Journal of Experimental and Theoretical Physics, Vol. 98, No. 1, 2004, pp. 1–13.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 1, 2004, pp. 5–18.

GRAVITATION, ASTROPHYSICS

Dark Matter in Galaxies and the Growth of Giant Black Holes

A. S. Ilyin*, K. P. Zybin, and A. V. Gurevich

P.N. Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 119991 Russia

**e-mail: asi.lpi.ru*

Received April 15, 2003

Abstract—The relationship of dark matter to giant black holes (BHs) in galactic nuclei is investigated. The simultaneous evolution of dark and baryonic matter under the effect of an averaged self-consistent gravitational field is considered. The distribution of dark matter is shown to remain spherically symmetric even if there is an appreciable asymmetry in the distribution of baryonic matter in the galaxy. A kinetic equation that describes the evolution of the distribution function for dark matter with gravitational scattering by stars is derived. A significant flux of dark matter on a seed BHe at the galactic center is shown to arise under these conditions. The law of growth of the seed BH via the absorption of dark matter has been established. The seed BH is shown to grow significantly, up to 10^7–$10^8 M_\odot$, in the lifetime of the galaxy. Observational data are briefly analyzed, and the presented theory has been found to be in reasonable agreement with experimental data. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, the intensive development of technology and the use of new methods for astrophysical observations (gas kinematics, maser kinematics, echo mapping, and stellar kinematics) have led to the reliable detection of a large number (more than 80) of giant black holes (BHs) [1, 2] with masses in the range $(2 \times 10^6$–$3 \times 10^9) M_\odot$ ($M_\odot = 2 \times 10^{33}$ g is the solar mass). The mass distribution is shown in Fig. 1. Analysis of observational data has revealed a relationship between the mass of a giant BH at the galactic center and the mass of the bulge.[1] The stellar rotation velocity in the bulge is generally much lower than the characteristic velocity dispersion σ. Therefore, the shape of the bulge does not differ too much from the spherical shape, and the number density of stars in it rapidly increases toward the galactic center. The characteristic bulge radius, r_b, increases with galactic mass: $r_b \sim 1$–30 kpc in the galaxies observed. The mass of giant BHs, M_{bh}, is lower than the bulge mass of the host galaxies by approximately three orders of magnitude.

Nondissipative dark matter is currently believed to constitute the bulk of the matter in the Universe. It is in dark matter that density fluctuations grow; these fluctuations form a large-scale gravitationally bound object, a galactic halo, at the nonlinear stage. An important property of this halo is the singular density distribution at its center [3]. The baryonic matter captured by the gravitational field of the halo gradually settles to the center and forms galaxies. The presence of a singularity in the distribution of dark matter leads to the possible formation of a primary BH [4] with a mass on the order of $10^3 M_\odot$ at the center of the future galaxy during the contraction of baryonic gas. Subsequently, the seed BH rapidly grows via the flow of baryonic and dark matter from the bulge into it. In this paper, we investigate the evolution of dark matter and its absorption by a BH at the galactic center.

The paper has the following structure. In Section 2, we consider the distribution of dark matter in a galaxy and its evolution under the effect of the averaged self-consistent gravitational field of baryonic matter that sinks to the central region of the galaxy due to energy losses and forms a bulge. We solve the problem in the adiabatic approximation by assuming that the baryonic matter is spherically symmetric.

In Section 3, we analyze the role of the deviations from spherical symmetry. The distribution of dark matter that adiabatically contracts under the effect of the

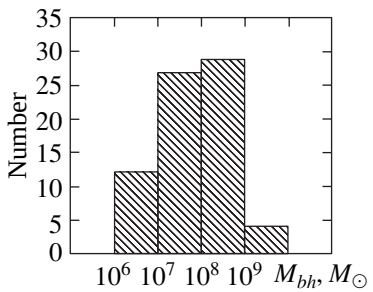

Fig. 1. The mass distribution of giant BHs (according to [1]).

[1] The bulge is the densest central region of a galaxy, which consists mostly of old stars.

combined large-scale self-consistent gravitational field of baryonic and dark matter is shown to remain spherically symmetric even if there is an appreciable asymmetry in the distribution of baryonic matter.

In Section 4, we study the growth of a seed BH at the galactic center via the direct capture of dark matter particles that move in an averaged self-consistent gravitational potential. We show that this mechanism is inefficient in real conditions and cannot cause any appreciable increase in the mass of the seed BH.

In Sections 5 and 6, we investigate the evolution of the distribution function for dark matter when the adiabaticity condition is violated. The main process of this kind is the gravitational scattering of particles during their collisions with stars in the bulge. An important property of this process is that the particle mean free path is much larger than the bulge scale length; therefore, collisions are rare. Taking this property into account, we derive an expression for the collision integral averaged over the oscillations of the dark matter particles captured in the self-consistent gravitational field and obtain a kinetic equation that describes the evolution of the distribution function for dark matter in Section 5. We show that diffusion in angular momentum space plays a major role here, because the initial distribution function is singular.

In Section 6, we solve the diffusion equation, determine the flux of dark matter on a BH, and establish the law of its growth.

Finally, in Section 7, we briefly analyze observational data. This analysis indicates that the growth of giant BHs via the absorption of dark matter is considerable.

In general, we may assert that the theory presented here is in reasonable agreement with the available observational data for giant BHs. Further development of the theory and observations and their detailed comparison are undoubtedly of considerable interest.

2. THE INFLUENCE OF BARYONIC MATTER ON THE DISTRIBUTION OF DARK MATTER IN THE CENTRAL REGIONS OF A GALAXY

Nondissipative dark matter plays a crucial role in the formation of the large-scale structure of the Universe—galaxies, clusters of galaxies, and superclusters. The evolution of small nonuniformities in the initial distribution of dark matter at the nonlinear stage leads to the gravitational contraction of dark matter and the formation of primary singularities of the density ρ after the passage of which a multistream flow is developed in the collisionless gas of dark matter. According to the analytical theory developed by Gurevich and Zybin [5], under very general assumptions about the shape of the initial perturbation, the development of flow oscillations gradually leads to the stirring of dark matter and the formation of stable, spherically symmetric self-captured objects with a singular density distribution at their centers:

$$\rho = Kr^{-\xi}, \quad \xi = 12/7. \tag{1}$$

Although this result was analytically obtained relatively long ago, for a long time only plane nonlinear structures had been obtained in numerical calculations. Navaro *et al.* [6] were able to also show the existence of stable, spherically symmetric clumps only by using special computational methods that actually reduced to the separation of individual density maxima. A further improvement in computational methods made it possible to separate a singular density profile in the central region of a clump with a singularity parameter ξ close to (1) [7]. Thus, now, numerical simulations may be considered to have confirmed the results of the analytical theory almost completely.

Following [5], let us consider a separate local perturbation of the dark matter density $\rho(\mathbf{r})$ in the following form near the maximum $\mathbf{r} = 0$:

$$\rho_{t=0} = \rho_0\left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2}\right), \quad a \geq b \geq c. \tag{2}$$

The growth of Jeans instability and the subsequent kinetic stirring of dark matter give rise to a nondissipative gravitational singularity (NGS) with scaling (1) and a spherically symmetric stationary distribution function

$$f(E, m) = f_0 I_R^{1/8} \delta(m^2 - l_0^2 I_R^2), \tag{3}$$

where

$$I_R(\dot{E}, m) = \frac{\sqrt{2}}{\pi}\int_{r_-}^{r_+} dr\left(E - \psi(r) - \frac{m^2}{2r^2}\right)^{1/2}$$

is the radial action, m is the magnitude of the angular momentum, $E = v^2/2 + \psi(r)$ is the energy, $\psi(r) \propto r^{2/7}$ is the NGS potential, $r_\pm$ are the turning points of dark matter particles, and f_0 and l_0 are constants that depend on the scale and shape of the initial perturbation. In particular,

$$l_0 = 0.16\epsilon, \tag{4}$$

where

$$\epsilon = (\epsilon_1^2 + \epsilon_2^2 - \epsilon_1\epsilon_2)^{1/2}, \quad \epsilon_1 = (a-b)/a,$$
$$\epsilon_2 = (a-c)/a$$

are the ellipticity parameters near the maximum of the initial density (2). If the parameters $\epsilon_{1,2}$ are assumed to

be generally random variables uniformly distributed in segment [0, 1], then the mean $\langle\epsilon\rangle = \sqrt{5/12}$, whence

$$\langle l_0\rangle \approx 0.1. \tag{5}$$

Thus, as follows from (3) and (4), the angular momenta of dark matter particles in the central region after NGS formation are low and proportional to the radial action I_R, while their orbits are highly elongated toward the NGS center.

Let us now consider the influence of the baryonic component on the structure of distribution function (3). Because of energy radiation during inelastic collisions, baryonic matter gradually sinks to the bottoms of potential wells produced by cold dark matter and forms galaxies, with the dark matter forming a giant galactic halo. The existence of a dark matter halo in galaxies is confirmed by rotation curves [8]. Other examples of such objects are clusters of galaxies. The dark matter halo shows up here via the gravitational confinement of a large amount of hot gas [9].

As follows from (1), the dark matter mass concentrated in the central NGS region tends to zero as the size of this region decreases as $r^{9/7}$. Therefore, although the total fraction of baryonic matter is small, its concentration in the central region greatly increases as it cools down and settles to the NGS center. This, in turn, can significantly affect the total gravitational potential Ψ produced by both dark, ψ_d, and baryonic, ψ_b, matter and thereby can change law (1).

In this section, we assume the distribution of the baryonic component to be spherically symmetric,

$$\psi_b = \psi_b(r).$$

The characteristic scale r_c on which deviations from universal law (1) would be expected due to the influence of baryons is defined by the relation $M_b(r_c) \geq M_d(r_c)$, where M_b is the mass of the baryonic matter and M_d is the mass of the dark matter. In general, this inequality holds in the bulge.

The main feature of the process under consideration is that the particles of baryonic matter lose their energy slowly—in a time comparable to the lifetime of the Universe. During this time, the particles of dark matter captured by the gravitational field in the central part of the halo oscillate many times. Thus, the change in total self-consistent potential Ψ is adiabatically slow. Under these conditions, the radial action is known [10] to be an integral of motion (an adiabatic invariant). Since the initial distribution function depends on the integrals of motion I_R and m, it retains its form (3) as a function of the radial action and the angular momentum throughout the slow evolution of the total potential from $\Psi = \psi_d$ to $\Psi = \psi_d + \psi_b$.

To determine the dark matter space density in the field $\Psi(r, t)$, it is convenient to pass from I_R and m to the variables E and m using the relation

$$I_R(E, m, t) = \frac{\sqrt{2}}{\pi}\int_{r_-}^{r_+} dr\left(E - \Psi(r, t) - \frac{m^2}{2r^2}\right)^{1/2}. \tag{6}$$

Since the radial action explicitly depends on time in (6), the distribution function of dark matter particles expressed in terms of energy and angular momentum, $f(E, m, t)$, also depends on time. In this case, the dark matter space density is given by

$$\rho(r, t) = \frac{\sqrt{2}\pi}{r^2}\int_0^\infty dm^2 \int_\Omega dE\, f(E, m, t) \times\left(E - \Psi(r, t) - \frac{m^2}{2r^2}\right)^{-1/2}, \tag{7}$$

where Ω is the energy range in which the following inequality holds:

$$E - \Psi(r, t) - m^2/2r^2 \geq 0.$$

Below, we will be interested in the region of space $r < r_c$ where, to a first approximation, we may disregard the influence of the dark mass on the total potential and assume that $\Psi = \psi_b$. For simplicity, we restrict our analysis to the case where the baryonic mass and potential in the final state have a power-law dependence on r:

$$M_b(r) \propto r^n, \quad \psi_b \propto r^{n-1}, \quad n \geq 0. \tag{8}$$

The radial action (6) may then be represented in factorized form:

$$I_R = I_0(r_+)C(\mu), \tag{9}$$

where

$$I_0 = G^{1/2}M_b^{1/2}(r_+)r_+^{1/2}, \quad \mu = \frac{m}{I_0},$$

$$C(\mu) = \frac{\sqrt{2}}{\pi}\int_{\beta_-}^1 d\beta\left(\int_\beta^1 \alpha^{n-2}d\alpha - \frac{\mu^2}{2}\left(\frac{1}{\beta^2} - 1\right)\right)^{1/2},$$

$$\beta_- = \beta_-(\mu) = \frac{r_-}{r_+},$$

G is the gravitational constant.

Using these relations and changing the variables $E, m \mapsto r_+, \mu$ in (7), we may also represent the dark matter space density as a power law:

$$\rho_d = K' r^{\frac{9}{16}n - \frac{39}{16}}. \quad (10)$$

It is important to note that, in the absence of the baryonic component, the exponent in (10) does not depend on the parameters f_0 and l_0 of the initial perturbation.

It follows from (10) and (1) that the densities of the baryonic and dark components at $n = 9/7$ change as $r^{-12/7}$. For an isothermal distribution of baryonic matter, $\rho_b \propto r^{-2}$ and $n = 1$. It then follows from (10) that the dark matter density increases as $r^{-15/8}$ when $r \longrightarrow 0$, i.e., slightly more slowly than the baryonic density. The largest contraction of dark matter would be expected at $n = 0$, when a compact massive baryonic object is formed at the NGS center. The dark matter distribution in its vicinity is

$$\rho_d \propto r^{-39/16}. \quad (11)$$

Note, however, that when a black hole serves as such an object, law (11) is violated due to the capture of dark matter particles by the BH. This question is considered in more detail in Section 6.

3. THE INFLUENCE OF ASYMMETRY IN THE DISTRIBUTION OF BARYONIC MATTER ON THE DISTRIBUTION FUNCTION FOR DARK MATTER

Above, we assumed the distribution of the baryonic component to be spherically symmetric. This restriction could actually be severe. Indeed, the baryonic potential usually contains a small spherically asymmetric part (associated, for example, with the asymmetry in the distribution of baryonic matter in the bulges of elliptical and spiral galaxies), while, in general, the angular momenta of the dark matter particles are not conserved as they move in an asymmetric potential. Therefore, at first glance, even a small asymmetric addition to the potential can significantly change the particle angular momenta in a time much longer than the characteristic oscillation period of the dark matter particles, thereby also changing the form of distribution function (3). Actually, for a small asymmetry and a slow change of the potential, the evolution of the radial action I_R, the magnitude of the angular momentum m, and the angular momentum component m_z consists only in small oscillations of these quantities about their initial values, while their mean values are virtually constant.

The generalization of the well-known theorem on the adiabatic invariance of action variables [10] proves this assertion. The collisionless motion of dark matter particles in a spherically symmetric, time-independent potential Ψ_0 is described by the kinetic equation

$$\frac{\partial f}{\partial t} + \{H_0, f\} = 0, \quad (12)$$

where

$$H_0 = \frac{p_r^2}{2} + \frac{p_\theta^2}{2r^2} + \frac{p_\varphi^2}{2r^2\sin^2\theta} + \Psi_0$$

is the Hamiltonian,

$$p_r = v_r, \quad p_\theta = \left(m^2 - \frac{m_z^2}{\sin^2\theta}\right)^{1/2}, \quad p_\varphi = m_z$$

are the canonical momenta, and {,} are the Poisson brackets. Next, let us make the canonical transformation to the action–angle variables. Using the always existing arbitrariness in constructing these variables [11], we choose I_R, m, and m_z as the action variables. Using the generating function

$$S(x, I) = \sqrt{2}\int dr\left(E(I, m) - \Psi_0 - \frac{m^2}{2r^2}\right)^{1/2} + \int d\theta\left(m^2 - \frac{m_z^2}{\sin^2\theta}\right)^{1/2} + m_z\varphi, \quad (13)$$

we define the corresponding angular variables ϕ in a standard way:

$$\phi_k = \frac{\partial S(x, I)}{\partial I_k} \quad (14)$$

(in what follows, x denotes the set of spherical coordinates r, θ, and φ; I denotes the set of action variables $\{I_k\} = \{I_R, m, m_z\}$; and φ denotes the set of angular variables.) Expressing the energy E in terms of the action variables and differentiating it with respect to I_k, we obtain the frequencies $\omega_k(I) = \partial E/\partial I_k$ that correspond to unperturbed motion in a time-independent, spherically symmetric potential. In this case, since E does not explicitly depend on m_z, frequency ω_3 is identically equal to zero. For this reason, the particle motion is degenerate in one coordinate and takes place in a plane perpendicular to the angular momentum vector. The kinetic equation in the new variables is

$$\frac{\partial f}{\partial t} + \omega_k\frac{\partial f}{\partial \phi_k} = 0. \quad (15)$$

Let us now consider the perturbed potential Ψ, which differs from Ψ_0 by the small asymmetric term

$\varepsilon\Psi_1$. We also assume that the potential explicitly depends on the slow time $t_1 = \varepsilon t$:

$$\Psi = \Psi_0(r, t_1) + \varepsilon\Psi_1(r, \theta, \varphi, t_1).$$

The perturbed kinetic equation can now be written as

$$\frac{\partial f}{\partial t} + \omega_k\frac{\partial f}{\partial \phi_k} = \varepsilon\left(\frac{\partial H_1}{\partial \phi_k}\frac{\partial f}{\partial I_k} - \frac{\partial H_1}{\partial I_k}\frac{\partial f}{\partial \phi_k}\right), \qquad (16)$$

where H_1 is defined according to general rules [10]:

$$H_1(I, \phi, t_1) = \Psi_1(x, t_1) + \frac{\partial S(x, I, t_1)}{\partial t_1}.$$

We solve Eq. (16) by a modified method of perturbation theory that allows the appearance of secular terms to be avoided. The idea of the method is to represent the function f as a series of successive approximations,

$$f = f_0 + \varepsilon f_1 + \varepsilon^2 f_2 + \ldots,$$

by formally assuming that f_k are functions of many times:

$$f_k = f_k(I, \phi, t_0, t_1, t_2, \ldots),$$

where

$$t_0 = t, \quad t_1 = \varepsilon t, \quad t_2 = \varepsilon^2 t, \ldots.$$

Accordingly, operator $\partial/\partial t$ can also be represented as a series:

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial t_0} + \varepsilon\frac{\partial}{\partial t_1} + \varepsilon^2\frac{\partial}{\partial t_2} + \ldots.$$

It follows from the equation of the zeroth approximation

$$\frac{\partial f_0}{\partial t_0} + \omega_k\frac{\partial f_0}{\partial \phi_k} = 0$$

and from the fact that initial function (3) does not depend on the angular variables ϕ and the third action variable m_z that f_0 does not depend on ϕ, m_z, and the fast time t_0 either:

$$f_0 = f_0(I_R, m, t_1, \ldots).$$

The equation of the first approximation is

$$\frac{\partial f_0}{\partial t_1} + \frac{\partial f_1}{\partial t_0} + \omega_k\frac{\partial f_1}{\partial \phi_k} = \frac{\partial H_1}{\partial \phi_k}\frac{\partial f_0}{\partial I_k}.$$

Averaging this equation over the angular variables ϕ and the fast time t_0 using the operations

$$\langle\ldots\rangle_\phi = \frac{1}{(2\pi)^3}\int d^3\phi\ldots,$$

$$\langle\ldots\rangle_T = \lim_{T\to\infty}\frac{1}{T}\int_0^T dt_0\ldots$$

yields

$$\frac{\partial f_0}{\partial t_1} = 0,$$

$$\frac{\partial f_1}{\partial t_0} + \omega_k\frac{\partial f_1}{\partial \phi_k} = \frac{\partial H_1}{\partial \phi_k}\frac{\partial f_0}{\partial I_k}.$$

This equation with the initial condition $f_1 = 0$ at $t = 0$ can be easily solved by the method of characteristics:

$$f_1 = \int_0^{t_0} dt_0'\frac{\partial H_1(I, \phi - \omega t_0', t_1)}{\partial \phi_k}\frac{\partial f_0}{\partial I_k}. \qquad (17)$$

The equality

$$\frac{\partial f_0}{\partial I_3} = \frac{\partial f_0}{\partial m_z} = 0$$

ensures that solution (17) is limited, although the frequency ω_3 is equal to zero. We write the equation of the second approximation as

$$\frac{\partial f_0}{\partial t_2} + \frac{\partial f_1}{\partial t_1} + \frac{\partial f_2}{\partial t_0} + \omega_k\frac{\partial f_2}{\partial \phi_k}$$
$$= \frac{\partial}{\partial I_k}\left(\frac{\partial H_1}{\partial \phi_k}f_1\right) - \frac{\partial}{\partial \phi_k}\left(\frac{\partial H_1}{\partial I_k}f_1\right),$$

whence after averaging over t_0, t_1, and ϕ, we obtain a diffusion equation in the form

$$\frac{\partial f_0}{\partial t_2} = \frac{\partial}{\partial I_k}\bar{R}_{kp}(I)\frac{\partial}{\partial I_p}f_0, \qquad (18)$$

where the diffusion coefficients are

$$\bar{R}_{kp}(I)$$
$$= \left\langle\int_0^{t_0} dt_0'\frac{\partial H_1(I, \phi, t_1)}{\partial \phi_k}\frac{\partial H_1(I, \phi - \omega t_0', t_1)}{\partial \phi_p}\right\rangle_{\phi, t_0, t_1}. \qquad (19)$$

Since Hamiltonian H_1 is now a 2π-periodic function of the angular variables ϕ, it may be expanded in a mul-

tiple Fourier series:

$$H_1(I, \phi, t_1) = \sum_n h_n(I, t_1)e^{in\phi}, \quad (20)$$

where

$$n \equiv \{n_1, n_2, n_3\}, \quad n\phi \equiv n_1\phi_1 + n_2\phi_2 + n_3\phi_3.$$

Substituting these expansions into (19) and averaging over ϕ, we obtain

$$\langle R_{kp}\rangle_\phi = \sum_n n_k n_p h_n h_{-n} \int_0^{t_0} dt_0' \exp(-in\omega t_0').$$

Next, we assume that frequencies $\omega_k(I)$ are independent in the sense that the equality

$$n_1\omega_1 + n_2\omega_2 = 0$$

implies $n_{1,2} = 0$. (The set of variables I for which this condition is not satisfied has a zero measure). Integration over t_0' then yields

$$\langle R_{kp}\rangle_\phi = \sum_{n_{1,2}\neq 0} n_k n_p h_n h_{-n} \frac{1}{in\omega}(1 + \exp(-in\omega t_0)) + \sum_{n_{1,2}=0} n_k n_p h_n h_{-n} t_0.$$

This expression contains the secular term

$$\sum_{n_{1,2}=0} n_k n_p h_n h_{-n} t_0.$$

However, this term contributes only to R_{33}; therefore, it may be discarded, because f_0 is independent of $I_3 = m_z$. We then obtain

$$\langle R_{kp}\rangle_\phi = \sum_{n_{1,2}\neq 0} n_k n_p h_n h_{-n} \frac{1}{in\omega}(1 - \exp(-in\omega t_0)) = -\sum_{n_{1,2}\neq 0} n_k n_p h_n h_{-n} \frac{1}{in\omega}\exp(-in\omega t_0).$$

Averaging this equality now over t_0 finally yields

$$\bar{R}_{kp} = 0, \quad \frac{\partial f_0}{\partial t_2} = 0.$$

Thus, we have shown that the presence of an asymmetric term on the order of ε in the gravitational potential does not change initial distribution function (3) for dark matter particles, to within small terms of order ε^2. A more detailed analysis based on consistent perturbation theory indicates that, actually, the deviations of the averaged action variables from their initial values are generally exponentially small if the adiabaticity conditions are satisfied. (For example, for the so-called Arnold diffusion, the mean rate of departure of the variables is on the order of $\exp(1/\sqrt{\varepsilon})$ [11].) This implies that the departure of the action variables cannot be recorded in any order of perturbation theory. Therefore, nonadiabatic processes play a much more important role in the dynamics of dark matter in real galactic systems. They will be considered in the following sections.

4. THE GROWTH OF A BLACK HOLE DUE TO THE DIRECT ABSORPTION OF DARK MATTER

Let a primary BH that begins to grow due to the absorption of dark matter be formed at the NGS center via the contraction of baryonic matter. The flux of dark matter on a BH with mass M_{bh} is composed of particles the angular momenta of which satisfy the inequality

$$m < m_g = 2cr_g, \quad (21)$$

where $r_g = 2GM_{bh}/c^2$ is the gravitational radius of the BH [12]. Their absorption causes the number of particles in the loss cone (21) to decrease. On the other hand, the BH growth causes m_g and, hence, the loss cone to increase. Thus, the law of BH growth is determined by these two competing processes.

Let us determine the conditions under which the BH growth stops and its mass reaches

$$M_{bh} = M_b + M_d,$$

where M_b is the mass of the seed baryonic BH and M_d is the mass of the absorbed dark matter. Since the particle angular momenta are conserved during spherically symmetric evolution, the BH will capture only those particles that are in the loss cone of the final BH with mass M_{bh}. The total mass of these particles can be determined from the initial distribution function f_i by using the equality

$$N(f_i; M_{bh}) = \int d^3r d^3v f_i(\mathbf{r}, \mathbf{v})\theta(m_g - m),$$

where

$$\theta = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases}$$

On the other hand, their total mass must determine M_d; hence

$$M_d = N(f_i; M_b + M_d). \quad (22)$$

Thus, if Eq. (22) has a solution $M_d > 0$, then the BH will stop growing at total mass M_{bh}. Otherwise, the BH will grow indefinitely. However, as we show below, Eq. (22) has a solution in almost all reasonable cases; moreover, the total mass of the captured dark matter particles M_d is much lower than the seed baryonic mass M_b.

As a model example, let us consider the growth of the seed BH for an initial isothermal distribution of dark matter. In this case, the distribution function, the density, and the potential are

$$f_i(E, m) = \frac{\rho_0}{(2\pi\sigma_d^2)^{3/2}}\exp\left(-\frac{E}{\sigma_d^2}\right), \quad (23)$$

$$\rho(r) = \frac{\rho_0}{r^2}, \quad (24)$$

$$\psi_d(r) = 4\pi G\rho_0 \ln r, \quad (25)$$

where σ_d is the dark particle velocity dispersion and $\rho_0 = \sigma_d^2/2\pi G$. The density of the particles with angular momenta lower than m_g determined from the initial distribution (23) is

$$\rho_g(r) = \frac{\rho_0}{r^2}\left(1 - \exp\left(-\frac{m_g^2}{2\sigma_d^2 r^2}\right)\right),$$

whence the total mass of the particle with angular momenta lower than m_g is

$$N_g = 2\pi\left(\frac{\rho_0}{G}\right)^{1/2} m_g. \quad (26)$$

Given (21) and (26), Eq. (22) takes the form

$$\frac{M_d}{M_b} = Q\left(1 + \frac{M_d}{M_b}\right), \quad (27)$$

where

$$Q = \frac{8\pi}{c}(G\rho_0)^{1/2}.$$

It is important to note that the dimensionless parameter Q is always small in real conditions. Indeed, since, for example, the mass of our galactic halo is $M_H \sim 10^{12}M_\odot$ and its radius is $R_H \sim 100$ kpc, we find from (24) and (27) that $Q \sim 10^{-3}$. A similar situation also takes place in other galaxies. Thus, the absorbed dark mass M_d is proportional to the baryonic mass and, according to (27), accounts for about 0.1% of the seed baryonic mass M_b.

