1)}\right)\chi_2$$

$$+\left(k^2h^2\frac{(\sigma^2-1)(-7\sigma^6-33\sigma^4-9\sigma^2+1)}{8\sigma^2(\sigma^2+1)}\right.$$

$$\left.+kh\frac{-\sigma^6+11\sigma^4-11\sigma^2+9}{4\sigma(\sigma^2+1)}+\frac{(\sigma^2-1)(9\sigma^2+5)}{8(\sigma^2+1)}\right)\chi_1$$

$$+k^4h^2\frac{(\sigma^2-1)(21\sigma^6-41\sigma^4-7\sigma^2+3)}{32\omega\sigma^2}$$

$$+k^3h\frac{(\sigma^2-1)(-\sigma^4+18\sigma^2-3)}{16\omega\sigma}+k^2\frac{-19\sigma^4+44\sigma^2-3}{32\omega},$$

$$P_{35}=\frac{1}{\omega}\tilde{\alpha}_1\beta_1+\frac{V_d^2-2\omega h\sigma V}{\omega^2}\tilde{\alpha}_1$$

$$+\left(-k^2h^2\frac{(\sigma^2-1)^2(3\sigma^2+1)}{8\sigma^2}+kh\frac{(\sigma^2-1)(3\sigma^2-1)}{4\sigma}\right.$$

$$\left.+\frac{-3\sigma^2+11}{8}\right)\gamma_2$$

$$+\left(k^2h\frac{(\sigma^2-1)(3\sigma^2+1)}{4\omega\sigma}+3k\frac{-\sigma^2+1}{4\omega}\right)\gamma_1\beta_1$$

$$+\left(-k^3h^3\frac{\sigma(\sigma^2-1)^2}{2k}+k^2h^2\frac{(\sigma^2-1)(2\sigma^4-3\sigma^2-1)}{2k\sigma^2}\right.$$

$$\left.+kh\frac{-\sigma^4+3\sigma^2-1}{2k\sigma}\right)\gamma_1+k^2\frac{\sigma^4-3}{2(\sigma^2+1)}\chi_{33}$$

$$+\left(k^2h\frac{-3\sigma^6+7\sigma^4-9\sigma^2-3}{8\sigma(\sigma^2+1)}+3k\frac{\sigma^4-5}{8(\sigma^2+1)}\right)\chi_2$$

$$+\left(k^2h^2\frac{-7\sigma^8+5\sigma^6+\sigma^4+11\sigma^2-2}{8\sigma^2(\sigma^2+1)}\right.$$

$$\left.+kh\frac{3\sigma^6-4\sigma^4+9\sigma^2+4}{4\sigma(\sigma^2+1)}+\frac{\sigma^4-\sigma^2+6}{8(\sigma^2+1)}\right)\chi_1$$

$$+k^4h^2\frac{(\sigma^2-1)(47\sigma^6-51\sigma^4-17\sigma^2-3)}{64\omega\sigma^2}$$

$$+k^3h\frac{(\sigma^2-1)(-27\sigma^4+28\sigma^2+9)}{32\omega\sigma}+k^2\frac{7\sigma^4-40\sigma^2+15}{64\omega},$$

$$\tilde{\rho}_{31}=P_{31}+s\left(\gamma_2\tilde{\alpha}_1-\gamma_1\frac{\tilde{\alpha}_{21}+\tilde{\alpha}_{22}}{2}\right),$$

$$\tilde{\rho}_{32}=P_{32}+s(\gamma_2\beta_1-\gamma_1\beta_2),$$

$$\tilde{\rho}_{33}=P_{33}+s(\gamma_2\beta_1-\gamma_1\beta_2),$$

$$\tilde{\rho}_{34}=P_{34}+s(\gamma_1\beta_2-2\gamma_2\beta_1),\quad \tilde{\rho}_{35}=P_{35},$$

$$\tilde{\alpha}_{31}=\tilde{\rho}_{31}+\tilde{\rho}_{12}\gamma_{31},\quad \tilde{\alpha}_{32}=\tilde{\rho}_{32}+\tilde{\rho}_{12}\gamma_{32},$$

$$\tilde{\alpha}_{33}=\tilde{\rho}_{33}+\tilde{\rho}_{12}\gamma_{32},$$

$$\tilde{\alpha}_{34}=\tilde{\rho}_{34}+\tilde{\rho}_{12}\gamma_{33},\quad \tilde{\alpha}_{35}=\tilde{\rho}_{35}.$$

Coefficients Used in Constructing the Complete Field

$$\tilde{r}_{01}=\sigma k\frac{k^2(\sigma^2-1)+4V\gamma_1}{4\omega^2},$$

$$\tilde{r}_{02}=k\frac{\sigma V}{\omega^2}\gamma_2+k\frac{\sigma}{\omega^2}\beta_1\gamma_1+k^2\sigma\frac{\omega-kV(1+\sigma^2)}{4\omega^3},$$

$$\tilde{r}_{21}=2k\frac{\sigma}{\omega}\chi_1+k^3\frac{\sigma(-3\sigma^2+1)}{4\omega^2},$$

$$\tilde{r}_{22}=-2k\frac{\sigma}{\omega}\chi_2+\left(kh\frac{-\sigma^4+8\sigma^2+1}{\omega(\sigma^2+1)}+\frac{\sigma}{\omega}\right)\chi_1$$

$$+k^3h\frac{(\sigma^2-1)(3\sigma^2+1)}{4\omega^2}+k^2\frac{\sigma(-3\sigma^2+1)}{4\omega^2},$$

$$\tilde{r}_{32}=k^3\frac{\sigma(-\sigma^4-6\sigma^2+1)}{\omega^2(\sigma^2+1)}\chi_1-3k\frac{\sigma}{\omega}\nu_1+3k^5\frac{\sigma^5}{8\omega^3}.$$

REFERENCES

1. V. I. Talanov, Pis'ma Zh. Éksp. Teor. Fiz. **2**, 218 (1965) [JETP Lett. **2**, 138 (1965)].
2. P. L. Kelley, Phys. Rev. Lett. **15**, 1005 (1965).
3. M. I. Rabinovich and D. I. Trubetskov, *Introduction to the Oscillation and Wave Theory* (Nauka, Moscow, 1984) [in Russian].

4. V. E. Zakharov, Prikl. Mekh. Tekh. Fiz. **2**, 86 (1968).
5. H. Hashimoto and J. Ono, J. Phys. Soc. Jpn. **33**, 805 (1972).
6. A. Davey, J. Fluid Mech. **53**, 769 (1972).
7. T. B. Benjamin and J. E. Feir, J. Fluid Mech. **27**, 417 (1967).
8. D. J. Benney and G. J. Roskes, Stud. Appl. Math. **48**, 377 (1969).
9. A. Davey and K. Stewartson, Proc. R. Soc. London, Ser. A **338**, 101 (1974).
10. E. M. Gromov and V. I. Talanov, Zh. Éksp. Teor. Fiz. **110**, 137 (1996) [JETP **83**, 73 (1996)].
11. K. B. Dysthe, Proc. R. Soc. London, Ser. A **369**, 105 (1979).
12. Yu. V. Sedletskiĭ, Zh. Éksp. Teor. Fiz. **124**, 200 (2003) [JETP **97**, 180 (2003)].
13. K. Trulsen, I. Kliakhandler, K. B. Dysthe, *et al.*, Phys. Fluids **12**, 2432 (2000).
14. R. S. Johnson, Proc. R. Soc. London, Ser. A **357**, 131 (1977).
15. T. Kakutani and K. Michihiro, J. Phys. Soc. Jpn. **52**, 4129 (1983).
16. E. J. Parkes, J. Phys. A: Math. Gen. **20**, 2025 (1987).
17. A. V. Slyunyaev and E. N. Pelinovskiĭ, Zh. Éksp. Teor. Fiz. **116**, 318 (1999) [JETP **89**, 173 (1999)].
18. A. V. Slyunyaev, Zh. Éksp. Teor. Fiz. **119**, 606 (2001) [JETP **92**, 529 (2001)].
19. R. S. Johnson, *A Modern Introduction to the Mathematical Theory of Water Waves* (Cambridge Univ. Press, Cambridge, 1997).
20. R. Grimshaw, D. Pelinovsky, E. Pelinovsky, *et al.*, Physica D (Amsterdam) **159**, 35 (2001).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 942–948.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 1078–1084.
Original English Text Copyright © 2005 by Mendonça, Bingham, Shukla.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

A Kinetic Approach to Bose–Einstein Condensates: Self-Phase Modulation and Bogoliubov Oscillations¶

J. T. Mendonça[a,b], R. Bingham[a,c], and P. K. Shukla[d,e]

[a]*Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 OQX, United Kingdom*
[b]*Instituto Superior Técnico, 1049-001 Lisboa, Portugal*
[c]*Department of Physics, University of Strathclyde, Glasgow G4 0NG, United Kingdom*
[d]*Institut für Theoretische Physik IV, Fakultät für Physik und Astronomie, Ruhr-Universität Bochum, D-44780 Bochum, Germany*
[e]*Department of Physics, Umeå University, SE-90187 Umeå, Sweden*
e-mail: T.Mendonca@rl.ac.uk; r.bingham@rl.ac.uk; ps@tp4.rub.de

Received April 22, 2005

Abstract—A kinetic approach to Bose–Einstein condensates (BECs) is proposed based on the Wigner–Moyal equation (WME). In the semiclassical limit, the WME reduces to the particle-number conservation equation. Two examples of applications are (i) a self-phase modulation of a BE condensate beam, where we show that part of the beam is decelerated and eventually stops as a result of the gradient of the effective self-potential, and (ii) the derivation of a kinetic dispersion relation for sound waves in BECs, including collisionless Landau damping. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Presently, Bose–Einstein condensates (BECs) provide one of the most active and creative areas of research in physics [1, 2]. The dynamics of BECs are usually described by a nonlinear Schrödinger equation (known in this field as the Gross–Pitaevskii equation (GPE) [3, 4]), which determines the evolution of a collective wavefunction of ultracold atoms in BECs, evolving in the mean field self-potential.

In this paper, we propose the use of an alternative but nearly equivalent approach to the physics of BECs, based on a kinetic equation for the condensate. We also show that this kinetic theory can lead to a more complete understanding of the physical processes occurring in BECs, not only by providing an alternative method for describing the system but also by improving our global view of the physical phenomena. It is our hope that this will also lead to the discovery of new aspects of BECs.

The key point of our approach is the use of the Wigner–Moyal equation (WME) for BECs, describing the spatiotemporal evolution of the appropriate Wigner function [5]. Wigner functions for BECs were discussed in the past [6, 7], and the WME has been used sporadically [8], but no systematic application of the WME to BECs has previously been considered. In the semiclassical limit, this equation reduces to the particle-number conservation equation, which is a kinetic equation formally analogous to the Liouville equation, but with a nonlinear potential. A description of BECs in terms of the kinetic equation is adequate in a series of problems, as is exemplified here, and can be seen as intermediate (in accuracy) between the GPE and the hydrodynamic equations usually found in the literature.

This paper is organized as follows. In Section 2, we establish the WME and discuss its approximate version as a kinetic equation for the Wigner function. We then apply the kinetic equation to two distinct physical problems. The first one, considered in Section 3, is the self-phase modulation of a BEC beam. A similar problem has been studied numerically in the past [9]. Here, we derive explicit analytical results and show that a part of the BEC beam is decelerated and eventually comes to a complete halt as a result of the collective forces acting on the condensate. The second example is considered in Section 4, where we establish a kinetic dispersion relation for sound waves in BECs, giving a kinetic correction to the usual Bogoliubov velocity of sound [10, 11] and predicting the occurrence of Landau damping [12, 13]. Our description of Landau damping is significantly different from that previously considered for transverse oscillations of BECs [14]. Finally, in Section 5, the virtues and limitations of the present kinetic approach are briefly discussed.

2. WIGNER–MOYAL EQUATION FOR THE BOSE CONDENSATE

It is known that, for an ultracold atomic ensemble and, in particular, for BECs, the ground-state atomic quantum field can be replaced by a macroscopic atomic

¶ The text was submitted by the authors in English.

wavefunction ψ. In a large variety of situations, the evolution of ψ is determined by the GPE

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi + (V_0 + V_{\text{eff}})\psi, \quad (1)$$

where $V_0 \equiv V_0(\mathbf{r})$ is the confining potential and V_{eff} is the effective potential that takes the interatomic interactions inside the condensate into account; in the simplest form,

$$V_{\text{eff}}(\mathbf{r}, t) = g|\psi(\mathbf{r}, t)|^2,$$

where g is a constant [3, 4].

We consider the situation where this wave equation can be replaced by a kinetic equation. To construct such an equation, we introduce the Wigner function associated with ψ via [5]

$$W(\mathbf{r}, \mathbf{k}, t) = \int\psi\left(\mathbf{r} + \frac{\mathbf{s}}{2}, t\right)\psi^*\left(\mathbf{r} - \frac{\mathbf{s}}{2}, t\right) \times \exp(-i\mathbf{k}\cdot\mathbf{s})d\mathbf{s}. \quad (2)$$

It is then possible to derive (see Appendix) the evolution equation for the Wigner function:

$$\left(\frac{\hbar^2}{2m}\mathbf{k}\cdot\nabla - i\hbar\frac{\partial}{\partial t}\right)W = -2V(\sin\Lambda)W, \quad (3)$$

where

$$\Lambda = \overleftarrow{\left(\frac{\partial}{\partial\mathbf{r}}\cdot\frac{\partial}{\partial\mathbf{p}}\right)}\overrightarrow{} \quad (4)$$

is a bidirectional differential operator that acts to the left on V and to the right on W [5]. In this equation, the potential is

$$V = V_0 + g\int W(\mathbf{r}, \mathbf{k}, t)\frac{d\mathbf{k}}{(2\pi)^3} + \delta V, \quad (5)$$

where

$$\delta V = g\left(|\psi(\mathbf{r}, t)|^2 - \int W(\mathbf{r}, \mathbf{k}, t)\frac{d\mathbf{k}}{(2\pi)^3}\right) \quad (6)$$

can be considered a noise term associated with the square mean deviations of the quasiprobability, determined by the Wigner function W with respect to the local quantum probability, which is determined by the wavefunction ψ.

Equation (3) can be seen as a WME describing the space and time evolution of BECs, and it is exactly equivalent to GPE (1). However, it is of little use in the above exact form, and it is convenient to introduce some simplifying assumptions. This is justified in the important case of slowly varying potentials. In this case, we can neglect the higher order spatial derivatives and introduce the approximation $\sin\Lambda \sim \Lambda$. This corresponds to the semiclassical approximation, where the quantum potential fluctuations can also be neglected, viz., $\delta V \longrightarrow 0$. Introducing these two simplifying assumptions, valid in the semiclassical limit, we reduce the WME to the much simpler form

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla + \mathbf{F}\cdot\frac{\partial}{\partial\mathbf{k}}\right)W = 0, \quad (7)$$

where $\mathbf{v} = \hbar\mathbf{k}/m$ is the velocity of the condensate atoms corresponding to the wavevector state $\mathbf{k}$ and $\mathbf{F} = -\nabla V$ is a force associated with the inhomogeneity of the condensate self-potential. The nonlinear term in GPE (1) is hidden inside this force $\mathbf{F}$. As we see in what follows, this nonlinear term looks very much like a ponderomotive force term, similar to radiation pressure.

We note that this new equation is a closed kinetic equation for the Wigner function W. In this semiclassical limit, W is just the particle occupation number for translational states with momentum $\mathbf{p} = \hbar\mathbf{k}$. Equation (7) is equivalent to a conservation equation, stating the conservation of the quasiprobability W in the six-dimensional classical phase space $(\mathbf{r}, \mathbf{k})$, and can also be written as

$$\frac{d}{dt}W(\mathbf{r}, \mathbf{k}, t) = 0. \quad (8)$$

This kinetic equation can then be used to describe physical processes occurring in a BEC, as long as the semiclassical approximation of slowly varying potentials is justified. The interest in such kinetic descriptions is illustrated with the aid of two simple and different examples, to be presented in the next two sections. Many other applications can be envisaged and will be explored in the future.

3. SELF-PHASE MODULATION OF A BEAM CONDENSATE

We first consider the kinetic description of self-phase modulation of a BEC gas moving with respect to the confining potential $V_0(\mathbf{r})$. Here, we can explore the similarity of this problem to that of self-phase modulation of short laser pulses moving in a nonlinear optical medium, which is well known in the literature [15]. To simplify our description, we consider the one-dimensional problem of a beam moving along the z axis and neglect the axial variation of the background potential, $\partial V_0/\partial z \approx 0$. The radial structure of the beam can easily be introduced later and does not substantially modify

the results obtained here. Kinetic equation (7) can then be written as

$$\left(\frac{\partial}{\partial t} + v_z\frac{\partial}{\partial z} + F_z\frac{\partial}{\partial k}\right)W(z, k, t) = 0, \quad (9)$$

with v_z and F_z given by

$$v_z = \frac{\hbar k}{m} + g\frac{\partial}{\partial t}I(z, t), \quad F_z = \frac{dk}{dt} = -g\frac{\partial}{\partial z}I(z, t), \quad (10)$$

where we have used the intensity of the beam condensate defined by

$$I(z, t) = \int W(z, k, t)\frac{dk}{2\pi}. \quad (11)$$

We assume that an ultracold atomic beam has the mean velocity $v_0 = \hbar k_0/m$. This suggests the use of the new space coordinate $\eta = z - v_0 t$. In terms of this new coordinate, the semiclassical equations of motion of a cold atom in the beam can be written as

$$\frac{d\eta}{dt} = \frac{\partial h}{\partial k} = \frac{1}{m}(k - k_0),$$
$$\frac{dk}{dt} = -\frac{\partial h}{\partial \eta} = -\frac{g}{\hbar}\frac{\partial}{\partial \eta}I(\eta, t), \quad (12)$$

where we have introduced the Hamiltonian function

$$h(\eta, k, t) = \omega(\eta, k, t) - kv_0 = \frac{k}{m}\left(\frac{k}{2} - k_0\right) + \frac{g}{\hbar}I(\eta, t). \quad (13)$$

Here, $\omega(\eta, k, t)$ is the Hamiltonian in the rest frame expressed in the new coordinate. A straightforward integration of the equations of motion leads to

$$k(t) = k_0 - \frac{g}{\hbar}\int_0^t \frac{\partial}{\partial \eta}I(\eta, t')dt'. \quad (14)$$

At this point, it is useful to introduce the concept of the beam energy chirp, $\langle \epsilon(\eta, t)\rangle$, in analogy with the frequency chirp of short laser pulses [15]. By definition, it is the beam mean energy at a given position and a given time,

$$\langle \epsilon(\eta, t)\rangle = \hbar\int W(\eta, k, t)\omega(\eta, k, t)\frac{dk}{2\pi}, \quad (15)$$

where the weight function $W(\eta, k, t)$ is the solution of one-dimensional kinetic equation (9). A formal solution of this equation can be written as

$$W(\eta, k, t) = W(\eta_0(\eta, k, t), k_0(\eta, k, t), t_0), \quad (16)$$

where η_0 and k_0 are the initial conditions corresponding to the observed values at time t, as determined by dynamical equations (12). With (16) used in Eq. (15), we obtain

$$\langle \epsilon(\eta, t)\rangle = \hbar\int W(\eta_0, k_0, t_0)\left[\frac{k^2}{2m} + \frac{g}{\hbar}I(\eta, t)\right]\frac{dk}{2\pi}. \quad (17)$$

From Eq. (14), we see that $dk = dk_0$. Neglecting higher order nonlinearities, we can then rewrite the above expression as [15]

$$\langle \epsilon(\eta, t)\rangle = \langle \epsilon(0)\rangle - \frac{k_0}{m}g\int_0^t \frac{\partial}{\partial \eta}I(\eta, t')dt', \quad (18)$$

where $\langle \epsilon(0)\rangle \equiv \langle \epsilon(\eta_0, t_0)\rangle$ is the initial beam energy chirp.

We first consider the case where the beam profile $I(\eta)$ is independent of time. This is, of course, only valid for very short time intervals where the beam velocity dispersion is negligible. In this simple case, we have

$$\langle \epsilon(\eta, t)\rangle = \langle \epsilon(0)\rangle - \hbar v_0 g\frac{\partial I}{\partial \eta}t. \quad (19)$$

The maximum energy shift is attained at some position inside the beam profile, $\eta = \eta_{max}$, determined by the stationarity condition

$$\frac{\partial}{\partial \eta}\langle \epsilon(\eta, t)\rangle = \frac{\partial^2 I}{\partial \eta^2} = 0. \quad (20)$$

To deduce more specific answers, we assume a Gaussian beam profile

$$I(\eta) = I_0\exp(-\eta^2/\sigma^2), \quad (21)$$

where σ determines the beam width. For this profile, we have $\eta_{max} = \pm\sigma/\sqrt{2}$, which leads to the maximum energy shift

$$\Delta\epsilon(t) \equiv \langle \epsilon(t)\rangle_{max} - \langle \epsilon(0)\rangle = \pm\frac{\hbar\sqrt{2}}{\sigma}gv_0 I_0\exp\left(-\frac{1}{2}\right)t. \quad (22)$$

This is similar to the well-known result in nonlinear optics that states that the maximum energy chirp due to a self-phase modulation is proportional to t, or to the distance traveled by the beam, $d = v_0 t$. This result clearly indicates that the initial beam eventually splits into two parts, one being accelerated to higher translational speeds and the other being decelerated. This corresponds to the redshift and blueshift observed in non-

linear optics. The decelerated beam eventually stops after a time $t \approx \tau$, such that $\Delta\epsilon(\tau) = \langle\epsilon(0)\rangle$. This determines the condition for translational beam freezing.

We note that the same result could also be obtained directly from GPE (1), but the present derivation is interesting, because it demonstrates the irrelevance of the phase of the wavefunction ψ, which was ignored in our kinetic calculation. Therefore, instead of the self-phase modulation, it would be more appropriate to call it the beam self-deceleration.