It is easy to verify that this conclusion will change only slightly if the distribution function for dark matter differs significantly from the isothermal distribution and is described by formula (3). Indeed, the total mass of the particles with angular momenta lower than m_g is now defined by the expression

$$N_g = \int d^3\phi \int_0^\infty dI_R \int_0^\infty dm \int_{-m}^{m} dm_z \times f_0 I_R^{1/8}\delta(m^2 - l_0^2 I_R^2)\theta(m_g - m);$$

hence

$$N_g = (2\pi)^3 \frac{8}{9} f_0 \left(\frac{m_g}{l_0}\right)^{9/8}. \quad (28)$$

Let us now rewrite Eq. (22) as

$$\frac{M_d}{M_b} = Q\left(1 + \frac{M_d}{M_b}\right)^{9/8}, \quad (29)$$

where

$$Q = Q' M_b^{1/8}, \quad Q' = (2\pi)^3 \frac{8}{9} f_0 \left(\frac{4G}{l_0 c}\right)^{9/8}.$$

The constant f_0 can be determined from the relation

$$M_h = (2\pi)^3 \int_0^\infty dI_R \int_0^\infty dm^2 \times f_0 I_R^{1/8}\delta(m^2 - l_0^2 I_R^2)\theta(m_g - m),$$

where, according to (9), the radial action of the boundary particles I_m is related to the halo radius R_H and mass M_H by

$$I_m \approx \frac{1}{\pi} G^{1/2} M_H^{1/2} R_H^{1/2}.$$

Hence, we obtain

$$f_0 = \frac{9}{8}\frac{\pi^{9/8}}{(2\pi)^3}\frac{M_H^{7/16}}{(GR_H)^{9/16}}. \quad (30)$$

Using this relation, we can show that, under the conditions of our galaxy, parameter Q is still small (on the order of 10^{-2}) and the mass of the dark matter captured by the BH is negligible.

We emphasize that a previous study of this question [13] by other authors led them to assert that the BH grows indefinitely via the direct onflow of dark matter. The cause of the error lay in the fact that the authors disregarded the change in the dark matter distribution

function due to the capture of the particles that fell within the loss cone. As a result, the loss cone was always filled, which caused the BH to grow indefinitely.

Thus, the dynamics of dark matter particles when the angular moments are conserved cannot be responsible for any appreciable BH growth, because the number of particles in the loss cone rapidly decreases and the loss cone itself grows too slowly to provide the influx of new particles. As follows from the results of Section 3, the presence of a small asymmetry in the potential does not lead to any significant change in the dark particle angular momenta and to the filling of the loss cone. The simplest nonadiabatic processes that can lead to the filling of the loss cone are collisions of dark matter particles with stars. This process is considered in following sections.

5. THE KINETIC EQUATION

The distribution function for dark matter particles, $f(\mathbf{r}, \mathbf{v}, t)$, with allowance made for their gravitational interaction with stars satisfies the following kinetic equation:

$$\frac{\partial f}{\partial t} + \{H_0, f\} = \mathrm{St}[f], \tag{31}$$

where H_0 is the Hamiltonian that corresponds to the particle motion in the averaged potential and $\mathrm{St}[f]$ is the collision term. Since the gravitational interaction of dark matter particles with individual stars is identical to the Coulomb interaction, the collision term may be written in Landau form [14]:

$$\mathrm{St}[f] = \frac{\partial}{\partial v_k} W_{kp} \frac{\partial}{\partial v_p} f, \tag{32}$$

where

$$W_{kp} = 2\pi G^2 M_\odot \Lambda \int d^3v' w_{kp} F(\mathbf{v}', r),$$

$$w_{kp} = (u^2 \delta_{kp} - u_k u_p)/u^3,$$

v_k are the components of the dark particle velocity $\mathbf{v}$, $F(\mathbf{v}', k)$ is the distribution function of the stars, $\mathbf{u} = \mathbf{v}' - \mathbf{v}$ is the relative velocity of the stars and dark particles, and $\Lambda \sim 10$ is the gravitational Coulomb logarithm [15].

Actually, as was noted above, the frequency of particle collisions with stars is much lower than the characteristic frequency of their orbital motion. Under these conditions, the kinetic equation (31) can be significantly simplified. To this end, as in Section 3, we rewrite it in the action–angle variables,

$$\frac{\partial f}{\partial t} + \omega_k \frac{\partial f}{\partial \phi_k} = \mathrm{St}_{I,\phi}[f], \tag{33}$$

and take into account the fact that initial distribution function (3) does not depend on the angular variables ϕ. Therefore, its change is attributable only to the collision term (32). However, as was noted above, collisions are rare and $\mathrm{St}[f] \propto \nu f$, where ν is the collision frequency. For this reason, we will again seek a solution of Eq. (33) in the form

$$f = f_0 + \nu f_1,$$

where νf_1 is a small correction to f_0, and both terms f_0 and f_1 depend on the fast, $t_0 = t$, and slow, $t_1 = \nu t$, times. In the zeroth approximation in ν,

$$\frac{\partial f_0}{\partial t_0} + \omega_k \frac{\partial f_0}{\partial \phi_k} = 0,$$

whence, as above, it follows that the main part of the distribution function f_0 does not depend on the angular variables and the fast time. The first approximation yields

$$\nu\left(\frac{\partial f_0}{\partial t_1} + \frac{\partial f_1}{\partial t_0} + \omega_k \frac{\partial f_1}{\partial \phi_k}\right) = \mathrm{St}_{I,\phi}[f_0].$$

Averaging this equation over the angular variables and the fast time and taking into account the fact that the second and third terms become zero after the averaging, we finally obtain

$$\frac{\partial f(I,t)}{\partial t} = \overline{\mathrm{St}}[f(I,t)]. \tag{34}$$

(Here, we omitted the subscript 0 in the distribution function and returned to ordinary time t.)

Thus, under the above assumptions regarding the collision frequency ν, we may seek the distribution function that depends only on the action variables as a solution of kinetic equation (34) with the averaged collision term:

$$\overline{\mathrm{St}}[f] = \frac{1}{(2\pi)^3} \int d^3\phi \, \mathrm{St}_{I,\phi}[f]. \tag{35}$$

The kinetic equation in averaged form (34) was first derived by Budker and Belyaev [16] in studying the dynamics of relativistic particles. In [17–19], this equation was used to study the dynamics of particles and stars in the field of a Coulomb center.

Our objective is to study the dynamics of particles in an arbitrary centrally symmetric field. It is first necessary to derive expression (35) for the averaged collision integral. Following [16], we consider the tensor differential form (23) formally in the six-dimensional space of velocities and coordinates $X = (\mathbf{v}, \mathbf{x})$:

$$\frac{\partial}{\partial v_k} W_{kp} \frac{\partial}{\partial v_k} = \frac{\partial}{\partial X_\mu} W_{\mu\nu} \frac{\partial}{\partial X_\nu}, \quad \mu, \nu = 1, \ldots, 6,$$

by assuming the spatial components of the tensor W to be zero. In this six-dimensional space, we make the canonical transformation to the action–angle variables:

$$X_\mu \mapsto Y_{\mu'}; \quad Y = \{I_R, m, m_z; \phi_1, \phi_2, \phi_3\}.$$

After this transformation, the differential form becomes

$$\frac{\partial}{\partial X_\mu} W_{\mu\nu} \frac{\partial}{\partial X_\nu} = \frac{1}{\sqrt{g}} \frac{\partial}{\partial Y_{\mu'}} \sqrt{g} R_{\mu'\nu'} \frac{\partial}{\partial Y_{\nu'}},$$

$$R_{\mu'\nu'} = \frac{\partial Y_{\mu'}}{\partial X_\mu} \frac{\partial Y_{\nu'}}{\partial X_\nu} W_{\mu,\nu},$$

where $\sqrt{g}$ is the Jacobian. However, the Jacobian of the canonical transformation is known to be unity. Therefore, the differential form in the new variables is

$$\frac{\partial}{\partial X_\mu} W_{\mu\nu} \frac{\partial}{\partial X_\nu} = \frac{\partial}{\partial Y_{\mu'}} R_{\mu'\nu'} \frac{\partial}{\partial Y_{\nu'}}.$$

It should be recalled that we seek a solution of kinetic equation (34) that does not depend on the angular variables ϕ. Hence,

$$\frac{\partial}{\partial Y} R \frac{\partial}{\partial \phi} = 0.$$

Next, when averaged over ϕ, the terms $(\partial/\partial\phi)R(\partial/\partial I)$ also become zero and the expression for the averaged collision integral again takes three-dimensional form:

$$\overline{\mathrm{St}}[f] = \frac{\partial}{\partial I_{k'}} \bar{R}_{k'p'} \frac{\partial}{\partial I_{p'}} f, \tag{36}$$

where

$$\bar{R}_{k'p'} = \frac{1}{(2\pi)^3} \int d^3\phi \frac{\partial I_{k'}}{\partial v_k} \frac{\partial I_{p'}}{\partial v_p} W_{kp}. \tag{37}$$

To obtain the specific differential form on the right-hand side of Eq. (36), we must specify the distribution function of the stars F. Observational data for the dynamics of stars in the bulge suggest that, to a first approximation, the distribution function of the stars may be assumed to be isotropic, i.e., dependent only on the energy

$$E' = \frac{v'^2}{2} + \Psi(r),$$

but not on the angular momentum. For simplicity, we also assume that the distribution function of the stars depends on energy as a power law:

$$F(v', r) = F_0 E'^{-\beta}. \tag{38}$$

It can be shown that the potential produced by stars with distribution (38) also depends on r as a power law:

$$\Psi(r) = \Psi_0 r^\alpha, \quad \alpha = \frac{4}{2\beta - 1}. \tag{39}$$

Distribution function (38) and potential (39) can be determined if we know the stellar velocity dispersion

$$\sigma = \sqrt{\langle v_k^2 \rangle} = \sqrt{\langle v^2/3 \rangle},$$

specified as a function of the distance to the galactic center:

$$\sigma(r) = \sigma_0 r^{\alpha/2}. \tag{40}$$

In this case, the velocity dispersion, together with the parameters σ_0 and α, is uniquely related to the distribution function (38) and potential (39) by

$$\Psi_0 = 3\sigma_0^2 \frac{\int_0^\infty x^2 (1 + x^2/2)^{-\beta} dx}{\int_0^\infty x^4 (1 + x^2/2)^{-\beta} dx},$$

$$F_0 = \frac{\alpha(1+\alpha)}{(4\pi)^2 G \int_0^\infty x^2 (1 + x^2/2)^{-\beta} dx} \Psi_0^{2/\alpha}.$$

For $\alpha = 0$, these relations formally become meaningless. However, as can be shown, stellar distribution function (38) for $\alpha \longrightarrow 0$ transforms from a power-law one to an isothermal one identical in form to (23) with a distance-independent velocity dispersion σ. In general, parameter α is small in the bulge.

Calculations show that the quadratic form W_{kp} from (32) for an isotropic stellar distribution function $F(E')$ becomes

$$W_{kp} = A(E, r)\delta_{kp} - B(E, r)\frac{v_k v_p}{v^2}, \tag{41}$$

where

$$A = \frac{16\pi^2}{3} G^2 M_\odot \Lambda \int_{\Psi(r)}^\infty dE' \times F(E') \begin{cases} 1, & E < E', \\ \frac{3}{2}\frac{v'}{v}\left(1 - \frac{v'^2}{3v^2}\right), & E > E', \end{cases}$$

$$A - B = \frac{16\pi^2}{3} G^2 M_\odot \Lambda \int_{\Psi(r)}^\infty dE' F(E') \begin{cases} 1, & E < E', \\ \frac{v'^3}{v^3}, & E > E'. \end{cases}$$

Another factor that simplifies our calculations is that, as follows from (9), the radial action of the particles moving in potential (39) can be accurately fitted by the following simple relation:

$$I_R(E, m) \approx J(E) - b_\alpha m, \tag{42}$$

where b_α is a positive constant and

$$J(E) = \begin{cases} \dfrac{\sqrt{2}\displaystyle\int_0^1 \sqrt{1-x^\alpha}\,dx}{\pi\Psi_0^{1/\alpha}} E^{1/\alpha+1/2}, & \alpha > 0, \\ \dfrac{\sigma}{\sqrt{\pi}} \exp\left(\dfrac{E}{2\sigma^2}\right), & \alpha = 0. \end{cases}$$

(Note that equality (42) is exact and $b = 1$ for Coulomb and oscillatory potentials [10].)

In view of (42), it seems natural to make a linear change of variables,

$$I_R, m, m_z \longrightarrow J, m, m_z.$$

In this case, expression (37) for the tensor will not change, because the transformation is linear. Therefore, below, by the variables I in (36) and (37) we mean J, m, and m_z.

Now, note that initial distribution function (3) has the same form in the new variables, because the coefficient l_0 (4) is small:

$$f(J, m) = f_0 J^{1/8} \delta(m^2 - l_0^2 J^2). \tag{43}$$

Since the initial function (43) does not depend on the variables m_z, a solution of Eq. (36) can also be sought in the form of a function of only J and m. Taking this fact into account and calculating the coefficients of quadratic form (37), we may write the collision term using (41) as

$$\begin{aligned} \overline{\mathrm{St}}[f] = {} & \frac{1}{m}\frac{\partial}{\partial m} m\left(\bar{R}_{22}\frac{\partial f}{\partial m} + \bar{R}_{12}\frac{\partial f}{\partial J}\right) \\ & + \frac{\partial}{\partial J}\left(\bar{R}_{12}\frac{\partial f}{\partial m} + \bar{R}_{11}\frac{\partial f}{\partial J}\right), \end{aligned} \tag{44}$$

where

$$\bar{R}_{11} = \left(\frac{\partial J}{\partial E}\right)^2 \langle (A-B) v^2 \rangle_\phi,$$

$$\bar{R}_{12} = \left(\frac{\partial J}{\partial E}\right) \langle (A-B) m \rangle_\phi,$$

$$\bar{R}_{22} = \left\langle Ar^2 - \frac{B}{v^2} m^2 \right\rangle_\phi.$$

We represent the operation of averaging over the angular variables as

$$\langle \ldots \rangle_\phi = \frac{2}{T(E, m)} \int_{r_-}^{r_+} \frac{dr}{v_r} (\ldots),$$

where

$$T = 2\int_{r_-}^{r_+} \frac{dr}{v_r}$$

is the radial oscillation period of the dark matter particles and

$$v_r = \sqrt{2}\left(E - \Psi(r) - \frac{m^2}{2r^2}\right)^{1/2}$$

is the radial velocity. Next, since the angular momenta of dark matter particles are low, we will calculate the coefficients $\bar{R}_{ab}(J, m)$ in (44) at $m = 0$ in the first approximation. In addition, we see from the form of initial distribution function (43) that for moderately long times, as long as the distribution function has a sharp maximum at $m = l_0 J$, $l_0 \ll 1$, the diffusion in J may be ignored compared to the diffusion in m. Thus, we write the averaged kinetic equation (34) as a diffusion equation only in the space of angular momenta m:

$$\frac{\partial f(J, m, t)}{\partial t} = R(J)\frac{1}{m}\frac{\partial}{\partial m} m \frac{\partial}{\partial m} f(J, m, t), \tag{45}$$

where the diffusion coefficient

$$R = \bar{R}_{22}\big|_{m=0} = \langle Ar^2 \rangle_\phi$$

calculated from the distribution of stars with the velocity dispersion (40) is

$$R(J) = 0.46 G M_\odot \Lambda \sigma_0^{2/(2+\alpha)} J^{\alpha/(2+\alpha)}. \tag{46}$$

For an isothermal distribution of stars in the bulge, $\alpha = 0$ and the diffusion coefficient R is a J-independent constant. We emphasize that the possibility of passing to the diffusion equation only in angular momentum space is attributable to the statement of the problem and to the singularity of initial distribution function (3) containing the small parameter l_0.

6. THE FLUX OF DARK MATTER ON A BLACK HOLE

Through the direct capture of the particles that fell into the loss cone of a seed BH, the distribution func-

tion of the dark matter particles in a time on the order of one period of their orbital motion takes the form

$$f(J, m, 0) = f_0 J^{1/8} \delta(m^2 - l_0^2 J^2)\theta(m - m_g), \quad (47)$$

which differs from (43) by the factor $\theta(m - m_g)$. The diffusion of dark matter particles in the angular momentum space followed by their fall into the loss cone (21) and absorption by the BH is described by diffusion equation (45) with initial condition (47) and the boundary condition

$$f|_{m = m_g} = 0. \quad (48)$$

The solution of the diffusion equation may be represented as

$$f(J, m, t) = \int_{m_g}^{\infty} dm_1 G(J, m, m_1, t) f(J, m, 0), \quad (49)$$

where G is the Green function of boundary-value problem (48):

$$G = \int_0^{\infty} d\lambda m_1 Z_\lambda(m_1, m_g) Z_\lambda(m, m_g) \exp(-\lambda R(J) t);$$

Z is the orthonormalized system of eigenfunctions for boundary-value problem (48):

$$Z_\lambda(m, m_g) = \frac{J_0(\sqrt{\lambda} m_g) N_0(\sqrt{\lambda} m) - N_0(\sqrt{\lambda} m_g) J_0(\sqrt{\lambda} m)}{(J_0^2(\sqrt{\lambda} m_g) + N_0^2(\sqrt{\lambda} m_g))^{1/2}};$$

and J_0 and N_0 are the zero-order Bessel and Neumann cylinder functions. Let us now calculate the flux of dark matter onto a BH. Let D be the region in phase space specified by the inequalities

$$I_R = J - b_\alpha m \geq 0, \quad m \geq m_g, \quad -m \leq m_z \leq m.$$

The total mass of the dark matter is defined by the integral

$$N(t) = \int_D d^3 I d^3 \phi f(J, m, t).$$

Next, it follows from the conservation of total mass that the flux through the boundary is

$$S = -dN/dt.$$

Using Eq. (32), we obtain

$$S(t) = -\int_D d^3 I d^3 \phi \frac{\partial}{\partial I_k} \bar{R}_{kp} \frac{\partial}{\partial I_p} f(J, m, t).$$

Transforming the volume integral into a surface integral using the Stokes formula and taking into account the fact that dark matter flows only through the $m = m_g$ surface, we obtain

$$S(t) = 2(2\pi)^3 \int_{b_\alpha m_g}^{\infty} dJ R(J) m_g \frac{\partial}{\partial m} f(J, m, t)\Big|_{m = m_g}. \quad (50)$$

Substituting the solution of diffusion equation (49) into (50) yields

$$S(t) = 2\frac{(2\pi)^3}{\pi} f_0 R_\alpha \int_{m_g/l_0}^{\infty} dJ J^{\gamma + \frac{1}{8}} \int_0^{\infty} d\lambda \times \exp(-\lambda R_\alpha J^\gamma t) \frac{Z_\lambda(l_0 J, m_g)}{(J_0^2(\sqrt{\lambda} m_g) + N_0^2(\sqrt{\lambda} m_g))^{1/2}},$$

where

$$\gamma = \frac{\alpha}{2 + \alpha}, \quad R_\alpha = 0.46 \Lambda G M_\odot \sigma_0^{2/(2+\alpha)}. \quad (51)$$

Next, it is convenient to introduce the normalized time

$$T = l_0^{-\gamma} R_\alpha t$$

and make the following changes

$$\lambda \longrightarrow \eta = \lambda T^{2/(2-\gamma)}, \quad J \longrightarrow y = l_0 J T^{-1/(2-\gamma)},$$

$$m_g \longrightarrow x = m_g T^{-1/(2-\gamma)}.$$

The expression for the flux then takes the form

$$S(t) = 2\frac{(2\pi)^3}{\pi} \frac{f_0}{l_0^{9/8+\gamma}} \frac{R_\alpha}{T^\zeta} \Phi_\gamma(x), \quad (52)$$

$$\zeta = \frac{7 - 8\gamma}{16 - 8\gamma}, \quad \Phi_\gamma(x) = \int_x^{\infty} dy H_\gamma(x, y),$$

$$H_\gamma(x, y) = y^{1/8+\gamma} \int_0^{\infty} d\eta \exp(-\eta y^\gamma) \times \frac{J_0(\sqrt{\eta} x) N_0(\sqrt{\eta} y) - N_0(\sqrt{\eta} x) J_0(\sqrt{\eta} y)}{J_0^2(\sqrt{\eta} x) + N_0^2(\sqrt{\eta} x)}.$$

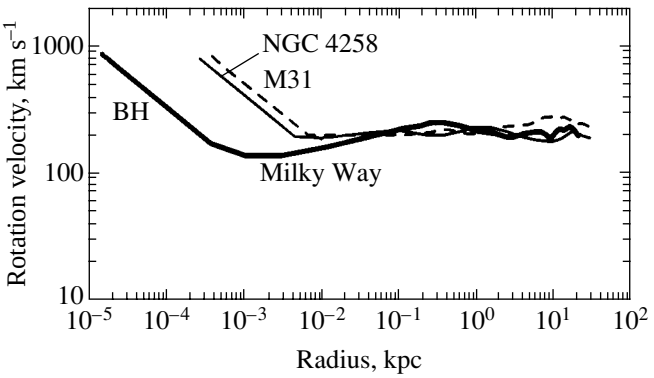

Fig. 2. Rotation curves for our galaxy, M 31, and NGC 4258 (according to [20]).

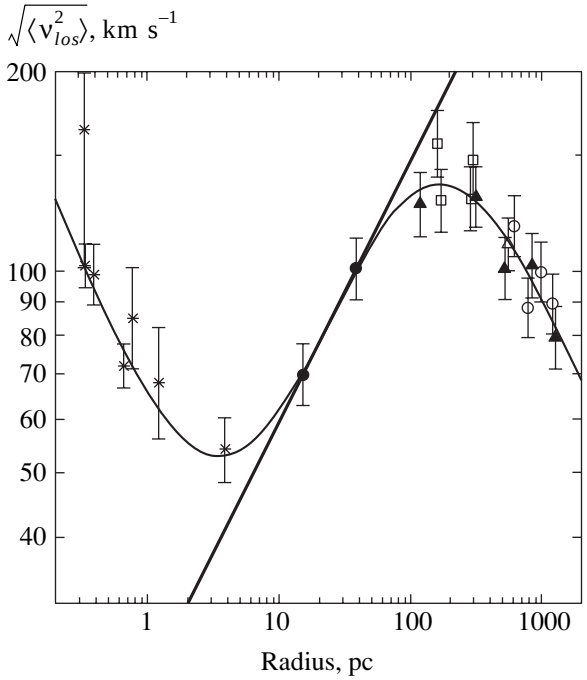

Fig. 3. Stellar velocity dispersion versus distance to the our galactic center (according to [21]).

It can be shown that $H_\gamma(x, y)$ is a finite, positively defined function that becomes exactly equal to zero at $y = x$ and exponentially small for $y > x + 4$. Therefore, we may assert that the flux of dark matter onto a BH at time t is composed of particles with a radial action in the range

$$\frac{m_g}{l_0} < J < \frac{m_g}{l_0} + \frac{4}{l_0} T^{1/(2-\gamma)}. \tag{53}$$

Thus, the region of space J from where dark matter flows grows with time as $t^{1/(2-\gamma)}$. The following estimate is valid for the function Φ in the total flux (52):

$$\Phi_\gamma(x) \approx x^{1/8+\gamma/2}.$$

Using this estimate, we finally obtain an expression for the flux of dark matter onto a BH through the $m = m_g$ boundary of the loss cone:

$$S(t) = 2\frac{(2\pi)^3}{\pi} \frac{f_0}{l_0^{9/8+\gamma}} R_\alpha^{1/2} \frac{m_g^{1/8+\gamma/2}}{t^{1/2}}, \tag{54}$$

Expression (54) was derived by assuming that the m_g boundary of the loss cone was time-independent. Actually, the boundary moves together with the total BH mass. However, it follows from (54) that the flux depends weakly on m_g. Therefore, expression (54) may be used to allow for the motion of the boundary as the BH grows by formally assuming that m_g is time-dependent:

$$m_g(t) = 4GM_{bh}(t)/c.$$

Thus, we write the BH growth law as

$$\frac{dM_{bh}}{dt} = 2\frac{(2\pi)^3}{\pi} \left(\frac{4G}{c}\right)^{1/8+\gamma/2} \frac{f_0}{l_0^{9/8+\gamma}} R_\alpha^{1/2} \frac{M_{bh}^{1/8+\gamma/2}}{t^{1/2}}.$$

Assuming the mass of the seed BH to be low, we obtain the following solution of this equation:

$$M_{bh} = Ct^{4/(7-4\gamma)}, \tag{55}$$

$$C = \left(\left(\frac{7-4\gamma}{2}\right)\frac{(2\pi)^3}{\pi}\left(\frac{4G}{c}\right)^{1/8+\gamma/2} \frac{f_0}{l_0^{9/8+\gamma}} R_\alpha^{1/2}\right)^{8/(7-4\gamma)},$$

where relation (30) may be used to estimate the parameter f_0. Thus, the BH grows via the absorption of dark matter scattered during collisions with stars as a power law:

$$M_{bh} \propto t^a, \quad a \approx 4/7.$$

7. CONCLUSIONS

In conclusion, let us compare the results of the theory presented above with observational data. In Fig. 2, stellar rotation velocities are plotted against distance to the centers of our galaxy and the galaxies M 31 and NGC 4258 [20]. We see that the distribution of stars in the bulges of M 31 and NGC 4258 is nearly isothermal; i.e., the stars have an almost constant, distance-independent orbital velocity (and, hence, dispersion), $\sigma = \sigma_0 \sim 200$ km s^{-1}, up to the region of BH influence ($r <$ 4–7 pc). Assuming the mass and size of the dark matter halos around these galaxies to be approximately equal, $10^{12} M_\odot$ and 100 kpc, respectively, we obtain $f_0 \approx 7.3 \times 10^8$ g s$^{9/8}$ cm$^{-9/4}$ and $l_0 \approx 0.1$ from (30) and (5). We have $\alpha = \gamma = 0$ and $R_\alpha \approx 6.1 \times 10^{33}$ cm^4 s^{-3} from (40) and (51). Next, assuming that the BH age t is comparable to the age of the Universe, $t \approx 3 \times 10^{17}$ s, we obtain the BH mass from (55), $M_{bh} \approx 5 \times 10^7 M_\odot$. Thus, the theoretical

BH masses for these galaxies are close to their observed values:

$$M_{bh} = 4.5(2.0\text{–}8.5) \times 10^7 M_\odot \text{ for M 31},$$

$$M_{bh} = 3.9(3.8\text{–}4.0) \times 10^7 M_\odot \text{ for NGC 4258}.$$

In the central region of the galactic bulge, the assumption of an isothermal stellar distribution at $r \leq 100$ pc is invalid. At the same time, as we see from (53), dark matter flows mainly from the central region. Stellar velocity dispersion is plotted against distance in Fig. 3 [21]. We see that on a scale from 10 to 100 pc, the dispersion may be roughly represented as

$$\sigma = 60 \text{ km/s}\left(\frac{r}{10 \text{ pc}}\right)^{0.3}.$$

It then follows from (40) that $\alpha \approx 0.6$. In this case, formula (55) yields $2 \times 10^7 M_\odot$ for the BH mass, which is much higher than the observed value of $M_{bh} \approx 2.6 \times 10^6 M_\odot$. More complete information about the stellar distribution function, its evolution, and the evolution of the dark matter distribution in the central region of the bulge is required to estimate the BH mass more accurately. The contribution from the BH itself to the total gravitational potential can also play an important role. Nevertheless, in general, we see that even in a rough approximation, allowance for the absorption of dark matter alone by a seed BH makes it possible to obtain a reasonable estimate for the observed masses of a large number of giant BHs (see Fig. 1). This fact may be considered as a further confirmation of the general theory for the large-scale structure of the Universe [5] based on the assumption about the corpuscular nature of dark matter.

Note also that the general kinetic theory developed in Section 5 allows us to also describe the absorption of baryonic matter or, more precisely, the capture of the stars themselves by the BH in the central region of the bulge via their gravitational scattering. This process is extremely important, in particular, for giant BHs with $M_{bh} \sim 10^9 M_\odot$ and active galactic nuclei.

It should be noted that another important process described by kinetic equation (34) is the possible decrease in the amount of dark matter in the bulge. The change in the energy of dark matter particles via their collisions with stars leads to their heating and expulsion from the bulge (a phenomenon similar to Fermi acceleration) and, hence, to a decrease in the dark matter density compared to the baryonic density. Qualitatively, this result agrees with the observations in recent years discussed in detail in [22, 23].