Another interesting aspect of our kinetic approach is that it can be easily refined, as is briefly shown here. We can improve the above calculation by considering the beam dispersion. It inevitably becomes relevant because of the linear velocity dispersion of the atomic beam. Such a dispersion decreases the chirping effect, because of the decrease of $\partial I/\partial\eta$ in time. To model it, we can assume a time-varying Gaussian beam shape, as described by

$$I(\eta, t) = I_0\left(\frac{\sigma_0}{\sigma(t)}\right)^{1/2}\exp\left(-\frac{\eta^2}{\sigma^2(t)}\right). \quad (23)$$

If we now assume that

$$\sigma(t) = \sigma_0(1 + \delta t^2),$$

where

$$\delta = \frac{2m}{\hbar^2}\frac{\Delta\epsilon_0}{\sigma_0}$$

is proportional to the initial energy spread $\Delta\epsilon_0$, we obtain a new expression for the maximum energy shift, of the form

$$\Delta\epsilon_d(t) = \frac{\ln t}{t}\Delta\epsilon(t), \quad (24)$$

where $\Delta\epsilon(t)$ is determined by Eq. (22). It is clear that the linear beam velocity dispersion decreases the maximum attainable chirp, by changing the linearity with time into a logarithmic law. However, this only occurs for very long times, $t \sim 1/\sqrt{\delta}$, which are not relevant for ultracold atomic beams with a very low translational energy dispersion $\Delta\epsilon_0$.

The other cause of the beam dispersion is the nonlinear process itself, which eventually breaks the initial pulse into two distinct pulses. In this case, the self-phase modulation process is not attenuated because the beam width is conserved, but the two secondary pulses suffer self-phase modulation themselves and eventually break up later, resulting in the formation of several secondary pulses with different mean energies. However, the nonlinear dispersion is also negligible whenever

$$\sigma_0^2 > \frac{4m}{\hbar^2}|\Delta\epsilon(t)|t^2.$$

A more complete description of all these dispersion regimes can be obtained by solving kinetic Eq. (9) numerically.

4. KINETIC DESCRIPTION OF BOGOLIUBOV OSCILLATIONS

The second example of an application of the kinetic equation for BECs deals with the dispersion relation of sound waves. For simplicity, we again consider the one-dimensional model and neglect the radial structure of the oscillations. This allows us to treat the lowest order oscillating modes of the condensate. We assume some given equilibrium distribution $W_0(z, k, t)$, for instance, corresponding to the Thomas–Fermi equilibrium solution in a given confining potential $V_0(\mathbf{r}_\perp, z)$ [16], and after linearization of the one-dimensional kinetic equation (9) with respect to the perturbation $\tilde{W}$, we obtain

$$\left(\frac{\partial}{\partial t} + v_z\frac{\partial}{\partial z}\right)\tilde{W}(z, k, t) + \tilde{F}\frac{\partial}{\partial k}W_0(z, k, t) = 0, \quad (25)$$

where the perturbed force is determined by

$$\tilde{F} = -\frac{g}{\hbar}\frac{\partial}{\partial z}\tilde{I}(z, t) = -\frac{g}{\hbar}\frac{\partial}{\partial z}\int\tilde{W}(z, k, t)\frac{dk}{2\pi}. \quad (26)$$

We now assume perturbations of the form $\tilde{W}, \tilde{I} \sim \exp(ikz - i\omega t)$. From the above equations, we then obtain a relation between the perturbation amplitude of the Wigner function $\tilde{W}$ and the perturbed beam intensity $\tilde{I}$,

$$\tilde{W} = -\frac{gk}{\hbar(\omega - kv')}\tilde{I}\frac{\partial}{\partial k'}W_0(k'), \quad (27)$$

where we now specify the particle wavenumber state with k' in order to avoid confusion with the wavenumber k of the oscillation that we intend to study. The velocity corresponding to this particle state is $v' = \hbar k'/m$. Integration over the momentum spectrum of the particle condensate then leads to the equation

$$1 + \frac{g}{\hbar}k\int\frac{\partial W_0(k')/\partial k'}{\omega - \hbar kk'/m}\frac{dk'}{2\pi} = 0. \quad (28)$$

This is the kinetic dispersion relation for axial perturbations in BECs. We illustrate this result by considering

the simple case of a condensate beam with no translational dispersion or with a translational temperature exactly equal to zero. The equilibrium state of the beam can then be described by

$$W_0(k') = 2\pi n_0 \delta(k' - k_0'), \tag{29}$$

where

$$n_0 = \frac{1}{2\pi}\int W_0(k')dk'$$

is the particle number density in the condensate. Replacing this in dispersion relation (28), we have

$$1 - \frac{gk^2}{m}\frac{n_0}{(\omega - kv_0')^2} = 0, \tag{30}$$

where $v_0' = \hbar k_0'/m = p_0'/m$ is the beam velocity. This can also be written as

$$(\omega - kv_0')^2 = k^2 c_s^2, \tag{31}$$

where

$$c_s = \sqrt{gn_0/m} \tag{32}$$

is nothing but the Bogoliubov velocity of sound. Obviously, Eq. (31) is the Doppler-shifted dispersion relation of sound waves in the BEC gas. In its reference frame, it reduces to $\omega = kc_s$.

We now consider a situation where, instead of distribution (29), we have a beam with a small translational velocity spread, such that the number of particles with a velocity $v' \sim c_s$ is small but nonzero. In this case, the resonant contribution in the integral of Eq. (28) has to be retained, although it is still possible to neglect the kinetic corrections in the principal part of the integral. The dispersion relation can then be written, in the condensate frame of reference, as

$$1 - \frac{k^2 c_s^2}{\omega^2} - \frac{i}{2}\frac{gm}{\hbar^2}\left(\frac{\partial W_0}{\partial k'}\right)_{k' = k_s'} = 0, \tag{33}$$

where $k_s' = mc_s/\hbar$ is the resonant momentum. The imaginary term in this equation can lead to damping of sound waves. Writing $\omega = kc_s + i\gamma$, with $|\gamma| \ll kc_s$, we obtain the expression for the damping coefficient

$$\gamma = \frac{\omega}{4}\frac{gm}{\hbar^2}\left(\frac{\partial W_0}{\partial k'}\right)_{k' = k_s'}. \tag{34}$$

This expression corresponds to the noncollisional Landau damping of Bogoliubov oscillations in BECs. The present approach can also be generalized in a straightforward way to higher order oscillations of the condensate, where the radial structure has to be taken into account [11, 17].

5. CONCLUSIONS

We have proposed a kinetic view of the Bose–Einstein condensate physics, based on the Wigner–Moyal equation. In the semiclassical limit, the latter can be reduced to a closed kinetic equation for the corresponding Wigner function. The kinetic approach to BECs can be seen as an intermediate step between the GPE and the hydrodynamic equations for the condensate gas, often found in the literature.

We have discussed two different physical problems, in order to illustrate the versatility of the kinetic theory. One is a self-phase modulation of a BEC beam. The other is the dispersion relation of the Bogoliubov oscillations in the condensate gas. The first example shows that due to the influence of its own inhomogeneous self-potential, nearly half of the beam is accelerated, while the other half is decelerated. Under certain conditions, the decelerated part of the beam tends to a complete halt. The second example shows that a kinetic dispersion relation for sound waves in BECs can be established where Landau damping is automatically included. The present results only involve the lowest order modes, but the same approach can be used to describe higher order oscillations of BECs, including their radial structures, as well as the coupling to a background thermal gas. This investigation is beyond the scope of the present work, however.

Several other different problems relevant to BECs can also be considered in the framework of the kinetic theory, such as modulational instabilities [18] and the wakefield generation. This indicates that the kinetic theory is a very promising approach to the physics of BECs, which will eventually allow introducing new ideas in this stimulating area of research and suggesting new configurations to the experimentalists. However, the present work also clearly states that the present theory is only valid in the semiclassical limit, and therefore, some relevant problems, where the phase of the BEC wavefunction plays an important role can only be treated by means of the GPE. Surprisingly, the self-phase modulation is not one of them, as demonstrated here.

ACKNOWLEDGMENTS

One of the authors (J.T.M.) appreciates the hospitality of the Rutherford Appleton Laboratory.

This work was supported, in part, the Centre for Fundamental Physics.

APPENDIX

Derivation of the Wigner–Moyal Equation

In this derivation, we follow a procedure already used in other cases, for instance, in the case of electromagnetic waves moving in a space- and time-dependent dielectric medium [19] (for a different but nearly equivalent derivation of the WME, see the Appendix in [8]). We consider two distinct sets of values for space and time coordinates, $(\mathbf{r}_1, t_1)$ and $(\mathbf{r}_2, t_2)$, and use the notation $\psi_j = \psi(\mathbf{r}_j, t_j)$ and $V_j = V(\mathbf{r}_j, t_j)$ for $j = 1, 2$. This allows us to write two versions of GPE (1) as

$$\left(\frac{\hbar^2}{2m}\nabla_j^2 - i\hbar\frac{\partial}{\partial t_j}\right)\psi_j = -V_j\psi_j. \tag{35}$$

Multiplying the $j = 1$ equation by ψ_2^* and the conjugate of the $j = 2$ equation by ψ_1, and subtracting the resulting equations, we obtain

$$\left[\frac{\hbar^2}{2m}(\nabla_1^2 - \nabla_2^2) - i\hbar\left(\frac{\partial}{\partial t_1} + \frac{\partial}{\partial t_2}\right)\right]C_{12} = -(V_1 - V_2)C_{12}, \tag{36}$$

where we set $C_{12} = \psi_1\psi_2^*$. The above equation suggests the use of two pairs of space and time variables,

$$\begin{aligned}\mathbf{r}_1 &= \mathbf{r} - \mathbf{s}/2, \quad t_1 = t - \tau/2,\\ \mathbf{r}_2 &= \mathbf{r} + \mathbf{s}/2, \quad t_2 = t + \tau/2.\end{aligned} \tag{37}$$

We can then rewrite the above equation as

$$\left[\frac{\hbar^2}{m}\frac{\partial}{\partial\mathbf{r}}\cdot\frac{\partial}{\partial\mathbf{s}} - i\hbar\frac{\partial}{\partial t}\right]C_{12} = -(V_1 - V_2)C_{12}. \tag{38}$$

It can also easily be shown, by expanding the potentials V_j around $V(\mathbf{r}, t)$, that

$$(V_1 - V_2) = 2\sinh\left(\frac{\mathbf{s}}{2}\cdot\frac{\partial}{\partial\mathbf{r}} + \frac{\tau}{2}\frac{\partial}{\partial t}\right)V(\mathbf{r}, t). \tag{39}$$

We now introduce the double Fourier transform of the function $C_{12} \equiv C(\mathbf{r}, \mathbf{s}, t, \tau)$ in the variables $\mathbf{s}$ and τ, defined by

$$W(\mathbf{r}, t, \omega, \mathbf{k}) = \int ds\int d\tau C(\mathbf{r}, \mathbf{s}, t, \tau) \times \exp(-i\mathbf{k}\cdot\mathbf{s} + i\omega\tau). \tag{40}$$

It can be rewritten in terms of the wavefunction ψ as

$$W(\mathbf{r}, t, \omega, \mathbf{k}) = \int ds\int d\tau\psi\left(\mathbf{r} + \frac{\mathbf{s}}{2}, t + \frac{\tau}{2}\right) \times \psi^*\left(\mathbf{r} - \frac{\mathbf{s}}{2}, t - \frac{\tau}{2}\right)\exp(-i\mathbf{k}\cdot\mathbf{s} + i\omega\tau). \tag{41}$$

Using this in Eq. (38), we obtain the equation

$$\left(\frac{\hbar^2}{m}\mathbf{k}\cdot\frac{\partial}{\partial\mathbf{r}} - \hbar\frac{\partial}{\partial t}\right)W = -2V(\sin\Lambda')W \tag{42}$$

for the Fourier transform, where we use the differential operator

$$\Lambda' = \frac{1}{2}\overleftarrow{\left(\frac{\partial}{\partial\mathbf{r}}\cdot\frac{\partial}{\partial\mathbf{k}} - \frac{\partial}{\partial t}\frac{\partial}{\partial\omega}\right)}\overrightarrow{} \tag{43}$$

acting to the left on the potential $V(\mathbf{r}, t)$ and to the right on W.

This is a formidable equation for W, which can be simplified by noting that the GPE implies the existence of a well-defined relation between energy and momentum. This means that ω must be equal to some function of $\mathbf{k}$, or $\omega = \omega(\mathbf{k})$. Hence, we can state that

$$W(\mathbf{r}, t, \omega, \mathbf{k}) = W(\mathbf{r}, \mathbf{k}, t)\delta(\omega - \omega(\mathbf{k})). \tag{44}$$

This leads to a much simpler evolution equation for $W(\mathbf{r}, \mathbf{k}, t)$. Before writing it, we also note that the nonlinear term in V depends on $|\psi|^2$ and not on the function W. Thus, we can finally write

$$\left(\frac{\hbar^2}{2m}\mathbf{k}\cdot\nabla - i\hbar\frac{\partial}{\partial t}\right)W = -2(V_0 + g|\psi|^2)(\sin\Lambda)W, \tag{45}$$

where Λ is the simpler differential operator

$$\Lambda = \overleftarrow{}\left(\frac{\partial}{\partial\mathbf{r}}\cdot\frac{\partial}{\partial\mathbf{p}}\right)\overrightarrow{}. \tag{46}$$

The function $W(\mathbf{r}, \mathbf{k}, t)$ can be seen as the Wigner function associated with the GPE, and Eq. (45), as the WME equation that describes its spatiotemporal behavior. This equation is equivalent to the initial wave equation (1), but it is not a closed equation for the quasiprobability function W. Therefore, some simplifying assumptions have to be introduced in order to make it more tractable, as explained in Section 2.

REFERENCES

1. A. J. Legget, Rev. Mod. Phys. **73**, 307 (2001).
2. F. Dalfovo, S. Giorgini, L. P. Pitaevskii, and S. Stringari, Rev. Mod. Phys. **71**, 463 (1999).

3. E. P. Gross, Nuovo Cimento **20**, 454 (1961); J. Math. Phys. **4**, 195 (1963).
4. L. P. Pitaevskiĭ, Zh. Éksp. Teor. Fiz. **40**, 646 (1961) [Sov. Phys. JETP **13**, 451 (1961)].
5. M. Hillary, R. F. O'Connel, M. O. Scully, and E. P. Wigner, Phys. Rep. **106**, 121 (1984).
6. M. J. Steel, M. K. Olsen, L. I. Plimak, *et al.*, Phys. Rev. A **58**, 4824 (1998).
7. C. W. Gardiner and M. J. Davis, J. Phys. B: At. Mol. Opt. Phys. **36**, 4731 (2003).
8. S. A. Gardiner, D. Jaksch, R. Dum, *et al.*, Phys. Rev. A **62**, 023612 (2000).
9. W. Zhang and D. F. Walls, Phys. Rev. A **49**, 3799 (1994).
10. M. R. Andrews, M.-O. Mewes, N. J. van Druten, *et al.*, Phys. Rev. Lett. **79**, 553 (1997).
11. E. Zaremba, Phys. Rev. A **57**, 518 (1998).
12. L. P. Pitaevskii, Phys. Lett. A **235**, 398 (1997).
13. B. Jackson and E. Zaremba, New J. Phys. **5**, 88 (2003).
14. M. Guilleumas and L. P. Pitaevskii, Phys. Rev. A **67**, 053607 (2003).
15. L. O. Silva and J. T. Mendonça, Opt. Commun. **196**, 285 (2001).
16. G. Byam and C. J. Pethick, Phys. Rev. Lett. **76**, 6 (1996).
17. S. Stringari, Phys. Rev. Lett. **77**, 2360 (1996).
18. V. V. Konotop and M. Salerno, Phys. Rev. A **65**, 021602 (2002).
19. J. T. Mendonça and N. L. Tsintsadze, Phys. Rev. E **62**, 4276 (2000).

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 949–961.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 1085–1098.
Original Russian Text Copyright © 2005 by Matveev.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

Evolution of a Langmuir Wave in a Weakly Inhomogeneous Plasma with a Positive Concentration Gradient

A. I. Matveev

Taganrog State University of Radio Engineering, Taganrog, 347928 Russia
e-mail: physics@EGF.tsure.ru

Received March 17, 2005

Abstract—Spatial evolution of a Langmuir wave excited by external sources in a weakly inhomogeneous electron plasma without external sources is considered for a small positive gradient of the plasma concentration in the direction of propagation of the wave. At the first state of the evolution, the dispersion of the wave is close to linear. When the phase velocity is doubled, the second stage of the evolution begins. The wave loses its individuality and becomes a hybrid of two waves. Its profile acquires the shape of an alternating sequence of fragments of these waves. The wave dispersion is determined by the dispersion of each fragment. In the course of evolution, the spacing between the equilibrium values of the wave fragments increases; as a result, the wave decays into two waves, which are also loaded by trapped electrons. Prior to decay, the humps of the wave become steeper; as a result, at the instant of the decay, the wave is transformed into a sequence of solitons with different polarities. *© 2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Nonlinear phenomena emerging during the evolution of a wave in an inhomogeneous plasma as a result of the resonant interaction with particles were considered by many authors [1–6]. This type of problems can be solved completely only with the complete system of the Vlasov–Maxwell equations. In most cases, this system was solved either after linearization of the equations, or using asymptotic methods [7–9]. In the case when the wave parameters change insignificantly over distances commensurate to the wavelength, the electron distribution is described with an exponential accuracy with the help of adiabatic invariants [10–12]. The application of such invariants leads to less cumbersome expressions for all possible moments of the distribution function as compared to asymptotic methods and makes it possible to separate the resonance and nonresonance components in these moments without solving the wave equation [1, 12]. The main difficulty encountered in solving the Vlasov–Maxwell equations is the construction of self-consistent solutions. However, the solution of a self-consistent problem using the adiabatic approach is simplified since it is not necessary to describe the motion of charged particles in detail; it is sufficient to find the function of distribution over adiabatic invariants or integrals of motion [13, 14].

We will describe the evolution of a Langmuir wave after its initiation by external sources in a weakly inhomogeneous plasma with a positive gradient of the electron concentration ($dN/dz > 0$). The evolution is characterized by three stages. At the first stage, anharmonism of the wave and the difference of its dispersion from the linear dispersion are taken into account by a nonlinear correction. With increasing concentration and phase velocity, the electrons trapped in potential wells of the wave are condensed at the bottom of these wells and the vacated region of the phase space is gradually filled with newly trapped electrons [2, 12]. The effect of trapped electrons on the wave profile is local since it results in deepening of the potential wells of the wave only in the vicinity of their minima. As a matter of fact, the Maxwell distribution function rapidly decreases exponentially; consequently, the number of electrons trapped at the initial instant of evolution as the phase velocity increases is considerably larger than the number of electrons trapped at any subsequent instant. The electrons trapped at the beginning of evolution are condensed at the bottom of the potential wells of the wave; as a result, a characteristic energy level H_0 is formed, such that most of such electrons are below this level. The current of electrons with an energy $H < H_0$ becomes large, exceeding the current of the remaining electrons. However, the phase interval of motion of these electrons is bounded by the walls of the potential wells. Consequently, the current of such electrons may change the wave profile, mainly in the phase interval $|\psi| < \theta_0$, $\theta_0 = \varphi^{-1}(H_0)$: the wave displays "sagging" of the potential in the region of its minima. The sag depth increases in the course of evolution and the sag is transformed into a fragment of the new wave. As a result, the Langmuir wave at the second stage of evolution can be represented as a sequence of alternating fragments of two waves with different wavelengths, which are continuously transformed into each other. The upper part of

the wave consists of positive humps of the initial wave, while the lower part consists of negative humps of the newly generated wave. Each sequence of fragments possesses its own dispersion (their wavelengths vary independently in the course of evolution). Consequently, a hybrid of two waves is in fact a wave with double dispersion in contrast to the classical Langmuir wave obtained by linearizing the Vlasov–Maxwell equations. The features of the evolution of the wave with double dispersion at the second stage are considered in Section 3.

The size of the fragments of both waves increases during evolution, and the formation of fragments is completed at the end of the second stage: each fragment has its own wavelength and amplitude. Consequently, when the difference between two equilibrium values of fragments of two waves becomes larger than the sum of their amplitude ($A_1 + A_2$), the hybrid decays into two waves loaded with trapped electrons. However, the nonlinearity of the wave produces a strong effect during the short period prior to the decay; as a result, the positive and negative humps of this wave become steeper. The wave profile acquires the shape of a sequence of alternating solitons with the positive and negative polarities. The evolution of such solitons is considered in Sections 4 and 5.

The phenomenon of unstable decay of a high-frequency plasma wave is well known [15, 16]. The decay of this wave into two low-frequency waves is possible under the synchronization condition $\omega_0 = \omega_1 + \omega_2$, $k_0 = k_1 + k_2$, where the subscripts 0, 1, and 2 correspond to the initial wave and the two waves resulting from the decay, respectively. In the decay of a wave in a weakly inhomogeneous plasma, considered below, the conditions for conservation of the mean densities of energy flux and current comprise these two conditions.

2. FORMULATION OF THE PROBLEM: CURRENTS OF TRANSIT AND TRAPPED ELECTRONS

Let us consider the time-independent self-consistent problem of generation and propagation of a longitudinal slow wave in a weakly inhomogeneous electron plasma whose concentration slowly varies along the z axis in the following formulation. We assume that external sources located in the region $z < 0$, where the plasma is homogeneous in zero field, slightly feed the slow wave being generated so that the amplitude of the wave propagating along the z axis increases from zero at $z \longrightarrow -\infty$ to $A(0)$ at $z = 0$. The unperturbed electron distribution function $f_0(v^2/2T)$ at $z \longrightarrow -\infty$ is known. External sources are intended for sustaining the self-consistent field of the wave of a preset type, taking into account its self-action. In the case of small amplitudes, we can show that, for a wave initiated in this way, the currents of resonance transit and trapped electrons compensate each other to within terms proportional to $\sqrt{A}$; consequently, the dispersion relation is close to linear at the stage of wave initiation. Since the variation of the amplitude (provided that it is small) does not affect the change in concentration N, the phase velocity of the wave in the course of initiation can be treated as constant and equal to u_0. In the region $z > 0$, where external sources are absent, the electron plasma concentration N increases very slowly along the z axis ($dN/dz > 0$). The evolution of the wave in the region of its initiation, as well as in the region $z > 0$ without external sources, occurs so slowly that it can be disregarded over distances comparable to the wavelength.

We will employ the dimensionless form of notation, in which time t and coordinate z are divided by ω^{-1} and k_0^{-1}, respectively; the phase velocity and the velocities of electrons are divided by $u_0 = \omega/k_0$; the distribution function $f_0(v^2/v_T^2)$, which is normalized to unity, is divided by k_0/ω; the electron concentration N is divided by $n_{cr} \equiv m\omega^2/4\pi e^2$; the current density j is divided by $e\omega n_{cr}/k_0$; the electron temperature $T = mv_T^2/2$ is divided by mu_0^2; and potential φ is divided by mu_0^2/e.