Further development of the consistent kinetic theory discussed above and its detailed comparison with observational data will undoubtedly provide insight into the main physical processes in galactic nuclei.

ACKNOWLEDGMENTS

We thank V.I. Sirota and M.I. Zelnikov for numerous and helpful discussions. This work was supported by the Program for Basic Research of the Presidium of the Russian Academy of Sciences "Mathematical Methods in Nonlinear Dynamics." K.Z. is grateful to the Foundation for Support of Science for financial support.

REFERENCES

1. A. M. Cherepashchuk, Usp. Fiz. Nauk **173**, 345 (2003) [Phys. Usp. **46**, 335 (2003)].
2. L. C. Ho, in *Observational Evidence for Black Holes in the Universe*, Ed. by S. K. Chakrabarti (Kluwer Academic, Dordrecht, 1999), Astrophys. Space Sci. Library, Vol. 234.
3. A. V. Gurevich and K. P. Zybin, Zh. Éksp. Teor. Fiz. **94**, 3 (1988) [Sov. Phys. JETP **67**, 1 (1988)].
4. A. V. Gurevich and K. P. Zybin, Zh. Éksp. Teor. Fiz. **97**, 20 (1990) [Sov. Phys. JETP **70**, 10 (1990)].
5. A. V. Gurevich and K. P. Zybin, Usp. Fiz. Nauk **165**, 723 (1995) [Phys. Usp. **38**, 687 (1995)].
6. J. F. Navaro, C. S. Frenk, and C. D. White, Astrophys. J. **462**, 563 (1996).
7. L. Fukshige and J. Makino, Astrophys. J. **477**, L9 (1997).
8. M. Persic and P. Salucci, Astrophys. J., Suppl. **99**, 501 (1995).
9. D. Puche and C. Carignan, Astrophys. J. **378**, 487 (1991).
10. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 1: *Mechanics*, 4th ed. (Nauka, Moscow, 1988; Pergamon, New York, 1988).
11. V. I. Arnol'd, *Mathematical Methods of Classical Mechanics*, 3rd ed. (Nauka, Moscow, 1989; Springer, New York, 1989).
12. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 6th ed. (Nauka, Moscow, 1973; Pergamon Press, Oxford, 1975).
13. J. D. MacMillan and R. N. Henriksen, astro-ph/0201153.
14. L. D. Landau, Zh. Éksp. Teor. Fiz. **7**, 203 (1937).
15. A. M. Fridman and V. L. Polyachenko, *Physics of Gravitating Systems* (Nauka, Moscow, 1976; Springer, New York, 1984).
16. G. I. Budker and S. T. Belyaev, in *Plasma Physics and Problems of Controlled Thermonuclear Reactions* (Gostekhizdat, Moscow, 1958), Vol. 2.
17. A. V. Gurevich, Geomagn. Aeron. **4**, 247 (1964).
18. V. I. Dokuchaev and L. M. Ozernoĭ, Zh. Éksp. Teor. Fiz. **73**, 1587 (1977) [Sov. Phys. JETP **46**, 834 (1977)].
19. A. P. Lightman and S. L. Shapiro, Astrophys. J. **211**, 244 (1977).
20. Y. Sofue and V. Rubin, astro-ph/0010594.
21. S. Tremaine *et al.*, astro-ph/0203468.
22. N. Napolitano, M. Arnaboldi, and M. Capaccioli, Astron. Astrophys. **383**, 791 (2002).
23. F. Combes, astro-ph/0206126.

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 98, No. 1, 2004, pp. 102–122.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 1, 2004, pp. 114–135.

SOLIDS
Structure

Transport of a Two-Component Mixture in One-Dimensional Channels

V. D. Borman, V. N. Tronin, I. V. Tronin*, and V. I. Troyan
Moscow Institute of Engineering Physics (State University), Moscow, 115409 Russia
**e-mail: ivt@rbcmail.ru*
Received June 25, 2003

Abstract—The transport of a two-component gas mixture in subnanometer channels is investigated theoretically for an arbitrary filling of channels. Special attention is paid to consistent inclusion of density effects, which are associated both with the interaction and with a finite size of particles. The analysis is carried out using the hard-sphere model, in which the interaction is manifested as the effective (dynamic) attraction of particles, leading to their correlation. The adsorption isotherm is calculated and the ground state of the mixture in one-dimensional channels is investigated. It is shown, using the density functional method, that the two-component mixture in channels with increasing degree of filling is transformed into a spatially inhomogeneous state. This gives rise to short-lived clusters with size and lifetime increasing with the degree of channel filling. The description of transport in subnanometer channels is reduced in this case to the description of diffusion in a spatially inhomogeneous high-density one-dimensional system. The transport of particles in a medium with short-lived clusters occurs as a collective effect of the barrier-free transfer of density excitation. It is shown that, for high fill factors, the two-component mixture acquires a new property: clusters with a definite size are stabilized in channels due to effective attraction emerging between particles. The lifetime of formed clusters increases exponentially in accordance with the Arrhenius law; at a low temperature, channels with such clusters might be blocked to transport of particles of the mixture. The dependences of fluxes on the mixture composition (degree of filling) and pressure obtained theoretically are in good agreement with the experimentally observed regularities. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Molecular transport in subnanometer channels attracts considerable attention both from the standpoint of fundamental science [1–8] and in connection with numerous applications of membrane technologies and nanotechnologies [9–13]. Among other things, this is due to the anomalously high coefficients of gas separation observed for synthesized ceramic membranes of complex oxides (zeolites); the channels formed in the atomic structure of these compounds have a diameter from 0.3 to 1.4 nm [10]. Such membranes form the basis of new technologies for separation, processing, and utilization of substances [13]. Transport of molecules in subnanometer channels of diameter $d < 1$ nm (to be more precise, smaller than double the diameter of molecules) with walls impenetrable to molecules can take place only along the channel axis and is one-dimensional [3].

It is well known that an increase in the density of particles (fill factor θ) in a channel in one-dimensional (1D) systems does not lead to a phase transition to the condensed state [14] and a nucleus of a new phase does not appear in the system. At the same time, the state of a high-density 1D system of particles is characterized by the emergence of strong density fluctuations, the lifetime and the size of short-lived clusters formed in the system increasing with θ. According to the results obtained in our previous publication [3], the transport of particles occurs as a collective process of barrier-free transport of density excitation. This leads to an increase in the flux j and the diffusion coefficient D upon an increase in the fill factor. Such a mechanism makes it possible to explain the dependences of D and j on the degree of channel filling with a one-component gas, which are observed for various molecules. It turned out, however, that the dependence of transport selectivity on the pressure in a two-component gas mixture is nonmonotonic: the selectivity attains its maximal value and then decreases instead of increasing. Thus, the mechanism of diffusion enhancement in a 1D channel, which was proposed in [3], is inapplicable for two-component mixtures.

The transport of a two-component mixture in subnanometer channels was considered earlier [15, 16] on the basis of the generalized phenomenological Stephan–Maxwell equation. The authors of these publications used the dependence of the chemical potential on the fill factor, taking into account the finite size of particles, but disregarding their interaction. The dependences obtained in [15, 16] also indicate a monotonic increase in the diffusion coefficients, fluxes, and selectivity upon an increase in θ and, hence, fail to describe the experimental data for two-component mixtures.

Here, we analyze the transport of a two-component gas mixture in subnanometer channels theoretically for an arbitrary degree of channel filling. The main problem in this case is to consistently take into account the density effects associated with both the interaction and the finite size of the particles. This is done in the hard-sphere model, in which the interaction is manifested as an effective (dynamic) attraction of particles, leading to their correlation [17]. It is well known [17, 18] that, applying this model to a 3D system, one can describe the density effects qualitatively and even quantitatively in some cases. Comparison with experimental data with the theoretical dependences obtained, which is carried out in Section 4, demonstrates the possibility of quantitative description of diffusion of a two-component mixture in a 1D system.

In Section 2, the adsorption isotherm of a mixture in 1D channels is calculated; this isotherm relates the fill factor and the concentration of particles in channels to the temperature, pressure, and composition of a mixture of gases whose diffusion has been studied in experiments [15, 16, 19]. The ground state of a mixture of particles in 1D channels is analyzed by the density functional method [20] generalized to the case of a two-component mixture. However, in contrast to [20], the free energy is obtained by directly calculating the correlation function and the response function using the method developed for 1D systems [21]. The density functional method is used to derive, from microscopic considerations, the equation of motion for the order parameter of the system, which is the Fourier component of the deviation of the particle concentration from its mean value. Analysis of this equation shows that the two-component mixture in channels is transformed to a spatially inhomogeneous state upon an increase in θ. As a result, short-lived clusters appear in channels, the cluster size and lifetime increasing with θ. A new and unexpected result obtained in this case is the emergence of a minimum in the fluctuating part of the free energy as a function of the wave vector and the Fourier component of the order parameter. Thus, at high filling levels, the two-component mixture acquires a new property: clusters of a definite size are stabilized by a potential barrier due to the emergence of effective attraction between particles in the channels. Such a situation is typical of transitions of the system to an inhomogeneous state [22]. The lifetime of clusters formed increases exponentially in accordance with the Arrhenius law; at a low temperature, channels with such clusters might be blocked to the transport of particles forming the mixture. It should be noted that the idea of stabilizing the clusters of one component by the other component in 1D systems was put forth in [7]. Thus, the description of transport in nanochannels is reduced to the description of diffusion in a spatially inhomogeneous high-density 1D system.

For a weakly nonequilibrium system, the problem of computing fluxes in a 1D channel is reduced to the calculation of relaxation frequency spectra for density fluctuations of mixture components (Section 3). Spectra $\omega_i(k)$ were determined by using the response functions derived in Section 2. Analysis of the dependences of ω_i on k and on fill factor θ proved that a hydrodynamic (acoustic) spectrum $\omega_i(k) = c_i k$ is typical of clusters ($\theta \sim 1$, finite values of k), while for transport of excitations over distances much longer than the characteristic size of clusters ($k \ll 1$), the spectrum is of the diffusion type, $\omega_i(k) \propto D_i k^2$, for an arbitrary fill factor. In accordance with the dependences of spectra and fluxes on the fill factor obtained here, three regimes of particle transport can be singled out. For $\theta \ll 1$, we have diffusion of solitary particles. As the value of θ increases, the flux and the diffusion coefficient increase due to the barrier-free (hydrodynamic) transport of particles along the increasing part of the length of a 1D channel filled with the short-lived clusters formed in it. As the value of θ increases further and the potential barrier (E_i) stabilizing the clusters arises, particle fluxes decrease exponentially since the value of E_i increases with θ. Thus, an increase in the fill factor gives rise to the new property of 1D two-component systems (transport being blocked by the clusters formed).

A comparison of the theory with experimental data and discussion of results can be found in Section 4. The obtained dependences of the fluxes on the mixture composition (fill factor) and of the selectivity on pressure successfully describe the experimental data known to the authors.

2. GROUND STATE OF A SYSTEM IN 1D CHANNELS

Let us consider the surface of a porous body in contact with a two-component gas mixture at temperature T and pressure P. Suppose that N adsorption centers are located on the surface. We assume that the particles on the outer surface do not interact with one another. We also assume that the energy of a gas molecule on the surface is equal to $\varepsilon_0^{(i)}$, $i = 1, 2$, depending on the species of the molecule. We also suppose that κ cylindrical channels ($\kappa \gg 1$) of diameter d and length L emerge at the surface. We assume that the diameter of a cylinder is comparable to the maximal diameter of the gas molecule. Let us assume that $\varepsilon_1^{(i)}$ is the binding energy of the ith particle at the mouth of a channel, $N_1^{(i)}$ is the number of particles of the ith species above the membrane, $q^{(i)}$ is the total number of particles of the ith species in the channel, $n^{(i)}$ is the total number of ith particles in the channel and on the surface, and N_0 is the number of "seats" in the channel. Then the partition function of the grand canonical ensemble taking into

account the interaction of gas particles in the channel has the form

$$\Theta = \sum_{i=1}^{2}\sum \frac{(N-k)!\exp[\beta\varepsilon_0^{(i)}(N_1^{(i)}-n^{(i)})]}{(N_1^{(i)}-n^{(i)})![N-k-(N_1^{(i)}-n^{(i)})]!} \times \frac{\kappa!\exp[\beta\varepsilon_1^{(i)}(n^{(1)})-q^{(1)}]}{(n^{(1)}-q^{(1)})![\kappa-(n^{(1)}-q^{(1)})]!} \times \frac{N_0!\exp(\beta\varepsilon_2^{(i)}q^{(i)})}{q^{(i)}!(N_0-q^{(i)})!}\exp(\beta\mu N_1^{(i)})Z_{\text{int}}(q^{(1)}, q^{(2)}), \quad N_1^{(i)} + q^{(i)} = n^{(i)}. \tag{2.1}$$

Here, the second sum corresponds to summation over configurations, $\beta = T^{-1}$, $\varepsilon_2^{(i)}$ is the binding energy of particles of the ith component in the channel, μ is the chemical potential, and $Z_{\text{int}}(q^{(1)}, q^{(2)})$ is the partition function corresponding to the inclusion of particle interaction in the channel. Considering that the main contribution to the partition function comes from states with a large number of particles ($q^{(i)} \gg 1$), we can replace partition function $Z_{\text{int}}(q^{(1)}, q^{(2)})$ by the partition function for particles interacting in the channel, calculated for the mean value $\bar{q}^{(i)}$ of the number of particles of each species in the channel. From the standpoint of physics, this corresponds to "averaging over channels," when a single channel (whose state is calculated by averaging the parameters of particles in κ channels) is considered instead of the large number of channels. It should be noted that this approximation is possible since the number of channels having a diameter of $d <$ 1 nm and emerging on 1 cm^2 of the zeolite membrane surface is large: $\kappa_1 \sim l_c^{-2} \sim 10^{14}$ cm$^{-2} \gg 1$, where l_c is the characteristic distance between the channels on the membrane surface [23]. Using relation (2.1) and following the method proposed in [3], we obtain for the adsorption isotherm

$$\frac{a}{\sigma_{\text{av}}}\theta = c\frac{p}{p+p_0^{(1)}} + (1-c)\frac{p}{p+p_0^{(2)}}, \quad \frac{a}{\sigma_{\text{av}}}c_1\theta = c\frac{p}{p+p_0^{(1)}}, \quad \frac{a}{\sigma_{\text{av}}}c_2\theta = (1-c)\frac{p}{p+p_0^{(2)}}, \quad p_0^{(i)} = \frac{1}{1-c}\left(\frac{1-c}{c}\right)^c\left(\frac{\alpha^{(2)}}{\alpha^{(1)}}\right)^c\frac{1}{\alpha^{(2)}}\exp[-\beta(\varepsilon_2^{(i)}+F_{\text{int}})], \quad i = 1, 2, \quad \alpha^{(i)} = \frac{1}{T}\left(\frac{2\pi\hbar^2}{m^{(i)}T}\right)^{3/2}, \quad i = 1, 2, \tag{2.2}$$

where $a = L/N_0$ is the mean distance between the seats in a channel, c is the concentration of the first component in the gas phase above the membrane, c_1 and c_2 are the concentrations of the components in a channel, σ_1 and σ_2 are the diameters of molecules of the first and second species, $\sigma_{\text{av}} \equiv (\sigma_1 + \sigma_2)/2$ is the average diameter, θ is the fill factor of the channel, and F_{int} has the meaning of the free energy of interaction per gas particle in the channel. The concentrations of the components in the channel are chosen so that the following relation is satisfied:

$$c_1 + c_2 = 1. \tag{2.3}$$

In accordance with relations (2.2), calculation of the adsorption isotherm is reduced to calculation of the value of F_{int}

$$F_{\text{int}}(\theta, T) = -T\ln(Q_N/L^N), \tag{2.4}$$

where

$$Q_N = \int\ldots\int \exp\left[-\beta\sum_{\substack{i=1\\ w=1,2}}^{N} U(x_i^w)\right]\prod_{\substack{i=1\\ w=1,2}}^{N} dx_i^w, \tag{2.5}$$

$U(x_i^w)$ is the potential energy of interaction of particles of species w separated by distance x from each other.

It is convenient for subsequent computations to introduce, instead of coordinates x_i of individual particles in a channel, "pair" coordinates $\zeta_i^{lm} \equiv |x_{i+1}^m - x_i^l|$, where $l = 1$, $m = 1$ corresponds to the location of two particles of the first species in the vicinity of point x_i; $l = 2$, $m = 2$ corresponds to the same for particles of the second species, and so on. Obviously, the coordinates are connected via the relation

$$\sum_{i=1}^{N_{11}}\zeta_i^{11} = L_1, \quad \sum_{i=1}^{N_{12}}\zeta_i^{12} = L_2, \quad \sum_{i=1}^{N_{22}}\zeta_i^{22} = L_3. \tag{2.6}$$

Here, N_{ij} has the meaning of the total number of pairs of closely spaced particles of species i and j and L_1 (L_2, L_3) is the effective length "occupied" by all pairs N_{11} (N_{12}, N_{22}) of particles. Considering that $L_1 + L_2 + L_3 = L$ and, as a consequence, the impossibility of mutual "hopping" of particles over one another, we can reduce integration with respect to coordinates x_i to integration with respect to coordinates ζ_i^{lm}, which corresponds to summation over all possible configurations of pairs of particles of the first and second species in the

channel. Consequently, we obtain the following expression for Q_N from Eq. (2.5):

$$Q_N = \int_0^\infty \dots \int_0^\infty \delta\left(\sum_{i=1}^{N_{11}} \zeta_i^{11} - L_1\right) \times \delta\left(\sum_{i=1}^{N_{12}} \zeta_i^{12} - L_2\right) \times \delta\left(\sum_{i=1}^{N_{22}} \zeta_i^{22} - L_3\right) \delta(L_1 + L_2 + L_3 - L) \times \exp\left[-\beta \sum_{\substack{i \\ k,m=1,2}} U(\zeta_i^{km})\right] \prod_i d\zeta_i^{11} d\zeta_i^{12} d\zeta_i^{22}. \quad (2.7)$$

Using the delta function representation in the form of a line integral [21] and carrying out integration with respect to ζ_i^{lm} independently for different values of indices l and m, we obtain

$$Q_n = \frac{1}{16\pi^4} \oint \dots \oint \exp(S_1 L_1) \times \left[\int_0^\infty \exp[\beta U(\zeta_i^{11}) - S_1 \zeta_i^{11}] \prod_i d\zeta_i^{11}\right]^{N_{11}} \times \exp(S_2 L_2) \left[\int_0^\infty \exp[\beta U(\zeta_i^{12}) - S_2 \zeta_i^{12}] \prod_i d\zeta_i^{12}\right]^{N_{12}} \times \exp(S_3 L_3) \left[\int_0^\infty \exp[\beta U(\zeta_i^{22}) - S_3 \zeta_i^{22}] \prod_i d\zeta_i^{22}\right]^{N_{22}} \times \exp\{S[L - (L_1 + L_2 + L_3)]\} dS_1 dS_2 dS_3 dS. \quad (2.8)$$

Integrating this relation with respect to ζ_i^{lm} for the simplest form of the intermolecular interaction potential of the hard-sphere type,

$$U(\xi_i^{11}) = \begin{cases} 0, & \zeta_i^{11} > \sigma_1, \\ \infty, & \zeta_i^{11} \le \sigma_1, \end{cases}$$

$$U(\xi_i^{12}) = \begin{cases} 0, & \zeta_i^{12} > \sigma_{\text{av}}, \\ \infty, & \zeta_i^{12} \le \sigma_{\text{av}}, \end{cases} \quad (2.9)$$

$$U(\xi_i^{22}) = \begin{cases} 0, & \zeta_i^{22} > \sigma_2, \\ \infty, & \zeta_i^{22} \le \sigma_2, \end{cases}$$

we obtain

$$Q_N = \frac{1}{4\pi N} \Gamma(Nc_1^2 + 1) \Gamma(Nc_1c_2 + 1) \Gamma(Nc_2^2 + 1) \times [L - N(\sigma_1 c_1^2 + \sigma_2 c_2^2 + 2\sigma_{\text{av}} c_1 c_2)]^N. \quad (2.10)$$

Here, we have used the definition $N_{ij} = Nc_ic_j$ (N is the total number of particles in a channel) and the representation of a factorial in terms of the gamma function, $N! = \Gamma(N + 1)$. Eliminating the factor corresponding to an ideal gas from relation (2.10) and substituting the result into Eq. (2.4), we obtain the following expression for F_{int}:

$$F_{\text{int}} = -T \ln(1 - C\theta), \quad C \equiv \frac{1}{\sigma_{\text{av}}}(\sigma_1 c_1^2 + \sigma_2 c_2^2 + 2\sigma_{\text{av}} c_1 c_2). \quad (2.11)$$

As $c_1 \longrightarrow 0$ or $c_2 \longrightarrow 0$, expression (2.11) is transformed into the corresponding expression for the free energy of a one-component gas in a 1D channel [3].

From relations (2.2) and (2.11) for the adsorption isotherm of a two-component gas, we finally derive the expressions

$$\frac{a}{\sigma_{\text{av}}}\theta = c\frac{p(1 - C\theta)}{p(1 - C\theta) + \tilde{p}_0^{(1)}} + (1 - c)\frac{p(1 - C\theta)}{p(1 - C\theta) + \tilde{p}_0^{(2)}},$$
$$\tilde{p}_0^{(i)} = \frac{1}{1 - c}\left(\frac{1 - c}{c}\right)^c \left(\frac{\alpha^{(2)}}{\alpha^{(1)}}\right)^c \frac{1}{\alpha^{(2)}} \exp[-\beta\varepsilon_2^{(i)}], \quad i = 1, 2. \quad (2.12)$$

Using these relations and specifying the external conditions, we can obtain the fill factor and the component concentration in the channel. For $c = 0$ and $c = 1$, isotherm (2.12) transforms into the isotherm for a one-component gas, which was obtained in [3].

Let us now analyze the ground state of a mixture in 1D channels. It is known [3] that short-lived clusters can be formed in 1D channels in the case of a one-component gas. The formation of clusters can apparently be expected in the case of a two-component gas as well. In order to describe possible transition to an inhomogeneous clustered state, we must calculate and analyze the free energy of the system taking into account fluctuations and the equation determining the amplitude of these fluctuations in the system (equation for the order parameter). Since molecular transport depends on the lifetime of clusters, the goal of analyzing the equation for the order parameter is to determine the characteristic lifetime of clusters as a function of the channel fill factor and, hence, on the pressure, temperature, and composition of the mixture above the membrane. According to [20], the free energy of particles in a channel taking into account the fluctuation component can be written in the form (here and below, we assume

summation over recurrent indices)

$$F_\Sigma = F_0 + \frac{1}{2}\int_0^1 d\lambda \int dx_1 dx_2 u_{\alpha\beta}(x_1 - x_2) \times g^{(2)}_{\alpha\beta}(x_1, x_2, \lambda) \equiv F^{(0)} + F_1, \quad (2.13)$$

where $F^{(0)}$ is the free energy calculated for a spatially homogeneous state of particles in the channel divided by temperature; F_1 has the meaning of the fluctuation correction to the free energy; $g^{(2)}_{\alpha\beta}(x_1, x_2, \lambda)$ is the pair distribution function ($\alpha, \beta = 1, 2$ are the species of particles); and λ is the interaction constant. For $\lambda \longrightarrow 0$, the system can be reduced to an ideal system, while the case with $\lambda \longrightarrow 1$ corresponds to a real system. In order to describe the ground state of the system and the kinetics of relaxation to this state, we must pass from free energy F to its mean value [20]

$$\tilde{\Delta} = \Delta_0 + \Delta, \quad \Delta_0 = \lim_{\tau\to\infty} \tau^{-1} \int_0^\tau F^{(0)}[n]dt, \quad (2.14)$$

$$\Delta = \lim_{\tau\to\infty} \tau^{-1} \int_0^\tau F_1[n(t, x)]dt.$$

Here, τ is the characteristic time of "coarsening," $\tau_{mom} \ll \tau \ll \tau_{dens}$, τ_{mom} being the characteristic time of momentum relaxation of the system and τ_{dens} the characteristic time of density relaxation of the system. In relations (2.14), $\tilde{\Delta}$ is an analog of the free energy of an equilibrium system and Δ is the fluctuation correction to the mean value $\tilde{\Delta}$. Functional $\tilde{\Delta}$ attains its minimal value for the equilibrium state of the system.

Quantity Δ can be expressed in terms of the response function $\beta_{\alpha\beta}(x, x', t, t')$ of the system, which is defined as

$$\delta n_\alpha(x, t) = \int \beta_{\alpha\beta}(x, x', t, t') e V^{ext}_\beta(x', t') dx' dt', \quad (2.15)$$

where $\delta n_\alpha(x, t)$ is the density fluctuation of particles of species α and $eV^{ext}_\beta(x, t)$ is a weak external field depending on time. Response function $\beta_{\alpha\beta}(x, x', t, t')$ can be derived by directly varying functional Δ over $\delta n(x, t)$ [20]:

$$\beta_{\alpha\beta}(x, x', t, t') = -(\delta^2\Delta/\delta n_\alpha(x, t)\delta n_\beta(x', t'))^{-1}. \quad (2.16)$$

Using this relation, we can obtain the expansion of functional (2.13) into a series in deviations δn of density from its mean value in the case of a multicomponent system. In the Fourier representation, the result has the form

$$\Delta = \frac{1}{2}\int d\mathbf{k} d\omega \beta^{-1}_{\alpha\beta}(\mathbf{k}, \omega)\delta n_\alpha(\mathbf{k}, \omega)\delta n^*_\beta(\mathbf{k}, \omega) + \frac{1}{3}\int d\mathbf{k} d\omega d\mathbf{k}' d\omega' \frac{\delta\beta^{-1}_{\alpha\beta}(\mathbf{k}, \omega)}{\delta n_\gamma(\mathbf{k}', \omega')} \times \delta n_\alpha(\mathbf{k}, \omega)\delta n^*_\beta(\mathbf{k}, \omega)\delta n_\gamma(\mathbf{k}', \omega') + \frac{1}{4}\int d\mathbf{k} d\omega d\mathbf{k}' d\omega' d\mathbf{k}'' d\omega'' \frac{\delta^2\beta^{-1}_{\alpha\beta}(\mathbf{k}, \omega)}{\delta n_\gamma(\mathbf{k}', \omega')\delta n^*_\delta(\mathbf{k}'', \omega'')} \times \delta n_\alpha(\mathbf{k}, \omega)\delta n^*_\beta(\mathbf{k}, \omega)\delta n_\gamma(\mathbf{k}', \omega')\delta n^*_\delta(\mathbf{k}, \omega) + \ldots. \quad (2.17)$$

Here, $\beta_{\alpha\beta}(\mathbf{k}, \omega)$ is the response function of the multicomponent system in the Fourier representation. It follows from the fluctuation-dissipative theorem [24] that response function $\beta_{\alpha\beta}(\mathbf{k}, \omega)$ is directly proportional to the pair distribution; this enables us to calculate $\beta_{\alpha\beta}(\mathbf{k}, \omega)$ for a 1D system with an arbitrary density, for which the pair distribution is exactly known [21].