It was shown in [1] that arbitrary functions $f^{\pm}(I^{\pm})$ for transit electrons and $f_T(J)$ for trapped electrons form a solution of the self-consistent system of Vlasov–Maxwell equations for a Langmuir wave in a weakly inhomogeneous plasma. Here,

$$I^{\pm} = \left\langle \frac{v^2}{2} \right\rangle = \frac{u^2}{2} + H \pm u\int_0^{2\pi} \frac{d\psi}{2\pi}\sqrt{2(H-\varphi)},$$

$$J = u\int_{\psi_1}^{\psi_2} \frac{d\psi}{2\pi}\sqrt{2(H-\varphi)}$$

are the adiabatic invariants of transit and trapped electrons [17]; the plus and minus signs correspond to leading and retarded transit electrons, respectively; ψ_1 and ψ_2 are turning points for which the radicand in the integrand vanishes; and

$$H = \frac{(v-u)^2}{2} + \varphi$$

is the electron energy in the noninertial reference frame. For a small wave anharmonism, we have

$$J = \frac{4}{\pi}u\sqrt{A}\kappa^2 B(\kappa), \qquad (2.1)$$

where $\kappa^2 = H/2A$ is the trapping parameter,

$$B(\kappa) = \frac{E(\kappa) - (1-\kappa^2)K(\kappa)}{\kappa^2},$$

and $K(\kappa)$ and $E(\kappa)$ are the first- and second order elliptic integrals.

In the region $z > 0$, most leading transit electrons whose velocity becomes equal to the phase velocity of the wave become retarded upon an increase in the phase velocity; however, a small part of these electrons are trapped by the wave. If $N_s = N(u_s)$ and u_s are the electron concentration and velocity at the instant of their trapping by the wave, $I_s^{\pm}(R) = G \pm R$ and $R = J(H = \varphi_m)$ are the values of the adiabatic invariants of transit and trapped electrons at the separatrix, $G = u^2/2 + \varphi_m$, and φ_m is the maximal value of φ, the electron distribution after their trapping can be written in the form [1]

$$N_T f_T(J) = \begin{cases} \dfrac{N_s u_s}{u} f_0\left(\dfrac{G(R)}{T}\right)\Big|_{R=J}, & J > R_0, \\ \dfrac{2N_0}{u} f_0\left(\dfrac{G_0(R_0)}{T}\right)\Big|_{R_0=J}, & J < R_0, \end{cases} \tag{2.2}$$

where $G \approx u_s^2/2$, $R_0 = R(z = 0)$, and $N_0 = N(z = 0)$. The concentration N_T of trapped electrons in formula (2.2) is defined with the help of the Liouville equation. Equating the value of adiabatic invariant (2.1) at the instant of electron trapping by the wave to its value at any point $z > 0$, at which the amplitude and phase velocity are equal to A and $u > 1$, while $B(\kappa) \approx 1$, we express G in terms of u and H and substitute it into relations (2.2),

$$N_T f_T(J) = \begin{cases} \dfrac{N_s u_s}{u} f_0(\gamma H^2), & H > H_0, \\ 2\dfrac{N_0}{u} f_0(v_T^{-2}), & H < H_0, \end{cases} \tag{2.3}$$

where $\gamma = u^2/8TAA_0$ and H_0 is the value of energy separating the electrons trapped during the initiation of the wave ($z < 0$) from the electrons trapped in the region $z > 0$, where external sources are absent ($H_0 \approx 2\sqrt{AA_0}/u$). Only leading transit electrons are trapped over the time interval during which the phase velocity increases; consequently, the distribution of trapped electrons is completely determined by the distribution of transit electrons in the region $v > u_0$. If $f_0(v^2/v_T^2)$ decreases monotonically, the peak value in Eq. (2.3) is attained for $H = H_0$; i.e., electrons are condensed in the vicinity of this energy level. In the case of the Maxwell function, this can easily be explained if we take into account the fact that number of electrons trapped in the potential wells in the course of evolution upon an increase in the phase velocity of the wave is much larger for low values of the phase velocity than for its larger values. Obviously, the largest number of electrons are trapped at the beginning of the wave evolution, especially in the case when the phase velocity of the wave is close to the thermal velocity of electrons.

Using Eq. (2.3), we can write the electron current in the region $z > 0$ in the form

$$j(\varphi) = \begin{cases} j_1(\varphi), & \varphi > H_0, \\ j_2(\varphi), & \varphi < H_0, \end{cases} \tag{2.4}$$

$$j_1(\varphi) = 2\int_{\varphi}^{\varphi_{\max}} \frac{N_s u_s f_0(\gamma H^2)}{2(H-\varphi)} dH + j_U,$$

$$j_2(\varphi) = 2N_0 \int_{\varphi}^{H_0} \frac{f_0(v_T^{-2})dH}{\sqrt{2(H-\varphi)}} + 2\int_{H_0}^{\varphi_{\max}} \frac{N_s u_s f_0(\gamma H^2)}{\sqrt{2(H-\varphi)}} dH + j_U, \tag{2.5}$$

where

$$j_U = uN\left(P(\varphi - \langle\varphi\rangle) + 2\int_{\varphi_{\max}}^{\sqrt{\varphi_{\max}}} \frac{f_0(u^2/v_T^2)dH}{\sqrt{2(H-\varphi)}}\right) - \langle j_T\rangle - \langle j_{T0}\rangle - \langle j_{Ur}\rangle, \quad P = \int_{-\infty}^{\infty} \frac{dv}{v-u}\frac{df_0}{dv}.$$

The first term in formula (2.5) is the current of electrons trapped during initiation of the wave and accelerated by the wave to a velocity equal to u; the second term in Eq. (2.5) and the first term in Eq. (2.4) describe the current of electrons trapped in region $z > 0$, while the second term in the expression for j_U is the current of resonance transit electrons. Their mean currents are given by

$$\langle j_{T0}\rangle = \frac{2J(H_0)N_0 f_0(v_T^{-2})}{u},$$

$$\langle j_T\rangle = 2\int_{G_0}^{G} N_s f_0\left(\frac{G}{T}\right)\frac{dR}{dG}dG,$$

$$\langle j_{Ur}\rangle = -2RN f_0(v_T^{-2}),$$

respectively, where

$$J(H_0) = u\langle\sqrt{2(H_0-\varphi)}\rangle.$$

In formulas (2.4) and (2.5), the integration domain $H > \varphi_{\max}$ is split by point $\sqrt{\varphi_{\max}}$ [1]. The mean current $\langle j_{T0} + j_T + j_{Ur}\rangle$ coincides with the recoil momentum of the plasma. It was noted earlier that resonance currents of transit and trapped electrons compensate each other

in the course of wave initiation (i.e., their mean current is zero). Therefore, the mean electron current density at any instant of evolution is preserved and is equal to zero, which can be easily verified directly. In the course of prolonged evolution of the wave ($f_0(u^2/v_T^2) \ll f_0(v_T^{-2})$), the current of resonance transit electrons becomes negligibly small and is hence disregarded in the further analysis. The condition $\varphi \le H_0$ is equivalent to the constraint on the phase $|\psi| \le \theta_0$, where θ_0 is the smallest root of the equation $H_0 = \varphi(\theta_0)$. Consequently, phase oscillations of electrons trapped in the region $z < 0$, as well as electrons trapped in the vicinity of point $z = 0$, are bounded by the interval $|\psi| \le \theta_0$. Their contribution to the total current in the form of both integrals (2.5) for a prolonged evolution of the wave becomes larger than the contribution from electrons whose phase oscillations are performed outside this interval.

3. EVOLUTION OF A WAVE WITH DOUBLE DISPERSION

The evolution of a wave in a weakly inhomogeneous plasma is described by the equation [1]

$$\frac{\partial^2\varphi}{\partial\psi^2} + uj(\varphi) = 0, \tag{3.1}$$

which is a consequence of the Vlasov–Maxwell equations and has the first integral

$$W = \left(\frac{\partial\varphi}{\partial\psi}\right)^2 + U(\varphi), \tag{3.2}$$

where

$$U(\varphi) = 2\int_0^\varphi uj(\varphi)d\varphi$$

is the effective potential. The dispersion equation for the wave after integration of Eq. (3.2) can be written as the condition of periodicity of potential φ in phase ψ,

$$\pi = \int_{\varphi_{\min}}^{\varphi_{\max}} d\varphi/\sqrt{W - U(\varphi)}, \tag{3.3}$$

where $\varphi_{\min}$ and $\varphi_{\max}$ are the roots of the radicand in the integrand. Let us integrate Eqs. (2.4) and (2.5) first with respect to φ and then with respect to H. Since, in the case of the Maxwell distribution, the function

$$f_0(\gamma H^2) = \frac{\exp(-\gamma H^2)}{\sqrt{2\pi T}}$$

decreases exponentially rapidly with increasing H, the main contribution in the second integrand in Eq. (2.5) comes from electrons with energy $H = H_0$ (for this reason, we take $\sqrt{H-\varphi}$ out of the integral during integration, setting $H = H_0$) and $N_s u_s \approx N_0$; the remaining integrals can be evaluated exactly [18],

$$U(\varphi) = 2u\int j(\varphi)d\varphi = \begin{cases} U_1(\varphi), & \varphi > H_0, \\ U_2(\varphi), & \varphi < H_0, \end{cases}$$

where

$$U_1(\varphi) = u^2NP\left\{(\varphi-\varphi_0)^2 - b_0\exp\left(-\frac{\gamma}{2}\varphi^2\right)D_{-3/2}(\sqrt{2\gamma}\varphi)\right\}, \tag{3.4}$$

$$U_2(\varphi) = u^2NP\{(\varphi-\varphi_0)^2 - b_1\sqrt{H_0-\varphi} - b_2(H_0-\varphi)^{3/2} - c_U\}. \tag{3.5}$$

Here,

$$b_0 = \frac{2N_0}{uNP\sqrt{T}(2\gamma)^{3/4}},$$

$$b_1 = \frac{2N_0}{uNP\sqrt{\gamma T}}(\Phi(\sqrt{\gamma}\varphi_{\max}) - \Phi(\sqrt{\gamma}H_0)),$$

$$b_2 = \frac{8\sqrt{2}N_0}{3uNP}f_0(v_T^{-2}),$$

$$\varphi_0 = \langle\varphi\rangle + \frac{\langle j_T + j_{T0} + j_{Ur}\rangle}{uNP},$$

$$c_U = b_0\exp\left(-\frac{\gamma}{2}H_0^2\right)D_{-3/2}(\sqrt{2\gamma}H_0),$$

$D_{-\nu}(x)$ is a function of a parabolic cylinder, $\Phi(x)$ is the error function, and c_U is a constant ensuring the continuity of $U(\varphi)$. The effective potential $U(\varphi)$ can be calculated exactly by expanding the integrand defined by relation (2.5). The result of this calculation is cumbersome and will not be given here; it is shown graphically in Fig. 1 for the following values of the plasma and wave parameters: $T = 10^6$ K, $v_T = 0.55 \times 10^7$ m/s, $v_T/u_0 = 0.7$, $u/u_0 = 2.5$ (*1*) and 3.5 (*2*).

At the first stage of evolution, the relief of $U(\varphi)$ is mainly determined by two points, viz., potential φ_0 at which the minimum of the effective potential is attained and a potential equal to H_0 (henceforth, $\varphi_{10} = \varphi_0$). Potential $\varphi = H_0$ defines the region $\varphi > H_0$, in which $U(\varphi)$ is very close to the effective potential in the linear case, from region $\varphi < H_0$, where qualitative changes caused by condensation of electrons at the bottom of the potential wave accumulate. With increasing phase velocity of the wave, the value of $H_0 = 2\sqrt{AA_0}/u$

decreases, while the value of φ_0 increases together with $\langle j_T \rangle \approx uNP\varphi_0$. Assuming that $H_0 = \varphi_0$, we can determine the value of the phase velocity at which the first stage terminates, $u_1 = 2\sqrt{AA_0}/\varphi_0 \approx 2$. The dispersion of the wave at the first stage is described in [1], where the nonlinear correction to the linear dispersion equation is determined.

Let us find out how the condensation of electrons trapped in potential wells of the wave affects the process of its evolution. If the wave evolution is long, $u > 2$, the value of H_0 decreases to $H_0 < \varphi_0$ and the negative contribution introduced by the terms proportional to b_0 to $U_1(\varphi)$, as well as the terms proportional to b_1 and b_2 to $U_2(\varphi)$, is shifted to the region $\varphi < \varphi_0$. For this reason, in accordance with relation (3.4), the effective potential $U_1(\varphi)$ first increases upon a decrease in φ (see Fig. 1) and then decreases after attaining its maximal value for the potential

$$\varphi_b = \frac{2A}{u}\left\{1 + \frac{v_T^2\sqrt{H_0}}{2uNP(\varphi_0 - H_0)}\exp\left(-\frac{1}{v_T^2}\right)\right\} \quad (3.6)$$

in the region $\varphi < \varphi_b$. Being a continuation of $U_1(\varphi)$, the effective potential $U_2(\varphi)$ also decreases with φ due to the negative contribution of the terms proportional to b_1 and b_2, but then increases after attaining its minimal value at

$$\varphi_{20} = H_0 - \frac{\sqrt{2}A_0 v_T}{\sqrt{\pi}\varphi_0 \exp(1/v_T^2)}.$$

Thus, after a long evolution of the wave, the effective potential $U(\varphi)$ acquires the second potential well on the left from the main minimum φ_0 due to the local increase in the current of trapped electrons within the phase interval $-\theta_0 < \psi < \theta_0$, $\theta_0 = \varphi^{-1}(H_0)$ (see Fig. 1). After the emergence of this well, the second stage of the wave evolution begins—the variation of potential φ is bounded not by the left branch of the parabola $U(\varphi) \sim (\varphi - \varphi_0)^2$, as at the first stage, but by the left wall of the new well (see Fig. 1). As a result, the lower part of the wave $\varphi < \varphi_b$ is lowered still further relative to its upper part in the interval $-\theta_2 < \psi < \theta_2$, $\varphi(\theta_2) = \varphi_b$ (Fig. 2). Condensation of electrons at the bottom of the potential wells of the wave leads to a peculiar effect, viz., sagging of its potential in the region of the minima of the potential wells. Since the shape of the sag is determined by the profile of the second well of the effective potential, the sag can be treated as a fragment of a new wave. After the formation of the second well in the effective potential $U(\varphi)$, variation of potential φ takes place alternately in both potential wells. Consequently, the Langmuir wave at the second state of evolution can be represented as a sequence of alternating fragments of two waves with different wavelengths, which are continuously transformed into each other. As the phase

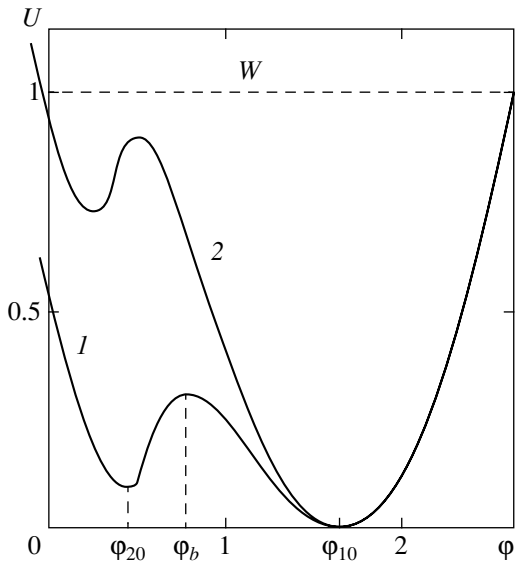


Fig. 1. Dependences of the effective potential on the potential of the wave; $u/u_0 = 2.5$ (*1*), 3.5 (*2*).

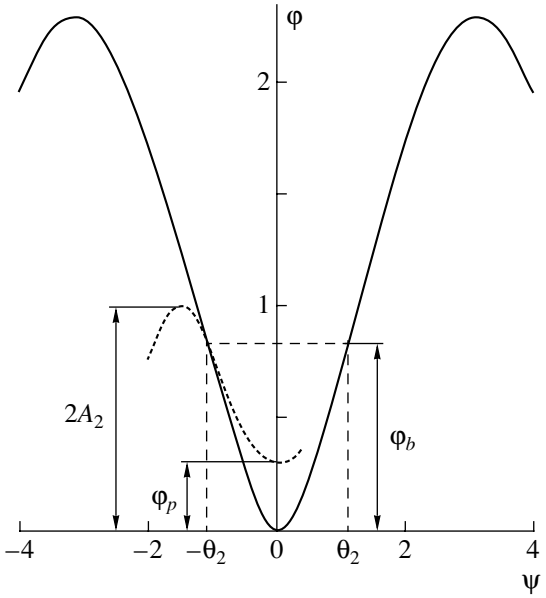


Fig. 2. Sagging of the wave potential.

velocity of the wave increases, potential (3.6) decreases, leading to an increase in the height of the barrier separating the wells and shift to the left (see Fig. 1). At a certain instant, the height of the effective potential $U(\varphi)$ becomes close to the value of the first integral $U(\varphi_b) \approx W$, remaining smaller than this integral. At this instant, the formation of wave profiles is completed; i.e., we can speak of the amplitudes of the first and second waves, which are $A \approx \varphi_{10} - \varphi_b$ and $A_2 \approx \varphi_b - \varphi_{20}$, where $\varphi_{10} \approx \varphi_0$ and $\varphi_{20} \approx H_0$ are the equilibrium values for the first and second waves, respectively.

We will use Eq. (3.1) for analytically describing the evolution of the wave taking into account the sagging of its potential. We evaluate the integrals in Eqs. (2.4) and (2.5) and expand the resulting expressions into series in $\tilde{\varphi}_1 = \varphi - \varphi_{10}$, $\tilde{\varphi}_2 = \varphi - \varphi_{20}$ to the leading order in $\tilde{\varphi}_1$ and $\tilde{\varphi}_2$. After substituting expansions into Eq. (3.1), in the reference frame associated with the wave, we obtain

$$\frac{\partial^2\tilde{\varphi}_1}{\partial z'^2} = -NP(1-g_1)\tilde{\varphi}_1, \quad \varphi_b < \varphi < \varphi_{\max},$$
$$\frac{\partial^2\tilde{\varphi}_2}{\partial z'^2} = -NP(1+g_2)\tilde{\varphi}_2, \quad \varphi < \varphi_b, \tag{3.7}$$

where $z' = z - ut$, $g_1 = g_1(\gamma, \varphi_0)$,

$$g_1(\gamma, \varphi) = \gamma b_0 \exp\left(-\frac{\gamma}{2}\varphi^2\right) D_{1/2}(\sqrt{2\gamma}\varphi),$$

$$g_2(\gamma) = \frac{\sqrt{2}\, f_0(\gamma H_0^2)}{uNP\sqrt{H_0 - \varphi_{20}}}\left\{\frac{Av_T^2}{2u(H_0 - \varphi_{20})} - 1\right\}.$$

It should be noted that system of equations (3.7) at the second state of the evolution is not self-consistent as yet. The difficulty encountered in solving any self-consistent problem is that the electron distribution is determined by the electric field potential; however, to find this potential, we must know this distribution. The distribution at the second stage of the evolution is not known. Nevertheless, the solution of the self-consistent problem is facilitated if we take into account the fact that the distribution of trapped electrons is formed in the course of capture of leading transit electrons, whose distribution is known. The distribution of trapped electrons is also determined by the potential well profile; however, this dependence is weaker. Indeed, the shape of the wave should be taken into account only in calculating adiabatic invariants such as (2.1). However, the value of an adiabatic invariant is primarily determined by the amplitude and phase velocity of the wave and depends on its shape to a smaller extent. Taking into account these remarks, we will solve the self-consistent problem of evolution of the wave at the second stage taking the distribution (2.3) of electrons trapped at the first stage as the initial distribution and assuming that A and u are the amplitude and phase velocity of the wave at the second stage.

An approximate solution of system (3.7) in both cases has the form

$$\varphi(\psi) = \begin{cases} A_1(1 + \cos(\psi_1 - \theta_1 - \theta_2)) + \varphi_p, & 2m(\theta_1 + \theta_2) + \theta_2 < \psi_1 < \theta_1 + (2m+1)(\theta_1 + \theta_2), \\ A_2(1 - \cos\psi_2), & 2m(\theta_1 + \theta_2) - \theta_2 < \psi_2 < \theta_2 + 2m(\theta_1 + \theta_2), \quad m = 0, 1, 2, \ldots, \end{cases} \tag{3.8}$$

where A_1 and A_2 are the amplitudes of wave fragments,

$$\psi_1 = \int k_1 dz - t, \quad \psi_2 = \int k_2 dz - t,$$

$$\theta_1 = 2\arcsin\sqrt{\frac{\varphi_{\max} - \varphi_b}{2A_1}}, \quad \theta_2 = 2\arcsin\sqrt{\frac{\varphi_b}{2A_2}},$$

and $\varphi_p = \varphi_{10} - A_1$ is the potential sag. The curve in Fig. 2 shows that, at the second stage of evolution, the value of φ_b decreases from $\varphi_0 \approx A_1$ to $2A_2$, while the value of $\varphi_{\max}$ increases from $2A_1$ to $2(A_1 + A_2)$ (we assume that $\varphi_{\min} = 0$). Consequently, if we take into account the fact that $A_2 \approx A_1$ at the beginning of the second stage, the values of θ_1 and θ_2 increase during this stage from $\pi/2$ to the maximal value equal to π. The period of potential (3.8) is $2(\theta_1 + \theta_2)$.

For the initial equations describing dispersion of wave fragments, we choose

$$k_1^2 = NP(1-g_1), \quad k_2^2 = NP(1+g_2). \tag{3.9}$$

Obviously, $\lambda_1 > \lambda_2$; $\lambda_1 = 2\pi/k_1$ and $\lambda_2 = 2\pi/k_2$ are the wavelengths of the first and second fragments. The corrections to wavelengths, which are determined by Eqs. (3.9), are quadratic in A and can be disregarded when $A \ll 1$.

Expressions (3.9) show that the dispersions of two mutually supplementing fragments are different; when the potential sag appears, the wavelength of the sequence of fragments forming the lower part of the wave (where electron condensation takes place) becomes smaller than the wavelength of the other sequence of fragments, which form the upper part of the wave.

Thus, an important feature of the steady-state nonlinear solution of the system of Vlasov–Maxwell equations for a Langmuir wave evolving in a plasma with a positive concentration gradient distinguishing it from other linearized solutions of this system of equations (e.g., the Van Kampen mode) is that solutions can exist in the form of double-dispersion waves, i.e., the waves formed by the sequence of alternating fragments of two waves, each of which has its own dispersion.

The wavelength is equal to the sum of the lengths of its fragments ($l_1 = 2\theta_1/k_1$ and $l_2 = 2\theta_2/k_2$). Taking into account formulas (3.8) and (3.9), we can write

$$\lambda = l_1 + l_2 = \frac{4}{\sqrt{NP}} \times\left(\frac{\arcsin\sqrt{(\varphi_{\max}-\varphi_b)/2A_1}}{\sqrt{1-g_1}} + \frac{\arcsin\sqrt{\varphi_b/2A_2}}{\sqrt{1+g_2}}\right). \quad (3.10)$$

At the beginning of the second stage, $\varphi_b \approx \varphi_0 \approx A_1 \approx A_2$, $\varphi_{\max} - \varphi_b \approx A_1$, $g_1 \approx 0$, and $g_2 \approx 0$; consequently, the dispersion is close to linear ($u^2NP \approx 1$). At the end of the second stage, $\varphi_b = 2A_2$ and $\varphi_{\max} = 2(A_1 + A_2)$; if we disregard g_1 and g_2, relation (3.10) leads to the dispersion equation $u\sqrt{NP} = 2$. Doubling of the right-hand side as compared to the case of a linear dispersion equation is due to the fact that the wavelength of the hybrid of two waves at the end of the second stage is twice as large as the wavelength of each fragment (In the linear approximation, the wavelengths λ_1 and λ_2 of the fragments are identical).

To make solution (3.8), (3.10) self-consistent, we must calculate the adiabatic invariant of electrons in the field of potential (3.8) and use it to find the electron distribution function at the second stage of the evolution. As a result of these calculations, we must replace γ in the expression for g_2 appearing in Eqs. (3.7) and (3.9) by the coefficient

$$\gamma_1 = \frac{u^2}{16v_T^2A_0A_2}.$$

The reason for the transformation of the Langmuir wave into the hybrid of two waves can be explained considering the plasma as a system with many vibrational degrees of freedom. At the first (quasi-linear) stage of evolution, the wave has only one vibrational degree of freedom; however, when the wave forms a flow of trapped electrons, the plasma system acquires one more degree of freedom determined by the flow parameters. The proposed decay scenario is possible for a wave loaded with trapped electrons with a distribution in which the majority of trapped electrons are condensed at the bottom of the potential wells of the wave. In this respect, the model of wave decay studied here differs from the models described in [15, 16, 19].

4. LANGMUIR WAVE PRIOR TO DECAY

The behavior of the wave and the variation of its profile prior to the decay are most interesting. At the end of the second stage, when the barrier height $U_1(\varphi_b)$ is close to W, we cannot state that anharmonism of wave fragments is small and that solution (3.8) can be used. If we expand the exact expressions for the effective potentials $U_1(\varphi)$ and $U_2(\varphi)$ into series in $\tilde{\varphi}_1 = \varphi - \varphi_{10}$ and $\tilde{\varphi}_2 = \varphi - \varphi_{20}$ in the vicinity of their minima φ_{10} and φ_{20}, retain the terms proportional to $\tilde{\varphi}_1^4$ and $\tilde{\varphi}_2^4$, and substitute the resulting expressions into Eq. (3.2), after integration we will obtain instead of (3.8) a solution containing elliptic cosines $\mathrm{cn}(\psi_1 - \theta_1 - \theta_2, \tau_1)$ and $\mathrm{cn}(\psi_2, \tau_2)$ instead of ordinary cosines, where

$$\tau_1 \approx A_1^{1/4}\left(\frac{u}{v_T}\right)\exp\left(-\frac{u^2}{4v_T^2}\right),$$

$$\tau_2 \approx \frac{A_2u^2}{A_1^{3/2}\sqrt{v_T}}\exp\left(-\frac{A_1}{2A_2v_T^2}\right).$$

It follows hence that anharmonism of the wave fragments with amplitudes A_1 and A_2 is noticeable only close to the thermal velocity (when $u \approx v_T$, $v_T \approx 1$). For the first wave, this can be explained by the fact that the nonlinear correction to term $(\varphi_1 - \varphi_0)^2$ in the expression for $U_1(\varphi)$ is exponentially small outside the region $\varphi \approx \varphi_b$. Consequently, formula (3.8) can be used for describing the shape of the pulses with positive and negative polarities in the region of equilibrium values of φ_{10} and φ_{20} right up to the wave decay. In the region $\varphi \approx \varphi_b$, in which potential $U_1(\varphi)$ attains its maximal value, the wave profile substantially differs from a sinusoid. To find out how the shape of the wave potential changes in the vicinity of $\varphi \approx \varphi_b$, let us consider curve *2* in Fig. 1. In a field whose potential has the same shape as curve *2*, a particle executes a finite motion, being reflected from the walls of a potential well and rapidly passes through the region containing these wells. Above the top of the barrier, the motion of the particles becomes slower. The change in potential φ corresponding to the motion of the particle in the case when $W - U(\varphi_b) \ll W$ is a sequence of alternating pulses of positive and negative polarity (Fig. 3). To determine the wave potential in the vicinity of $\varphi \approx \varphi_b$, we confine our analysis to the search for effective potentials $U_1(\varphi)$ and $U_2(\varphi)$ in this neighborhood. First, we obtain the distribution of electrons with an energy of $H \approx \varphi_b$ in the field of potential (3.8). Prior to the decay, $\varphi_b \approx 2A_2$; consequently, trapped electrons with an energy $\varphi_b < H < \varphi_{\max}$ will be transit electrons relative to wave fragments with amplitude A_2, while electrons with an energy $H_0 < H < \varphi_b$ will be trapped. The adiabatic invariant for the former electrons is

$$J_1 \approx \frac{4}{\pi}u\sqrt{A_2}\kappa_2E(\kappa_2^{-1}),$$

where $\kappa_2^2 = H/2A_2$, while for the latter electrons,

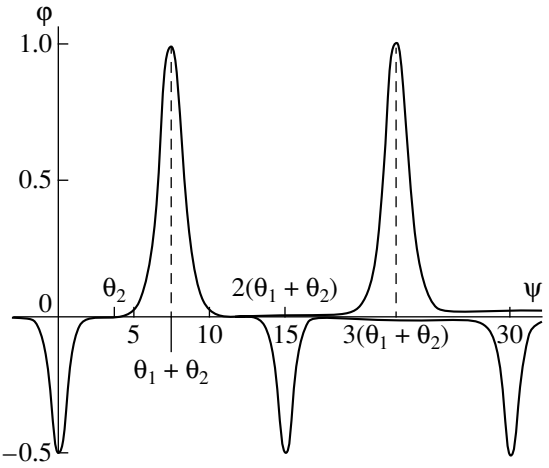


Fig. 3. Decay of a sequence of solitons with different polarities into two sequences of positive and negative solitons.

we have

$$J_2 = \frac{4}{\pi} u\sqrt{A_2}\kappa_2^2 B(\kappa_2).$$

Electrons with an energy $H \approx \varphi_b$ are trapped at the beginning of the second stage after the emergence of the second potential well. Consequently, having determined the value of function (2.1) at this instant and equating the result to the values of adiabatic invariants J_1 and J_2 in the approximation $E(\kappa_2^{-1}) \approx 1$, $B(\kappa_2) \approx 1$, we obtain the distribution of trapped electrons with energy $H \approx \varphi_b$ prior to decay:

$$f_{Tb} = \begin{cases} f_0(\beta_1 H), & \varphi_b < H < \varphi_{\max}, \\ \beta_1 = \dfrac{u^2}{2v_T^2 A_0}, & \\ f_0(\beta_2 H^2), & H_0 < H < \varphi_b, \\ \beta_2 = \dfrac{u^2}{4v_T^2 A_0 A_2}. & \end{cases} \quad (4.1)$$

Let us calculate the effective potential

$$U_1(\varphi) = u^2 NP(\varphi - \varphi_0)^2 - 4u\int_{\varphi}^{\varphi_{\max}} N_s u_s f_{Tb}\sqrt{2(H-\varphi)}\,dH, \quad \varphi > H_0, \quad (4.2)$$

and expand it into power series in $\tilde{\varphi} = \varphi - \varphi_b$ in the vicinity of $\varphi = \varphi_b$. Retaining the terms up to those proportional to $\tilde{\varphi}^3$, we obtain

$$U_1(\tilde{\varphi}) = U_1(\varphi_b) - u^2 NP\left\{(q_1 - 1)\tilde{\varphi}^2 - \frac{q_2}{3}\tilde{\varphi}^3\right\}, \quad \varphi \approx \varphi_b, \quad (4.3)$$

where

$$q_1 = q_1(\beta_1) = N_0 \frac{\exp(-\beta_1\varphi_b)}{NPv_T^2\sqrt{A_0}}$$

for $\varphi_b < H < \varphi_{\max}$,

$$q_1 = q_1(\beta_2) = g_2(\beta_2, \varphi_b)$$

for $H_0 < H < \varphi_b$. Under the condition

$$\frac{u^2}{v_T^2} > \sqrt{A_0}\exp\frac{1}{v_T^2},$$

the coefficient of $\tilde{\varphi}^2$ in formula (4.3) is negative, which is required for the emergence of a soliton solution. Having determined approximately the roots of the equation $W - U_1(\tilde{\varphi}) = 0$, we factorize the expression $W - U_1(\tilde{\varphi})$ in the region of $\varphi > \varphi_b$,

$$W - U_1(\tilde{\varphi}) = u^2 NP\frac{q_2}{3}\{(\tilde{\varphi}^2 + c^2)(a - \tilde{\varphi})\}, \quad \tilde{\varphi} = \varphi - \varphi_b, \quad (4.4)$$

where

$$c^2 = \frac{\delta w}{q_1 - 1}, \quad q = 3\frac{q_1 - 1}{q_2}, \quad \delta w = \frac{W - U_1(\varphi_b)}{u^2 NP}.$$

Integrating Eq. (3.2) in the vicinity of $\varphi \approx \varphi_b$ with the radicand determined above with the help of the substitution

$$\tilde{\varphi} = a - \sqrt{a^2 + c^2}\tan^2\frac{\xi}{2}, \quad (4.5)$$

we obtain

$$u\sqrt{NP\frac{q_2}{3}}\psi = \frac{F(\xi_0, \tau_s) - F(\xi, \tau_s)}{(a^2 + c^2)^{1/4}}, \quad (4.6)$$

where

$$\tau_s^2 = \frac{a + \sqrt{a^2 + c^2}}{2\sqrt{a^2 + c^2}},$$

while the value of ξ_0 can be determined from the equation

$$\tan^2\frac{\xi_0}{2} = \frac{a}{\sqrt{a^2 + c^2}}.$$

As $U(\varphi_b) \longrightarrow W$, we set $F(\xi_0, \tau_s) \approx K(\tau_s)$ in formula (4.6). After inversion of the incomplete elliptical integral in Eq. (4.6) and substitution of the result into Eq. (4.5), we obtain

$$\tilde{\varphi} = \pm\left\{a - \sqrt{a^2 + c^2}\left(\frac{1 - \mathrm{cn}(k_s z', \tau_s)}{1 + \mathrm{cn}(k_s z', \tau_s)}\right)\right\}, \qquad (4.7)$$

where

$$k_s = \sqrt{\frac{NPq_2\sqrt{a^2 + c^2}}{3}}.$$

In this expression, $z' = z - ut$ is counted from the point at which $\tilde{\varphi}$ attains its maximal value. Since expression (3.8) is a good approximation for the potential in the vicinity of φ_{10} and φ_{20}, and expression (4.7) successfully describes the potential at the foot of solitons, these solutions should be joined. Equating relations (3.8) and (4.7), we obtain the points at which the half-wave of amplitude A_1 is joined with the foot of the positive soliton $\varphi \approx \varphi_b$: $\pm l_+ \approx \pm\pi/k_1$, as well as the points at which the half-wave of amplitude A_2 is joined with the foot of the negative soliton: $\pm l_- \approx \pm\pi/k_2$,

$$k_{1,2} = \sqrt{NP(1 \mp g_{1,2})}.$$

At the end of the second stage, the wave potential in the reference frame attached to the wave in the interval $-\lambda_s/2 < z' < \lambda_s/2$ assumes the form

$$\tilde{\varphi} = \begin{cases} A_1\left\{1 + \cos\left(k_1\left(z' \pm \frac{\lambda_s}{2}\right)\right)\right\}, & -\lambda_s/2 < z' < -\lambda_s/2 + l_+, \\ & \lambda_s/2 - l_+ < z' < \lambda_s/2, \\ \pm\sqrt{\frac{\delta w}{q_1 - 1}}\,\mathrm{sn}\left(k_s\left(z' \pm \frac{\lambda_s}{4}\right), \tau_s\right)/\mathrm{dn}\left(k_s\left(z' \pm \frac{\lambda_s}{4}\right), \tau_s\right), & -\lambda_s/2 + l_+ < z' < -l_-, \\ & l_- < z' < \lambda_s/2 - l_+, \\ -A_2(1 + \cos(k_2 z')), & -l_- < z' < l_-, \end{cases} \qquad (4.8)$$

where λ_s is the soliton period; the minus sign in front of $\sqrt{\delta w/(q_1 - 1)}$ corresponds to $-\lambda_s/2 + l_+ < z' < -l_-$, while the plus sign corresponds to $l_- < z' < \lambda_s/2 - l_+$. The origin of the coordinate system coincides with one of the minima of the potential.

Solution (4.8) is a periodic function with a period (wavelength) $\lambda_s = 4K(\tau_s)/k_s$. In the limit $U(\varphi_b) \longrightarrow W$, the wavelength increases in accordance with the logarithmic law [20]

$$\lambda_s \approx \frac{4\sqrt{3}}{\sqrt{aNPq_2}}\ln\frac{4}{\sqrt{1 - \tau_s^2}}, \quad \tau_s \longrightarrow 1. \qquad (4.9)$$

5. DECAY OF A LANGMUIR WAVE

The evolution of plasma oscillations $U_1(\varphi_b) \approx W$ at the instant of wave decay should be considered separately since the distance between the positive and negative pulses increases indefinitely. This obviously does not affect the distribution function of trapped electrons within the negative pulse. The lower part of potential $\varphi \le \varphi_b$ is determined by formula (4.8) as before. In the limit $\delta w \longrightarrow 0$, $c \longrightarrow 0$, potential (4.7) becomes a soliton [20, 21]; it was noted above, however, that the self-consistent solution can be written only for negative solitons:

$$\varphi = \begin{cases} \varphi_b - A_2\{1 + \cos(k_2(z - ut))\}, \\ |z - ut| < l_-, \\ \varphi_b - \dfrac{a}{\cosh^2(k_s(z - ut)/2)}, \\ l_- < |z - ut| < \dfrac{\lambda_s}{2}. \end{cases} \qquad (5.1)$$

Here,

$$a = \frac{3(q_1(\beta_2) - 1)}{q_2},$$

$\pm l_- \approx \pm\pi/k_2$ are the points at which the potential corresponding to the foot of the soliton is joined with a negative pulse.

To construct a self-consistent solution for a sequence of positive solitons, we first find the distribution of electrons trapped by the field of the sequence of these solitons. We assume that the potential of positive solitons differs from zero only in a very small interval $\Delta z' \ll \lambda_s$, where λ_s is the spacing between solitons. In this case, the adiabatic invariant describing the motion

of electrons trapped by the field of a periodic sequence of solitons is given by

$$J = \frac{uk_s}{2\pi}\int_0^{\lambda_s}\sqrt{2(H-\varphi)}dz' \approx \frac{2}{\pi}K(\tau_s)u\sqrt{2H_b},$$

$$H_b = H - \varphi_b.$$

This adiabatic invariant has sense if the motion of electrons is periodic, which is possible for a finite distance λ_s between solitons. It will be shown below that, although the distance between the solitons at the instant of decay of the wave $U(\varphi_b) = W$ increases unlimitedly upon a decrease in the growth rate of the barrier height, it remains finite if this rate does not vanish. The distribution of electrons trapped by a sequence of positive solitons, which is calculated using this adiabatic invariant, has the form

$$f_{Tb} = f_0\left(\frac{u_s^2}{v_T^2}\right) = f_0(\beta_b H_b), \quad \beta_b = \frac{u^2K^2(\tau_s)}{2v_T^2A_0}.$$

Comparing this distribution with (4.1), we note that effective potential $U_1(\tilde{\varphi})$ remains in the previous form (4.3) except that β_1 must be substituted for β_b in coefficients q_1. Consequently, formula (4.7) can be used for describing potential $\varphi \approx \varphi_b$ at the foot of positive solitons. In the limit $\delta w \longrightarrow 0$, $c \longrightarrow 0$, this potential becomes a soliton; consequently, for $\varphi > \varphi_b$, we can write

$$\varphi = \begin{cases} \varphi_b + A\{1 + \cos(k_1(z-ut))\}, \\ |z-ut| < l_+, \\ \varphi_b + \dfrac{a}{\cosh^2(k_s(z-ut)/2)}, \\ l_+ < |z-ut| < \dfrac{\lambda_s}{2}, \end{cases} \qquad (5.2)$$

where $a = 3(q_1(\beta_b) - 1)/q_2$ and $l_+ \approx \pi/k_1$. It should be noted that the solution obtained in the form of alternating negative and positive solitons (5.1) and (5.2) is possible if the plasma becomes homogeneous for $U(\varphi_b) \longrightarrow W$. In this case, the wave does not experience a decay and the distance between solitons increases indefinitely.

The proposed scenario of the evolution of the Langmuir wave does not lead to the dispersion relation at the instant of wave decay. In the case when $U(\varphi_b) \longrightarrow W$, the unlimited increase in the distance between solitons can be easily explained using the analogy between mechanical motion of a particle and variation of the potential defined by Eq. (4.3). When the particle moves above the top of the barrier and its total energy is equal to the barrier height, the time of flight of the particle above the barrier increases indefinitely. Since the distance between solitons and their phase velocity increase indefinitely at the instant of wave decay, the question arises: does the decay occur at all? To answer this question, we must take into account the fact that the time of flight of a particle above the barrier (or, to be more precise, the time of particle transition from the trapped to the transit state and vice versa) becomes finite due to a change in the barrier height (even if it is very slow) [22, 23]. Analogously, if the barrier height $U(\varphi_b)$ increases (even very slowly), the distance between solitons remains finite.

Let us consider the case of a linear increase in the barrier height

$$W - U(\varphi_b) = \varepsilon_d(z_d - z), \quad \varepsilon_d = \left|\frac{d(W - U(\varphi_b))}{dz}\right|,$$

where z_d is the wave decay point. Let us analyze the evolution immediately before the wave decay over a distance on the order of several λ_s to the point z_d of decay. Having for convenience chosen the point at which a positive soliton attains its maximum value at $t = 0$ as the origin, we denote in relation (4.4)

$$c^2 \approx \varepsilon z = \varepsilon(z' + ut), \quad \varepsilon = \frac{\varepsilon_d}{u^2NP(q_1 - 1)}.$$

In this case, Eq. (3.1) for the right slope of a positive soliton in its own reference frame assumes the form

$$\frac{d\varphi_r}{dz'} = -\frac{ks}{\sqrt{a}}\sqrt{(a-\varphi_r)(\varphi_r^2 + \varepsilon(z'+ut))}, \quad 0 < z < z_d, \qquad (5.3)$$

where $\varphi_r = \varphi_r(k_s z') = \varphi(k_s z') - \varphi_b$. The minus sign in front of the radical is chosen because $d\varphi_r/dz' < 0$ on the right slope of the soliton. For $\varphi \approx \varphi_b$, we can write solution (5.3) in the highest order in ε under the condition

$$k_s z' \gg 1, \quad \cosh(k_s z') \approx \frac{\exp(k_s z')}{2}$$

in the form

$$\varphi \approx \varphi_b + 4ae^{-k_s z'} - \frac{\varepsilon z' \exp(k_s z')}{16a}, \quad 0 < z' < z_d. \qquad (5.4)$$

Setting $\varphi = \varphi_b$, in the approximation $\ln z' \ll k_s z'$, we obtain

$$\lambda_s \approx \frac{1}{k_s}\ln\frac{8a}{\sqrt{\varepsilon}}. \qquad (5.5)$$

Thus, prior to decay, the wave acquires the form of a sequence of alternating solitons with different polarities (see Fig. 3). The spacing between the solitons

increases in the course of wave evolution and attains the value λ_s (5.5) at the instant of decay (i.e., the spacing remains finite if the rate of increase in $U(\varphi_b)$ differs from zero). Consequently, the application of the adiabatic invariant is justified at the instant of decay also provided that $\lambda_s \ll L$, where L is the characteristic size of inhomogeneity. It should also be noted that

$$\lim_{\varepsilon \to 0} \varepsilon\lambda_s = 0;$$

therefore, the approximation $c^2 \approx \varepsilon z$ holds in spite of the very large value of λ_s. Since $\Delta W \sim \varepsilon$, the value of W has no time to change in the course of decay over distances comparable to the wavelength:

$$\lim_{\varepsilon \to 0} (\Delta W/W) = 0.$$

Let us find the potential of waves after the decay ($U(\varphi_b) > W$). This can be done within the adiabatic approximation if the phase velocities of the sequence of solitons before and after the decay are identical. The question arises: what will happen after the instant of $U(\varphi_b) = W$ upon a small change in the effective potential profile $U(\varphi)$ if this change only involves an increase in the barrier height $U(\varphi_b) > W$? The height of the effective potential barrier separating potential wells increases with the phase velocity of the wave; consequently, a gap appears between two sequences of positive and negative solitons (see Fig. 3). We can assume that the two waves are formed at the instant $U(\varphi_b) = W$. In the linear approximation, the currents of these waves are determined by the distribution of transit electrons, which does not change in the course of decay. This means that the change in the relief of effective potentials $U_1(\varphi_1)$ and $U_2(\varphi_2)$ at a distance from the top of the barrier is reflected only in corrections in A_1 and A_2 of an order higher than second. Thus, the shape of the shape of positive and negative pulses in the linear approximation in the amplitude coincides with the shape defined by expression (4.8) in the course of decay. At the feet of these pulses, we can put $\varphi \approx \varphi_b$; in this case, the distribution of trapped electrons in the given approximation at the instant of decay remains unchanged. As in the case (4.2), this enables us to write the effective potential immediately after the decay in the vicinity of $\varphi \approx \varphi_b$ for the first and second waves,

$$\begin{aligned} U_1(\tilde{\varphi}_1) &= u^2 NP\tilde{\varphi}_1^2 \\ &- 4u\int_{\tilde{\varphi}_1}^{2A_1} N_s u_s f_0(\beta_b H)\sqrt{2(H-\tilde{\varphi}_1)}\,dH, \\ U_2(\tilde{\varphi}_2) &= u^2 NP\tilde{\varphi}_2^2 \\ &+ 4u\int_{\tilde{\varphi}_2}^{\varphi_{b2}} N_s u_s f_0(\beta_2 H^2)\sqrt{2(H-\tilde{\varphi}_2)}\,dH, \end{aligned} \tag{5.6}$$

where $\tilde{\varphi}_1 = \varphi_1 - \varphi_{b1}$, $\tilde{\varphi}_2 = \varphi_2 - \varphi_{b2}$, and φ_{b1} and φ_{b2} are the minimal and maximal values of the potentials of the first and second waves. The expansion of effective potentials $U_1(\varphi_1)$ and $U_2(\varphi_2)$ into power series in $\tilde{\varphi}_1$ and $\tilde{\varphi}_2$ in the vicinity of $\varphi \approx \varphi_b$, in which we retain only the terms containing $\tilde{\varphi}_1$ and $\tilde{\varphi}_2$ to a power not higher than third, coincides with expression (4.3). However, the results of factorization of $W - U_1(\varphi_1)$ and $W - U_2(\varphi_2)$ differ from expression (4.4) in the sign of δw. Therefore, we have

$$\begin{aligned} &W - U_1(\tilde{\varphi}_{1,2}) \\ &= u^2 NP\frac{q_2}{3}\{(\tilde{\varphi}_{1,2}^2 - c^2)(a_d - \tilde{\varphi}_{1,2})\}, \end{aligned} \tag{5.7}$$

where

$$c^2 = \frac{\delta w}{q_1 - 1}, \quad \delta w = \frac{U_1(\varphi_b) - W}{u^2 NP}, \quad a_d \approx 3\frac{q_1 - 1}{q_2}.$$