To calculate the value of Δ, we use the local approximation that makes it possible to replace functional derivatives by ordinary derivatives [20]:

$$\frac{\delta^2\varphi}{\delta\rho(\mathbf{x}_1)\delta\rho(\mathbf{x}_2)} \approx \delta(\mathbf{x} - \mathbf{x}_1)\delta(\mathbf{x} - \mathbf{x}_2)\frac{d^2\varphi}{d\bar{\rho}^2}. \quad (2.18)$$

Using this relation, we obtain from Eq. (2.17)

$$\Delta = \frac{1}{2}\int d\mathbf{k} d\omega \beta^{-1}_{\alpha\beta}(\mathbf{k}, \omega)\delta n_\alpha(\mathbf{k}, \omega)\delta n^*_\beta(\mathbf{k}, \omega) + \frac{1}{3}\int d\mathbf{k} d\omega d\mathbf{k}' d\omega' \frac{\partial\beta^{-1}_{\alpha\beta}(\mathbf{k}, \omega)}{\partial n_\gamma(\mathbf{k}', \omega')} \times \delta n_\alpha(\mathbf{k}, \omega)\delta n^*_\beta(\mathbf{k}, \omega)\delta n_\gamma(\mathbf{k}', \omega') + \frac{1}{4}\int d\mathbf{k} d\omega d\mathbf{k}' d\omega' d\mathbf{k}'' d\omega'' \frac{\partial^2\beta^{-1}_{\alpha\beta}(\mathbf{k}, \omega)}{\partial n_\gamma(\mathbf{k}', \omega')\partial n^*_\delta(\mathbf{k}'', \omega'')} \times \delta n_\alpha(\mathbf{k}, \omega)\delta n^*_\beta(\mathbf{k}, \omega)\delta n_\gamma(\mathbf{k}', \omega')\delta n^*_\delta(\mathbf{k}, \omega) + \ldots. \quad (2.19)$$

Thus, the evaluation of functional (2.19) in the 1D case can be reduced to the calculation of response functions. In the case of two-component systems, the equa-

tions for determining partial response functions β have the form [20]

$$\hat{\beta}(k, \omega) = \hat{\beta}^{(0)}(k, \omega) + \hat{\beta}^{(0)}(k, \omega)\hat{R}(k, \omega)\hat{\beta}(k, \omega),$$

$$\hat{\beta}(k, \omega) \equiv \beta_{ik}(k, \omega) = \begin{pmatrix} \beta_{11}(k, \omega) & \beta_{12}(k, \omega) \\ \beta_{21}(k, \omega) & \beta_{22}(k, \omega) \end{pmatrix},$$
$$\hat{R}(k, \omega) \equiv R_{ik}(k, \omega) = \begin{pmatrix} R_{11}(k, \omega) & R_{12}(k, \omega) \\ R_{21}(k, \omega) & R_{22}(k, \omega) \end{pmatrix}, \quad (2.20)$$

$$\hat{\beta}^{(0)}(k, \omega) \equiv \beta_{ik}^{(0)}(k, \omega) = \begin{pmatrix} \beta_{11}^{(0)}(k, \omega) & \beta_{12}^{(0)}(k, \omega) \\ \beta_{21}^{(0)}(k, \omega) & \beta_{22}^{(0)}(k, \omega) \end{pmatrix}.$$

Here, $\beta_{ik}^{(0)}(k, \omega)$ and $\beta_{ik}(k, \omega)$ are the partial response functions for noninteracting and interacting particles, respectively. The quantity $R_{ik}(k, \omega)$ describes the effective dynamic interaction between particles and is defined as

$$R_{ik}(k, \omega) = V_{ik}(k) - \frac{1}{2}\frac{\delta^2}{\delta n_i(k, \omega)\delta n_k(k, \omega)} \times \int V_{i'k'}(x - x')\beta_{i'k'}(x, x', t, t')dxdx'dtdt'. \quad (2.21)$$

Premultiplying expression (2.20) by $(\hat{\beta}^{(0)}(k, \omega))^{-1}$ and postmultiplying it by $\hat{\beta}^{-1}(k, \omega)$, we obtain

$$(\hat{\beta}^{(0)}(k, \omega))^{-1} = \hat{\beta}^{-1}(k, \omega) + \hat{R}(k, \omega). \quad (2.22)$$

Over long time periods ($t \gg \tau_{\text{mom}}$), function $R_{ik}(k, \omega)$ can be represented in the form

$$\hat{R}(k, 0) = (\hat{\beta}^{(0)}(k, 0))^{-1} - \hat{\beta}^{-1}(k, 0). \quad (2.23)$$

In view of homogeneity and the absence of correlations in a system of noninteracting particles, we have $\hat{\beta}^{(0)}(k, 0) = \hat{\beta}^{(0)}(0, 0) = b_{ij}$, where

$$b_{ij} \equiv -\frac{\partial^2 S}{\partial c_i \partial c_j}, \quad (2.24)$$

S being the entropy of the two-component gas divided by temperature in the absence of the intermolecular interaction. In this case, we obtain the following expression for the effective interaction from Eq. (2.28):

$$\hat{R}(k, 0) = (\hat{\beta}^{(0)}(0, 0))^{-1} - \hat{\beta}^{-1}(k, 0). \quad (2.25)$$

Substituting this expression into Eq. (2.22), we obtain the following relation for response function $\hat{\beta}(k, \omega)$:

$$\hat{\beta}^{-1}(k, \omega) = (\hat{\beta}^{(0)}(k, \omega))^{-1} + \hat{\beta}^{-1}(k, 0) - (\hat{\beta}^{(0)}(0, 0))^{-1}. \quad (2.26)$$

In order to calculate the response function $\hat{\beta}(k, 0)$ for a system of interacting particles, we can use the fluctuation-dissipative theorem connecting response function $\hat{\beta}(k, 0)$ to the pair correlation function for a system of interacting particles:

$$\beta_{ij}(k, 0) = -[b_{ij}^{-1} + c_i c_j \theta^2 \nu_{ij}(k)]. \quad (2.27)$$

Here, $\nu_{ij}(k)$ is the pair correlation function divided by temperature [17]. Relation (2.27) is a generalization of the known relation $\beta(k, 0) = -\theta[1 + \theta\nu(k)]$ [3, 20] for one-component to two-component systems.

Thus, the calculation of the response function can be reduced to the calculation of the pair correlation function or function g_{ik}, called a pair distribution [17] and in terms of the pair correlation function through the relation

$$\nu_{ij}(k) = g_{ij}(k) - \delta(k). \quad (2.28)$$

We will calculate the pair distribution using the method described in [21] and determine $g_{11}(k)$. Functions $g_{12}(k)$ and $g_{22}(k)$ are calculated similarly. Pair distribution $g_{11}(x)$ can be written in the form [21]

$$g_{11}(k) = \sum_{m=1}^{\infty} \Psi_m^{11}(k), \quad (2.29)$$

where the function $\Psi_m^{11}(k)$ is the Fourier transform of the function $\Psi_m^{11}(\zeta)$, which is equal to the probability of finding two particles of the first species separated from each other by m other particles and by a distance ζ. Function $\Psi_m^{11}(k)$ can be written in the form

$$\Psi_m^{11}(k) = \int d\zeta e^{ik} \frac{\sum_{n=0}^{N_{12}} \sum_{l=0}^{N_{22}} \overline{Q}_{m-n-l, n, l}(\zeta)\overline{Q}_{N_{11}-(m-n-l), N_{12}-n, N_{22}-l}(L-\zeta)}{\overline{Q}_{N_{11}, N_{12}, N_{22}}(L)}, \quad (2.30)$$

where $\bar{Q}_{m,n,l}(\zeta)$ is the configuration integral corresponding to the presence of m pairs of particles of the first species, n pairs of particles of the first and second species, and l pairs of particles of the second species in a channel. For $m = N_{11}$, $n = N_{12}$, and $l = N_{22}$, the expression for configuration integral $\bar{Q}_{m,n,l}(\zeta)$ coincides with formula (1.6) in [21]:

$$\bar{Q}_{m,n,l}(\zeta) = \oint dS[\phi_{11}(S)]^m[\phi_{12}(S)]^n[\phi_{22}(S)]^l e^{S\zeta},$$
$$\phi_{ij}(S) = \int_0^\infty \exp(-Sx)\exp\left(-\frac{U_{ij}(x)}{T}\right)dx. \tag{2.31}$$

Here, ζ has the meaning of a coordinate and $\phi_{ij}(S)$ is the Laplace transform of function $f_{ij}(x) \equiv \exp(-U_{ij}(x)/T)$. In the case of an intermolecular interaction potential in the form of that in the hard-sphere model, we have

$$U_{11}(x) = \begin{cases} 0, & x > \sigma_1, \\ \infty, & x \le \sigma_1, \end{cases}$$
$$U_{12}(x) = \begin{cases} 0, & x > \sigma_{av}, \\ \infty, & x \le \sigma_{av}, \end{cases} \tag{2.32}$$
$$U_{22}(x) = \begin{cases} 0, & x > \sigma_2, \\ \infty, & x \le \sigma_2. \end{cases}$$

Substituting these relations into Eq. (2.31), we obtain the following expression for the configuration integral:

$$\bar{Q}_{m,n,l}(\xi) = \exp\left(\frac{p\xi}{T}\right)\left[\varphi_{11}\left(\frac{p}{T}\right)\right]^m\left[\varphi_{12}\left(\frac{p}{T}\right)\right]^n\left[\varphi_{22}\left(\frac{p}{T}\right)\right]^l,$$
$$\varphi_{11}\left(\frac{p}{T}\right) = \frac{T}{p}\exp\left(-\frac{p\sigma_1}{T}\right),$$
$$\varphi_{12}\left(\frac{p}{T}\right) = \frac{T}{p}\exp\left(-\frac{p\sigma_{av}}{T}\right), \tag{2.33}$$
$$\varphi_{22}\left(\frac{p}{T}\right) = \frac{T}{p}\exp\left(-\frac{p\sigma_2}{T}\right).$$

Here, p has the meaning of one-dimensional "pressure". Substituting relations (2.33) into (2.30), we obtain the following expression for pair distribution $g_{11}(k)$:

$$g_{11}(k) = \frac{p}{T}\exp(ik\sigma_1)\left[\frac{p}{T} - ik - \frac{p}{T}\exp(ik\sigma_1)\right]^{-1}$$
$$\times\{1 - \exp[ik(\sigma_{av} - \sigma_1)]\}^{-1} \tag{2.34}$$
$$\times\{1 - \exp[ik(\sigma_2 - \sigma_1)]\}^{-1}.$$

In the long-wave approximation ($0 \le k \le 2\pi/\max\{\sigma_1, \sigma_2\}$), expression (2.34) contains no poles and we obtain

$$g_{11}(x) = \frac{1}{\theta_{eff}}\sum_{m,\,x-ma>0}\frac{\left(y - m\frac{\sigma_1}{a_{eff}}\right)^{m-1}}{(m-1)!(1/\theta_{eff} - 1)^m}$$
$$\times\exp\left(-\frac{y-m}{1/\theta_{eff} - 1}\right),$$
$$y = \frac{x}{a_{eff}(c_1, c_2)}, \quad \theta_{eff} = \frac{Na_{eff}(c_1, c_2)}{L}, \tag{2.35}$$
$$a_{eff}(c_1, c_2) = \sigma_1 c_1^2 + \sigma_2 c_2^2 + 2\sigma_{av}c_1c_2.$$

Functions $g_{12}(x)$ and $g_{22}(x)$ can be calculated similarly. Substituting expressions (2.35) and (2.27) into (2.26) and carrying out the normalization procedure (see Section 3 below), we obtain the final expression for response function $\hat{\beta}(k, 0)$:

$$\frac{1}{\hat{\beta}(k, 0)} = \begin{pmatrix} -\frac{a_{11}}{1 + a_{11}\theta^2 c_1^2\Delta\nu_{11}(k)} & -\frac{a_{12}}{1 + a_{12}\theta^2 c_1c_2\Delta\nu_{12}(k)} \\ -\frac{a_{21}}{1 + a_{21}\theta^2 c_1c_2\Delta\nu_{21}(k)} & -\frac{a_{22}}{1 + a_{22}\theta^2 c_2^2\Delta\nu_{22}(k)} \end{pmatrix}. \tag{2.36}$$

Here,

$$\Delta\nu_{ij}(k) \equiv \nu_{ij}(k) - \nu_{ij}(0), \quad a_{ij} \equiv -\frac{\partial^2 F}{\partial c_i \partial c_j}, \tag{2.37}$$

F is the free energy of a two-component gas with interaction per unit temperature, which can be calculated by differentiating configuration integral $\bar{Q}_N$ twice with respect to concentration. Function $\hat{\beta}^{(0)}(k, \omega)$ can be obtained by generalizing the response function of the one-component gas,

$$\frac{1}{\beta_0(k, \omega)} = -\frac{1}{\theta}\left(1 + \frac{\omega}{\omega_0}\right) = \frac{\partial^2 F}{\partial\theta^2}\left(1 + \frac{\omega}{\omega_0}\right),$$

derived in [3], to the case of a two-component mixture:

$$\frac{1}{\hat{\beta}^{(0)}(k, \omega)} = \begin{pmatrix} \left(1 + \frac{\omega}{\omega_1}\right)b_{11} & \left(1 + \frac{\omega}{\omega_1}\right)b_{12} \\ \left(1 + \frac{\omega}{\omega_2}\right)b_{21} & \left(1 + \frac{\omega}{\omega_2}\right)b_{22} \end{pmatrix}, \tag{2.38}$$
$$\omega_1^{(p)} = -iD_1k^2, \quad \omega_2^{(p)} = -iD_2k^2.$$

Here, D_1, D_2, $\omega_1^{(p)}$, and $\omega_2^{(p)}$ are the diffusion coefficients and the relaxation frequency spectra for pure components.

Substituting relations (2.36) and (2.38) into Eq. (2.26), we obtain the following expression for response function $\beta(k, \omega)$:

$$\frac{1}{\beta(k,\omega)} = \begin{pmatrix} \frac{\omega}{\omega_1^{(p)}} b_{11} - \frac{a_{11}}{1 + a_{11}\theta^2 c_1^2 \Delta\nu_{11}(k)} & \frac{\omega}{\omega_1^{(p)}} b_{12} - \frac{a_{12}}{1 + a_{12}\theta^2 c_1 c_2 \Delta\nu_{12}(k)} \\ \frac{\omega}{\omega_2^{(p)}} b_{21} - \frac{a_{21}}{1 + a_{21}\theta^2 c_1 c_2 \Delta\nu_{21}(k)} & \frac{\omega}{\omega_2^{(p)}} b_{22} - \frac{a_{22}}{1 + a_{22}\theta^2 c_2^2 \Delta\nu_{22}(k)} \end{pmatrix}. \quad (2.39)$$

It is convenient for subsequent analysis to pass to new variables. For a two-component system, the relation $c_1 + c_2 = 1$ holds and quantity δn can be represented in the form

$$\delta\mathbf{n}(\mathbf{k}, \omega) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \xi(\mathbf{k}, \omega), \quad (2.40)$$

where ξ can be interpreted as an order parameter. Substituting Eq. (2.40) into (2.19), we obtain the expression for functional Δ,

$$\Delta = \frac{1}{2} d\mathbf{k} d\omega \Omega |\xi_{k,\omega}|^2 + \frac{2}{3} d\mathbf{k} d\omega \frac{\partial\Omega}{\partial n} |\xi_{k,\omega}|^2 \xi_{k,\omega} + \int d\mathbf{k} d\omega \frac{\partial^2\Omega}{\partial n^2} |\xi_{k,\omega}|^4 + \ldots, \quad (2.41)$$

where

$$\Omega \equiv \beta_{11}^{-1}(\mathbf{k}, \omega) - \beta_{12}^{-1}(\mathbf{k}, \omega) - \beta_{21}^{-1}(\mathbf{k}, \omega) + \beta_{22}^{-1}(\mathbf{k}, \omega). \quad (2.42)$$

Series (2.41) can be summed and written in the form

$$\Delta = \frac{1}{2} \int dk d\omega |\xi|_{k,\omega}^2 \Omega(n, c + \xi_{k,\omega}, k, \omega). \quad (2.43)$$

The formal proof of the correctness of representing quantity Δ in form (2.43) without analyzing the convergence of the corresponding series follows from the expansion of expression (2.43) into a power series of order parameter ξ. This expansion exactly reproduces series (2.41) (at least to within terms on the order of ξ^5).

Using relations (2.39), (2.42), and (2.43), we can obtain the dependence of functional Δ on the wave vector and order parameter of the system. The minimum of functional Δ determines the ground state of the system. Substituting relation (2.39) into (2.43), evaluating the derivative of Δ with respect to the order parameter, and applying the inverse Fourier transformation in frequency ω to the resulting relation, we obtain

$$\frac{\partial\xi_k}{\partial t} = -\tilde{D} k^2 \frac{\partial(F_1(\xi_k, k))}{\partial\xi_k}, \quad (2.44)$$

where

$$\tilde{D} = \left(\frac{b_{11} - b_{12}}{D_1} + \frac{b_{22} - b_{21}}{D_2} \right)^{-1},$$

$$F_1(\xi_k, k) \equiv \frac{1}{2} \Omega(c + \xi_k, k) \xi_k^2,$$

$$\Omega(c, k) = -\frac{a_{11}}{1 + a_{11}\theta^2 c_1^2 \Delta\nu_{11}(k)} + \frac{2a_{12}}{1 + a_{12}\theta^2 c_1 c_2 \Delta\nu_{12}(k)} - \frac{a_{22}}{1 + a_{22}\theta^2 c_2^2 \Delta\nu_{22}(k)}.$$

Equation (2.44) has the same form as the equation usually used for the order parameter. The role of the order parameter is played in this case by the quantity defined by relation (2.40). It should be noted that the order parameter here could be either positive or negative. The cases when $\xi > 0$ and $\xi < 0$ correspond to the density fluctuation of the first and the second component, respectively. The role of the diffusion coefficient is played by quantity $\tilde{D}$ depending on the fill factor and the concentration of components in channels via coefficients a_{ij} and b_{ij} defined by relations (2.24) and (2.36). It should be noted that the concentration of the components in a channel (and its fill factor) could be determined unambiguously from isotherm (2.12) proceeding from the pressure, temperature and composition of the gas mixture.

Analysis of function $F_1(\xi_k, k)$ makes it possible to investigate the possibility for a transition of the system to an inhomogeneous state. For example, when the minimum of function F_1 is attained for $\xi \neq 0$, $k = 0$, a conventional phase transition to a homogeneous state

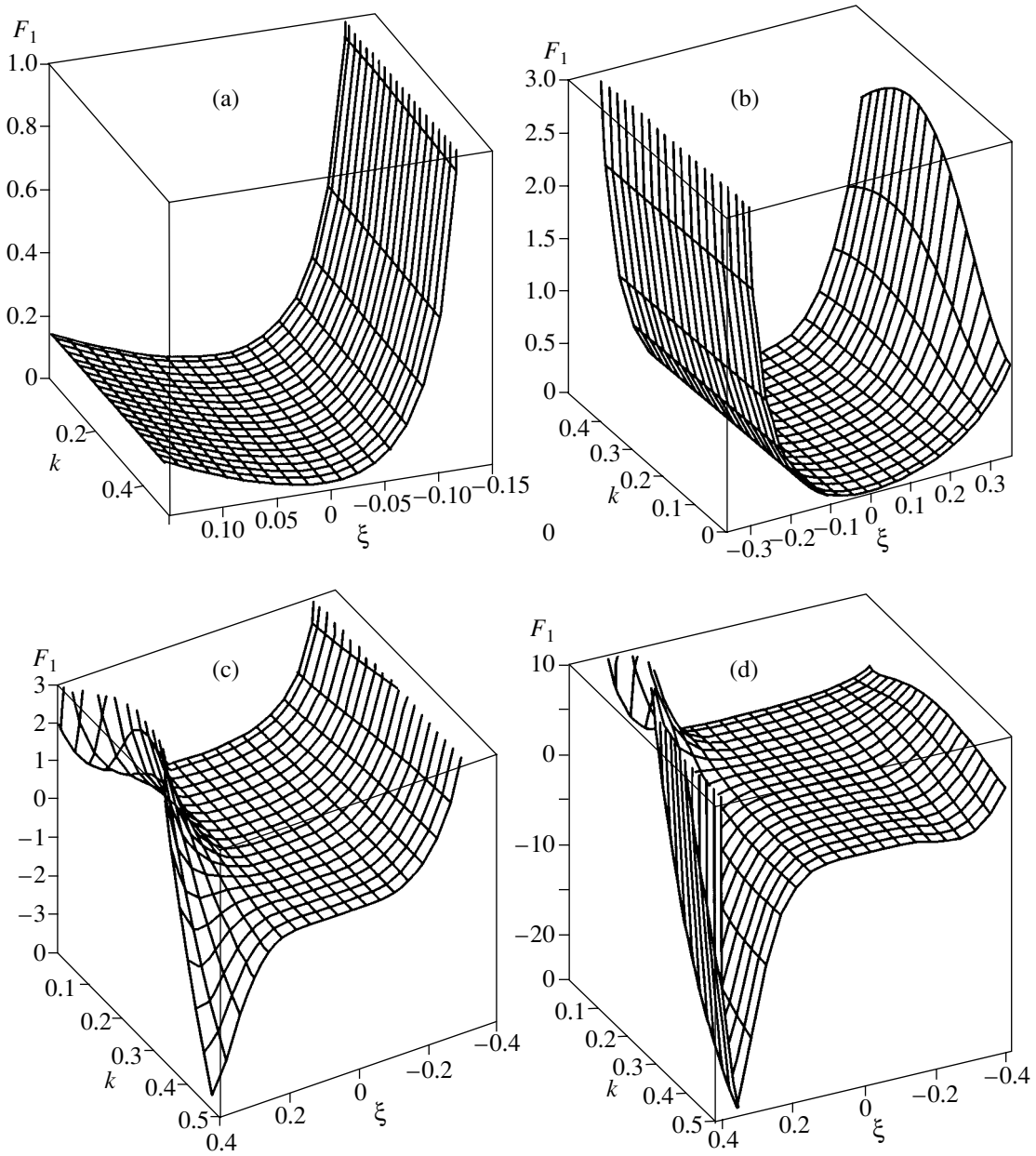

Fig. 1. Dependence of free energy F_1 on the order parameter and wave vector for various concentrations of the first component in the mixture: (a) $c = 0.1$ ($\theta = 0.67$); (b) $c = 0.25$ ($\theta = 0.71$); (c) $c = 0.3$ ($\theta = 0.74$), and (d) $c = 0.45$ ($\theta = 0.75$).

takes place [22]. The minimum at $\xi = 0$, $k \equiv k_c \neq 0$ corresponds to the propagation of a density wave over a distance of $r \sim k_c^{-1}$. In the case when the minimum is attained at $\xi \neq 0$ and $k \neq 0$, a transition to an inhomogeneous state with clusters formed in the system is realized [22]. Local minima here determine metastable states. Figure 1 shows the graphs illustrating the dependence of F_1 on order parameter ξ and wave vector k for a mixture in which one of the gases (first) is a strong sorbate, while the other gas is a weak sorbate. The chosen parameters ($a = 3.8$ Å, $\sigma_1 = 4.3$ Å, $\sigma_2 = 3.6$ Å, $\varepsilon_1 = 0.61$ eV, $\varepsilon_2 = 0.38$ eV, $P = 2000$ kPa, and $T = 300$ K) correspond to a methane–butane molecular mixture [15] and make it possible to describe adsorption and fluxes of these one-component gases in zeolite membranes. It should be noted that the choice of parameters P, T, and c unambiguously determines fill factor θ and concentrations c_1 and c_2 of gases in a channel (see Eqs. (2.14)).

For low concentrations of the first component in a gas mixture and for fill factors $\theta \leq 0.7$, function F_1 has only one minimum at $\xi = 0$, $k = 0$ (Figs. 1a and 1b). This corresponds to a homogeneous state of the system. An increase in the concentration of the first component leads to the emergence of local minima in function F_1.

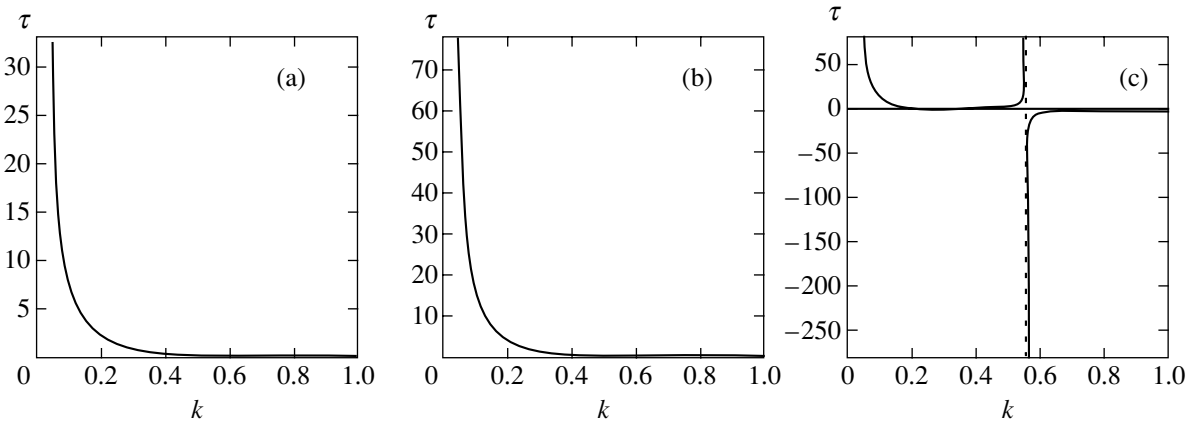

Fig. 2. Dependence of lifetime τ on wave vector k for $c = 0.1$ (a), 0.3 (b), and 0.45 (c).

In this case, the global minimum is shifted to point $\xi = \xi_c > 0$, $k = k_c$. In this case, the ground state of the system becomes spatially inhomogeneous, which corresponds to the formation of clusters in the system. This situation will be analyzed in greater detail below.

Equation (2.44) for the order parameter also allows us to calculate the characteristic relaxation times for the emerging density fluctuation depending on external conditions. For this purpose, we expand function $F_1(\xi_k, k)$ into a series in the vicinity of $\xi_k = 0$ to within the first nonvanishing term:

$$F_1(\xi_k, k) \sim \lambda(k)\xi_k^2. \tag{2.45}$$

Substituting Eq. (2.45) into (2.44), we obtain

$$\frac{\partial \xi_k}{\partial t} = -\frac{\xi_k}{\tau}, \quad \tau = \frac{1}{2\lambda(k)\tilde{D}k^2}. \tag{2.46}$$

The quantity τ is the lifetime of the kth mode of the density fluctuation emerging in the vicinity of $\xi_k = 0$. The curves describing the lifetime of the emerging fluctuation as a function of the wave vector for different mixture compositions are shown in Fig. 2.

It can be seen from Fig. 2 that, for a low concentration of the highly adsorbed component, the lifetime decreases monotonically with increasing wave vector, while the free energy has a minimum at $\xi = 0$. From the standpoint of physics, this means that the state of the system in a channel for a given composition of the mixture is homogeneous and the growth of clusters is energetically disadvantageous. It should be noted that the fill factor of the channel is small in this case, but short-lived clusters with a lifetime of

$$\tau \sim \frac{r^2}{D} = \frac{r^2}{D_0}\exp\left(\frac{E_a}{T}\right) \tag{2.47}$$

can be formed due to density fluctuations. Here, $D = D_0\exp(-E_a/T)$ is the diffusion coefficient of solitary particles, E_a is the diffusion activation energy, and r is the characteristic size of fluctuations. Characteristic size r is, determined, as in [3], from the position of the extremum of the imaginary part of the spectrum divided by the square of the wave vector. It should be noted that the value of r increases with the fill factor of the channel [3]. Expression (2.48) can be derived by expanding order parameter (2.44) into a series in the vicinity of $\xi = 0$. For low concentrations, expression (2.47) coincides with the characteristic diffusion decay time for density fluctuations.

An increase in the concentration of the first component of the mixture to $c = 0.3$ elevates the fill factor of the channel to $\theta = 0.74$ (see Fig. 1c), while F_1 acquires two more minima for $k \sim 1$. Figure 3 shows the dependences of the free energy on the order parameter for different values of the wave vector at $c = 0.3$. It can be seen from the figure that the free energy minimum is attained at point $\xi = 0$ for values of the wave vector $0 < k < 0.4$. For $k_c \approx 0.5$, function F_1 has two minima (at $\xi_c^{(0)} = 0$ and $\xi_c^{(1)} = 0.35$) and the state with $\xi_c^{(1)}$ is separated from the state with $\xi_c^{(0)}$ by a potential barrier. Since the global minimum is attained at $\xi = \xi_c^{(1)}$, the state with $\xi_c^{(0)}$ is metastable. The ground state of the system is attained at $\xi = \xi_c^{(1)}$ and is clustered. It should be noted that, in accordance with formula (2.40), clusters of the first component are formed in the channel since $\xi_c^{(1)} > 0$. The lifetime of the clusters can be determined by passing from Eq. (2.44) for the order parameter to a stochastic differential equation (Langevin equation) and the corresponding Fokker–Plank equation via the introduction of additive noise. Steady-state solutions to the Fokker–Planck equation define the probability of the

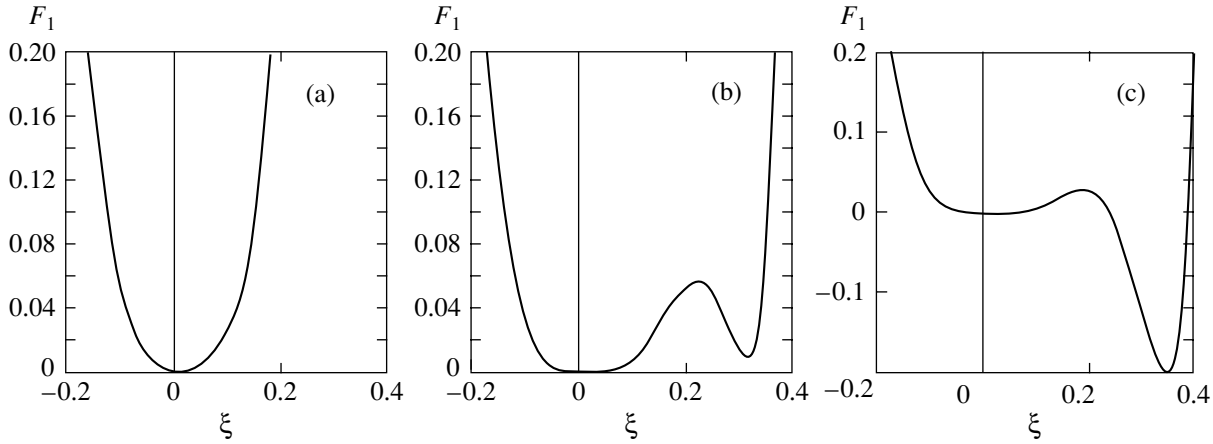

Fig. 3. Dependence of free energy F_1 on the order parameter for different values of the wave vector for $c = 0.3$: $k = 0.1$(a), 0.3 (b), and 0.5 (c).

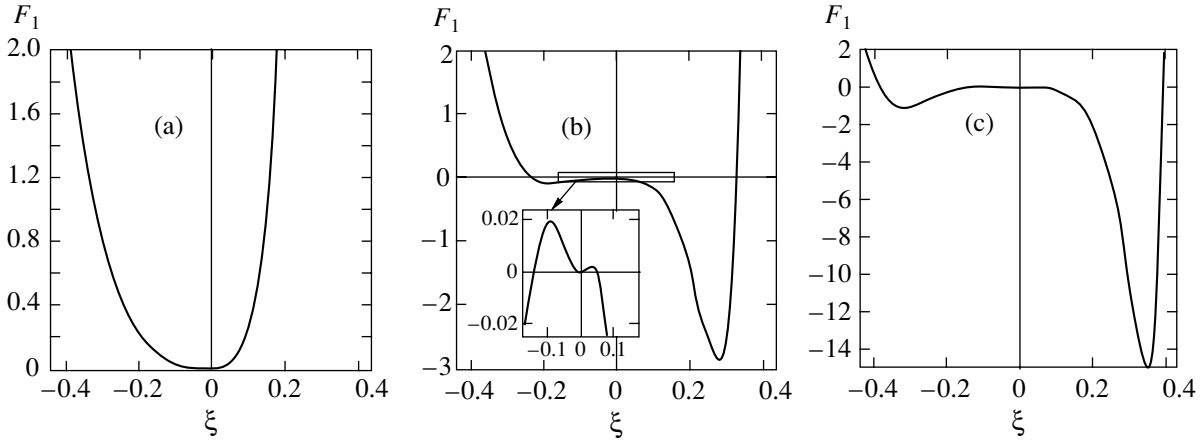

Fig. 4. Dependence of free energy F_1 on the order parameter for different values of the wave vector for $c = 0.45$: $k = 0.1$(a), 0.3 (b), and 0.5 (c).

system being in the state described by order parameter $\xi_c^{(1)}$. For the lifetime, we have

$$\tau \sim \frac{r^2}{D_0}\exp\left(\frac{E_a}{T} - \delta F\right), \qquad \delta F = F_1(c + \xi_c^{(1)}, k_c)T. \tag{2.48}$$

The decay of a cluster formed in the system occurs via the overcoming of an energy barrier of height δF. From the standpoint of physics, this can be interpreted as an effective increase in the diffusion activation energy for the particles in the cluster. However, for $c = 0.3$, the barrier height is small ($\delta F \sim 0.2T$; see Fig. 1c) and lifetime (2.48) is comparable to time (2.47). A further increase in the first component concentration in the mixture leads to negative values of the lifetime for a certain value of $k = k_c < 1$ (see Fig. 2c). In this case, clusters with a size of $r \sim k_c^{-1}$ are formed in the channel; for $k < k_c$, fluctuations decay over a finite time, while for $k > k_c$, the emerging density fluctuations evolve. Such a situation is typical of transitions to an inhomogeneous condensate state [22].

The curves describing the dependence of the free energy on the order parameter for different values of the wave vector are shown in Fig. 4. Since function F_1 has a single minimum at $\xi = 0$ for $k \sim 0$, the state of particles in the channel is homogeneous. However, as the wave vector increases, curve $F_1(\xi)$ acquires two more minima separated from the state with $\xi = 0$ by potential barriers. A further increase in the wave vector leads to disappearance of the barriers between the state with $\xi = 0$ and the states corresponding to two other minima of F_1, and the state with $\xi = 0$ becomes unstable.

Thus, for a low concentration of the highly adsorbed component, the fill factor of the channel is small and the state of the system in the channel is homogeneous. As the concentration of the highly adsorbed component increases, the fill factor of the channel increases and

short-lived clusters of the highly adsorbed component can be formed in the channel. A further increase in the concentration (and the fill factor of the channel) leads to an increase in the lifetime of the clusters formed in the channel. At a certain concentration of the highly adsorbed component, the clusters in the channel become long-lived and the ground state of the system becomes clustered.

The formation of clusters of one of the components in a channel may strongly affect the transport of the gas through the membrane [3]. In the next section, partial gas fluxes through the membrane will be calculated and the effect of clusters on the mechanisms of transport will be analyzed.

3. TRANSPORT IN A TWO-COMPONENT GAS IN A HIGH-DENSITY 1D SYSTEM

In order to calculate partial fluxes and to analyze the transport, we will use the approach proposed in [3]. It will be shown below that the evaluation of the flux can be reduced to the calculation of the relaxation frequency spectra $\omega(\mathbf{k})$ for density fluctuations of the components. We can judge the mechanism of particle transport in a subnanometer channel from the type of dependence of the spectrum on the wave vector of the system.

We can write the expression determining the relaxation of the Fourier component $n(k, t)$ of the number density of particles in a channel in the case of an arbitrary density $n(k, t)$ under conditions of slight deviation from equilibrium [3, 25]:

$$\dot{n}(k, t) = i\omega(k)n(k, t). \tag{3.1}$$

It follows from this relation that the equation for fluctuation amplitude δn can be written in the form

$$\delta\dot{n}(k, t) = i\omega(k)\delta n(k, t), \tag{3.2}$$

where $\omega(k)$ is the relaxation frequency spectrum for the system under study; in the case of diffusion of noninteracting particles, this spectrum has the form [24]

$$\omega = \omega_0(k) = -iDk^2. \tag{3.3}$$

Equation (3.2) describes relaxation of the kth component of density fluctuation for an arbitrary value of the wave vector. In particular, this equation for $k \neq 0$ makes it possible to describe the relaxation of density fluctuations and the propagation of a perturbation over a finite-size cluster in the case of its formation. For $k \longrightarrow 0$, Eq. (3.2) describes the relaxation of density fluctuation on a large spatial scale. This quantity is associated with macroscopic fluxes. In order to calculate the fluxes, we write the continuity equation

$$\dot{n} + \operatorname{div}\mathbf{j} = 0 \tag{3.4}$$

and apply the Fourier transformation to this equation:

$$ikj(k, t) = -\dot{n}(k, t). \tag{3.5}$$

Substituting Eq. (3.1) into (3.5), we obtain the following relation for flux $j(k)$:

$$j(k, t) = \frac{n(k, t)\omega(k)}{k}. \tag{3.6}$$

For partial fluxes, we have

$$j_i(k, t) = \frac{n_i(k, t)}{k}\omega_i(k), \tag{3.7}$$

where n_i is the density of the ith component and ω_i is the corresponding spectrum. The total gas flux is defined as the sum of partial fluxes of the components:

$$\sum_{i=1}^{2} j_i(k, t) = j(k, t). \tag{3.8}$$

It follows from Eq. (3.7) that partial fluxes are determined by relaxation frequency spectra $\omega_i(k)$. Thus, the problem is reduced to computing these spectra. The relaxation frequency spectra can be determined from the condition of the existence of nonzero density fluctuations for each component in an arbitrarily weak external field. Consequently, applying the Fourier transformation to Eq. (2.15), we find that the relaxation frequency spectrum can be determined by solving the following system of homogeneous equations:

$$\beta^{-1}(k, \omega)\begin{pmatrix} 1 \\ -1 \end{pmatrix} = 0. \tag{3.9}$$

Here, matrix $\beta^{-1}(k, \omega)$ of the response functions is defined by relation (2.39). In view of δ-function singularities for $k \longrightarrow 0$ that emerge in pair correlation functions (2.28) appearing in Eq. (2.39), we must calculate spectra $\omega_i(k)$ and the values of $\omega_i(k = 0)$, after which the renormalization procedure must be carried out. Using Eq. (2.39) and passing to the limit $k \longrightarrow 0$, we obtain for spectra $\omega_i(k = 0)$

$$\beta^{-1}(0, \omega)\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} \dfrac{\omega}{\omega_1^{(p)}}b_{11} - \dfrac{\omega}{\omega_1^{(p)}}b_{12} + a_{11} - a_{12} \\ \dfrac{\omega}{\omega_2^{(p)}}b_{22} - \dfrac{\omega}{\omega_2^{(p)}}b_{21} + a_{22} - a_{21} \end{pmatrix} = 0, \tag{3.10}$$

where quantities a_{ij}, b_{ij}, $\omega_1^{(0)}$, and $\omega_2^{(p)}$ are defined by relations (2.24), (2.37), and (2.38). For $\omega_i^{(0)} \equiv \omega_i(k=0)$, we obtain from relation (3.10)

$$\omega_1^{(0)} = -\frac{\omega_1^{(p)}(a_{11}-a_{12})}{b_{11}-b_{12}},$$
$$\omega_2^{(0)} = -\frac{\omega_2^{(p)}(a_{22}-a_{21})}{b_{22}-b_{21}}. \tag{3.11}$$

On the other hand, substituting Eq. (2.39) into (3.10), we obtain the following expressions for relaxation frequency spectra $\omega_i(k)$ of the system under investigation:

$$\tilde{\omega}_1(k) = \frac{iD_1k^2}{b_{11}-b_{12}} \times\left[\frac{1}{\theta^2c_1^2\nu_{11}(k)+b_{11}^{-1}}-\frac{1}{\theta^2c_1c_2\nu_{12}(k)+b_{12}^{-1}}\right],$$
$$\tilde{\omega}_2(k) = \frac{iD_2k^2}{b_{22}-b_{21}} \times\left[\frac{1}{\theta^2c_2^2\nu_{22}(k)+b_{22}^{-1}}-\frac{1}{\theta^2c_1c_2\nu_{21}(k)+b_{21}^{-1}}\right]. \tag{3.12}$$

Tildes indicate that $\tilde{\omega}_i(k=0) \neq \omega_i^{(0)}$ (quantities $\omega_i^{(0)}$ are defined in relations (3.11)).

We will carry out the renormalization procedure for spectra (3.12) taking into account relations (3.11), i.e., imposing the requirement

$$\tilde{\omega}_i = \omega_i^{(0)}(k=0). \tag{3.13}$$

Taking into account the expressions for $\omega_i^{(p)}$, we find from relations (3.11) and (3.12) that

$$a_{11}-a_{12} = \frac{b_{11}}{b_{11}\theta^2c_1^2\nu_{11}(k)+1}-\frac{b_{12}}{b_{12}\theta^2c_1c_2\nu_{12}(k)+1},$$

or

$$a_{11} = \frac{b_{11}}{b_{11}\theta^2c_1^2\nu_{11}(k)+1}, \quad a_{12} = \frac{b_{12}}{b_{12}\theta^2c_1c_2\nu_{12}(k)+1};$$

similarly, for a_{22}, we have

$$a_{22} = \frac{b_{22}}{b_{22}\theta^2c_2^2\nu_{22}(k)+1}. \tag{3.14}$$

In order to separate singularities of pair correlation function $\nu_{ij}(k)$ for $k \longrightarrow 0$, we expand it into a series in the vicinity of $k=0$ to within first-order terms. Using relations (3.12), we then obtain

$$\tilde{\omega}_1(k) = \frac{iD_1k^2}{b_{11}-b_{12}}\left[\frac{1}{\theta^2c_1^2[\nu_{11}(0)+k\nu'_{11}(k)]+b_{11}^{-1}}-\frac{1}{\theta^2c_1c_2[\nu_{12}(0)+k\nu'_{12}(k)]+b_{12}^{-1}}\right]$$
$$= \frac{iD_1k^2}{b_{11}-b_{12}}\left[\frac{a_{11}}{1+ka_{11}\theta^2c_1^2\nu'_{11}(k)}-\frac{a_{12}}{1+ka_{12}\theta^2c_1c_2\nu'_{12}(k)}\right]. \tag{3.15}$$

Taking into account the long-wave approximation and the smallness of k, replacing the derivative in the last expression by the difference $k\nu'_{ij}(k) \approx \nu_{ij}(k)-\nu_{ij}(0)$, omitting primes, and carrying out similar calculations for ω_2, we finally obtain

$$\omega_1(k) = \frac{iD_1k^2}{b_{11}-b_{12}}\left[\frac{a_{11}}{1+a_{11}\theta^2c_1^2[\nu_{11}(k)-\nu_{11}(0)]}-\frac{a_{12}}{1+a_{12}\theta^2c_1c_2[\nu_{12}(k)-\nu_{12}(0)]}\right],$$
$$\omega_2(k) = \frac{iD_2k^2}{b_{22}-b_{21}}\left[\frac{a_{22}}{1+a_{22}\theta^2c_2^2[\nu_{22}(k)-\nu_{22}(0)]}-\frac{a_{21}}{1+a_{21}\theta^2c_1c_2[\nu_{21}(k)-\nu_{21}(0)]}\right]. \tag{3.16}$$

Using relations (2.12), (3.7), and (3.16) and passing to the limit for $k \longrightarrow 0$, we can obtain the dependence of partial fluxes of the components on the pressure, temperature and composition of the mixture. For $k \neq 0$, expressions (3.2) and (3.16) enable us to analyze the relaxation mechanism of the emerging density fluctuation with a characteristic size of $r \approx 2\pi/k$.

In order to analyze the experimentally observed partial fluxes, it is convenient to pass to the coordinate representation. Spectra (3.16) have the real and imaginary parts since pair correlation function $\nu_{ij}(k)$ is a complex quantity in accordance with relations (2.28) and (2.33). Separating the real and imaginary parts of the spectra and applying the inverse Fourier transformation to expression (3.7), we obtain (summation over recurring indices is not carried out)

$$j_i = n_i\psi_i - D_i\frac{\partial n_i}{\partial x},$$
$$\psi_i = \mathrm{Re}\,\omega_i\left(k \longrightarrow \frac{2\pi}{L}\right),$$
$$D_i = \mathrm{Im}\left.\frac{\omega_i(k)}{k^2}\right|_{k\to 2\pi/L}. \tag{3.17}$$

Here, D_i is the diffusion coefficient of a component, L is the channel length, and ψ is the term emerging due to the mutual effect on the mixture components in a channel for large fill factors. It should be noted that, since $L \gg r_c$, where r_c is the characteristic size of clusters, passing to the limit for $k \longrightarrow 2\pi/L$ in relation (3.17) corresponds to averaging over characteristic scales of inhomogeneities in the case when the ground state of the system is clustered. Separating the terms linear in the wave vector from the real part of the spectrum, we obtain the following expression for ψ:

$$\psi_i = -D_i \frac{1}{T}\frac{\partial U_i}{\partial x} + V_i. \tag{3.18}$$

Here, the first term describes the transport of molecules of the ith component, which is induced by the effective intermolecular interaction, while the second term has the meaning of the "drag" effect familiar in the kinetics of mixtures. Substituting Eq. (3.18) into (3.17), we obtain the final expression for partial fluxes:

$$j_i = -D_i \frac{\partial n_i}{\partial x} - D_i \frac{n_i}{T}\frac{\partial U_i}{\partial x} + V_i n_i. \tag{3.19}$$

Thus, it follows from this equation that the partial flux is the sum of three terms. The first term corresponds to diffusion transport, while the second term emerges due to field diffusion. These two terms in the k representation can be combined into one by introducing the effective diffusion coefficient

$$D_i = D_i^{(0)}\left(1 + \frac{n_i}{T}U_i(k)\right). \tag{3.20}$$

It was shown in the previous section that, for large fill factors, clusters are formed in the system. The mechanism of density relaxation may be different for the case of transport over the characteristic scale L, when $2\pi\sigma_{av}/L \ll 1$ and $k \longrightarrow 0$ (which corresponds to macroscopic transport of gas components through the channel), and for the case when $k \neq 0$, which corresponds to relaxation over distances comparable to the size of clusters formed in the system. To prove this, we must analyze dependences $\Delta\nu_{ij}(k)$. In order to avoid cumbersome formulas, we consider this problem in greater detail in the limiting case of $c_1 = 1$, which corresponds to a one-component mixture.

Passing to the limit for $c_1 \longrightarrow 1$ in Eq. (3.16), we obtain the relaxation frequency spectrum for a one-component system,

$$\omega(k) = -\frac{iD_0k^2}{1 + \nu(k)}, \tag{3.21}$$

where $\nu(k)$ is the pair relaxation function, which is connected to the pair distribution via the relation

$$\nu(k) = g(k) - \delta(k). \tag{3.22}$$

Pair distribution $g(k)$ can be derived from Eq. (2.34) by passing to the limit for $c_1 \longrightarrow 1$:

$$g(k) = \frac{1}{\theta}e^{ik\sigma}\left[1 - ik\sigma\left(\frac{1}{\theta} - 1\right) - e^{ik\sigma}\right]^{-1}. \tag{3.23}$$

It should be noted that, in accordance with formula (3.21), the pair distribution and, hence, the pair correlation function associated with it depend not on the wave vector k, but on the product ik:

$$\nu(k) \equiv \nu(ik). \tag{3.24}$$

Let us now write relation (3.21) in the form

$$\begin{aligned}\omega(k) &= -\frac{iD_0k^2}{1 + \nu(0) + \nu(k) - \nu(0)} \\ &= -\frac{iD_0k^2}{1 + \nu(0) + \Delta\nu} = \frac{iD_0k^2}{1 + \nu(0) + \mathrm{Re}\Delta\nu + i\mathrm{Im}\Delta\nu}.\end{aligned} \tag{3.25}$$

Separating the real and imaginary parts from relation (3.25), we obtain

$$\begin{aligned}\omega(k) &= -\frac{iD_0k^2[(1-\theta)^2 + \mathrm{Re}\Delta\nu]}{[(1-\theta)^2 + \mathrm{Re}\Delta\nu]^2 + (\mathrm{Im}\Delta\nu)^2} \\ &\quad - \frac{D_0k^2\mathrm{Im}\Delta\nu}{[(1-\theta)^2 + \mathrm{Re}\Delta\nu]^2 + (\mathrm{Im}\Delta\nu)^2}.\end{aligned} \tag{3.26}$$

Here, we have used the fact that

$$1 + \nu(0) \approx (1-\theta)^2. \tag{3.27}$$

This relation follows from the known relation [17]

$$\frac{\partial p}{\partial\theta} \sim \frac{1}{1 + \nu(0)}$$

describing the increase in the compressibility with density. Using relation (3.24), we now expand $\Delta\nu$ to within first-order terms:

$$\Delta\nu \equiv \left.\frac{\partial\nu(k)}{\partial(ik)}\right|_{k=0} ik \equiv i\nu'(0)k. \tag{3.28}$$

The value of $\nu'(0)$ is finite for $\theta \neq 0$. For example, using direct expansion into the Taylor series, we can obtain from Eqs. (3.22) and (3.23)

$$\nu'(0) = -\theta + \frac{4}{3}\theta^2 - \frac{1}{2}\theta^3. \tag{3.29}$$

Substituting relation (3.28) into Eq. (3.26), we get

$$\begin{aligned}\omega(k) &\approx -\frac{iD_0k^2(1-\theta)^2}{(1-\theta)^4 + \nu'^2(0)k^2} \\ &\quad - \frac{D_0k^3\nu'(0)}{(1-\theta)^4 + \nu'^2(0)k^2}.\end{aligned} \tag{3.30}$$

To calculate the fluxes for high densities ($\theta \sim 1$), we must consider two limiting transitions in relation (3.30):

for $\theta \longrightarrow 1$ and $k \longrightarrow 0$. From the standpoint of physics, the transition for $\theta \longrightarrow 1$ and $k \neq 0$ corresponds to analysis of transport in a dense cluster. The transition for $k \longrightarrow 0$ and arbitrary θ is equivalent to analysis of diffusion in a channel of length $L \gg r_c$, where r_c is the characteristic size of a cluster. For $\theta \longrightarrow 1$, we obtain, instead of relation (3.30),

$$\omega(k) = -\frac{D_0 k}{\nu'(0)}. \tag{3.31}$$

This relation shows that spectrum $\omega(k)$ corresponds to the hydrodynamic mode [17], for which quantity $D_0/\nu'(0)$ is the effective velocity of sound.

Passing in relation (3.30) to the limit for $k \longrightarrow 0$ and retaining the lowest order in wave number k, we obtain

$$\omega(k) = -\frac{iD_0 k^2}{(1-\theta)^2}. \tag{3.32}$$

This relation shows that the spectrum in this case is of the diffusion type with the diffusion coefficient

$$D \approx \frac{D_0}{(1-\theta)^2}, \tag{3.33}$$

which increases indefinitely as $\theta \longrightarrow 1$.

It should be noted that, in accordance with Eq. (3.30), both relaxation mechanisms (hydrodynamic and diffusion) operate in the system when $k \neq 0$ and $\theta \neq 1$. From the standpoint of physics, this corresponds to the diffusion transport between clusters (diffusion mode), over which density perturbations propagate (hydrodynamic mode). In spite of different interpretations of the sequence of limiting transitions, the calculated dependences of fluxes on external conditions are found to be equivalent from the standpoint of physics. This is due to the fact that the hydrodynamic component is manifested in the second case in the increase in the effective diffusion coefficient $D \approx D_0/(1-\theta)^2$ for high values of fill factor θ.

In the two-component case, the transition for $k \longrightarrow 0$ corresponds to the transition from relation (3.16) to (3.11). It can be seen from relation (3.11) that the spectra obtained are of the diffusion type since $\omega_j^{(p)} = -iD_j k^2$, $j = 1, 2$, while quantities a_{ij} and b_{ij} are independent of k. The limiting transition for $\theta \longrightarrow 1$ in the two-component case involves considerable computational difficulties. However, the above analysis shows that the relaxation frequency spectra for the components also acquire a hydrodynamic mode in view of the dependence of pair correlation function (2.28) on factor ik for $\theta \longrightarrow 1$.

Depending on the sequence of the limiting transition, a decisive role in relation (3.19) is played either by the term corresponding to diffusion or the nongradient (hydrodynamic) part of the flux. Since the gas flux for arbitrary fill factors is measured in experiments on the penetrability of the membrane over large distance ($L \gg r_c$, $k \longrightarrow 0$), the limiting transition for $k \longrightarrow 0$ should be performed first. In this case, the effects associated with the formation of clusters are taken into account in the dependence of the diffusion coefficient on the concentration and fill factor of the channel. Numerical calculations show that the last term in relation (3.16) is insignificant in the entire range of concentrations and fill factors of the channel. Thus, the relaxation frequency spectrum of a two-component system can be treated as a diffusion spectrum in the entire range of concentrations and fill factors, where the role of the diffusion coefficient is played by a quantity taking into account clustering of the components in the channel.

It should be noted that, for calculating response function (2.39), relaxation frequency spectra (3.16) and, as a consequence, partial fluxes (3.19), we used the fluctuation-dissipative theorem (2.27) that presumes a homogeneous ground state of the system. In the case when the ground state of the system is clustered (see Fig. 1d), the kinetic version of the fluctuation-dissipative theorem [24] should be used,

$$\langle \delta n(x, t)\delta n(x', t)\rangle = -T\int_0^\infty \beta(x, x', t, \tau)[f(x, \tau) - f'(x', \tau)]d\tau, \tag{3.34}$$

where $f(x, \tau)$ is the probability of a particle being located at point x. In accordance with the detailed balancing principle, in the case when states x and x' differ from the equilibrium state insignificantly, we have

$$f(x', t) \propto f(x, t)\exp\left(\frac{\Delta E}{T}\right), \tag{3.35}$$

where ΔE is the energy difference between states x and x'. Using this relation, we obtain the following expression for the fluctuation-dissipative theorem:

$$\langle \delta n(k, \omega)\delta n(k', \omega)\rangle = -T\int dx dx' dt \exp(ikx)\exp(ik'x')\exp(i\omega t) \times \int_0^\infty \beta(x, x', t, \tau)\exp\left(\frac{\Delta E}{T}\right)d\tau. \tag{3.36}$$

Applying the Fourier transformation to this relation and using Eq. (3.9), we obtain the relaxation frequency spectrum in the form

$$\omega^{(\text{new})}(k) = \omega(k)\exp\left(-\frac{\Delta E}{T}\right). \tag{3.37}$$

Thus, the application of relation (3.34) instead of (2.39) leads to the emergence of an additional exponential term in relation (3.19),

$$j_i(k, t) = \frac{n_i(k, t)}{k}\omega_i(k)\exp\left(-\frac{\Delta E}{T}\right), \qquad (3.38)$$

where ΔE is the depth of the potential well occupied by the system. In the case of small fill factors of the channel, the state of the system is homogeneous and $\Delta E = 0$ (see Fig. 1a). Then relation (3.38) transforms into (3.19). As the fill factor of the channel increases, the state of the system becomes clustered. For $\Delta E \ll T$, the exponential in relation (3.38) can be disregarded, and the flux virtually coincides with the flux calculated for a homogeneous state. This is due to the fact that expression (2.48) for the lifetime of clusters practically coincides with expression (2.47) for the lifetime of clusters in a homogeneous state. In this case, as noted above, density excitation propagates via a cluster, leading to an increase in the effective diffusion coefficient. A further increase in the fill factor of the channel increases the value of ΔE. The flux thereby decreases, which can be explained by an increase in the lifetime of the clusters formed in the channel. As a result, the mechanism of excitation transport via a cluster is not realized.