We assume that the amplitudes $2A_1$ and $2A_2$ of positive and negative pulses do not change in the course of decay. Substituting expression (5.7) into (3.2) and integrating the result, we can write the potentials of the waves after the decay in the form

$$\varphi_{1,2}(z-ut) = \begin{cases} \pm A_{1,2}(1 + \cos(k_{1,2}(z-ut))) \pm c, \\ |z-ut| < l_{\pm}, \\ \pm \dfrac{a_d}{\cosh^2(k_d(z-ut)/2)} \pm c, \\ l_{\pm} < |z-ut| < \dfrac{\lambda_d}{2}. \end{cases}$$

Here,

$$k_d = \sqrt{\frac{(a_d + c)NPq_2}{3}}, \quad \tau_d = \sqrt{\frac{a_d - c}{a_d + c}};$$

the plus and minus signs correspond to positive and negative solitons, respectively; and λ_d is the soliton period. The solutions obtained above show that the gap

$$2c \sim \sqrt{U(\varphi_b) - W}$$

between the feet of negative and positive pulses increases with the barrier height. The phase velocities of the waves after the decay can be determined from the condition of conservation of the mean energy flux density in the course of the decay:

$$S_w(\varphi) + S_e = S_{w1}(\varphi_1) + S_{w2}(\varphi_2) + S_{e1} + S_{e2}.$$

Here [1],

$$S_w(\varphi) = \frac{3NT(\langle\varphi_+^2\rangle + \langle\varphi_-^2\rangle)}{u^3},$$

$$S_{w1,2}(\varphi_{1,2}) = \frac{3NT\langle\varphi_{1,2}^2\rangle}{u^3}$$

are the mean energy flux densities of the field before and after the decay ($\varphi_+ = \varphi$ for $\varphi > 0$ and $\varphi_- = \varphi$ for $\varphi < 0$);

$$S_e = \langle j(\varphi_+) + j(\varphi_-)\rangle u^2/2,$$

$$S_{e1,2} = \langle j_{1,2}(\varphi_{1,2})\rangle u_{1,2}^2/2$$

are the mean density of vibrational energy fluxes for trapped electrons before and after the decay; and

$$\langle j_1\rangle = \frac{8}{\pi}\left\{\int_{u_b}^{u_d} N_s f_0 \frac{d}{dv}(v\sqrt{A})dv\right\},$$

$$\langle j_2\rangle \approx \frac{8}{\pi}\left\{\int_1^{u_b} N_s f_0 \frac{d}{dv}(v\sqrt{A})dv - \frac{H_0}{2\sqrt{A_0}} N_0 f_0(v_T^{-2})\right\},$$

where $f_0 = f_0(v^2/2T)$ and u_b is the phase velocity of the wave, at which the electrons having energy $H = \varphi_b$ at the instant of decay were trapped. If we disregard the difference between a and a_d, we can assume that $\varphi_+ = \varphi_1$ and $\varphi_- = \varphi_2$ at the instant of decay; consequently, the energy balance equation implies that the phase velocities before and after the decay are close ($u \approx u_1 \approx u_2$). It should be noted that their values differ only by $\sqrt{2(A_1 + A_2)}$. The wavelength after the decay is given by

$$\lambda_d \approx \frac{4\sqrt{3}}{\sqrt{a_d NPq_2}} \ln\left(2\sqrt{\frac{2a_d}{c}}\right), \quad \tau_d \longrightarrow 1. \qquad (5.8)$$

Comparing formulas (4.9) and (5.8), we note that the wavelengths before and after the decay are identical if $a_d \approx a$. Comparison of the potentials of the waves before and after the decay shows that the shape of positive and negative pulses changes insignificantly (mainly at their feet). The shape of positive and negative solitons after they pass the decay point changes only in the way that the edges of positive solitons, which were formerly connected with the edges of negative solitons, are now connected to each other; the same is observed for negative solitons. Such a change in the soliton shape does not lead to any appreciable change in the distribution function of trapped electrons (the more so, of transit electrons).

6. CONCLUSIONS

The simplest example of decay is the generation of new harmonics in a nonlinear medium [16]. The harmonics are formed immediately after the decay and their increase begins from zero intensity. In the case of parametric instability, the existence of small perturbation waves against the background of a pumping wave is presumed [15]. Decay instability with two perturbation waves forms a particular case. The proposed model of the decay differs from the known models since it can be applied only to a wave loaded with trapped electrons. The main distinguishing feature of our model is that the wave experiences a qualitative rearrangement prior to the decay. In the bulk of the wave, fragments of a new wave are generated. Prior to the decay, a hybrid of two waves possessing its own dispersion is formed. After the complete formation of wave fragments, the difference between their equilibrium states becomes equal to the sum of the amplitudes $A_1 + A_2$. Beginning from this instant, the new wave can separate itself from the original wave. If, however, the waves remained periodic at the instant of decay, their potentials before and after the decay could not be joined. Indeed, the phases of the waves at the decay point will be different at different instants in view of the difference in the phase velocities of the waves before and after the decay. This problem can be solved if we take into account the fact that the positive and negative humps become steeper prior to the decay and that the distance between the humps increases indefinitely. In other words, to give rise to two new waves loaded with trapped electrons, the original wave is transformed prior to the decay into a sequence of solitons with different polarities, which is not periodic at the instant of the decay. For a single event of decay, an increase in the phase velocity by a factor of $2\sqrt{A_1/A_2}$ is sufficient, which can easily be realized. It follows hence that new waves loaded with trapped electrons will "gemmate" from the initial Langmuir wave after the decay described above. Thus, the evolution of a Langmuir wave in a weakly inhomogeneous plasma with a positive concentration gradient is accompanied by the emergence of weak turbulence with a scenario of evolution analogous to that proposed by Landau for turbulence of a hydrodynamic flow of a viscous liquid.

REFERENCES

1. V. L. Krasovskiĭ, Zh. Éksp. Teor. Fiz. **107**, 741 (1995) [JETP **80**, 420 (1995)].

2. V. I. Karpman and D. R. Shklyar, Zh. Éksp. Teor. Fiz. **67**, 103 (1975) [Sov. Phys. JETP **40**, 53 (1975)].

3. Ya. I. Istomin, V. I. Karpman, and D. R. Shklyar, Zh. Éksp. Teor. Fiz. **64**, 2072 (1973) [Sov. Phys. JETP **37**, 1045 (1973)].
4. E. Asseo, G. Laval, R. Pellat, and A. Roux, J. Plasma Phys. **8**, 341 (1972).
5. G. Laval and R. Pellat, J. Geophys. Res. **75**, 3255 (1970).
6. N. S. Erokhin, N. N. Zol'nikova, V. L. Krasovskiĭ, *et al.*, Zh. Éksp. Teor. Fiz. **100**, 832 (1991) [Sov. Phys. JETP **73**, 460 (1991)].
7. V. S. Beskin, A. V. Gurevich, and Ya. I. Istomin, Zh. Éksp. Teor. Fiz. **92**, 1277 (1987) [Sov. Phys. JETP **65**, 715 (1987)].
8. V. I. Karpman and D. R. Shklyar, Zh. Éksp. Teor. Fiz. **62**, 944 (1972) [Sov. Phys. JETP **35**, 500 (1972)].
9. V. P. Milant'ev, Zh. Éksp. Teor. Fiz. **85**, 132 (1983) [Sov. Phys. JETP **58**, 78 (1983)].
10. Yu. A. Gurevich, Zh. Éksp. Teor. Fiz. **53**, 925 (1967) [Sov. Phys. JETP **26**, 560 (1968)].
11. A. S. Bakaĭ and Yu. P. Stepanovskiĭ, *Adiabatic Invariants* (Naukova Dumka, Kiev, 1981) [in Russian].
12. V. Ya. Davydovskiĭ and A. I. Matveev, Fiz. Plazmy **11**, 1368 (1985) [Sov. J. Plasma Phys. **11**, 784 (1985)].
13. A. S. Bakaĭ, Vopr. At. Nauki Tekh., Ser.: KhFTI (1976).
14. V. Ya. Davydovskiĭ and A. I. Matveev, Fiz. Plazmy **13**, 443 (1987) [Sov. J. Plasma Phys. **13**, 251 (1987)].
15. L. A. Artsimovich and R. Z. Sagdeev, *Plasma Physics for Physicists* (Atomizdat, Moscow, 1979) [in Russian].
16. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 2nd ed. (Nauka, Moscow, 1982; Pergamon, Oxford, 1984).
17. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 1: *Mechanics*, 2nd ed. (Nauka, Moscow, 1963; Pergamon, Oxford, 1965).
18. A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series. Elementary Functions* (Nauka, Moscow, 1981; Gordon and Breach, New York, 1986).
19. A. I. Akhiezer, I. A. Akhiezer, R. V. Polovin, *et al.*, *Plasma Electrodynamics* (Nauka, Moscow, 1974; Pergamon, New York, 1975).
20. E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon, Oxford, 1981).
21. B. B. Kadomtsev, *Collective Phenomena in a Plasma* (Nauka, Moscow, 1988) [in Russian].
22. A. V. Timofeev, Zh. Éksp. Teor. Fiz. **75**, 1303 (1978) [Sov. Phys. JETP **48**, 656 (1978)].
23. V. Ya. Davydovskiĭ and A. I. Matveev, Zh. Tekh. Fiz. **53**, 2125 (1983) [Sov. Phys. Tech. Phys. **28**, 1302 (1983)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 101, No. 5, 2005, pp. 962–977.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 128, No. 5, 2005, pp. 1099–1116.
Original Russian Text Copyright © 2005 by Sitnikov, Nechaev, Taran.

STATISTICAL, NONLINEAR, AND SOFT MATTER PHYSICS

A Quantitative Mean-Field Theory of the Hydrophobic Effect of Neutral and Charged Molecules of Arbitrary Geometry

G. V. Sitnikov[a,b,*], S. K. Nechaev[c,d], and M. D. Taran[e]

[a]*Moscow Institute of Physics and Technology, Institutskaya ul. 9, Dolgoprudnyĭ, Moscow oblast, 141700 Russia*
[b]*Algodign LLC, Bol'shaya Sadovaya ul. 8(1), Moscow, 123379 Russia*
[c]*LPTMS, Université Paris Sud, Orsay Cedex, 91405 France*
[d]*Landau Institute for Theoretical Physics, Chernogolovka, Moscow oblast, 142438 Russia*
[e]*Russian State Scientific Center, Troitsk Institute for Innovation and Thermonuclear Research, Troisk, Moscow oblast, 142190 Russia*
**e-mail: goshik91mipt@list.ru; gsitnikov@algodign.com*

Received July 8, 2005

Abstract—A self-consistent two-length scale theory of the interaction between a hydrophobic molecule and a water environment is considered. This theory allows the width of the hydrophobic layer to be calculated for molecules of arbitrary geometry by explicitly taking into account the water structure through the correlation function of a pure liquid. This approach is used to calculate the density profile $\rho(r)$ around a molecule of arbitrary geometry and the solvation free energy $\Delta G(R)$ related to the transport of the molecule from a vacuum to a liquid. The model parameters are adjusted by comparing the results of numerical Monte Carlo simulations taken from the literature with predictions of the model for molecules of spherical geometry. The free energy of the interaction $\Delta G(D)$ between two spheres of radius R separated by distance D is also determined using the developed approach. The model is generalized to electrostatic interactions within the framework of a self-consistent scheme in which water is modeled by a gas of point dipoles. Analysis of the derived equations shows that this theory coincides with the electrostatic theory of a continuous medium with an effective permittivity in the limit of weak electric fields. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

This work should be considered as a preliminary step in constructing a general theory of the interaction between solute molecules and a solvent. The ultimate goal of the study is to calculate the binding constants between biologically active molecules in a water environment at a level of accuracy comparable to that of molecular-dynamical calculations. Specifically, we consider the narrower problem of quantitatively determining the solvation free energy of hydrophobic molecules of arbitrary geometry in water.

There are many approaches to solving this problem in the literature. In our view, they can be classified as follows.

Explicit approaches. These include explicit molecular-dynamical calculations and Monte Carlo simulations of the behavior of a solute molecule in an environment of solvent molecules. The shortcoming of these approaches is that they allow the dynamics of the system to be traced on time scales that are several orders of magnitude shorter than those required to reach thermodynamic equilibrium and to determine the corresponding equilibrium parameters. In addition, the computational time increases greatly as the number of molecules in the system increases.

Empirical approaches. The main idea behind these approaches is to choose an appropriate parameterization of the solvation energy. The relation between the surface area of a solute molecule accessible to a solvent and the solvation energy or, more precisely, the part of it associated with the formation of a cavity of the corresponding size in the solvent is most commonly assumed to be linear. Studies based on the idea of an effective renormalization of the interaction in a vacuum using solubility, evaporation, and other data to allow for the influence of water [1–3] also belong to this type of works.

Semiempirical approaches. These approaches are based on the solution of the exact statistical equations that describe the microscopic properties of the corresponding systems. Unfortunately, these equations can be solved only in certain special cases by introducing simplifying assumptions (see, e.g., [4]). Nevertheless, there is hope that the parameterization arising in these approaches is more consistent with the physical picture than the parameterization of "empirical" models. In addition, these approaches retain many advantages of "explicit" models.

The assumption that the solvation energy is proportional to the molecular surface area, which is commonly used in empirical approaches [5, 6], appears plausible in the case where the size of the solute mole-

cule is much larger than that of the water molecule. In contrast, for small molecules and molecules of complex geometry, the corresponding surface area is often determined using rough empirical assumptions, for example, by "rolling" a water molecule around a solute molecule. This kind of calculations can hardly be considered satisfactory from both theoretical and practical points of view. Modifications of this approach can be found in [7]. Another solution of this problem can be obtained in terms of the Gaussian approximation suggested, for example, in [8].

In our view, the following approaches are most consistent and constructive among the theoretical approaches devoted to the solute–solvent interactions.

1. Directly calculating the correlation function of a solvent in the presence of a solute (with or without electrostatic interactions). This scheme suggests two possible ways: (i) introducing special collective variables, weighted densities, and constructing an appropriate functional whose form is completely determined by physical constraints [4] or (ii) choosing an empirical closure condition for the correlation function that satisfies the hierarchy of integral equations [9]. Once the correlation functions have been calculated, the remaining thermodynamic parameters can be determined easily.

2. Calculating the mean density profile of a fluctuating field in the presence of a solute molecule (with or without electrostatics). The discrete structure of the liquid on small scales is taken into account via the correlation function of a pure solvent unperturbed by the solute. The solute molecule is taken into account by the requirement that the total solvent density be expelled from the volume occupied by the solute [10–13]. The solvation energy and the density profile near the solute molecule can then be calculated by minimizing the free energy.

3. The SPT [14, 15] and IT [16] theories. Both approaches use the relationship between the probability of spontaneous formation of a cavity of a given shape and volume containing a fixed number of particles and the solvation energy of the molecule that forms this cavity. The cavity formation probability can be calculated from combinatorial considerations.

Our work may be placed into the second group of the above list. We mainly use the ideas set out in [10–12], in which a two-length scale model of the interaction between hydrophobic molecules and a solvent was consistently developed and modified. In this approach, the solvent density is described by a fluctuating field with a separated component that changes greatly at distances on the order of the size of the water molecule; it may be interpreted as a fluctuation of the medium. The remaining part of the density is assumed to change slowly, with the definition of a mean observed solvent density being assigned to it. In our view, this approach is optimal for the following reasons: on the one hand, being "semimicroscopic," it is physically clear; on the other hand, being computationally much simpler and faster than the corresponding explicit approaches, it allows the detailed structure of the solvent to be effectively taken into account. Thus, our objective is to construct a quantitative theory of the solute–solvent interaction for molecules of arbitrary geometry. We take into account the correlations of the medium (water) by specifying a correlation function and adjust the remaining independent model parameters by comparing them with available published data of numerical or actual experiments for molecules of the simplest (spherical) geometry.

In our view, the approach being developed here has the following advantages over other phenomenological theories.

1. The solvation energy of a molecule of arbitrary geometry and the interaction energy of two or more solvated objects are calculated in terms of a unified approach. The derived equations require no special modification to allow for several solvated objects.

2. The discrete water structure on small scales is taken into account.

3. No empirical recipes like the rolling of the water molecule around the solvent surface are required to determine the geometrical parameters of the solute. The geometry of the solute molecule can be uniquely related to the features of the interaction potential between the latter and the solvent molecules. Although some freedom in choosing the surface remains, it can be removed by properly adjusting the free parameters.

4. The model contains a small number of free parameters that can in principle be determined from (numerical or real) experiments.

The paper is structured as follows. In Section 2, we describe the model and derive the basic equations for the equilibrium density and the solvation free energy of molecules of arbitrary geometry. Here, we also compare the numerical solution of the derived equations by Monte Carlo simulation in [13] and adjust the free model parameters. In Section 3, we calculate the solvation free energy of alkanes and the free energy of the interaction between two spherical molecules in water. In Section 4, we discuss the results and derive the equations that allow the electrostatic interactions in the system to be taken into account.

2. MODEL

Following the general scheme set out in [10–12], we consider a solvent whose density profile is described by a field $\rho(\mathbf{r})$. The part related to the fluctuations whose length scale is specified by the correlation function of a pure solvent, $\chi(\mathbf{r}, \mathbf{r}')$, which below is denoted by $\omega(\mathbf{r})$, can be separated in the field $\rho(\mathbf{r})$. The density $n(\mathbf{r}) = \rho(\mathbf{r}) - \omega(\mathbf{r})$ is the spatially averaged component of the field $\rho(\mathbf{r})$ that changes greatly at distances that are large compared to the size of the water molecule. Since the solvent cannot penetrate into the volume V_{in} occupied by the solute, we must impose the condition $\rho(\mathbf{r}) = 0$ on the field $\rho(\mathbf{r})$ for $\mathbf{r} \in V_{in}$. The equality to zero of the

density $\rho(\mathbf{r})$ inside the volume occupied by the solute results in efficient interaction between $n(\mathbf{r})$ and $\omega(\mathbf{r})$. In general, however, direct interaction between $n(\mathbf{r})$ and $\omega(\mathbf{r})$ should be added to it.

The thermodynamic properties of a pure solvent (water) are determined by the free energy functional in terms of the Ginzburg–Landau (GL) expansion, which describes the liquid–vapor phase transition. The density $n(\mathbf{r})$ acts as an order parameter. The interaction between the small-scale and large-scale densities, $\omega(\mathbf{r})$ and $n(\mathbf{r})$, can shift the state of the system into the evaporation region. This, in turn, can cause a large decrease in the water density near the solute surface, an effective drying-out of the surface. By minimizing the free energy functional, we calculate the density profile and the solvation free energy. It should be noted that the presence of two length scales in our theory is the result of the mean-field description in terms of the GL interaction functional between water molecules. In a completely microscopic theory, the division of the density into small-scale and large-scale components is a redundant procedure—all must be described in terms of the density $\rho(\mathbf{r})$ alone. In this case, the theory must contain a description of the interaction between water molecules on a microscopic level. An attempt to construct a completely microscopic theory was made in [4]. In our view, the computational difficulties in [4] severely limit the applicability of this theory to the solvation of molecules with large sizes and complex geometry.

Before we turn to a detailed description of the approach, let us consider the differences between our model and similar (in content) models discussed in [10–12]. In [12], the large-scale density n is described in terms of a lattice model in which the entire space is divided into cells. The description of the densities by continuous functions used in our paper allows this procedure to be avoided. In addition, in our paper, all densities are calculated in a self-consistent way, without invoking additional averaging over the fluctuations, as, for example, in [11]. Thus, our model describes more accurately the solvation of objects of complex geometry and the interaction between solvated objects at small distances than the models in [11, 12]. In addition, we generalize the mean-field description of the solvation of hydrophobic molecules to charged molecules.

A. Free Energy Functional

In our model, the state of the liquid in the absence of a solute is described by the functional

$$\mathcal{F}_0[\omega, n] = \frac{1}{2}\int \omega(\mathbf{r})\chi^{-1}(\mathbf{r}, \mathbf{r}')\omega(\mathbf{r}')d\mathbf{r}d\mathbf{r}' + \int\left\{\frac{a}{2}(\nabla n(\mathbf{r}))^2 + W(n(\mathbf{r}))\right\}d\mathbf{r}. \quad (1)$$

Here, $n(\mathbf{r})$ is the slowly changing (mean) density of the liquid, which is considered in our case as an order parameter; $\omega(\mathbf{r})$ is the field that corresponds to the small-scale density fluctuations; $\chi(|\mathbf{r} - \mathbf{r}'|)$ is the correlation function of the unperturbed liquid; a is the phenomenological parameter that must be defined by the microscopic theory (see below); $W(n(\mathbf{r}))$ is the self-consistent potential specified in a standard way as the Ginzburg–Landau expansion in terms of the order parameter $n(\mathbf{r})$ by a fourth-degree polynomial:

$$W(n) = \frac{b}{2}(n - n_1)^2(n - n_2)^2 \quad (0 \le n_1 \le n_2 \le 1), \quad (2)$$

where n_1 and n_2 are the values of the order parameter n in vapor and water, respectively (below, unless stated otherwise, it is assumed that $n_1 = 0$); b is the coupling constant that, together with a, defines the surface tension (see (10)). Functional (1) consists of two (so far independent) parts. The first term describes the nonlocal small-scale Gaussian fluctuations of the solvent whose spatial correlation is defined by the correlation function $\chi(|\mathbf{r} - \mathbf{r}'|)$,

$$\chi(|\mathbf{r} - \mathbf{r}'|) = \langle\omega(\mathbf{r})\omega(\mathbf{r}')\rangle. \quad (3)$$

The large-scale part with the gradient term in (1) is the excess of free energy with respect to a homogeneous liquid when the vapor–liquid interface is formed and, thus, describes the density fluctuations related to the vapor–liquid phase transition.

To some extent, the functional form of potential (2) is chosen arbitrarily. In general, the functional form must satisfy the requirements that directly follow from the physical formulation of the problem. As was mentioned above, functional (1) with potential (2) must describe the liquid–vapor transition near the solute surface. The presence of a solute shifts the parameters of a pure liquid into the vapor–liquid phase equilibrium region [11]. On the other hand, this requirement may be considered as an effective allowance for the repulsive part of the solute–water interaction potential, thereby fixing the mechanism of the hydrophobic effect. The general form of the functional that describes the behavior of the solvent in the phase coexistence region is known; this is a potential with two minima that is usually parameterized by the GL expansion (see Fig. 1). The specific form of the parameterization is not exhausted by the polynomial form of the GL theory; the main requirement is that the corresponding parameterization describe correctly the vapor–liquid transition region under nearly normal conditions. In general, the GL expansion is enough to elucidate the main features of the physical picture. One possible way of improving the model in the future is to choose a more realistic form of potential (2) that appears in the large-scale part of the functional. Figure 1b displays an asymmetric potential that belongs to a broader class of potentials than that described by Eq. (2). The figure is shown for $n_1 = 0$.

The potential $W(n)$ in the GL expansion depends on four parameters (a, b, n_1, n_2), two of which (n_1 and n_2) are fixed by the solvent density in the coexisting phases. The parameters a and b are considered as free adjustable parameters of the theory. Assuming the more general form of potential (2) shown in Fig. 1b, more specifically,

$$\tilde{W}(n) = W(n) + \mu(n - n_1),$$

we may consider the relative difference between the values at the minima of the potential, $\mu(n_2 - n_1)$, as an additional parameter. This quantity is closely related to the asymptotic behavior of the solvation energy in the limit of large molecules. The asymmetry of the potential contributes to the volume-dependent part of the solvation energy. This contribution tends to PV_{sol} as the solute volume V_{sol} increases [14] (here, P is the liquid pressure). Simple estimates show that the contribution from the volume part to the solvation energy on nanoscales is negligible compared to the part proportional to the solute surface area [13]. Thus, we may disregard the asymmetric form of the potential in Fig. 1b and consider a potential that is symmetric relative to the phase change, as in Fig. 1a. It is this form of the potential that is given in (2).

In the absence of interaction between the fields n and ω, minimizing functional (1) leads to the system of equations

$$\int \chi(\mathbf{r} - \mathbf{r}')\omega(\mathbf{r}')d\mathbf{r}' = 0, \tag{4a}$$

$$-a\Delta n(\mathbf{r}) + \frac{\delta W(n(\mathbf{r}))}{\delta n(\mathbf{r})} = 0. \tag{4b}$$

The solution of Eq. (4a) corresponds to the state of a homogeneous liquid, $\omega(\mathbf{r}) = \text{const}$, while the solutions of Eq. (4b) complemented by appropriate boundary conditions describe the following: (i) the plane vapor–liquid interface in the one-dimensional case, $n(-\infty) \equiv n_1 = 0$ ($n(\infty) = n_2$) and (ii) the density profile near an extended surface of macroscopic sizes in the two- and three-dimensional cases. In this case, $n = 0$ on the solute surface and $\nabla n = 0$ for $|\mathbf{r}| \longrightarrow \infty$. In this section, for the simplicity, we restrict our analysis to objects of spherical geometry (discussion of the solvation of objects with a complex shape is deferred for Section 3).

The solutions obtained can be qualitatively analyzed by calculating the first integrals of motion by analogy with classical mechanics [17]. Writing the Laplace operator in spherical coordinates,

$$\Delta n(r) = n''(r) + \frac{2}{r}n'(r),$$

multiplying (4b) by

$$n'(r) \equiv \frac{dn(r)}{dr}$$

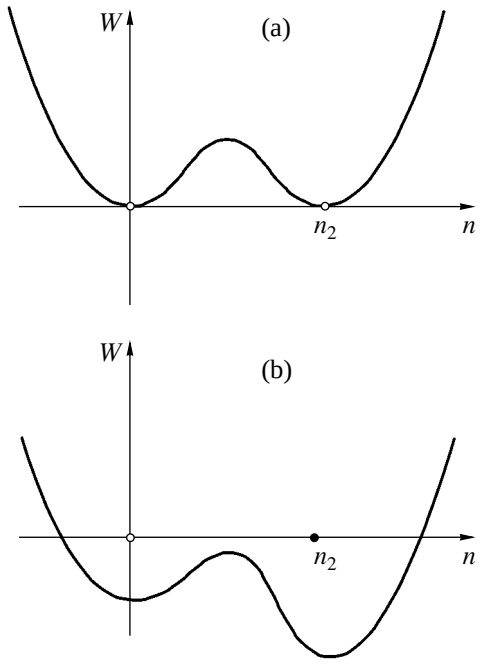


Fig. 1. Possible forms of the potential $W(n)$. Case (a) corresponds to Eq. (2) and is considered in this paper.

and integrating from the sphere radius r_{in} to ∞, we obtain

$$-a\int_{r_{in}}^{\infty}\frac{d}{dr}\left[\frac{(n'(r))^2}{2}\right]dr - a\int_{r_{in}}^{\infty}\frac{2}{r}(n'(r))^2 dr + \int_{r_{in}}^{\infty}\frac{dW(r)}{dr}dr = 0, \quad n'(\infty) = 0. \tag{5}$$

The integration in (5) yields

$$(n'(r))^2\big|_{r = r_{in}} = \int_{r_{in}}^{\infty}\frac{4}{r}(n'(r))^2 dr + \frac{2}{a}[W(n(r_{in})) - W(n(\infty))]. \tag{6}$$

Using (6) and the explicit form of potential (2), we can calculate $n'(\mathbf{r})$ at $r = r_{in}$. The value of $n'(\mathbf{r})$ determines the qualitative behavior of the density profile. Since the potential-related part in our case disappears, the density monotonically increases from n_1 in $\mathbf{r}_{in}$ to n_2 in $|\mathbf{r}| = \infty$. In practice, the density profile reaches its equilibrium value at a certain distance from the solute surface. The asymmetric potential can be analyzed in a similar way.

The meaning of the parameters a and b becomes clear in the limit of large radii, $r_{in} \gg \delta$, where δ is the

characteristic size of the water–vapor transition layer (see below). Substituting (2) into (4b) and replacing the interaction of $n(\mathbf{r})$ with the solute molecule by the boundary condition, we obtain

$$\begin{aligned} -a\Delta n(\mathbf{r}) + 2b(n(\mathbf{r}) - n_{\text{avr}}) \\ \times (n(\mathbf{r}) - n_1)(n(\mathbf{r}) - n_2) = 0, \\ n(|\mathbf{r}| = r_{\text{in}}) = n_1, \\ n'(\infty) = 0, \end{aligned} \tag{7}$$

where $n_{\text{avr}} = (n_2 + n_1)/2$. Assuming that $r \gg \delta$ in (7) and expanding the Laplace operator in a Taylor series,

$$\Delta = \frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr}\bigg|_{r \gg \delta} = \frac{d^2}{dr^2} + O\left(\frac{\delta}{r}\right),$$

we derive an equation that defines the liquid–vapor interface in the one-dimensional case:

$$\begin{aligned} -a\tilde{n}''(r) + 2b(\tilde{n}(r) - n_{\text{avr}}) \\ \times (\tilde{n}(r) - n_1)(\tilde{n}(r) - n_2) = 0, \\ \tilde{n}(r_{\text{in}}) = n_{\text{avr}}, \\ \tilde{n}'(\infty) = 0. \end{aligned} \tag{8}$$

The solutions that describe the density profile near the solute are

$$\begin{aligned} \tilde{n}(r) = \frac{n_2 - n_1}{2}\tanh\left(\frac{r - r_{\text{in}}}{\delta}\right) + \frac{n_2 + n_1}{2}, \\ \delta = \frac{2}{n_2 - n_1}\sqrt{\frac{a}{b}}. \end{aligned} \tag{9}$$

It is easy to see that in the case of a symmetric potential parameterized by the GL expansion, the large-scale part of the functional contains only one parameter. This parameter, δ, fixes the width of the liquid–vapor interface. Calculating the functional yields the following solvation energy per unit area of a spherical molecule [18]:

$$\begin{aligned} \frac{\tilde{G}}{4\pi r_{\text{in}}^2} = \frac{1}{4\pi r_{\text{in}}^2} \\ \times \int_{-\delta}^{\delta}\left\{\frac{a}{2}(\nabla\tilde{n})^2 + \frac{b}{2}(\tilde{n} - n_1)^2(\tilde{n} - n_2)^2\right\}4\pi r^2 dr \\ = \frac{\sqrt{ab}}{6}(n_2 - n_1)^3 + O\left(\frac{\delta}{r_{\text{in}}}\right), \end{aligned} \tag{10}$$

where

$$\sigma = \frac{\sqrt{ab}}{6}(n_2 - n_1)^3$$

has the meaning of surface tension. When deriving Eq. (10), we substituted the exact boundary condition $n(r_{\text{in}} - \delta) = 0$ with the approximate $n(r_{\text{in}}) = n_{\text{avr}}$, which clearly can be done for $0 < \delta/r_{\text{in}} \ll 1$.

B. Correlation Function

An important initial component of our theory is the bulk correlation function of a pure solvent, $\chi(|\mathbf{r} - \mathbf{r}'|)$. By definition,

$$\chi(|\mathbf{r} - \mathbf{r}'|) = \bar{n}\delta(\mathbf{r} - \mathbf{r}') + \bar{n}^2 h(|\mathbf{r} - \mathbf{r}'|). \tag{11}$$

Here, $\bar{n}$ is the mean solvent density in the bulk, $h(\mathbf{r} - \mathbf{r}') = g(\mathbf{r} - \mathbf{r}') - 1$, and the function $g(|\mathbf{r} - \mathbf{r}'|)$ has a clear physical meaning:

$$n^2(\mathbf{r})g(\mathbf{r}) = V^{-1}\left\langle \sum_{i \neq j}\delta(\mathbf{r} - \mathbf{r}_i + \mathbf{r}_j)\right\rangle. \tag{12}$$

The function $g(\mathbf{r})$ can be either directly determined from experimental data, or derived by numerical simulation, or calculated from a self-consistent integral with an appropriate closure condition [19].

The following relation directly results from the definition of $\chi(\mathbf{r} - \mathbf{r}')$:

$$\begin{aligned} \bar{n}\int(g(|\mathbf{r} - \mathbf{r}'|) - 1)d\mathbf{r}' \\ = 4\pi\bar{n}\int h(r)r^2 dr \approx -1. \end{aligned} \tag{13}$$

The latter equality may be considered exact, because the compressibility of the liquid is infinitesimal. We define the inverse correlation function $\chi_{\text{in}}^{-1}(\mathbf{r}, \mathbf{r}')$ as

$$\begin{aligned} \int_{V_{\text{in}}} d\mathbf{r}''\chi_{\text{in}}^{-1}(\mathbf{r}, \mathbf{r}'')\chi(|\mathbf{r}'' - \mathbf{r}'|) = \delta(\mathbf{r}' - \mathbf{r}), \quad \mathbf{r} \in V_{\text{in}}, \\ \chi_{\text{in}}^{-1}(\mathbf{r}, \mathbf{r}'') = 0, \quad \mathbf{r} \in V_{\text{out}}. \end{aligned} \tag{14}$$

It should be noted that the integration in (14) is only over the volume V_{in} occupied by the solute; therefore, the inverse correlation function possesses no translational invariance.

In this paper, we use two correlation functions: (i) the correlation function of water derived from numerical molecular-dynamical simulations [20] and (ii) the correlation function of hard spheres derived from the solution of the self-consistent Percus–Yevic integral equation [21–23]. The functions $h(\mathbf{r})$ for water and hard spheres are shown in Fig. 2. Both functions correspond to the dimensionless water bulk density $n_2 = 0.7$. Other approximate equations of the statistical theory of liquids can be found, for example, in [19].

Having the correlation functions of water and hard spheres at our disposal, we can compare the sensitivi-

ties of the density profile and the solvation energy to the detailed structure of the liquid and to check the convergence of the numerical scheme: the computational result must be virtually independent of the specific choice of the correlation function far from the solute surface.

The correlation function naturally defines the length scale ξ in the theory. Below, we fix $\xi = 2.78$ Å, the distance from the coordinate origin to the first peak of the correlation function.

C. Partition Function

Let us now turn to the solvent–solute interactions. We take into account the solvation of the molecule, first, by requiring the absence of a solvent within the volume occupied by the solute, v_{in}:

$$\rho(\mathbf{r}) = n(\mathbf{r}) + \omega(\mathbf{r}) \equiv 0, \tag{15}$$

and, second, by introducing a direct interaction between the small-scale and large-scale solvent densities in the simplest form,

$$V[\omega, n] = c\int \omega(\mathbf{r})n(\mathbf{r})d\mathbf{r}, \tag{16}$$

where c is the coupling constant. Below, all analytical calculations are performed for an arbitrary c; the results of our numerical calculations are presented only for $c = 0$. The case of $c \neq 0$ requires a separate analysis.

We define the solute–solvent partition function as

$$Z = \int \mathcal{D}\{\omega(\mathbf{r})\} \prod_{\mathbf{r} \in V_{\text{in}}} \delta[\rho(\mathbf{r})] e^{-\mathcal{F}}, \tag{17}$$

where

$$\mathcal{F} = \mathcal{F}_0[\omega, n] + V[\omega, n] + \int d\mathbf{r}\varphi(\mathbf{r})\rho(\mathbf{r}), \tag{18a}$$

$$\rho(\mathbf{r}) = n(\mathbf{r}) + \omega(\mathbf{r}). \tag{18b}$$

We write the constraints imposed by the Dirac δ function in (17) using the functional Fourier transform as

$$\prod_{\mathbf{r} \in v_{rm}} \delta[\rho(\mathbf{r})] = \frac{1}{(2\pi)^{\mathcal{N}}}\int \mathcal{D}\{\psi(\mathbf{r})\} \times \exp\left\{ i \int_{v_{\text{in}}} \rho(\mathbf{r})\psi(\mathbf{r})d\mathbf{r} \right\}. \tag{19}$$

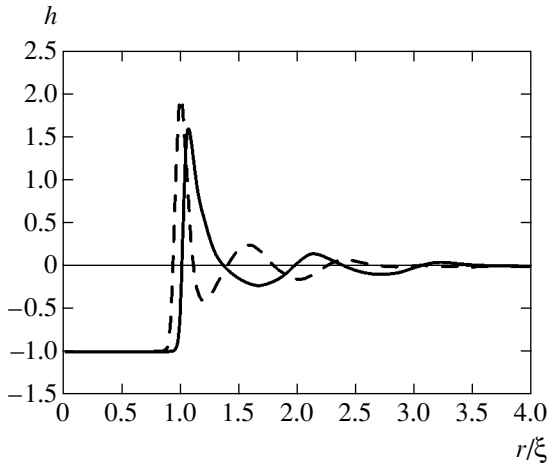


Fig. 2. The correlation function of water and the Percus–Yevic correlation function of hard spheres. The distance is in dimensionless units r/ξ ($\xi = 2.78$ Å, the dimensionless bulk density of water in these units is $n_2 = 0.7$; the solid and dashed lines represent hard spheres and water, respectively.

Here, $\mathcal{N}$ is the number of space points in the product $\prod_{\mathbf{r} \in v_{\text{in}}}(\ldots)$, and the integration is over all points of all fields $\psi(\mathbf{r})$:

$$\mathcal{D}\{\psi(\mathbf{r})\} \equiv \prod_{\mathbf{r} \in v_{\text{in}}} d\psi(\mathbf{r}).$$

This representation allows the partition function in (17) to be written as a functional integral over the fields $\psi(\mathbf{r})$:

$$Z = \Xi^{-1}\int \mathcal{D}\{\omega(\mathbf{r})\}\mathcal{D}\{\psi(\mathbf{r})\} e^{-S[\omega, n, \psi]},$$
$$S[\omega, n, \psi] = \mathcal{F}[\omega, n] + i\int_{v_{\text{in}}} \rho(\mathbf{r})\psi(\mathbf{r})d\mathbf{r}, \tag{20}$$

where Ξ is a constant that depends on the normalization of the partition function Z.

The partition function Z with the action $S[\omega, n, \psi]$ defined by (20) is a key object of the theory. Such basic quantities as the total, $\rho(\mathbf{r})$, and mean, $n(\mathbf{r})$, densities, the solvation free energy ΔG, and the interaction free energy of several solvated objects, can be determined using the generating functional Z.

D. The Mean Density

Having first integrated all fields in (20), we determine the mean (large-scale) density $n(\mathbf{r})$ by requiring that

$$\frac{\delta S[n(\mathbf{r})]}{\delta n(\mathbf{r})} = 0 \tag{21}$$

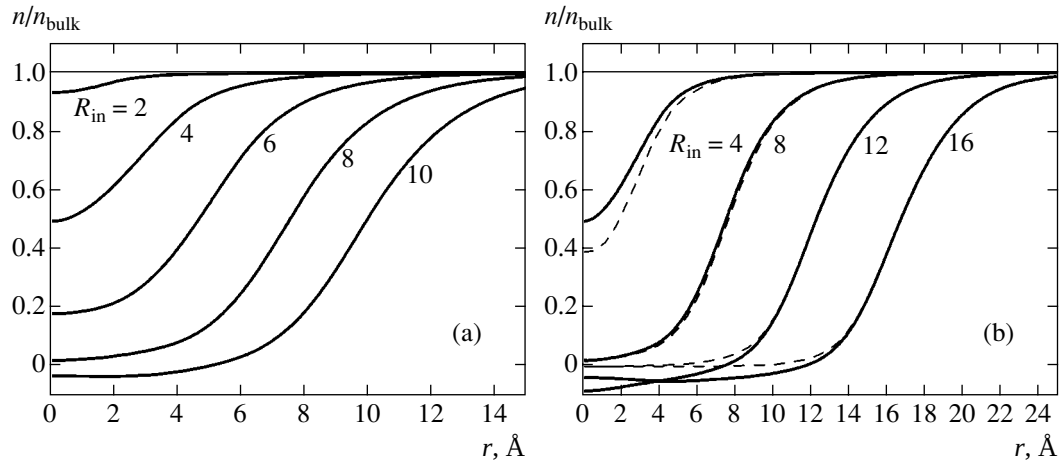


Fig. 3. Typical profiles of the large-scale density $n(r)$ around a spherical molecule for various radii R_{in}: the solid and dashed lines represent water and hard spheres, respectively.

or, in explicit form,

$$\begin{aligned} &-a\Delta n(\mathbf{r}) + \frac{\delta W(n(\mathbf{r}))}{\delta n(\mathbf{r})} - 2cn(\mathbf{r}) \\ &-c\int\limits_{V_{in}} d\mathbf{r}' \int\limits_{V_{out}} d\mathbf{r}''\chi_{in}^{-1}(\mathbf{r},\mathbf{r}')\chi(|\mathbf{r}'-\mathbf{r}''|)n(\mathbf{r}'') \\ &+\int\limits_{V_{in}} d\mathbf{r}'\chi_{in}^{-1}(\mathbf{r},\mathbf{r}')n(\mathbf{r}') = 0, \\ &\mathbf{r}\in V_{in} \end{aligned} \tag{22a}$$

and

$$\begin{aligned} &-a\Delta n(\mathbf{r}) + \frac{\delta W(n(\mathbf{r}))}{\delta n(\mathbf{r})} - c^2\int\limits_{V_{out}} d\mathbf{r}'\chi(|\mathbf{r}-\mathbf{r}'|)n(\mathbf{r}') \\ &+c^2\int\limits_{V_{in}} d\mathbf{r}'\int\limits_{V_{in}} d\mathbf{r}''\int\limits_{V_{out}} d\mathbf{r}'''\chi(|\mathbf{r}-\mathbf{r}'|) \\ &\times\chi_{in}^{-1}(\mathbf{r}',\mathbf{r}'')\chi(|\mathbf{r}''-\mathbf{r}'''|)n(\mathbf{r}''') \\ &-c\int\limits_{V_{in}} d\mathbf{r}'\int\limits_{V_{in}} \mathbf{r}''\chi(|\mathbf{r}-\mathbf{r}'|)\chi_{in}^{-1}(\mathbf{r}',\mathbf{r}'')n(\mathbf{r}'') = 0, \\ &\mathbf{r}\in V_{out}. \end{aligned} \tag{22b}$$

Thus, condition (21) leads to a system of integro-differential equations that relate the density $n(\mathbf{r})$ in the inner region, V_{in}, occupied by the solute and the density $n(\mathbf{r})$ in the outer region, V_{out}.

In this paper, we consider only the case of $c = 0$, i.e., when there is no direct coupling between the small-scale and large-scale fluctuations. In this case, the equations are simplified significantly:

$$\begin{aligned} &-a\Delta n(\mathbf{r}) + \frac{\delta W(n(\mathbf{r}))}{\delta n(\mathbf{r})} \\ &+\int\limits_{V_{in}} d\mathbf{r}'\chi_{in}^{-1}(\mathbf{r},\mathbf{r}')n(\mathbf{r}') = 0, \quad \mathbf{r}\in V_{in}, \end{aligned} \tag{23a}$$

$$-a\Delta n(\mathbf{r}) + \frac{\delta W(n(\mathbf{r}))}{\delta n(\mathbf{r})} = 0, \quad \mathbf{r}\in V_{out}. \tag{23b}$$

The following technique is used to numerically solve system (23a), (23b). When Eq. (14) is solved, the determination of $\chi_{in}^{-1}(\mathbf{r}, \mathbf{r}')$ takes the bulk of the computational time. To speed up this procedure, let us multiply (23a) by $\chi(|\mathbf{r}-\mathbf{r}'|)$ and, integrating over the entire space V_{tot} ($V_{tot} = V_{in} \cup V_{out}$), make use of (14). We then obtain

$$\begin{aligned} &n(\mathbf{r}) + \int\limits_{V_{in}} d\mathbf{r}'\chi(|\mathbf{r}-\mathbf{r}'|) \\ &\times\left(-a\Delta n(\mathbf{r}') + \frac{\delta W(n(\mathbf{r}'))}{\delta n(\mathbf{r}')}\right) = 0, \quad \mathbf{r}\in V_{in}, \\ &-a\Delta n(\mathbf{r}) + \frac{\delta W(n(\mathbf{r}))}{\delta n(\mathbf{r})} = 0, \quad \mathbf{r}\in V_{out}. \end{aligned} \tag{24}$$

These equations no longer explicitly contain $\chi_{in}^{-1}(\mathbf{r}, \mathbf{r}')$; therefore, (14) may not be solved at all, thereby reducing significantly the computational time.