Thus, for small fill factors of the channel, the transport in the system follows the diffusion mode. This can be demonstrated by passing to the limit of small fill factors θ in relation (3.16). Then the second terms in the denominator become insignificant and the spectra are reduced to diffusion spectra (3.11). An increase in the fill factor results in the formation of clusters. It was shown in the previous section that, depending on the fill factor of the channel, the clusters formed may be either short-lived or stable in the case when the ground state of the system is clustered. For short-lived clusters (see Fig. 1b), the transport between clusters occurs via diffusion, while the transport over clusters occurs via a rapid barrier-free transfer of density excitation, and the arrival of a particle from one side of a cluster leads to the emergence of a particle from the other side [3]. In this case, the spectra contain both the diffusion and the hydrodynamic mode, and the effective diffusion coefficient increases in accordance with Eq. (3.33). An increase in the fill factor of the channel reduces the distance between clusters and increases their size. Since the transport over a cluster is faster than the diffusion transport, the effective diffusion coefficient and the partial flux increase. However, an increase in the fill factor also increases the lifetime of the clusters formed in the channel; as a result, the rate of excitation transfer over a cluster decreases. In the case when a cluster is stable (see Fig. 1d), the transport of excitation over the cluster is ruled out and the presence of such a cluster in the channel leads to blockage of transport in the system.

Using relations (3.16) and (3.38), we can calculate the partial gas fluxes proceeding from the data on the penetrability for pure components. Figures 5a and 5b show the dependence of partial fluxes on the mixture composition in the case when both gases are slightly adsorbed. The energies of interaction and the diameters of particles correspond to a methane–argon mixture (a = 3.8 Å, σ_1 = 3.6 Å, σ_2 = 3.0 Å, ε_1 = 0.38 eV, ε_2 = 0.25 eV [15], P = 100 kPa, and T = 300 K).

It can be seen from Fig. 5a that an increase in the concentration of the first component for fixed pressure and temperature reduces the degree of channel filling with the second component, while the degree of channel filling with the first component increases. A decrease in the total fill factor of the channel in the concentration range $0 < c < 0.4$ of first component takes place due to depletion of the channel in the second component (Fig. 5b). An increase in the total fill factor of the channel for $c > 0.4$ is associated with preferred enrichment of the channel in the first component (see Fig. 5b). The partial flux of the first component in this case increases monotonically due to an increase in the degree of channel filling with the first component (see Fig. 5a), while the partial flux of the second component decreases due to a decrease in the degree of channel filling with the second component. The transport is of the diffusion type, and clusters are not formed in the channel in view of small fill factors of the channel ($\theta \le 0.25$ for any composition of the mixture).

Figures 5c and 5d show the dependence of partial fluxes on the mixture composition in the case when one of the gases (first) is a strong sorbate, while the other gas is a weak sorbate. The energy of interaction and the diameter of particles correspond to a butane–methane mixture (a = 3.8 Å, σ_1 = 4.3 Å, σ_2 = 3.6 Å, ε_1 = 0.61 eV, ε_2 = 0.38 eV [15], P = 2000 kPa, and T = 300 K). The behavior of the total and partial fill factors of the channel upon an increase in the concentration of the first (highly adsorbed) component in this case coincides qualitatively with the case of slightly adsorbed gases described above: the partial degree of channel filling with the first (highly adsorbed) component increases monotonically against the background of a monotonic decrease in the degree of channel filling with the second (slightly adsorbed) component (see Fig. 5c). The behavior of partial fluxes in this case is less trivial. It can be seen from the figure that the partial flux of the first component increases in the concentration interval $0 < c < 0.2$. This is due to the fact that, in accordance with relation (2.47), short-lived clusters (see Fig. 1a) whose size increases with concentration are formed for such concentrations and fill factors of the channel. The transport over clusters occurs via a barrier-free transfer of density excitation, which increases the effective diffusion coefficient. This process is similar to the transport in one-component systems for high fill factors [3]. It should be noted, however, that the formation of clusters and the mechanism of transfer of density excitation

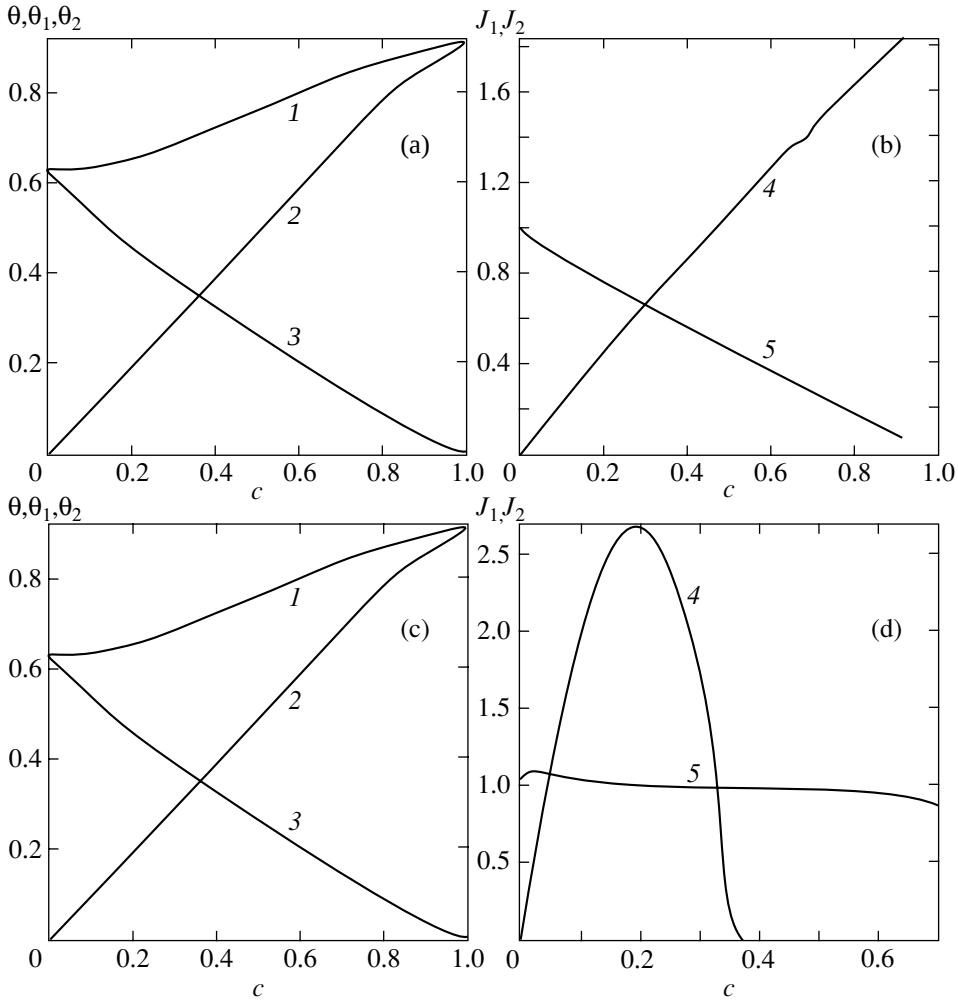

Fig. 5. Dependence of the total and partial fill factors of a channel (a, c) and partial fluxes (b, d) on the composition of methane–argon (a, b) and butane–ethane mixtures (c, d): *1*—total fill factor of a channel; *2*, *3*—degrees of channel filling with the first and second components, respectively; *4*, *5*—partial fluxes of the first and second components, respectively; *c* is the concentration of the first component in the mixture.

through a cluster in a two-component mixture can be realized for smaller fill factors ($\theta = 0.71$; see Figs. 1b, 5c, and 5d) as compared to the one-component case ($\theta \geq 0.8$). An increase in the concentration of the first component increases the lifetime of the clusters formed in the system (see Figs. 2b and 3c). As a result of this increase, the transport over clusters becomes slower and the flux decreases. A further increase in the concentration leads to the formation of stable clusters in the channel (see Figs. 2c and 4c); in accordance with relation (2.28) the lifetime of these clusters is much longer than the lifetime (2.47) of short-lived clusters since $\delta F \gg T$. In this case, the transfer of density excitation over clusters is ruled out and the partial flux vanishes.

Thus, three mechanisms of particle transport in a channel are possible in the two-component case. For small fill factors, the transport occurs via diffusion. An increase in the fill factor leads to the formation of short-lived clusters in a channel. As in the one-component case, this leads to an increase in the effective diffusion coefficient due to barrier-free transport of density excitation over a cluster. As the fill factor increases further, the characteristic lifetime of clusters increases, leading to a decrease in the flux. A further increase in the fill factor makes the clusters stable, and the so-called blocking effect takes place, when the partial flux of one of the components vanishes.

A decrease in pressure does not lead to a qualitative change in the behavior of the dependences of partial fluxes on the mixture composition.

4. COMPARISON WITH EXPERIMENT

The transport of two-component gas mixtures in MFI zeolite membranes (Silicalite, ZSM-5) with a pore diameter of ~ 0.6 nm has been studied by the technique widely used for one-component gases [15]. However, in contrast to rich variety of experiments with one-com-

ponent gases, a consistent analysis of a two-component mixture has not been reported in the literature. The most detailed information on penetrability for two-component mixtures is given in [15, 16].

The experiments [15, 16] on penetration of two-component mixtures were made in a chamber with a porous stainless steel substrate with a thickness of approximately 3 mm. A polycrystalline layer of zeolite ZSM-5 with a thickness of ~ 50 μm and a random orientation of crystals was deposited on the substrate. Zeolites have a complex crystalline structure based on silicon oxide with admixtures of sodium and aluminum [15]. The crystalline structure of zeolite ZSM-5 is formed by straight channels with an elliptical cross section (0.57×0.52 nm^2) intersecting sinusoidal channels with a circular cross section of diameter 0.54 nm [23]. The measured adsorption capacity and the calculated fraction of molecules in the channel intersections make it possible to treat the channels as one-dimensional for the molecular mixtures under study [23]. The experiments [15, 16] were carried out at temperatures from 300 to 700 K under pressures from 25 to 500 kPa on the external side of the membrane. The purity of the mixture components was higher than 99.95%. The mixture composition at the membrane exit was measured on a quadrupole mass spectrometer with a sensitivity of approximately 25 ppm when a Faraday cell was used for recording.

The selectivity measured in the experiments was determined as the relative change in concentration c of n-butane molecules and concentration $1 - c$ of methane molecules at the membrane entrance (F) and exit (P):

$$\alpha = \left(\frac{c}{1-c}\right)_F \Big/ \left(\frac{c}{1-c}\right)_P. \qquad (4.1)$$

It should be noted that, for a constant composition of the mixture at the entrance to the membrane, the selectivity is proportional to the ratio of partial fluxes of the mixture components at the membrane exit:

$$\alpha \propto J_1/J_2. \qquad (4.2)$$

It was found experimentally [15] that the partial flux of the highly adsorbed gas n-C_4H_{10} at $T = 300$ K under a total pressure of $P = 100$ kPa changes considerably (by a factor of several units) in the presence of the slightly adsorbed component CH_4 as compared to the flux of a pure gas, while the partial flux of the slightly adsorbed gas changes by two orders of magnitude in the presence of the strongly absorbed component. Depending on the mixture composition, selectivity α varies and attains its maximum value $\alpha_{max} = 380$ for a concentration ratio of 5 : 95 at the entrance to the membrane.

It was shown [16] that the dependences of partial fluxes on the mixture composition and external pressure for the mixtures studied (C_2H_6–CH_4 and C_3H_8–CH_4) are monotonic in the range of experimental pressures. The partial flux of methane decreases in the presence of the second component (ethane or propane), while the partial flux of the second component increases with its concentration in the mixture. The selectivity of both mixtures increases monotonically with the concentration of the more highly adsorbed component in the mixture. The dependences of the partial fluxes on the total pressure of the mixture are also monotonic, but the pressure dependence of selectivity in ethane for the mixture C_2H_6–CH_4 has a peak at a pressure of $P \approx 300$ kPa [16]. Similar dependences of partial fluxes were observed in experiments with another mixture of slightly adsorbed gases (CO_2–N_2) [19] (Fig. 6).

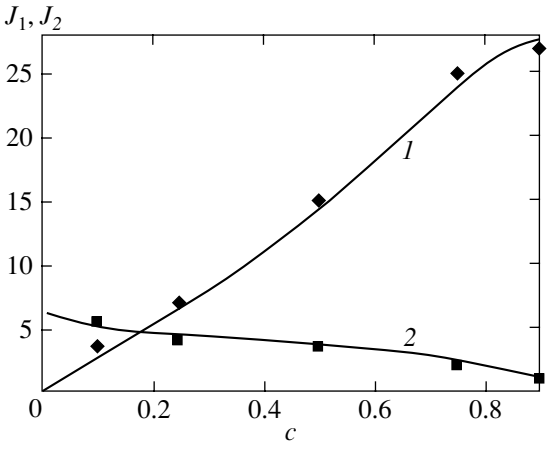

Fig. 6. Dependence of partial fluxes on the mixture composition for slightly adsorbed gases: *1*, *2*—theoretical dependences for CO_2 and N_2 fluxes, respectively; c is the CO_2 concentration above the membrane; squares correspond to experimental data from [19]; $a = 3.8$ Å, $\sigma_1 = 3.8$ Å, $\sigma_2 = 3.6$ Å [15]; $P = 100$ kPa, $T = 300$ K.

The theory developed in the previous sections enables us to describe peculiarities in the transport of binary gas mixtures in zeolite membranes. In the case of a mixture of slightly adsorbed gases, when the fill factor of a channel is small ($\theta \ll 1$) for any mixture composition, the system relaxes in accordance with the diffusion mechanism, corresponding to diffusion of solitary particles without the formation of clusters in a channel. In this case, in accordance with Eqs. (2.12), (3.16), and (3.38), the partial flux of the first component decreases monotonically upon an increase in the concentration of the second component, while the partial flux of the second component increases with its concentration in the mixture.

Figure 6 shows that the theoretical curves calculated by formulas (2.12), (3.16), and (3.38) are in good agreement with the experimental data borrowed from [19]. The mean distance a between the seats in a channel was estimated proceeding from the value cor-

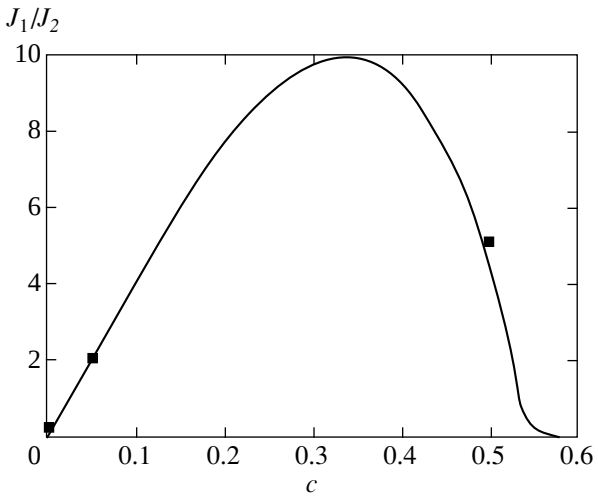

Fig. 7. Ratio of the partial fluxes of butane and methane as a function of the butane concentration; c is the concentration of n-C_4H_{10} above the membrane; squares correspond to experimental data obtained in [15]; a = 3.8 Å, σ_1 = 4.3 Å, σ_2 = 3.6 Å [15]; P = 100 kPa, T = 300 K.

responding to close packing of a unit cell of a ZSM-5 membrane with nitrogen [23].

The degree of channel filling for mixtures with a strongly absorbed component depends on the mixture composition; it is significant and can be as high as $\theta \sim 1$. In this case, in accordance with the arguments put forth in Section 3, the dependence of fluxes on the mixture composition is nonmonotonic due to the formation of clusters in a channel. Figure 7 shows how the ratio of partial fluxes of gases for the mixture n-C_4H_{10}–CH_4 depends on the butane concentration in the mixture under a pressure of P = 100 kPa at temperature T = 300 K. It can be seen from the figure that the ratio of partial fluxes increases with the concentration of the highly adsorbed component (butane) in the concentration range $0 < c < 0.35$. Analysis of experimental data [15] and calculations based on formulas (2.12), (3.16), and (3.38) show that the methane flux for a butane concentration of $c > 0.05$ changes insignificantly upon an increase in the butane concentration in the mixture. In this case, the behavior of the flux ratio is completely determined by the behavior of the butane flux. The increase in the butane flux is due to the fact that, in accordance with formula (2.47), short-lived butane clusters are formed for concentrations of $0 < c < 0.35$ and for the fill factors corresponding to these concentrations (see Fig. 1a); the size of these clusters increases with the concentration of the highly adsorbed component. The transport over clusters (see Fig. 1b) occurs via barrier-free transfer of density excitation, leading to an increase in the effective diffusion coefficient for butane, which is observed for $c < 0.35$. The increase in the butane concentration in the mixture increases the lifetime of the clusters formed in the system (see Figs. 2b and 3c). With increasing lifetime, the transport over the clusters slows down and the butane flux decreases for concentrations of $c > 0.35$. A further increase in the concentration results in the formation of stable clusters in a channel (see Figs. 2c and 4c), leading to a rapid (exponential) decrease in the butane flux upon an increase in its concentration in the mixture.

Thus, it follows from the theory that the butane flux becomes exponentially small for a butane concentration of $c > 0.6$ in the mixture and the transport of particles must be blocked. It should be noted that calculations based on formulas (2.12), (3.16), and (3.38) for butane concentrations in the interval $0 < c < 0.05$ show that the methane flux decreases approximately by an order of magnitude upon an increase in the butane concentration, which is also in qualitative agreement with the experimental data [15]. The numerical difference between the predictions of the theory developed above

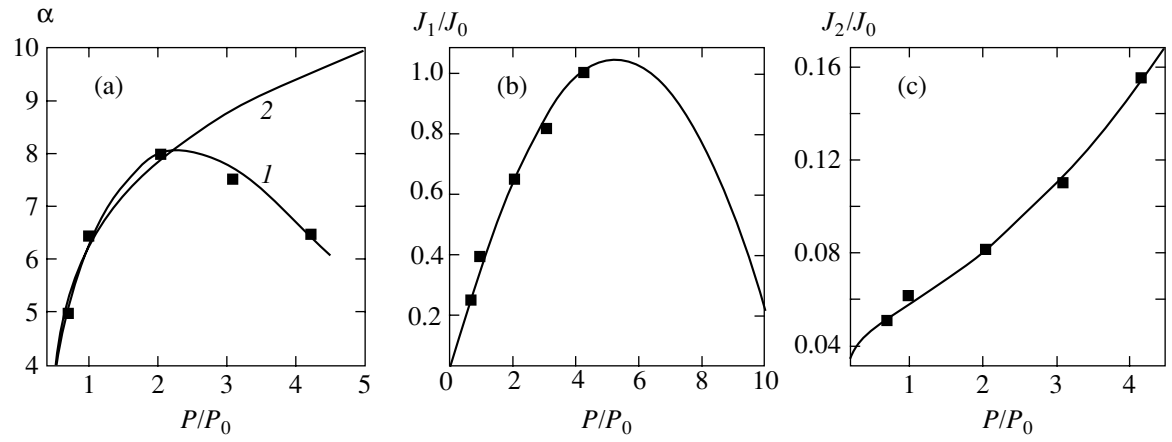

Fig. 8. Dependence of the ethane selectivity (a) and of the partial fluxes of C_2H_6 (b) and CH_4 (c) for the C_2H_6–CH_4 mixture on the total pressure of the mixture: *1*—calculations based on formulas (2.12), (3.16), (3.38), and (4.1); *2*—calculations based on the generalized Maxwell–Stefan equation [16]; squares correspond to experimental data obtained in [16]; a = 3.8 Å, σ_1 = 3.8 Å, σ_2 = 3.6 Å [15]; J_0 = 63 mmole/(m^2 s), P_0 = 100 kPa.

and the experimental data in this range of butane concentrations is due to the fact that relations (2.12), (3.16), and (3.38) were derived for the model potential of intermolecular interaction of the type of interaction (2.9) between hard spheres. It is well known that the potentials of hard spheres correctly describe the behavior of real systems for high densities [17, 18]. In order to describe the behavior of the system in the entire range of fill factors, potentials of the Lennard–Jones type, which take into account the attraction between particles even in the zeroth order in density, should be used. However, although the application of this type of potentials of intermolecular interaction does not qualitatively change the results, it leads to the formation of butane clusters at lower degrees of channel filling and, hence, to a sharper decrease in the methane flux upon an increase in the butane concentration in the concentration range under investigation.

Relations (2.12), (3.16), (3.38), and (4.1) also make it possible to derive the dependence of the selectivity on the pressure, temperature, and composition of the mixture. Figure 8a shows how the ethane selectivity for the C_2H_6–CH_4 mixture with a concentration ratio of 50 : 50 at the entrance depends on the total pressure of the mixture [16]. It can be seen that the dependence is nonmonotonic. This is due to the fact that the diffusion coefficient first increases due to an increase in the pressure and fill factor of the channel as a result of the formation of short-lived clusters and that the partial ethane flux increases (Fig. 8b). The increase in the total fill factor of the channel is determined by the increase in the partial fill factors for both components. This process continues until the channel filling with ethane slows down. In this case, the fill factor of the channel increases mainly due to channel filling with methane. At such pressures, the increase in the ethane flux slows down (see Fig. 8b), while the methane flux continues to increase (Fig. 8c). The theoretical dependences plotted in the figures show that a further increase in pressure due to an increase in the lifetime of ethane clusters formed in the channel must lead to a decrease in the ethane flux (Fig. 8b). Thus, the analysis of experimental data proves that the theory constructed here describes the experimental results satisfactorily.

One of the main results of the proposed theory is the prediction of nonmonotonic dependences of partial fluxes and selectivities on the mixture composition and pressure (see Figs. 7 and 8), while conventional models [15, 16, 19] (in particular, the generalized Maxwell–Stefan equation [15, 16]) lead to monotonic dependences of these quantities on the same parameters (see Fig. 8a). The qualitative discrepancy between the dependences in question is due to the fact that conventional models [15, 16, 19] take into account only the finite size of particles, while the interaction between particles in a channel is disregarded. It is well known [17, 18, 26], however, that the interaction between particles in dense systems plays a decisive role both in the construction of the equations of state of the system and in the description of transport. For example, the inclusion of interaction between particles of the hard-sphere type leads to the emergence of peaks in the pair correlation function at distances equal to one, two, three, etc., particle diameters [17, 18, 27]. This fact indicates the existence of "effective" attraction between particles and necessitates the inclusion of cluster formation in the description of the behavior of the system at high densities, when it passes to a spatially inhomogeneous state. It was demonstrated above that this leads to the emergence of nonmonotonic dependences of fluxes and selectivities on external parameters.

It should be noted that the stabilization of clusters in 1D systems by the second component at high fill factors, which has been investigated here, was obtained numerically in a recent work [7]. It should be borne in mind, however, that, in accordance with our previous results [3], clusters with a large but finite lifetime can be formed even in a one-component system with high degrees of channel filling. It was 1D gold clusters of this type that were apparently observed in recent experiments under the ultrahigh vacuum conditions [9].

ACKNOWLEDGMENTS

The authors are grateful to L.A. Maksimov, N.I. Laguntsov, V.V. Volkov, M.V. Tsodikov, V.V. Teplyakov, F. Kapteijn, and L. Gora for fruitful discussions of the results obtained in this study. One of the authors (I.V.T.) thanks the NWO (grant no. 03-03-89002) for partial financial support.

REFERENCES

1. S. Yu. Krylov, A. V. Prosyanov, and J. J. M. Beenakker, J. Chem. Phys. **107**, 6970 (1997).
2. J. J. M. Beenakker, V. D. Borman, and S. Yu. Krylov, Phys. Rev. Lett. **103**, 4622 (1995).
3. V. D. Borman, V. V. Teplyakov, V. N. Tronin, *et al.*, Zh. Éksp. Teor. Fiz. **117**, 1094 (2000) [JETP **90**, 950 (2000)].
4. S. Vasenkov and J. Karger, Phys. Rev. E **66**, 052601 (2002).
5. K. Hahn, H. Jobic, and J. Karger, Phys. Rev. E **59**, 6662 (1999).
6. C. Rodenbeck, J. Karger, and K. Hahn, Phys. Rev. E **57**, 4382 (1998).
7. N. V. Skorodumova and S. I. Semak, Phys. Rev. B **67**, 121404 (2003).
8. V. D. Borman, S. Yu. Krylov, and A. V. Prosyanov, Zh. Éksp. Teor. Fiz. **97**, 1795 (1990) [Sov. Phys. JETP **70**, 1013 (1990)].
9. V. Rodrigues and D. Ugarte, Phys. Rev. B **63**, 073405 (2001).
10. K. K. Sirkar, Chem. Eng. Commun. **157**, 145 (1997).
11. C. L. Flanders, V. A. Tuan, R. D. Noble, and J. L. Falconer, J. Membr. Sci. **176**, 43 (2000).
12. T. C. Bowen, H. Kalipcilar, J. L. Falconer, and R. D. Noble, J. Membr. Sci. **215**, 235 (2003).

13. S.-T. Hwang and K. Kammermeyer, *Membranes in Separations* (Wiley, New York, 1975; Khimiya, Moscow, 1981).
14. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon Press, Oxford, 1980), Part 1.
15. W. J. W. Bakker, G. Zheng, F. Kapteijn, *et al.*, in *Precision Process Technology: Perspectives for Pollution Prevention*, Ed. by M. P. C. Weijnen and A. A. H. Drinkenburg (Kluwer Academic, Dordrecht, 1993), p. 425.
16. J. M. Van de Graaf, F. Kapteijn, and J. A. Moulijn, AIChE J. **45**, 497 (1999).
17. R. Balescu, *Equilibrium and Non-Equilibrium Statistical Mechanics* (Wiley, New York, 1975; Mir, Moscow, 1978), Vols. 1 and 2.
18. G. N. Sarkisov, Usp. Fiz. Nauk **169**, 625 (1999) [Phys. Usp. **42**, 545 (1999)].
19. L. J. P. Van den Broeke, W. J. W. Bakker, F. Kapteijn, and J. A. Moulijn, AIChE J. **45**, 976 (1999).
20. Yu. N. Devyatko and V. N. Tronin, Zh. Éksp. Teor. Fiz. **98**, 1570 (1990) [Sov. Phys. JETP **71**, 880 (1990)].
21. I. Z. Fisher, *Statistical Theory of Liquids* (Fizmatlit, Moscow, 1961; Univ. of Chicago Press, Chicago, 1964).
22. D. N. Voskresenskiĭ, in *Multiparticle Effects in Solids* (Énergoatomizdat, Moscow, 1983), p. 25.
23. R. E. Richards and L. V. C. Rees, Langmuir **3**, 335 (1987).
24. Yu. L. Klimontovich, *Statistical Physics* (Nauka, Moscow, 1982; Harwood Academic, New York, 1986).
25. O. G. Bakunin, Usp. Fiz. Nauk **173**, 317 (2003) [Phys. Usp. **46**, 309 (2003)].
26. P. Resibois and M. de Leener, *Classical Kinetic Theory of Fluids* (Wiley, New York, 1977; Mir, Moscow, 1980).
27. J. M. Ziman, *Models of Disorder: the Theoretical Physics of Homogeneously Disordered Systems* (Cambridge Univ. Press, Cambridge, 1979; Mir, Moscow, 1982).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 98, No. 1, 2004, pp. 123–134.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 1, 2004, pp. 136–149.

SOLIDS
Electronic Properties

The Mechanism of Interlayer Exchange Coupling in Iron–Chromium Type Nanostructures

V. N. Men'shov* and V. V. Tugushev**
Kurchatov Institute, Russian State Scientific Center, Moscow, 123182 Russia
**e-mail: vnmenshov@mail.ru;*
***e-mail: vvtugushev@mail.ru*
Received May 15, 2003

Abstract—A mechanism of the interlayer exchange coupling in layered structures of the Fe/Cr(001) type with rough interfaces is proposed. The theory is based on a model of the charge-induced spin density wave (SDW) formed in the chromium layer. It is shown that the effective magnetic coupling between thick ferromagnetic layers arises due to variations of the SDW vector orientation in the antiferromagnetic layer over a characteristic length ζ determined by the exchange stiffness of chromium. A general expression for the effective magnetic coupling energy $E(\psi)$ as a function of the angle ψ between magnetic moments of the ferromagnetic layers is obtained and numerically analyzed for an arbitrary value of the parameter $\rho\zeta$, where ρ is the density of monoatomic steps on the interface. For $\rho\zeta \gg 1$, the form of $E(\psi)$ is typical of a model with the "biquadratic" interaction, while in the case of $\rho\zeta \ll 1$, the dependence obtained differs significantly. The proposed mechanism is used to interpret the results of measurements of the interlayer exchange coupling in Fe/Cr(001) structures.

1. INTRODUCTION

It is well known that the quality of interfaces between ferromagnetic (FM) and antiferromagnetic (AFM) layers in magnetic multilayer structures of the Fe/Cr type is a factor determining the magnitude and character of the effective exchange coupling between the neighboring FM layers [1–6]. For elucidating the mechanisms of this influence, it is necessary to study in detail both the morphology of interfaces and the process of redistribution of the charge and spin densities of quasiparticles at these interfaces.

A comparative analysis of the properties of interfaces in the structures obtained using different technologies must take into account the different scales of fluctuations of the surface relief on the Fe/Cr boundaries. First, any method used for the growth of such multilayer structures leads to unavoidable small-scale fluctuations within several boundary atomic layers, caused by the mutual diffusion of atoms of the two metals in contact and by the breakage (frustration) of regular interatomic bonds [1, 2, 6]. Second, any technology involves more or less pronounced large-scale ("geometric") fluctuations of the surface relief, referred to in the literature as the interface "roughness" [1–3]. This roughness is usually described [1, 4, 5] in terms of two statistical characteristics, the dispersion $\sigma = \sqrt{\langle h^2 \rangle}$ (vertical roughness) and the correlation length R (lateral roughness) of fluctuations in the relief height h. The values of σ characterizing the degree of roughness in the direction perpendicular to the interface plane usually vary within 2–6 Å, while the lateral parameter R can vary within three orders of magnitude: from several nanometers (in the case of molecular beam epitaxy of a three-layer Fe/Cr/Fe(001) structure on an Ag(001)/Fe/GaAs(001) substrate [4]) to several microns (for the epitaxial layer growth of chromium on Fe(001) whiskers [1, 5]).

Special magnetic properties of Fe/Cr multilayer systems are related to a specific AFM order inherent in chromium, taking the form of a spin density wave (SDW). In a layered structure, this SDW is highly sensitive both to the presence of interfaces as such and to the structure of these boundaries. Previously [7, 8], we proposed a model of the magnetic ordering in layered structures of the Fe/Cr type with perfectly smooth interfaces. The order appears in these structures below a certain temperature T_0 that is significantly higher than the Néel temperature (T_N) of bulk chromium. According to this model (called the charge-induced SDW model), a short-range AFM order appears in the chromium layer due the charge density redistribution at the interfaces between iron and chromium. The results obtained in [8] showed that the SDW formed in the chromium layer aligns the magnetic moments of the neighboring Fe layers, rendering them collinear (parallel or antiparallel, depending on the number N of monolayers in the intermediate Cr layer).

We also considered a model problem of determining the energy-optimized configuration of a charge-induced SDW at the Fe/Cr interface containing one isolated monoatomic step (referred to below as a monostep) [9]. It was found that, provided the iron layers are sufficiently thick to exclude the formation of fer-

romagnetic domain walls, the most energetically favorable configuration is that with a 90° AFM domain wall formed inside the chromium layer. The character of the mutual orientation of magnetic moments in the neighboring iron layers changes to noncollinear, whereby the angle between these moments has a certain value different from 0 or π. It was naturally suggested that this very change of the SDW configuration in the Fe/Cr type structures with rough interfaces may account in many cases for the noncollinear ordering of magnetic moments of the neighboring Fe layers—a phenomenon extensively discussed in recent years [1–3].

In this study, aimed at verification of the above hypothesis, we have generalized the model [9] to the case of an arbitrary density of monosteps and a relatively thin AFM layer (with a thickness below the doubled "amplitude" correlation length $\xi(T)$, on which scale the SDW amplitude varies across the AFM layer). For such systems, it can be naturally assumed that the AFM order in the chromium layer is characterized by a considerable longitudinal magnetic stiffness (i.e., the SDW amplitude is almost constant across the AFM layer). On the other hand, our previous analysis [9] showed that the direction of the SDW polarization vector in the AFM layer changes along the interface over a characteristic "angular" correlation length $\zeta(T)$ and is highly sensitive to fluctuations in the exchange coupling at the Fe/Cr interface. In this paper, the interface structure is modeled by a system of flat regions (of an average length R) separated by monosteps, whereby the sign of the surface exchange potential exhibits a jumplike alternation upon crossing every step. By varying the ratio of the characteristic lengths $\zeta(T)$ and R, it is possible to consider the properties of structures with interfaces possessing various degrees of roughness. This analysis will show how and to what extent the equilibrium SDW configuration in the AFM layer determines the mutual (in the general case, noncollinear) orientation of magnetic moments in the neighboring iron layers.

By decreasing the FM layer thickness, it is possible to obtain an inhomogeneous magnetic configuration of the system other than that suggested in [9]. This is related to the formation of a 180° FM domain wall in the iron layer, whereas a 90° AFM domain wall in the chromium layer is not formed. Below we will consider this situation in more detail to show that the system geometry, together with the quality of interfaces, determines to a considerable extent the criteria of applicability of the proposed model to real layered structures of the Fe/Cr type.

2. THE MODEL OF ANTIFERROMAGNETIC DOMAIN WALLS IN STRUCTURES WITH HOMOGENEOUS MAGNETIZATION OF FERROMAGNETIC LAYERS

Using the approach developed previously [8, 9], we will consider the simplest structure-forming element of the Fe/Cr system—a three-layer structure (trilayer) comprising two FM layers (Fe) separated by an AFM spacer (Cr). Technological Fe/Cr interfaces are parallel to the $\mathbf{n}_y\mathbf{n}_z$ plane; the normal $\mathbf{n}_x$ to this plane coincides with the growth direction and is parallel to the [100] cubic axis ($\mathbf{n}_x$, $\mathbf{n}_y$, $\mathbf{n}_z$ are the unit basis set vectors). We consider the temperature interval T corresponding to a short-range AFM order in the chromium layer: $T_N < T < T_0$ ($T_0 \ll T_C$, where T_C is the Curie temperature for the iron layers). The FM layers are assumed to be sufficiently thick, so that the magnetization S inside these layers at $T_0 \ll T_C$ can be considered as homogeneous and independent of temperature. The AFM spacer thickness L can be varied within rather broad limits, but so that $L > 2\xi_0$, where ξ_0 is the coherence length ranging, according to various estimates, from seven to ten Cr monolayers. In the temperature interval under consideration, the sublattice magnetization $\boldsymbol{\sigma}(\mathbf{r})$ in the AFM layer can be significantly inhomogeneous (depending on the layer thickness L) and strongly dependent on temperature [7, 8].

Let us introduce an order parameter (with the dimension of energy) describing the SDW envelope, $\boldsymbol{\Delta}(\mathbf{r}) = U\boldsymbol{\sigma}(\mathbf{r})$, where U is the effective SDW potential (the explicit form of which is not discussed here; for this see, e.g., review [10]). The considerations will be restricted to the case of a transversely polarized SDW, whereby $\boldsymbol{\sigma}(\mathbf{r}) \perp \mathbf{n}_x$ (this situation, corresponding to an experimental situation with not too thick AFM spacers, $L \le 100$ Å, has been studied most frequently [1–6]). In this case, the order parameter can be written as

$$\boldsymbol{\Delta}(\mathbf{r}) = \mathbf{n}_y\Delta_y(\mathbf{r}) + \mathbf{n}_z\Delta_z(\mathbf{r}), \tag{1}$$

where $\mathbf{r} = (x, y, z)$; $|x| \le l$ (l is a half of the AFM spacer thickness, $L = 2l$); $|y|, |z| \le l_\perp$ ($2l_\perp$ is the spacer size in the $\mathbf{n}_y$ and $\mathbf{n}_z$ directions, $l_\perp \gg l$). Assuming that the $\boldsymbol{\Delta}(\mathbf{r})$ value is small ($|\Delta| \ll \pi T$) and slowly varying in space ($|\partial\boldsymbol{\Delta}/\partial\mathbf{r}| \ll \pi T/\xi_0$), let us write an expression for the thermodynamic potential $F[\boldsymbol{\Delta}]$ of the AFM layer in the form of a Ginzburg–Landau expansion in powers of the $\boldsymbol{\Delta}(\mathbf{r})$ function and its derivatives. Since this approach was thoroughly justified in [8, 9], we omit the comments and present the final expression for the functional $F[\boldsymbol{\Delta}]$:

$$F = F_v + F_s, \tag{2}$$

$$F_v = \frac{1}{2}\int f_v(\mathbf{r})dxdydz, \tag{3}$$

$$f_v = c_1\boldsymbol{\Delta}^2 + c_2 v_F^2\left(\frac{\partial\boldsymbol{\Delta}}{\partial\mathbf{r}}\right)^2 + c_2\boldsymbol{\Delta}^4, \tag{4}$$

$$\begin{aligned} F_s = {} & \frac{\nu}{4}\int dydz(\boldsymbol{\Delta}^2(l, y, z) + \boldsymbol{\Delta}^2(-l, y, z)) \\ & + \frac{A}{2}\int dydz(\eta(l, y, z)\mathbf{m}(l, y, z)\boldsymbol{\Delta}(l, y, z) \\ & + \eta(-l, y, z)\mathbf{m}(-l, y, z)\boldsymbol{\Delta}(-l, y, z)). \end{aligned} \tag{5}$$

Here, the integration is performed within the AFM spacer boundaries as denoted above; the values of $x = \pm l$ in Eq. (5) correspond to the right- and left-hand interfaces. The quantities F_v and F_s have the meaning of the volume and surface parts of the total thermodynamic potential. Expressions for the coefficients c_1, c_2, ν, and A are given in [9, 10]; these values were calculated previously (see, e.g., [10]); in what follows, $c_1, c_2 > 0$, $\nu < 0$, and $A > 0$ (these conditions correspond to the AFM exchange at the Fe/Cr interface). The parameter A is proportional to the magnetization of the FM layer. The value of this magnetization is assumed to be constant over each iron layer, while its orientation $\mathbf{m}(\pm l, y, z)$ ($|\mathbf{m}| = 1$) in the general case is a function of the coordinates. The quantity v_F in Eq. (4) is the projection of the electron velocity onto the growth direction $\mathbf{n}_x$ in flat regions of the Fermi surface of chromium (these regions are responsible for the SDW formation in the octahedral model [11]).

The expansion according to Eqs. (3) and (4) is valid, strictly speaking, in almost the entire region of $|x| < l$, except for the AFM spacer regions with a width on the order of ξ_0 at the interfaces, where the local approximation (2) for the functional $F[\mathbf{\Delta}]$ becomes incorrect. Fine details of the charge and spin distributions on such a scale are not described within the framework of our approach: these details can be considered as implicitly taken into account by the coefficients ν and A. The "exchange" term linear in $\mathbf{\Delta}$ appearing in Eq. (5) is directly related to the exchange interaction between spins of the FM and AFM layers. The "Coulomb" term quadratic in $\mathbf{\Delta}$ reflects the charge transfer between the layers of different metals (Fe vs. Cr) and the resulting contact potential difference arising between the FM and AFM layers. According to the estimates obtained in [8, 9], the Coulomb term predominates over the exchange term in a broad temperature range $T > T_N$ and determines both the characteristic temperature T_0 for the short-range AFM order formation and the SDW amplitude. The SDW is induced due to an increase in the electron spin susceptibility of chromium near the interface. The ratio of $\xi/D = \tanh(l/\xi)$, where $\xi = v_F\sqrt{c_2/c_1}$ is the AFM correlation length and $D = 2c_2 v_F^2/|\nu|$ is the spatial scale of the charge density redistribution at the interface, characterizes the critical temperature $T_0(L)$ as dependent on the AFM spacer thickness. This dependence shows a quite good correlation with a real phase diagram of the Fe/Cr(001) structure [1, 2]. No difficulties are encountered when the proposed model is used for the interpretation of variations of the critical temperatures T_0 and T_N caused by the introduction of dopants into the chromium layer [12].

Despite a relatively less pronounced influence on the SDW amplitude, the exchange term determines to a considerable extent details of the spatial distribution of the SDW and its orientation relative to the magnetization $S\mathbf{m}(\pm l, y, z)$ in the FM layers. This term plays an important role in the description of a layered structure with not perfectly smooth (i.e., rough) interfaces at which both small- and large-scale fluctuations of the charge and spin potentials unavoidably take place. Averaging over the small-scale fluctuations can be, in principle, performed within the framework of a standard model describing the SDW interaction with point impurities [10, 11], but allowance for the large-scale fluctuations presents a more complicated problem. The results of experiments [4] showed that a shortwave (with a period of two monolayers) component of the magnetic coupling between Fe layers is determined to a considerable extent by compact regions of the Cr spacer of a constant thickness, with a characteristic lateral size on the order of 3–4 nm. Indeed, the exchange contribution to the surface energy F_s of a structure with the ideal flat interfaces sharply changes (in contrast to the Coulomb contribution) its sign when the Cr spacer thickness is varied by only one monolayer [1–3]. This dependence of the energy on the evenness of the number N of monolayers in the spacer allows the long-range fluctuations in the Cr layer thickness to be simply modeled by introducing the random factors $\eta(\pm l, y, z)$ into Eq. (5). According to this model, the interface consists of ideal flat terraces on which η has a constant value of +1 or –1. The boundaries between adjacent terraces represent monoatomic steps; crossing such a step changes the sign of η to opposite. The steps are randomly distributed over the yz plane, while being oriented along $\mathbf{n}_y$ and $\mathbf{n}_z$ axes coinciding with the easy magnetization axes of the Fe/Cr(100) structure possessing a bcc crystal lattice. This model structure is consistent with empirical data on the Fe/Cr(100) interface morphology [4].

Previously [9], we have performed a self-consistent calculation of the thermodynamically equilibrium states of the functional determined by Eqs. (2)–(5) in the temperature range of $T > T_N$ both for the ideal flat Fe/Cr interface and for the interface with isolated monosteps. In the former case, the SDW polarization vector in the Cr spacer and the magnetic moments of both Fe layers always lie in the same (e.g., xz) plane–in other words, the collinear state is preferred. In the latter case, the density of monosteps on the interface was assumed to be small enough to ignore the contribution to the total system energy related to the inhomogeneity-induced transverse (relative to $\mathbf{n}_x$) SDW deformations. It was established that such AFM spacer thickness fluctuations give rise to a noncollinear configuration of the magnetic moments of Fe layers. Below, these results are generalized so as to include the case of a relatively thin AFM layer ($\xi_0 < l \ll \xi(T)$) and an arbitrary density of monosteps in the yz plane.

Finding equilibrium three-dimensional configurations of the system described by the functional according to Eqs. (2)–(5) in the general case is an extremely difficult (if not desperate) task. Therefore, some assumptions have to be made in order to simplify the

problem to an acceptable level while retaining the physical meaning of the results. We will consider only coplanar magnetic configurations with both $\mathbf{\Delta}(\mathbf{r})$ and $\mathbf{m}(\pm l, y, z)$ vectors lying in the interface plane. Of most interest is the limiting case of thick FM layers, whereby both these layers can be considered as homogeneously magnetized, so that $\mathbf{m}(\pm l, y, z) = \mathbf{m}(\pm l)$, although in the general case $\mathbf{m}(l) \neq \pm\mathbf{m}(-l)$. Note that, in the absence of the exchange interaction at the interfaces ($A = 0$), the ground state of the system is strictly described by a scalar SDW with a one-dimensional symmetric envelope $\Delta(\mathbf{r}) = \Delta_+(x)$ [9]. Switching on a weak exchange ($A \ll 1$) can influence the structure and parameters of the ground state in a system with perfectly smooth interfaces provided only that the AFM spacer is sufficiently thick ($l > D$) and/or the temperature is sufficiently high ($T \geq T_0$). Otherwise, in the region $\{l < D, T < T_0\}$, only a small correction (proportional to A^2) to the $\Delta_+(x)$ value appears. In this context, we will restrict the consideration of the Fe/Cr structures with rough interfaces to the region of Cr spacer thicknesses and temperatures $\{l < D, T < T_0\}$, in which the SDW amplitude can be considered as independent of the parameter A and almost constant across the Cr spacer ($|x| < l$) [9]:

$$\Delta(\mathbf{r}) = \Delta_+(x) = \frac{v_F}{\sqrt{2}}\left(\frac{1}{lD} - \frac{1}{\xi^2}\right)^{1/2}, \quad l/D \ll 1. \quad (6)$$

On the other hand, as was pointed out above, the SDW orientation is highly sensitive to jumps of the exchange potential related to the Cr spacer thickness variations. In the approximation adopted, variation of the transverse coordinates y, z is accompanied over the entire spacer thickness only by rotation of the vector $\mathbf{\Delta}(\mathbf{r})$ (1) without a change in its magnitude. The thermodynamic potential (2)–(5) reduces to a one-parametric functional for the static SDW orientation fluctuations:

$$F = F_A + F_\phi, \quad (7)$$

$$F_A = -c_2 v_F^4/4lD^2, \quad (8)$$

$$F_\phi = lc_2 v_F^2 \Delta^2 \int\limits_{\{\text{total}\}} dydz\left(\left(\frac{\partial\phi}{\partial y}\right)^2 + \left(\frac{\partial\phi}{\partial z}\right)^2\right) - A\Delta\left[\cos\left(\frac{\Psi}{2}\right)\int\limits_{\{\text{odd}\}} dydz\cos\phi + \sin\left(\frac{\Psi}{2}\right)\int\limits_{\{\text{even}\}} dydz\sin\phi\right]. \quad (9)$$

Here, F_A is a part of the functional F that is independent of the angle of rotation $\phi = \phi(y, z)$ of the SDW vector. The function $\phi(y, z)$ is introduced as follows:

$$\Delta_x(r) = 0, \quad \Delta_z(r) = -\Delta\cos\phi(y, z), \quad \Delta_y(r) = -\Delta\sin\phi(y, z), \quad (10)$$

where Δ is determined by formula (6). The vector $\mathbf{m}(\pm l)$ is defined as

$$m_x(\pm l) = 0, \quad m_y(\pm l) = \pm\sin\left(\frac{\Psi}{2}\right), \quad m_z(\pm l) = \cos\left(\frac{\Psi}{2}\right), \quad (11)$$

where Ψ is the angle between the magnetization directions in the FM layers. In Eq. (9), integration in the first term is performed over the whole interface, while integrals in the second and third terms are taken over the fragments containing odd and even numbers N of Cr monolayers, respectively. Figure 1 schematically shows the geometry and the magnetic structure of a Fe/Cr/Fe(001) trilayer in the vicinity of a monostep.

3. THE SHORT-RANGE ANTIFERROMAGNETIC ORDER STRUCTURE AT AN ARBITRARY DENSITY OF MONOSTEPS ON THE INTERFACE

Writing equations for minimization of the functional (7)–(9) is quite a simple task. However, even if it were possible to construct the exact solutions for a given configuration of $\eta(\pm l, y, z)$ determining the fragments of interfaces with odd and even N, the procedure of averaging of the interlayer exchange coupling over a random distribution $\{\eta(\pm l, y, z))\}$ cannot be performed in practice and requires making additional simplifying assumptions. We will assume that (i) one of the two interfaces is perfectly smooth, (ii) the Cr spacer thickness exhibits variations only in one direction ($\mathbf{n}_z$), and (iii) these variations are periodic; that is, for

$$n(l_o + l_e) < z < l_o + n(l_o + l_e),$$

we have

$$\eta(-l, y, z) = 1, \quad \eta(l, y, z) = 1,$$

and for

$$l_o + n(l_o + l_e) < z < n(l_o + l_e),$$

we have

$$\eta(l, y, z) = -1.$$

Here, n is an integer and l_o and l_e are the lengths of the fragments with odd and even N, respectively. The above assumptions render the problem one-dimensional, restrict the domain of the variable to the segment $-l_e \leq$

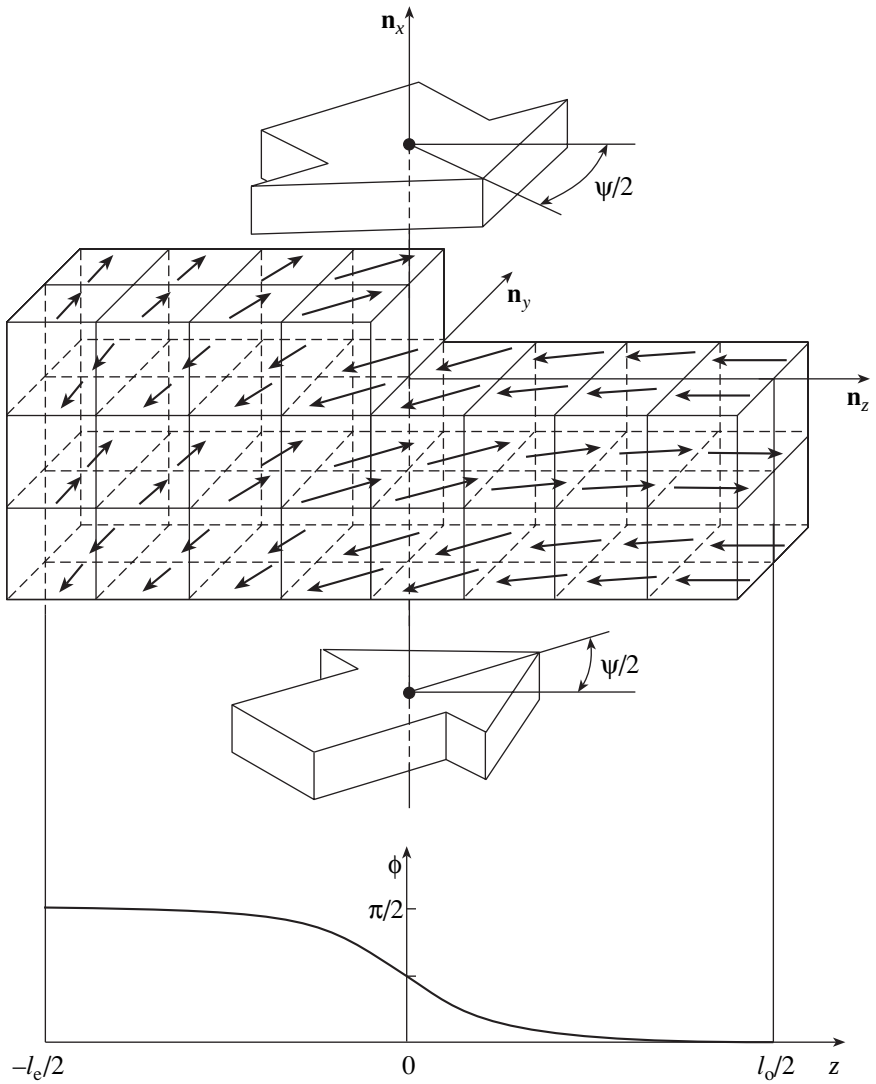

Fig. 1. A schematic diagram illustrating the geometry and magnetic structure of a Fe/Cr/Fe(001) trilayer in the vicinity of a monostep. Large ("3D") top and bottom arrows indicate the orientation of magnetic moments in the FM layers. The rows of small thin arrows show a change in the local magnetization direction in the AFM spacer (all vectors are in the *yz* plane). The bottom curve qualitatively shows the variation of the SDW polarization angle $\phi(z)$ in the wide terrace limit.

$z \le l_o$, and limit variation of the function $\phi(z)$ to the interval $0 < \phi < \pi/2$; the lengths l_o and l_e play the role of fluctuating quantities. Note that, if both interfaces were rough, the limits of variation of the angle ϕ in the general case would be different.

Variation of the functional F_ϕ in Eq. (9) leads to the sine-Gordon equations,

$$\frac{d^2\phi}{dz^2} - \frac{\sin\phi}{\zeta_0^2} = 0, \quad \frac{d^2\phi}{dz^2} + \frac{\cos\phi}{\zeta_e^2} = 0 \qquad (12)$$

(where the first and second equations describe the behavior of the angular phase $\phi(z)$ of the order parameter for the Cr spacer fragments with odd and even *N*, respectively), with the condition of continuity of the derivative $d\phi/dz$ on the boundaries of fragments; we supplement this by the natural condition of continuity for the function $\phi(z)$ possessing a period of $l_o + l_e$. In Eqs. (12), we introduced the quantities

$$\zeta^2 = \frac{2lc_2 v_F^2 \Delta}{A}, \quad \left(\frac{\zeta}{\zeta_o}\right)^2 = \cos\left(\frac{\psi}{2}\right), \quad \left(\frac{\zeta}{\zeta_e}\right)^2 = \sin\left(\frac{\psi}{2}\right), \qquad (13)$$

where ζ is called the "angular" correlation length (in contrast to the "amplitude" correlation length ξ); this value characterizes the SDW orientation fluctuations in a thin intermediate AFM layer of the Fe/Cr/Fe trilayer with thick FM layers.

Solutions of Eqs. (13) have the following form:

$$\phi(z) = \begin{cases} 2\arcsin\left[\mathrm{dn}\left(\dfrac{z+z_o}{\zeta_o}, k_o\right)\right], & 0 \le z \le l_o, \\ \dfrac{\pi}{2} - 2\arcsin\left[\mathrm{dn}\left(\dfrac{z-z_e}{\zeta_e}, k_e\right)\right], & -l_e \le z \le 0, \end{cases} \quad (14)$$

where dn is the elliptic Jacobian function with modulus k. The boundary conditions determine the unknown parameters $\{z_o, z_e, k_o, k_e\}$ via the following system of equations:

$$\frac{2z_o + l_o}{\zeta_o} = 2K(k_o), \quad \frac{2z_e + l_e}{\zeta_e} = 2K(k_e),$$

$$k_o k_e \mathrm{cn}\left(\frac{l_o}{2\zeta_o}, k_o\right)\mathrm{cn}\left(\frac{l_e}{2\zeta_e}, k_e\right) \quad (15)$$

$$-\frac{1}{\sqrt{2}}\mathrm{dn}\left(\frac{l_o}{2\zeta_o}, k_o\right)\mathrm{dn}\left(\frac{l_e}{2\zeta_e}, k_e\right) = k_o' k_e',$$

$$\frac{k_o k_o'}{\zeta_o}\frac{\mathrm{sn}\left(\dfrac{l_o}{2\zeta_o}, k_o\right)}{\mathrm{dn}\left(\dfrac{l_o}{2\zeta_o}, k_o\right)} = \frac{k_e k_e'}{\zeta_e}\frac{\mathrm{sn}\left(\dfrac{l_e}{2\zeta_e}, k_e\right)}{\mathrm{dn}\left(\dfrac{l_e}{2\zeta_e}, k_e\right)}, \quad (16)$$

where sn and dn are the Jacobian elliptic functions, $K(k)$ is the complete elliptic integral of the first kind, and $k' = \sqrt{1-k^2}$ is the complementary modulus [13].

Equations (14)–(16) completely determine the spin density distribution in the AFM layer for a given set of $\{l_o, l_e\}$ and ψ. Now it would be important to determine the scale of roughness R that accounts for the maximum contribution to the interlayer exchange or, in other words, to find the optimum values of l_o and l_e and the corresponding angle ψ_0.