The general form of the density profile $n(\mathbf{r})$ for a spherical molecule at certain (so far arbitrary) constants a and b is shown in Fig. 3. The vapor (n_1) and water (n_2)

densities are $n_1 = 0$ and $n_2 = 0.7$, respectively. The numerical solutions of (24) were obtained by using the correlation functions of hard spheres (dashed line) and water (solid line).

The presence of regions with a negative density should not be considered an erroneous result, since only the total density $\rho(\mathbf{r})$ should be positive, while the density $n(\mathbf{r})$ is auxiliary in nature, defining the remaining quantities in a self-consistent way, and has no literal physical meaning. That the averaged large-scale density $n(\mathbf{r})$ has a negative mean value within the solute just implies that the density fluctuations are asymmetric in this region, so locally $\langle\omega\rangle \neq 0$.

The total density $\rho(\mathbf{r}) = n(\mathbf{r}) + \omega(\mathbf{r})$ that corresponds to the minimum of functional (20) satisfies the system of equations

$$\begin{aligned}
&\rho(\mathbf{r}) = 0, \quad \mathbf{r} \in V_{\text{in}},\\
&\rho(\mathbf{r}) = n(\mathbf{r}) - \int_{V_{\text{in}}} d\mathbf{r}' \int_{V_{\text{in}}} d\mathbf{r}''\chi(|\mathbf{r}-\mathbf{r}'|)\\
&\times \chi_{\text{in}}^{-1}(\mathbf{r}', \mathbf{r}'')n(\mathbf{r}'')\\
&+ c\int_{V_{\text{in}}} d\mathbf{r}' \int_{V_{\text{in}}} d\mathbf{r}'' \int_{V_{\text{in}}} d\mathbf{r}'''\chi(|\mathbf{r}-\mathbf{r}'|)\\
&\times \chi_{\text{in}}^{-1}(\mathbf{r}', \mathbf{r}'')\chi(|\mathbf{r}-\mathbf{r}'''|)n(\mathbf{r}''')\\
&- c\int_{V_{\text{out}}} d\mathbf{r}'\chi(|\mathbf{r}-\mathbf{r}'|)n(\mathbf{r}'),\\
&\mathbf{r} \in V_{\text{out}}.
\end{aligned} \tag{25}$$

As in the case of system (23), Eqs. (25) at $c = 0$ can be written in a form that does not explicitly contain $\chi_{\text{in}}^{-1}(\mathbf{r}, \mathbf{r}')$ [10].

E. Solvation Free Energy

We define the solvation free energy ΔG as the energy that should be spent to transport the solute from the solvent to a vacuum (we make no difference between ΔF and ΔG, since $PV_{\text{sol}} \ll \Delta G$). The partition function of the solvent containing the solute molecule can be written by directly using (17) as

$$Z_{\text{solv–sol}} = \int_{V_{\text{solv}}} \mathcal{D}\{\omega(\mathbf{r})\} \prod_{\mathbf{r} \in V_{\text{sol}}} \delta[\rho(\mathbf{r})] e^{-\mathcal{F}}, \tag{26}$$

where $\mathcal{F}$ is defined by Eq. (18a). The solvation free energy ΔG can be calculated in a standard way as the logarithm of the ratio of the corresponding partition functions,

$$\Delta G = -\ln \frac{Z_{\text{solv–sol}}}{Z_{\text{solv}}}. \tag{27}$$

Here, Z_{solv} is the partition function of a pure solvent in the absence of a solute. At first glance, it seems natural to write Z_{solv} as

$$Z_{\text{solv}} = \int_{V_{\text{solv}}} \mathcal{D}\{\omega(\mathbf{r})\} e^{-\mathcal{F}}. \tag{28}$$

However, the divergences that are formally related to the emergence of an uncompensated product of the Gaussian integrals in the ratio $Z_{\text{solv–sol}}/Z_{\text{solv}}$ arise when (28) is directly used to calculate (27). Physically, these divergences are associated with the difference between the statistical ensembles in which the corresponding partition functions are calculated. When we represent $Z_{\text{solv–sol}}$ in form (26) by imposing constraints in the form of a δ function, we fix a certain total density ρ within the solute volume V_{in}; thus, with regard to the field ρ, we are dealing with a canonical ensemble. At the same time, calculating Z_{solv} using (28), we take into account all the possible configurations of the field ρ within the volume occupied by the solute, which corresponds in meaning to a large canonical ensemble.

Equation (28) can be regularized by using the relation between the solvation free energy and the probability of spontaneous formation of a cavity with a given volume in the solvent [11]. Let us represent ΔG in (27) as

$$\Delta G = -\ln \frac{Z_V(0)}{\sum_{N \geq 0} Z_V(N)}, \tag{29}$$

where $Z_V(N)$ is the partition function of the solvent, provided that N solvent molecules are in the volume V_{in}. Let us pass from the summation in (29) to integration by introducing the corresponding continuous variable $\tilde{n} = N/V_{\text{in}}$, the mean density of the solvent molecules within the volume V_{in}. Z_{solv} can then be written in regularized form as

$$\begin{aligned}
Z_{\text{solv}} &= \int_0^\infty d\tilde{n} \int_{V_{\text{solv}}} \mathcal{D}\{\omega(\mathbf{r})\}\\
&\times \prod_{\mathbf{r} \in V_{\text{sol}}} \delta[\rho(\mathbf{r}) - \tilde{n}] e^{-\mathcal{F}}.
\end{aligned} \tag{30}$$

In this paper, we use a different expression for Z_{solv} that can be derived from the above expression via integration over the variable $\tilde{n}$ by the stationary phase method,

$$Z_{\text{solv}} = \int_{V_{\text{solv}}} \mathcal{D}\{\omega(\mathbf{r})\} \prod_{\mathbf{r} \in V_{\text{sol}}} \delta[\rho(\mathbf{r}) - \bar{n}] e^{-\mathcal{F}}. \tag{31}$$

Thus, definition (27) with the partition functions (26) and (31) naturally arises from simple physical consid-

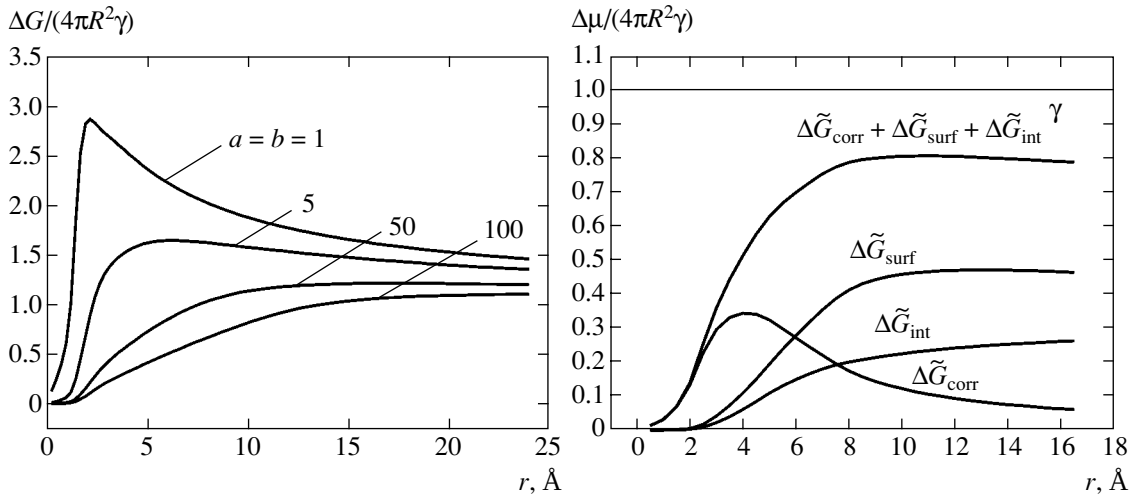


Fig. 4. Qualitative behavior of the normalized solvation free energy for various parameters a and b in the large-scale functional of the excess free energy (1) (left panel); the contributions $\Delta\tilde{G}_{corr}$, $\Delta\tilde{G}_{surf}$, and $\Delta\tilde{G}_{int}$ to the normalized solvation free energy of a spherical molecule, $\Delta\tilde{G}_{solv}(R_{in})$ at $a = 0$ and $b = 30$ (right panel).

erations and no longer contains any divergences. It is important to note that the equilibrium density profile n minimizing Z_{solv} in (30) coincides with the equilibrium bulk density of a pure solvent, $\bar{n}$; therefore, the effective action in (31) that corresponds to this density profile is exactly equal to zero.

The solvation free energy ΔG can be calculated by substituting the equilibrium density profile $n_{eq}(\mathbf{r}) \equiv n(\mathbf{r})$ derived by solving system (22a), (22b) into functional (18a). The solvation free energy $G \equiv \mathcal{F}\{n_{eq}\}$ can then be represented as the sum of three terms:

$$\Delta G_{solv}(V_{in}) = \Delta G_{corr} + \Delta G_{surf} + \Delta G_{int}, \qquad (32)$$

where

$$\Delta G_{corr} = \frac{1}{2}\int_{V_{in}} n_{eq}(\mathbf{r})\chi_{in}^{-1}(\mathbf{r}, \mathbf{r}')n_{eq}(\mathbf{r}')d\mathbf{r}d\mathbf{r}',$$
$$\Delta G_{surf} = \frac{a}{2}\int(\nabla n_{eq}(\mathbf{r}))^2 d\mathbf{r}, \qquad (33)$$
$$\Delta G_{int} = \int W(n_{eq}(\mathbf{r}))d\mathbf{r}.$$

As in the calculations of the mean density $n(r)$, we consider only the case of $c = 0$. Using (23a) and (23b), we can rewrite (32) without the inverse correlation function $\chi_{in}^{-1}(\mathbf{r}, \mathbf{r}')$ as

$$\Delta G_{solv}(V_{in}) = \int\left(W(n) - \frac{n}{2}\frac{\delta W}{\delta n}\right)d\mathbf{r}. \qquad (34)$$

In Fig. 4, the normalized solvation energy

$$\Delta\tilde{G}_{solv} = \Delta G_{solv}/4\pi R_{in}^2 \qquad (35)$$

for a spherical molecule is plotted against its radius R_{in}. Also shown here are the corresponding specific contributions $\{\Delta\tilde{G}_{corr}, \Delta\tilde{G}_{surf}, \Delta\tilde{G}_{int}\} = \{\Delta G_{corr}, \Delta G_{surf}, \Delta G_{int}\}/(4\pi R_{in}^2)$.

The solvation free energy of a spherical hydrophobic molecule [10–12] is proportional (as expected) to the volume of the solute molecule for fairly small R_{in} and to its surface area for fairly large R_{in}. We clearly see a nonmonotonic behavior of the solvation free energy as a function of the size of the solute molecule, which has already been discussed previously (see, e.g., [13]).

F. Adjusting the Model Parameters

The free parameters a and b of functional (1), (2) were chosen to satisfy the following requirements

The theoretically calculated normalized solvation free energy $\Delta\tilde{G}_{solv} = \Delta G_{solv}/(4\pi R^2)$ for a spherical molecule of radius R (see (35)) should closely agree with the results of Monte Carlo simulations [13].

The theoretically calculated total density $\rho(r|R)$ for several radii R of a spherical molecule (25) should reproduce the behavior $\bar{n}g(r)$ of the correlation function near a spherical cavity of radius R derived from Monte Carlo simulations [13].

The corresponding results are shown in Fig. 5 for $a = 9$ and $b = 30$.

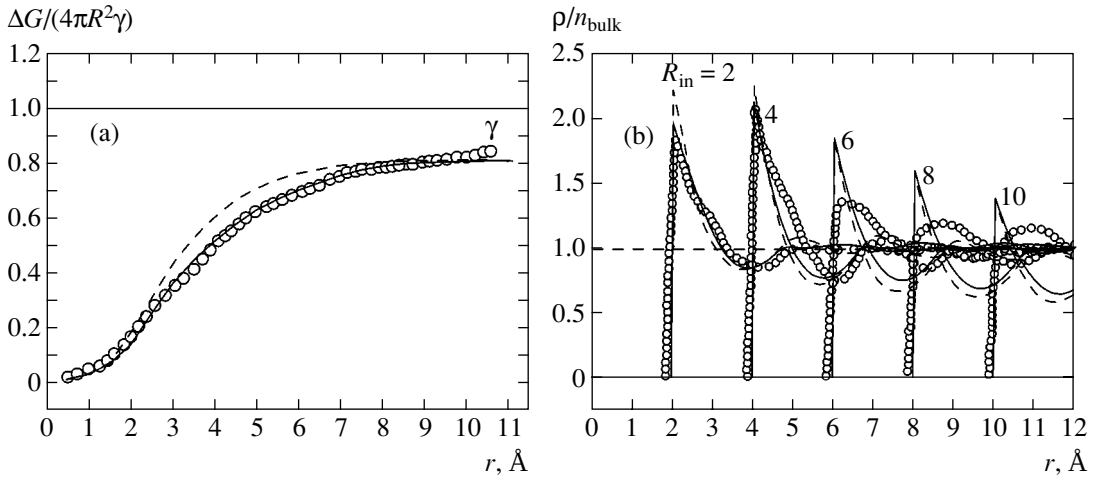


Fig. 5. (a) Normalized solvation free energy (34) of a spherical molecule in comparison with the results of Monte Carlo simulations [13]; (b) the total density profile $\rho(r)/\bar{n}$ around a spherical cavity of various radii R_{in} in comparison with the results of Monte Carlo simulations [13] for the correlation function $g(r)$; the dots represent Monte Carlo simulations, the solid and dashed lines correspond to water and hard spheres, respectively.

The dashed and solid curves represent the results obtained using the correlation functions of hard spheres and water, respectively, under normal conditions. The solvation free energy $\Delta\tilde{G}_{solv}$ is presented in Fig. 5 in dimensionless units $\Delta\tilde{G}_{solv}/\gamma$, where γ is the surface tension of a plane liquid–vapor interface. The numerical value of $\gamma \approx 72$ mJ m^{-2} was taken from [13].

The solvation free energies ΔG_{solv} and $\Delta\tilde{G}_{solv}$ calculated by directly using (34) and (35) can be converted to standard units using the scaling relations

$$\begin{cases} \Delta G_{solv}[\text{dimensionless units}] \times 0.2479 = \Delta G_{solv}, \text{ kcal/mol}, \\ \Delta\tilde{G}_{solv}[\text{dimensionless units}] \times \dfrac{0.2479}{(2.78)^2} = \Delta\tilde{G}_{solv}, \text{ kcal/(mol Å}^2). \end{cases} \quad (36)$$

In all of the subsequent calculations, these scaling factors are used without modifications.

3. RESULTS AND DISCUSSION: THE SOLVATION OF ARBITRARILY SHAPED OBJECTS

Using our theory, we calculated the solvation free energy of neutral molecules (linear alkanes) and the interaction energy of several objects of simple geometry (spheres).

A. Solvation of Alkanes

As noted above, the model parameters and the scaling factors were adjusted in a such a way that the numerically computed solvation free energy and solvent profile near the solute agreed closely with the Monte Carlo simulations for hard spheres (see the previous section); subsequently, they no longer changed. We used the same values of the parameters to compute the solvation free energy of alkanes. As we show below, the results obtained are in excellent agreement with predictions of the scaled particle theory (SPT), whose parameters were specially adjusted to alkanes [24].

This close agreement requires elucidations. There is the opinion [25–27] supported by numerical simulations that the solute–solvent interaction energy can be divided into two parts: (i) the cavitation component of the cavity formation free energy, ΔG_{cav}, and (ii) the dispersion (attractive) part corresponding to the tail of the Van der Waals (VdW) potential, ΔG_{disp}. The numerical simulation methods based on thermodynamic integration allow each of the contributions, ΔG_{cav} and ΔG_{disp}, to be calculated separately [27]. They can also be calculated separately in the SPT suggested in [14, 15]. The corresponding calculations were performed in [24], whose authors reported excellent agreement between the sum $\Delta G_{cav} + \Delta G_{disp}$ and the experimental data. Extreme caution should be exercised with regard to the latter assertion. The point is that the experimental solvation free energies for alkanes are extracted from data

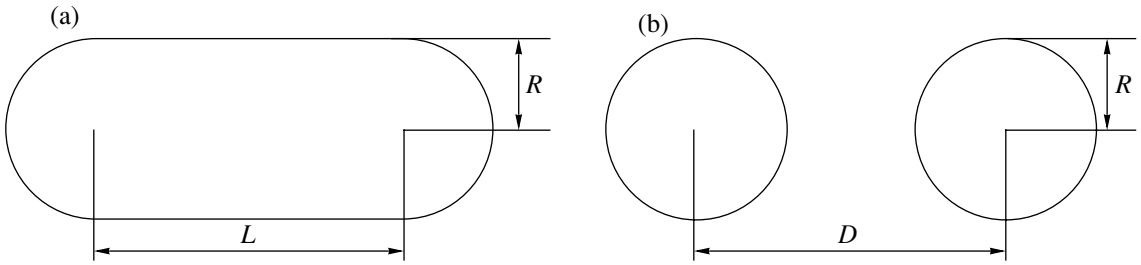


Fig. 6. Schematic view of a solvated cylinder with hemispherical edges (a) and two spherical solute molecules (b).

on their solubility in different solvents, in particular, in water. The universally accepted procedure for determining the solvation free energy is described in [28]. Some of the authors attempt to apply various kinds of corrections to the solvation free energies obtained in this way (more specifically, to the dispersion tails of the VdW potential) [29], which can significantly change the corresponding quantities. The "corrected" experimental data from [29] are given in [24]. In general, the cavitation component remains unchanged. Therefore, we will focus our attention on the determination of this quantity.

In the model described above, the solute is considered as an object surrounded by a hard wall; thus, only the cavitation part of the solvation free energy is taken into account. The attractive part of the VdW potential has not yet been considered. In our view, there are no fundamental difficulties of including this part of the interaction in our model. In addition, the δ-function constraint in (19) can be modified to describe the VdW potential in the region V_{in} more accurately. The corresponding computations are being performed and will be the subject of a separate analysis. The cavitation part of the alkane solvation free energy calculated in our model and the results obtained in [24] and [30] are given in the table.

Recall once again that the parameters $a = 9$ and $b = 30$ in our model and the scaling factors (36) were adjusted to agree with the results of Monte Carlo simulations without any additional adjustment for alkanes. The volume V_{in} occupied by the alkane molecule was calculated as follows. Having taken the coordinates of the centers of all atoms from a standard database, we surrounded them by spheres of radius 3.5 Å for all carbon atoms (the length of the C–O bond in water) and 3.05 Å for all hydrogen bonds (the length of the O–H bond in water) by defining V_{in} as the union of the corresponding spherical volumes. We determined the lengths of the C–O and O–H bonds by calculating the pair correlation function [C(methane)–O(water] and [H(methane)–O(water)] of a system composed of 216 water molecules and one methane molecule by the Monte Carlo method [31]. Similar values of the parameters can be extracted from the force fields suggested in [32].

Comparison of the calculated cavitation part of the solvation energy for alkanes, ΔG_{cav}, for our model with the results obtained in [24] and [30]

Alkanes	ΔG_{cav}, kcal/mol (our model)	ΔG_{cav}, kcal/mol [24]	ΔG_{cav}, kcal/mol [30]
CH_4	5.69	5.61	5.36
C_2H_6	7.54	7.53	7.15
C_3H_8	9.07	9.13	
C_4H_{10}	10.52	10.8	
C_5H_{12}	11.96	12.8	
C_6H_{14}	13.24	14.8	14.22

As was mentioned above, the dispersion part of the VdW potential can be easily included in the model. The simplest way is to add the following term to the original GL functional:

$$U_{\text{disp}} = \int U(\mathbf{r})n(\mathbf{r})d\mathbf{r},$$

where $U(\mathbf{r})$ is the attractive part of the VdW potential. With this modification, Eqs. (23a) and (23b) for the equilibrium large-scale mean density change somehow. More specifically, Eq. (23a) remains unchanged, while an additional term related to the presence of potential $U(\mathbf{r})$ in the region $\mathbf{r} \in V_{\text{out}}$ appears in Eq. (23b):

$$-a\Delta n(\mathbf{r}) + \frac{\delta W(n(\mathbf{r}))}{\delta n(\mathbf{r})} + U(\mathbf{r}) = 0. \qquad (37)$$

B. Solvation of Cylindrical Objects and the Interaction between Two Distant Spheres

In this section, we consider the solvation free energy of a cylindrical cavity and the interaction energy of two spherical cavities in a fluctuating medium (i.e., water). Both systems are schematically shown in Fig. 6.

The analysis of these questions has a double meaning.

(1) First, we analyzed the skin depth of the solvated object. A direct calculation of the solvation free energy

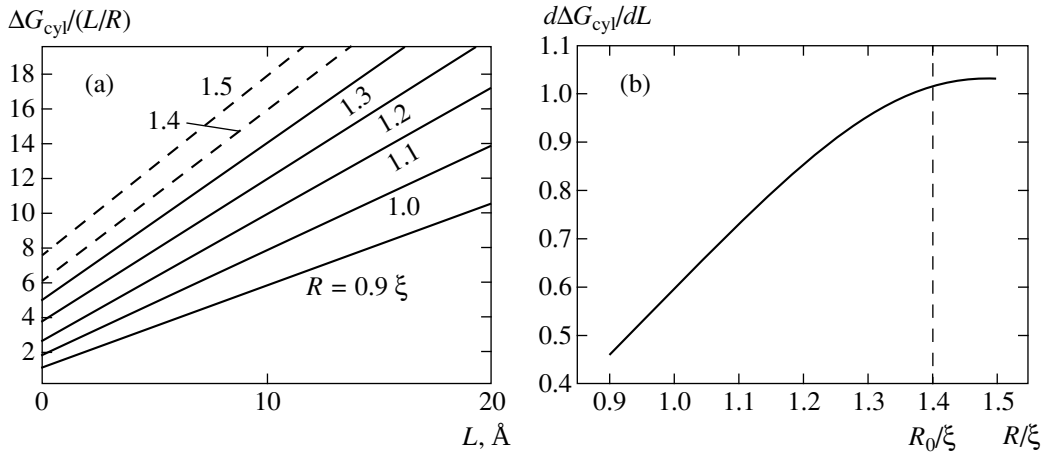


Fig. 7. (a) Solvation free energy of a cylinder, $\Delta G_{cyl}(L|R)$ versus its length for various R; (b) $d\Delta G_{cyl}/dL$ versus R at $L = 12$ Å.

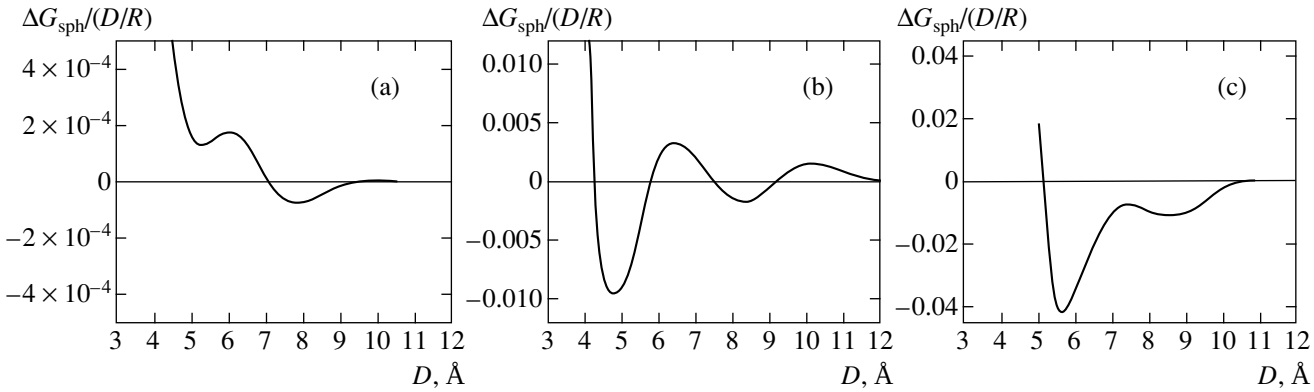


Fig. 8. Interaction free energy of two spherical solute molecules, ΔG_{sph}, versus distance between their centers D for $R = 0.4\xi$ (a), 0.7ξ (b), and 0.9ξ (c).

for biological systems such as ligand-proteins take significant computational time. Indeed, according to (24), the large-scale density $n(\mathbf{r})$ must be taken into account at all points of the solute within the volume V_{in}. On the other hand, it is clear that the hard-wall constraint prevents the density $n(\mathbf{r})$ from penetrating deep into this region. The penetration depth of the density $n(\mathbf{r})$ into the volume V_{in} can be estimated by considering the solvation of cylindrical molecules of various radii and lengths.

The solvation free energy $\Delta G_{cyl}(L|R)$ is plotted against the length L for several radii R in Fig. 7a.

As we see from Fig. 7b, the slope of the curves in Fig. 7a is constant for $R > R_0 \approx 1.4\xi = 3.9$ Å. This implies that the fluctuating density field for $R > R_0$ penetrates into the solute to distances up to a certain fixed value (i.e., to the skin depth). Thus, the part of the inner volume located at a distance larger than the skin length, $R_0 \approx 3.9$ Å, inward from the solute surface may be disregarded.

(2) Second, we analyzed the interaction energy $\Delta G_{sph} = \Delta G_{sph}(D|R) - \Delta G_{sph}(D \longrightarrow \infty|R)$ of two spherical molecules of radius R as a function of the distance D between their centers. The interaction free energy minus the solvation free energy of the separate spheres, or the mean-force potential [33, 34] $\Delta G_{sph}(D|R)$, was calculated for the potential of a hard wall $U(D)$,

$$U(D) = \begin{cases} \infty, & D \le 2R, \\ 0, & D > 2R; \end{cases}$$

i.e., we again considered only the cavitation part of the interaction energy by disregarding the dispersion (attractive) part of the VdW potential between the impenetrable spheres of radius R. Even for this simplified model situation, we found the behavior of $\Delta G_{sph}(D|R)$ with distance to be qualitatively different for different R with respect to the scale length ξ. It should be noted once again that ΔG_{sph} is the interaction free energy of spherical cavities minus the solvation

free energy of the two spheres dissolved separately, $\Delta G_{sph}(D \longrightarrow \infty|R)$. The qualitative behavior of this quantity is shown in Fig. 8.

For all values of the radius, $R = 0.4\xi$ Å, 0.7ξ Å, and 0.9ξ Å, we reproduce the nonmonotonic D dependence $\Delta G_{sph}(D)$. The nonmonotonic behavior of this type is related to the oscillations of the water correlation function. A qualitatively similar behavior was found earlier using direction simulations consistent with SPT predictions (see [33] and references therein) and calculations based on the concept of solvent surface accessible area [34]. Including the attractive part of the VdW potential can significantly change the behavior of $\Delta G_{sph}(D)$; nevertheless, the nonmonotonic behavior of $\Delta G_{sph}(D)$ will remain.

Comparison of Figs. 8a and 8c shows that the dependence $\Delta G_{sph}(D)$ is sensitive (even on a qualitative level) to the radius R of the interacting spheres. The effective attraction between smaller spheres is weaker than that between larger spheres. Qualitatively, the emerging picture can be understood from the following simple considerations. For fairly large sphere radii, $R \gtrsim \xi$, the fact that the molecules are closely spaced is attributable to the loss of free energy, which can be interpreted as a classical manifestation of the Casimir effect. At the same time, small (with $R \lesssim \xi$) objects fit into the spontaneously forming (through fluctuations) cavities in water with sizes of the order of ξ; as a result, the loss of energy that corresponds to the spheres being closely spaced is much smaller than that for larger objects. Moreover, it can be noticed (Fig. 8a) that the mean-force potential has the attractive (negative) part that corresponds to the second molecular water shell. Thus, two fairly small spherical objects can form a spontaneous bound state, being separated by a monomolecular water shell. The detected effect is in qualitative agreement with the results of [23], in which the authors analyzed the mean-force potential acting between two spherical molecules.

4. ELECTROSTATICS IN A FLUCTUATING DIPOLAR MEDIUM

There are several alternative ways of including the electrostatic interactions in our model. In the simplest standard scheme, the permittivity of the solvent, ε, is assumed to depend linearly on its density $n(\mathbf{r})$, i.e., $\varepsilon(\mathbf{r}) = 1 + \alpha n(\mathbf{r})$. Here, we do not use this assumption; instead, we treat the solvent as a gas of dipoles, obtaining a system of self-consistent equations as a result. Our approach is ideologically close to the model of Debye screening in an electrically neutral plasma [35]. In contrast to the plasma, this model imposes holonomic constraints on the motion of charges of opposite signs, resulting in the formation of dipoles. Many papers are devoted to the screening in systems composed of charged extended objects. The theory that is most suitable for our purposes was developed in [36].

Below, we derive the basic equations that include both the electrostatic and hydrophobic interactions in a fluctuating medium.

Let us introduce the following quantities: $\rho^q_{mol}(r, \tau)$ is the charge distribution in the solvent molecule with a given orientation τ and the center of mass located at the coordinate origin ($\rho^q_{mol}(\mathbf{r}, \tau) = 0$ outside the molecule); $\rho^q_{sol}(\mathbf{r})$ is the charge distribution in the solute; and $\phi_{eff}(\mathbf{r})$ is the effective mean electrostatic field.

According to the logic of the mean-field approach [36], the molecules of the medium are in an effective electrostatic field $\phi_{eff}(\mathbf{r})$. Averaging over the spatial charge distribution in the molecule yields the mean energy of a molecule:

$$\phi(\mathbf{r}, \tau) = \int \phi_{eff}(\mathbf{r}')\rho^q_{mol}(\mathbf{r}' - \mathbf{r}, \tau)d\mathbf{r}'. \tag{38}$$

The electrostatic potential $\phi_{eff}(\mathbf{r})$ is defined by the Poisson equation

$$\Delta\phi_{eff}(\mathbf{r}) = -\frac{4\pi}{\varepsilon_{sol}}\rho^q_{sol}(\mathbf{r}) - \mu(\mathbf{r}), \tag{39}$$

where $\mu(\mathbf{r})$ is the polarization charge of the medium associated with the orientation of the molecules along the electric field,

$$\mu(\mathbf{r}) = \iint (e^{-\beta\phi(\mathbf{r}', \tau)} - 1)\rho^q_{mol}(\mathbf{r} - \mathbf{r}', \tau) \times \rho(\mathbf{r}')\frac{d\Omega_\tau}{4\pi}d^3\mathbf{r}', \tag{40}$$

and $\rho(\mathbf{r}) = n(\mathbf{r}) + \omega(\mathbf{r})$ is the fluctuating density of the medium. Disregarding the small-scale fluctuating part $\omega(\mathbf{r})$, we can write the following in the mean-field approximation:

$$\mu(\mathbf{r}) = \iint (e^{-\beta\phi(\mathbf{r}', \tau)} - 1)\rho^q_{mol}(\mathbf{r} - \mathbf{r}', \tau)n(\mathbf{r}') \times \frac{d\Omega_\tau}{4\pi}d^3\mathbf{r}'. \tag{41}$$

The free energy functional for the system can be written as

$$\mathcal{F}[n(\mathbf{r}), \phi_{eff}(\mathbf{r})] = \int d\mathbf{r}\left(\frac{a}{2}(\nabla n(\mathbf{r}))^2 + W(n)\right) + \frac{1}{2}\int_{V_{in}} d\mathbf{r}d\mathbf{r}'n(\mathbf{r})\chi^{-1}_{in}(\mathbf{r}, \mathbf{r}')n(\mathbf{r}') + \lambda\int d\mathbf{r}(\nabla\phi_{eff}(\mathbf{r}))^2, \tag{42}$$

where λ is a coefficient that depends on the chosen system of units; it is commonly assumed to be equal to $1/(8\pi)$.

According to [37], the minimum of the functional corresponds to equilibrium charge density and electric potential distributions, provided that the relation between the charge density and the potential in the form of the Poisson equation is retained. It thus follows that we must vary the functional

$$\begin{aligned}\tilde{\mathcal{F}}[n(\mathbf{r}), \phi_{\text{eff}}(\mathbf{r})] &= \int d\mathbf{r}\left(\frac{a}{2}(\nabla n(\mathbf{r}))^2 + W(n)\right)\\ &+ \frac{1}{2}\int_{V_{\text{in}}} d\mathbf{r}\,d\mathbf{r}'\,n(\mathbf{r})\chi_{\text{in}}^{-1}(\mathbf{r},\mathbf{r}')n(\mathbf{r}')\\ &+ \lambda\int d\mathbf{r}(\nabla\phi_{\text{eff}}(\mathbf{r}))^2\\ &+ \int d\mathbf{r}\varphi(\mathbf{r})\left(\Delta\phi_{\text{eff}}(\mathbf{r}) + \frac{4\pi}{\varepsilon_{\text{sol}}}\rho_{\text{sol}}^q(\mathbf{r}) + \mu(\mathbf{r})\right)\end{aligned} \tag{43}$$

in which the field φ appears as the Lagrange multiplier, for the constraint specified by the Poisson equation.

Minimizing $\tilde{F}$ in fields $\varphi(\mathbf{r})$, $\phi_{\text{eff}}(\mathbf{r})$, and $n(\mathbf{r})$, we obtain a system of self-consistent equations whose solution yields the corresponding equilibrium field and density distributions in our problem.

So far, the charge density distribution in the solvent molecule has been assumed to be arbitrary. Let us illustrate the application of the theory outlined above for a liquid of point dipoles. In this case, the charge density ρ_{mol}^q is given by

$$\begin{aligned}\rho_{\text{mol}}^q(\mathbf{r}) &= e\delta\left(\mathbf{r} - \frac{\mathbf{l}}{2}\right) - e\delta\left(\mathbf{r} + \frac{\mathbf{l}}{2}\right)\\ &\approx -e\mathbf{l}\nabla\delta(\mathbf{r}) = -\mathbf{d}\nabla\delta(\mathbf{r}),\end{aligned} \tag{44}$$

where $\mathbf{d}$ is the dipole moment of the molecule. Then,

$$\phi(\mathbf{r}, \mathbf{d}) = -\int d\mathbf{r}'\phi_{\text{eff}}(\mathbf{r}')\mathbf{d}\nabla_{\mathbf{r}'}\delta(\mathbf{r}' - \mathbf{r}) = \mathbf{d}\nabla\phi_{\text{eff}}(\mathbf{r}). \tag{45}$$

This equation, together with the linearized form (40), leads to the following expression for the polarization charge density $\mu(\mathbf{r})$:

$$\begin{aligned}\mu(\mathbf{r}) &= \int\int d\mathbf{r}'\frac{d\Omega_{\mathbf{d}}}{4\pi}\beta\mathbf{d}\nabla_{\mathbf{r}'}\phi_{\text{eff}}(\mathbf{r}')\mathbf{d}\nabla_{\mathbf{r}}\delta(\mathbf{r} - \mathbf{r}')\\ &= \frac{d^2\beta}{3}\nabla(n(\mathbf{r})\nabla\phi_{\text{eff}}(\mathbf{r})).\end{aligned} \tag{46}$$

Substituting the latter relation into functional (43) and minimizing it yields

$$\begin{aligned}\frac{\delta\tilde{\mathcal{F}}}{\delta n} &= -a\Delta n(\mathbf{r}) + \frac{\delta W}{\delta n} + \int_{V_{\text{in}}} d\mathbf{r}'\chi_{\text{in}}^{-1}(\mathbf{r},\mathbf{r}')n(\mathbf{r}')\\ &\quad - \frac{d^2\beta}{3}\nabla\varphi(\mathbf{r})\nabla\phi_{\text{eff}}(\mathbf{r}) = 0,\\ \frac{\delta\tilde{\mathcal{F}}}{\delta\phi_{\text{eff}}} &= \frac{d^2\beta}{3}\nabla(n(\mathbf{r})\nabla\varphi(\mathbf{r}))\\ &\quad - \lambda\Delta\phi_{\text{PA}} + \Delta\varphi(\mathbf{r}) = 0,\\ \frac{\delta\tilde{\mathcal{F}}}{\delta\varphi} &= \Delta\phi_{\text{eff}} + \frac{4\pi}{\varepsilon}\rho_{\text{O}\}\backslash\S.}^q(\mathbf{r})\\ &\quad + \frac{d^2\beta}{3}\nabla(n(\mathbf{r})\nabla\phi_{\text{eff}}(\mathbf{r})) = 0.\end{aligned} \tag{47}$$

This system of equations should be complemented by physically justified boundary conditions such as

$$\begin{aligned}\frac{\partial\varphi(\mathbf{r})}{\partial\mathbf{r}} &\longrightarrow 0, \quad |\mathbf{r}| \longrightarrow \infty;\\ \phi_{\text{eff}}(\mathbf{r}) &\longrightarrow 0, \quad |\mathbf{r}| \longrightarrow \infty;\\ n(\mathbf{r}) &\longrightarrow \bar{n}, \quad |\mathbf{r}| \longrightarrow \infty.\end{aligned} \tag{48}$$

It is easy to verify that Eqs. (47) transform into the standard equations for a continuous medium with an effective permittivity dependent on the density of the medium in the limit of a weak electric field. Indeed, when the field ϕ_{eff} is weak enough, in view of the second equation of system (47),

$$\phi_{\text{eff}} \sim \varphi, \tag{49}$$

which implies that

$$\nabla\varphi\nabla\phi_{\text{eff}} \sim \phi_{\text{eff}}^2. \tag{50}$$

Given relations (49) and (50), it can be seen that Eqs. (47) split in the first order in ϕ_{eff}. Therefore, they can be written as

$$\begin{aligned}-a\Delta n(\mathbf{r}) + \frac{\delta W}{\delta n} + \int_{V_{\text{in}}} d\mathbf{r}'\chi_{\text{in}}^{-1}(\mathbf{r},\mathbf{r}')n(\mathbf{r}') &= 0,\\ \nabla(\varepsilon(n(\mathbf{r}))\nabla\varphi(\mathbf{r})) + \frac{4\pi}{\varepsilon}\rho_{*\dagger\ddagger}^q(\mathbf{r}) &= 0,\end{aligned} \tag{51}$$

where

$$\varepsilon(n(\mathbf{r})) \approx 1 + \frac{d^2\beta}{3}n(\mathbf{r}). \tag{52}$$

The closed system of equations (51), (52) describes the electrostatics in a continuous medium with a permittivity proportional to the density of the medium.

It should be noted that our model suggests the absence of a correlation between the orientational degrees of freedom of the water molecules in the bulk and in the immediate vicinity of the solute surface. We can preaverage $\mu(\mathbf{r})$ in (41) over the corresponding orientational degrees of freedom only in this case. Our assumption is based on the conclusions drawn in [38, 39] from numerical simulations: the effects related to the reorientation of water molecules near a hydrophobic surface play a secondary role in solvation thermodynamics (see [27] for details).

5. CONCLUSIONS

Below, we briefly present our results.

(1) The method described here provides a good basis for a consistent description of the solvation effects for objects with an arbitrarily shaped hard surface. The well-known cavitation contributions to the solvation free energy of alkanes can be described using the two-length scale fluctuational continuous model with free parameters adjusted to agree with the results of Monte Carlo simulations on the solvation of hard spheres without any additional adjustment to the parameters.

(2) Two different problems can be analyzed in terms of a unified approach: (i) calculating the solvation free energy of cylindrical objects and (ii) calculating the free energy of two spheres separated by a certain distance. Analysis of the first problem leads us to conclude that the skin depth (i.e., the penetration length of the fluctuating field into the solute) is $\sim 1.4\xi$ Å, where ξ is the characteristic length scale in the problem (the distance to the first maximum of the solvent correlation function). Analysis of the second problem leads us to conclusions about the mean-force potential between two hard spheres in a fluctuating medium.

(3) The developed approach can be easily generalized to include the dispersion part of the solute–solvent interaction and the electrostatic interactions between charged molecules. The main steps needed for the corresponding modifications were described in this paper.

ACKNOWLEDGMENTS

We are grateful to I. Erukhimovich, I. Bodrenko, E. Tkalya, O. Khoruzhy, V. Zosimov, A. Danchev, M. Subbotin, M. Olevanov, and V. Ozrin for fruitful discussions. We also thank M. Lewitt and C. Queen for valuable comments and suggestions. This work was supported in part by grant no. ACI-NIM-2004-243 “Nouvelles interfaces des mathématiques” (France).

REFERENCES

1. P. George, C. W. Bock, and M. Trachtman, J. Comput. Chem. **3**, 283 (1982).
2. A. Gavezzotti, Modell. Simul. Mater. Sci. Eng. **10**, R1 (2002).
3. A. Finkelstein and L. Pereyaslavets (in press).
4. Y. Rosenfeld, J. Chem. Phys. **98**, 8126 (1993); Y. Rosenfeld and P. Tarazona, Mol. Phys. **95**, 141 (1998).
5. T. Ooi, M. Oobatake, G. Nemethy, and H. A. Sheraga, Proc. Natl. Acad. Sci. USA **84**, 3086 (1987).
6. G. Makhatadze and P. L. Privalov, Adv. Protein Chem. **47**, 307 (1995).
7. M. L. Connolly, Science **221**, 709 (1983).
8. T. Lazaridis and M. Karplus, J. Mol. Biol. **288**, 477 (1999).
9. Q. Du, D. Beglov, and B. Roux, J. Phys. Chem. B **104**, 796 (2000).
10. D. Chandler, Phys. Rev. E **48**, 2898 (1993).
11. K. Lum, D. Chandler, and J. D. Weeks, J. Phys. Chem. B **103**, 4570 (1999).
12. P. R. Wolde, S. X. Sun, and D. Chandler, Phys. Rev. E **65**, 011201 (2002).
13. D. M. Huang, Ph. L. Geissler, and D. Chandler, J. Phys. Chem. B **105**, 6704 (2001).
14. H. Reiss, H. L. Frisch, and J. L. Lebowitz, J. Chem. Phys. **31**, 369 (1959).
15. R. A. Pierotti, Chem. Rev. **76**, 717 (1976).
16. G. Hummer, S. Garde, A. García, *et al.*, Proc. Natl. Acad. Sci. USA **93**, 8951 (1996).
17. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 1: *Mechanics*, 3rd ed. (Nauka, Moscow, 1973; Pergamon, Oxford, 1976).
18. D. M. Huang and D. Chandler, J. Phys. Chem. B **106**, 2047 (2002).
19. G. N. Sarkisov, Usp. Fiz. Nauk **169**, 625 (1999) [Phys. Usp. **42**, 545 (1999)].
20. A. K. Soper, Chem. Phys. **258**, 121 (2000).
21. J. K. Percus and G. J. Yevic, Phys. Rev. **110**, 1 (1958).
22. M. S. Wertheim, Phys. Rev. Lett. **10**, 321 (1963).
23. J. Throop and R. J. Bearman, J. Chem. Phys. **42**, 2408 (1965).
24. M. Irisa, K. Nagayama, and F. Hirata, Chem. Phys. Lett. **207**, 430 (1993).
25. B. Lee, Biopolymers **24**, 813 (1985).
26. B. Widom, Chem. Phys. **86**, 869 (1982).
27. E. Gallicchio, M. M. Kubo, and R. M. Levy, J. Phys. Chem. **104**, 6271 (2000).
28. A. Ben-Naim and Y. Marcus, J. Chem. Phys. **81**, 2016 (1984).

29. K. A. Sharp, A. Nicholls, R. Friedman, and B. Honig, Biochemistry **30**, 9686 (1991).

30. A. A. C. C. Pais, A. Sousa, M. E. Eusébio, and J. S. Redinha, Phys. Chem. Chem. Phys. **3**, 4001 (2001).

31. V. Ozrin, private communication.

32. T. A. Halgren, J. Comput. Chem. **17**, 490 (1996); **17**, 520 (1996); **17**, 553 (1996); **17**, 587 (1996); **17**, 616 (1996).

33. N. T. Southall and K. A. Dill, Biophys. Chem. **101**–**102**, 295 (2002).

34. S. Shimizu and H. S. Chan, Proteins: Struct., Funct., Genet. **48**, 15 (2002).

35. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 2nd ed. (Nauka, Moscow, 1964; Pergamon, Oxford, 1980), Part 1.

36. A. R. Khokhlov and K. A. Khachaturian, Polymer **23**, 1793 (1982).

37. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 2nd ed. (Nauka, Moscow, 1982; Pergamon, Oxford, 1984).

38. B. Madan and B. Lee, Biophys. Chem. **51**, 279 (1994).

39. S. Durell and A. Wallquist, Biophys. J. **71**, 1695 (1996).

Translated by V. Astakhov