In the limit of wide terraces (l_o, $l_e \gg \zeta_o$, ζ_e and k_o, $k_e \longrightarrow 1$), the SDW has a constant phase ϕ over almost the entire Cr layer ($\phi = 0$ for an odd N and $\phi = \pi/2$ for an even N) and changes at the rare monosteps with the formation of thin domain walls on the scale of the angular correlation length ζ. This state can be considered as a lattice of independent topological kinks of the type

$$\phi(z) = \begin{cases} 2\arcsin\left[\mathrm{sech}\left(\dfrac{z+z_o}{\zeta_o}\right)\right], & z > 0, \\ \dfrac{\pi}{2} - 2\arcsin\left[\mathrm{sech}\left(\dfrac{z_e - z}{\zeta_e}, k_e\right)\right], & z < 0, \end{cases} \quad (17)$$

$$\sinh\left(\frac{z_o}{\zeta_o}\right) = 1 + \sqrt{2\cot\left(\frac{\psi}{2}\right)},$$

$$\sinh\left(\frac{z_e}{\zeta_e}\right) = 1 + \sqrt{2\tan\left(\frac{\psi}{2}\right)},$$

each bearing a topological charge of $Q = 1/4$ [14]. The size ζ of the region of inhomogeneous SDW polarization determined by formula (13) is a characteristic of the Fe/Cr structure. The parameters of matching of the solution (17) depend only on the value of ψ, whose deviation from $\pi/2$ leads to a shift in the center of gravity (distortion) of the domain wall (17) relative to the point $z = 0$. In this limit of negligibly small contribution from domain walls to the total system energy [9], we had previously calculated the effective exchange energy and justified the existence of noncollinear states in Fe/Cr structures with a certain angle ψ_0 (such that $\cos(\psi_0/2) = l_o/\sqrt{l_o^2 + l_e^2}$) between the magnetic moments of the FM layers.

Our approach is fundamentally different from that developed by Slonczewski in the well-known torsional model (proximity magnetism model) [15]. Fishman [16], applying that model to the Fe/Cr type systems, explained noncollinear magnetization of the neighboring FM layers by assuming a strong exchange interaction at the interface between components of the multilayer structure. This interaction led to separation of the AFM layer along the monostep into domains featuring SDWs of helicoidal configuration with opposite orientations. However, estimates based on the band theory of antiferromagnets with SDWs are indicative of a small exchange ($A \ll 1$) in Fe/Cr multilayer structures [7, 8]. Therefore, the states with helicoidal structures are energetically less favorable than the states described by Eq. (14) [9]. Note that, in the case of wide terraces featuring state (17), the angle between the magnetic moments of iron layers and those of the interfacial monolayers of chromium by no means amounts to 180°. For example, at $l_o = l_e$, we have $\psi_0 = 90°$, whereas far from the domain wall this angle is equal to 135°. Thus, the bonds involved in the exchange at the interface are partly frustrated (see Fig. 1). Below, we will calculate the energy of exchange coupling between the neighboring FM layers for an arbitrary density of monosteps.

4. THE ENERGY OF EXCHANGE COUPLING BETWEEN NEIGHBORING FERROMAGNETIC LAYERS

Let the above assumptions concerning the morphology of Fe/Cr interfaces still be valid. Substituting solutions (14) into thermodynamic potential (9) and performing integration, we obtain an expression for the

total system energy (effective exchange energy per unit area $(2l_\perp)^2$) as a function of the angle ψ:

$$\begin{aligned} E(\psi) &= -A\Delta\Bigg\{\Lambda\cos\left(\frac{\psi}{2}\right) + (1-\Lambda)\sin\left(\frac{\psi}{2}\right) \\ &+ 2\left(\Lambda(k_0')^2\cos\left(\frac{\psi}{2}\right) + (1-\Lambda)(k_e')^2\sin\left(\frac{\psi}{2}\right)\right) \\ &- 4\Bigg[\Lambda\frac{\zeta_o}{l_o}\cos\left(\frac{\psi}{2}\right)\left(E\left(\frac{l_o}{\zeta_o}, k_o\right)\right. \\ &\left. - k_o^2\,\mathrm{sn}\left(\frac{l_o}{\zeta_o}, k_o\right)\mathrm{cd}^2\left(\frac{l_o}{2\zeta_o}, k_o\right)\right) \\ &+ (1-\Lambda)\frac{\zeta_e}{l_e}\sin\left(\frac{\psi}{2}\right)\left(E\left(\frac{l_e}{\zeta_e}, k_e\right)\right. \\ &\left.\left. - k_e^2\,\mathrm{sn}\left(\frac{l_e}{\zeta_e}, k_e\right)\mathrm{cd}^2\left(\frac{l_e}{2\zeta_e}, k_e\right)\right)\Bigg]\Bigg\}, \end{aligned} \tag{18}$$

where

$$\Lambda = \frac{l_o}{l_o + l_e}, \quad 1-\Lambda = \frac{l_e}{l_o + l_e},$$

and $E(u, k)$ is the incomplete elliptic integral of the second kind [13]. In the course of transformations, we excluded the parameters z_o and z_e by using Eqs. (15); the remaining parameters k_o and k_e in Eq. (18) can be calculated using Eqs. (16).

Expression (18) can be simplified only in some limiting cases. For example, in the case of almost isolated monosteps, the final result is as follows (lengthy algebra is omitted):

$$\begin{aligned} E(\psi) &= -A\Delta\left[\Lambda\cos\left(\frac{\psi}{2}\right) + (1-\Lambda)\sin\left(\frac{\psi}{2}\right)\right] \\ &+ 4A\Delta\zeta\rho\left[\sqrt{\cos\left(\frac{\psi}{2}\right)} + \sqrt{\sin\left(\frac{\psi}{2}\right)}\right. \\ &\left. - \sqrt{\cos\left(\frac{\psi}{2}\right) + \sqrt{\sin\psi} + \sin\left(\frac{\psi}{2}\right)}\right]. \end{aligned} \tag{19}$$

This formula is valid under the conditions

$$\frac{l_o}{\zeta_o} \gg 1, \quad \frac{l_e}{\zeta_e} \gg 1,$$

which imply a sufficiently smooth interface and angles ψ not very close to 0 or π (see (13)). The first term in the dependence of the exchange energy on angle ψ in Eq. (19) coincides with the relation obtained previously [9], while the second term (a correction on the order of $\rho\zeta \ll 1$) represents a positive energy (17) of a lone 90° domain wall in chromium, multiplied by the linear density of monosteps $\rho = 2/(l_o + l_e)$. The next small term proportional to

$$(k'_{o,e})^2 \propto \exp(-l_{o,e}/\zeta_{o,e}),$$

representing overlapped tails of the neighboring domain walls, was neglected. It should be noted that (since $\zeta \sim \sqrt{\Delta}$, see (13)), the conditions of applicability of the wide terrace approximation are somewhat improved with increasing temperature or decreasing AFM spacer thickness, although the magnitude of the effective exchange energy (19) tends to decrease:

$$|E(\psi)| \propto A\Delta.$$

Thus, formula (19) can be used for the interpretation of experimental data on the effective coupling in Fe/Cr structures with high-quality interfaces and relatively thick iron layers.

In the limit of closely spaced monosteps, the SDW phase $\phi(z)$ in the AFM layer exhibits small oscillations about the angle $\phi(0)$:

$$\phi(z) = \phi(0) + \delta(z), \quad |\delta| \propto (\rho\zeta)^{-2},$$

$$\phi(0) = 2\arcsin\sqrt{\frac{1}{2}\left\{1 - \left[1 + \frac{l_e}{l_o}\tan\left(\frac{\psi}{2}\right)\right]^2\right\}^{-1/2}}.$$

Retaining terms to within the second order of smallness with respect to the ratio $l_{o,e}/\zeta \ll 1$, we obtain (cumbersome transformations are omitted)

$$\begin{aligned} E(\psi) &= -\frac{A\Delta}{2}[1 + (2\Lambda-1)^2 + 2(2\Lambda-1)\cos\psi]^{1/2} \\ &- \frac{A\Delta}{6\zeta^2\rho^2}\left[\left(\Lambda\cos\left(\frac{\psi}{2}\right)\right)^{-2} + \left((1-\Lambda)\sin\left(\frac{\psi}{2}\right)\right)^{-2}\right]^{-1}. \end{aligned} \tag{20}$$

If the interfacial relief represents a plane with separate narrow hills of monoatomic height (such that $(\rho\zeta)^2 \gg 1$ and, e.g., $l_o \gg l_e$ or $\Lambda \longrightarrow 1$), the first term in expression (20) predominates. This implies that the magnetic moments of the neighboring FM layers in the Fe/Cr structure in equilibrium are parallel ($\psi = 0$); in contrast, for $l_o \ll l_e$ or $\Lambda \longrightarrow 0$, these moments are antiparallel ($\psi = \pi/2$).

In a more interesting situation of strong roughness (representing a Fe/Cr interface cut with a network of monoatomic steps), the integral characteristic of fluctu-

ations in the Cr spacer thickness is $\Lambda \approx 1/2$ (or $l_o \approx l_e$). In the case of

$$2\sqrt{6}|2\Lambda - 1|\zeta\rho \ll 1,$$

the angular dependence (20) of the interlayer magnetic coupling energy takes a form corresponding to the effective exchange with a "biquadratic" interaction [1–3]:

$$E(\psi) - E(\pi/2) = J_1\cos\psi + J_2\cos^2\psi, \qquad (21)$$

$$J_1 = -A\Delta\left(\Lambda - \frac{1}{2}\right), \quad J_2 = \frac{A\Delta}{96\rho^2\zeta^2},$$

$$E\left(\frac{\pi}{2}\right) = -\frac{A\Delta}{2} - J_2.$$

The coefficients in this expression for the interlayer interaction energy have a rather simple form and can be readily evaluated as functions of the Cr spacer thickness L and the temperature T. Note that $J_1(L, T) \sim \Delta(L, T)$, while the coefficient J_2 is not explicitly related to the SDW amplitude (because $\zeta^2 \sim \Delta$). In this respect, the behavior of the biquadratic coupling coefficient, $J_2(L, T) \sim (c_2(T)L)^{-1}$, exhibits a universal character independent of the particular model of the AFM order parameter $\Delta(L, T)$ (such as, e.g., (6)) in a thin chromium layer.

The condition of $|J_1| < 2J_2$ corresponds to an equilibrium noncollinear state with an angle between the magnetic moments of the FM layers dependent on the interface morphology parameters,

$$\cos\psi_0 = 24(2\Lambda - 1)\rho^2\zeta^2,$$

and the effective exchange energy

$$E(\psi_0) - E(\pi/2) = -6A\Delta(2\Lambda - 1)^2\rho^2\zeta^2.$$

In particular, for

$$24(2\Lambda - 1)\rho^2\zeta^2 \longrightarrow 0,$$

the magnetic moments of the FM layers are mutually perpendicular ($\psi_0 = \pi/2$) and the SDW polarization vector is virtually not affected by the spacer thickness fluctuations ($\psi(z) \approx \pi/4$).

Formula (21) indicates that impaired technological quality of the interface ($L \longrightarrow 1/2$, $(\rho\zeta)^2 \longrightarrow \infty$) must lead to a significant decrease in the interlayer coupling energy in Fe/Cr structures as compared to a value (about $A\Delta$) characteristic of the wide terrace limit. This conclusion was confirmed by numerical analysis of $E(\psi)$ according to Eq. (18). This analysis was performed by simultaneously solving Eqs. (16) with respect to k_o and k_e for the interface roughness parameters ($\rho\zeta$, Λ) varied within broad limits. Figure 2a shows the results of our numerical calculations for $\Lambda = 1/2$ and $(\rho\zeta)^{-1} = 0.1, 1, 2, 3, 4, 8, \infty$; the dotted curve corresponds to the calculations using approximate formula (19) with $(\rho\zeta)^{-1} = 8$ (the data are plotted for $0 \le \psi \le \pi/2$, since the function $E(\psi)$ for $\Lambda = 1/2$ is mirror symmetric relative to $\psi = \pi/2$).

Figure 2b presents the results of our numerical calculations for $\Lambda = 1/4$ and $(\rho\zeta)^{-1} = 1, 2, 4, 8, 16, \infty$; the dotted curve refers to the calculations using approximate formula (19) with $(\rho\zeta)^{-1} = 16$ (the pattern of exchange energy $E(\psi)$ for $\Lambda = 3/4$ can be obtained as a mirror reflection of Fig. 2b relative to the vertical axis $\psi = \pi/2$). As can be seen, the function $E(\psi)$ for $\rho\zeta \ge 1/8$ exhibits only a trivial minimum at $\psi = \pi$. The tendency to the transition from noncollinear to collinear magnetic configuration for the FM layers with impaired interface quality is general for $\Lambda \ne 1/2$. A comparison of the families of $E(\psi)$ curves constructed for various values of the parameter Λ shows that the case of $\Lambda = 1/2$ corresponds to the minimum amplitude of oscillations in the exchange coupling energy. This amplitude significantly decreases with increasing disorder on the interfaces. However, the amplitude of oscillations in the exchange coupling energy becomes weakly dependent on the $\rho\zeta$ value already for $\Lambda = 1/4$.

5. AN ALTERNATIVE APPROACH: THE MODEL OF FERROMAGNETIC DOMAIN WALLS

The model of AFM domain walls considered above has certain limitations of both physical and geometric character, which hinders description of the magnetic order in some real structures. In particular, we assumed that the FM layers are homogeneously magnetized (that is reasonable for the layers of sufficiently large thickness). However, since iron as such possesses a finite magnetic stiffness γ, it can be energetically favorable for thin FM layers to split into domains—and this will significantly modify the SDW structure in the chromium layer. A self-consistent calculation of the spin density distribution in the entire structure for arbitrary thicknesses and geometries of the Fe and Cr layers is practically impossible. For this reason, below we only consider a situation that is, in a certain sense, opposite to the case analyzed above and is apparently closer to the limit of thin FM layers. Note, however, that it is necessary to strictly define a particular geometry (rather than only a magnetic configuration) of the system under consideration.

Let us consider, for example, an asymmetric trilayer (frequently studied in experiments) comprising a thin ("upper") and thick ("lower") FM iron layer separated by an AFM spacer of chromium. We retain the assumptions made in Sections 2 and 3 concerning the Fe/Cr interface morphology. However, now we will additionally assume that the order parameter in the AFM spacer

has a fixed direction (corresponding to the maximum gain in the exchange energy on the lower interface that is assumed to be perfectly smooth) and that the roughness of the upper ($x = l$) interface causes an inhomogeneous redistribution of magnetization in a thin FM layer, so that the latter separates into FM domains along the monosteps.

If the AFM spacer features a homogeneously polarized SDW with an amplitude of $\mathbf{\Delta}(\mathbf{r}) = -\mathbf{n}_z\Delta(x)$, the magnetic moments in the upper FM layer experience the action of the exchange field $\pm\Delta(l)$ changing sign upon crossing each monostep. As a result, the orientation of the vector $\mathbf{m}(l, y, z)$ in the FM layer gradually varies over a characteristic length ζ, so as to adjust to the external field $\pm\Delta(l)$. In the lower FM layer, the magnetization is constant and equal to $\mathbf{m}(-l) = \mathbf{n}_z$. The thermodynamic potential (2)–(5) acquires the form of the effective functional

$$F = F_A + F_\varphi, \tag{22}$$

$$F_\varphi = d\gamma \int\limits_{\{\text{total}\}} dydz\left(\left(\frac{\partial\varphi}{\partial y}\right)^2 + \left(\frac{\partial\varphi}{\partial z}\right)^2\right) - A\Delta\left[\int\limits_{\{\text{odd}\}} dydz\cos\varphi - \int\limits_{\{\text{even}\}} dydz\cos\varphi\right] - A\Delta(2l_\perp)^2. \tag{23}$$

Note that, in the limit under consideration ($l/D \longrightarrow 0$), the energy F_A and the SDW amplitude Δ are given by expressions (8) and (6), respectively. The quantity F_φ equals the sum of the functional of static magnetization orientation fluctuations in the upper FM layer and the exchange energy on the lower Fe/Cr interface (the last term in (23)). The function $\varphi = \varphi(y, z)$ is defined as follows:

$$m_x(l, y, z) = 0, \quad m_z(l, y, z) = \cos\varphi(y, z), \quad m_y(l, y, z) = \sin\varphi(y, z). \tag{24}$$

In the approximation of a one-dimensional periodic structure of monosteps on the upper interface, variation of the functional F_φ in Eq. (23) leads to the sine-Gordon equations

$$\frac{d^2\varphi}{dz^2} - \frac{\sin\varphi}{\zeta^2} = 0, \quad \frac{d^2\varphi}{dz^2} + \frac{\sin\varphi}{\zeta^2} = 0. \tag{25}$$

Here, the first and second equations describe the behavior of the angular phase $\varphi(z)$ of the magnetization in parts of the FM layer that come into contact with the Cr spacer fragments with odd and even N, respectively. We also introduce the natural conditions of continuity of the function $\varphi(z)$ and its derivative $d\varphi/dz$ on the monosteps. Here, the value of the correlation length ζ for the magnetization orientation fluctuations in the iron layer of thickness d in the Fe/Cr structure is defined as

$$\zeta^2 = \frac{2d\gamma}{A\Delta}. \tag{26}$$

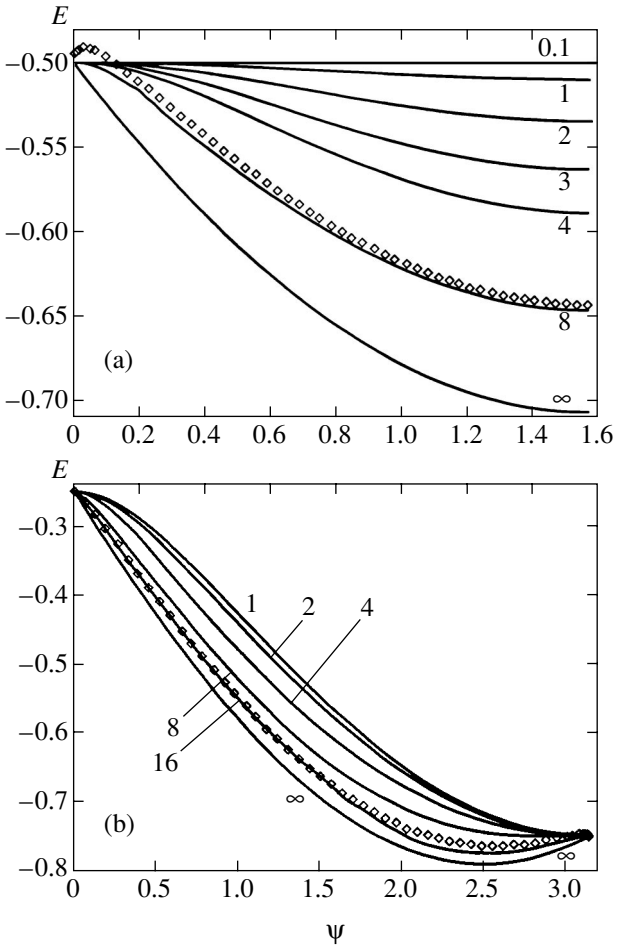

Fig. 2. The angular dependences of the effective interlayer exchange coupling energy $E = E(\psi)/A\Delta$ (ψ is the angle in radians) calculated for $\Lambda = 1/2$ (a) and 1/4 (b) and various values of the parameter $(\rho\zeta)^{-1}$ (indicated at the curves). Solid curves show the results of numerical analysis of the exact formulas (16) and (18); dotted curves are calculated by the approximate formula (19) for $(\rho\zeta)^{-1} = 8$ (a) and 19 (b).

The structure and energy of states of the functional (22)–(23), as well as their variation depending on the parameters l_o/ζ and l_e/ζ, can be calculated and studied in detail as was done above for functional (9). We will not dwell on this case here; we will only mention the characteristic features of these states.

In the case of a relatively small distance between monosteps (l_o, $l_e < \zeta$), the domain of variation of the function determined by Eqs. (25) is restricted to the interval $0 < \varphi(z) < \pi$. As can be shown for a sufficiently high density of monosteps, such that ($\rho^2\zeta^2 \gg 1$ and

$\Lambda \approx 1/2$, the direction of magnetization of the FM layer weakly oscillates about the average value

$$\varphi\left(\pm\frac{l_o}{2}\right) = \frac{\pi}{2} \mp \left(\frac{l_o}{4\zeta}\right)^2. \quad (27)$$

In other words, the magnetization of the upper iron layer is almost completely oriented 90° relative to that in the lower layer,

$$\langle m_y \rangle = 1, \quad (m_z) = 0.$$

The exchange coupling energy per unit area of the interface is

$$E = -A\Delta - \frac{A\Delta}{24\rho^2\zeta^2}. \quad (28)$$

In the limit of wide terraces, such that

$$k'_{o,e} \approx 2\exp(-l_{o,e}/2\zeta) \longrightarrow 0,$$

the pattern is different. Far from a monostep, the magnetization is almost antiparallel relative to spins in the chromium monolayer adjacent to the interface. For definiteness, let us assume that $\varphi = 0$ for an odd N and $\varphi = \pi$ for an even N; in the vicinity of a monostep, vector $\mathbf{m}(x = l, z)$ exhibits rotation by 180° over the angular correlation length ζ (26). This state can be considered as a system of almost isolated topological domain walls of the type

$$\varphi(z) = \begin{cases} \arcsin\left[\operatorname{sech}\left(\frac{z + z_o}{\zeta}\right)\right], & z > 0, \\ \pi - \arcsin\left[\operatorname{sech}\left(\frac{z_e - z}{\zeta}\right)\right], & z < 0. \end{cases} \quad (29)$$

The condition of matching for this solution at $z = 0$ is as follows:

$$\sinh\left(\frac{z_o}{\zeta}\right) = \sinh\left(\frac{z_e}{\zeta}\right) = 1. \quad (30)$$

As can be seen, the vector of magnetization $\mathbf{m}(x = l, z)$ according to Eq. (24) for the upper FM layer in the vicinity of a monostep exhibits a noncollinear local orientation relative to the homogeneous magnetization of the lower FM layer. Therefore, the lone domain wall (29)–(30) can be assigned a magnetization component along the $\mathbf{n}_y$ axis:

$$\int_{-\infty}^{\infty} m_y(z)dz = 2\sqrt{2}\zeta. \quad (31)$$

The description of a domain wall with the opposite (negative) magnetization is obtained by replacing $\varphi(z)$ by $\varphi(z) \pm \pi$ in Eq. (25). In the case of an ultimately low density of monosteps ($\rho\zeta \longrightarrow 0$), the system represents a set of isolated 180° domain walls of different signs with the energy

$$E = -2A\Delta\left(1 - 4\rho\zeta\left(1 - \frac{1}{\sqrt{2}}\right)\right). \quad (32)$$

With allowance for an exponentially small (proportional to $\exp(-l_{o,e}/\zeta)$) interaction between the neighboring domain walls, it is possible to conclude that the most energetically favorable magnetic structure is that with monotonically increasing (or decreasing) angular function $\varphi(z)$ having the shape of the "devil's staircase" type. This structure is characterized by wide (almost over the entire terrace width) plateaus with a constant magnetization orientation along $\mathbf{n}_z$: $\varphi(z) = (2n + 1)\pi$ for an even N and $\varphi(z) = 2n\pi$ for an odd N (n is an integer). These plateaus are separated by thin FM domain walls of strictly alternating polarization signs along the $\mathbf{n}_y$ axis. Averaging over the entire area of the interface yields

$$\langle m_z \rangle = 2\Lambda - 1, \quad \langle m_y \rangle = 0.$$

In the general case, when the relations between model parameters are not as simple (e.g., $l_o < \zeta < l_e$ or, vice versa, $l_e < \zeta < l_o$), even the simplified model of a rough interface adopted above requires a complicated analysis far beyond the framework of this paper. It can be shown that there is a transition with respect to the parameter $\rho\zeta$, Λ between two qualitatively different regimes of magnetization distribution in the thin FM layer, as briefly described above in the limits of $l_o, l_e \ll \zeta$ and $l_o, l_e \gg \zeta$.

As was noted above, the assumption of a thick "upper" FM layer made for the calculations in Sections 3 and 4 was opposite to that made in Section 5. Therefore, expressions for the exchange energy obtained in the models of AFM and FM domain walls cannot be compared from the standpoint of the relative energy gain. A more general theoretical analysis of the possible magnetic configurations in the system with an arbitrary geometry of the iron and chromium layers, which is beyond the limits of the one-dimensional approximation, remains to be conducted. Nevertheless, below we will attempt, based on the results, to make certain inferences concerning the character of the effective interlayer exchange coupling in real structures of various geometry and quality of the Fe/Cr interface.

6. CONCLUSIONS

The presence of structural defects on the boundaries between FM and AFM layers leads both to the frustration of bonds involved in the interlayer exchange and to

the formation of inhomogeneous spin configurations inside the layers. This statement seems to be valid for a broad class of magnetic nanostructures of the type under consideration (Fe/Cr, Co/Cr. Fe/Mn, etc.), while the mechanisms involved in the development of the inhomogeneity and frustration for each particular system have to be specified. In this paper, we have proposed a theoretical scheme taking into account both factors, intended for description of the effective exchange coupling between FM layers in the Fe/Cr type structure. This structure is commonly accepted as a model system and, at the same time, considered as highly promising for practical use in superhigh-density magnetic recording devices.

Experimental investigations of a magnetic configuration of the Fe–Cr system even of the simplest geometry, such as a chromium film on a Fe(001) whisker [1, 5] or a thin iron film on a massive chromium single crystal [17, 18], revealed a high sensitivity of this system with respect to structural parameters of the Fe/Cr interface and the temperature. The magnetic coupling between FM layers in more complicated systems (trilayers, superlattices, etc.) is also an important probe of the quality of interfaces as determined by the material's preparation technology. For symmetric Fe/Cr/Fe trilayer structures with a wedge-shaped AFM spacer and relatively thick (5 nm) FM layers epitaxially grown on GaAs/Fe/Ag(001) substrates, Schmidt *et al.* [4] studied the interlayer coupling using the magneto-optical Kerr effect and characterized the interface morphology by scanning tunneling microscopy (STM). The magneto-optical data showed evidence of a significant (severalfold) variations in the amplitude of the shortwave component of the effective interlayer exchange potential depending on the temperature regime of the layer growth. On the other hand, a thorough statistical analysis of the STM images of the growth front in the same samples revealed a direct correlation between the amplitude of the shortwave oscillations of the effective interlayer exchange potential and a longitudinal roughness parameter of the interface (the R value in the best case was about 22 nm). It was established that the dominant role determining the effective exchange coupling between the FM layers was played by special regions of the Cr spacer with a constant thickness (N = const) and a lateral size of not less than 3–4 nm ("pillars" in terms of [4]), whereas the other regions ("edges") of the Cr spacer featuring frequent monolayer-order thickness fluctuations were barely involved in the interlayer exchange.

The approach developed in the main sections of this paper provides a theoretical basis for the empirical conclusions following from the results obtained in [4]. A more rigorous description of the system would require a self-consistent (generally speaking, three-dimensional) calculation of the order parameter $\mathbf{\Delta}(\mathbf{r})$ in a random field of the AFM layer thickness fluctuations, followed by averaging over these fluctuations. Nevertheless, even the simplest model of the interface morphology using a minimum number of parameters (Λ and $\rho\zeta$) and the approximation of magnetic inhomogeneities by one-dimensional variations of the SDW vector (analogous to the domain walls) allowed us to reveal the main features of the interlayer exchange. According to formulas (19)–(21) and Fig. 2, the long-wave components of the roughness fluctuations must make the main contribution to the interlayer exchange; on the other hand, it is also evident that this contribution is limited by the small statistical weight of these components. Therefore, there must exist an optimum fluctuation with a certain characteristic length R_0 that can be quite naturally related to the "pillars" observed in [4].

It should be noted that expression (21) is formally analogous to the traditional phenomenological model with bilinear and biquadratic terms. We may suggest that the interlayer exchange behavior of the Fe/Cr structures grown in the optimum regime can be expected to exhibit deviations from (21) toward (19). Based on the magnetization hysteresis observations, Schreyer *et al.* [19] stated that a biquadratic model cannot explain the residual 50° coupling between the neighboring FM layers in a [Fe(52 Å)/Cr(17 Å)] superlattice with rather broad terraces ($R \geq 100$ Å). There are many other experimental facts (see, e.g., [6, 12, 20] and the discussion in review [3]) that cannot be given satisfactory interpretation within the framework of the biquadratic model either. According to our rough estimates, the correlation length ζ on which the SDW vector orientation changes in the chromium layer amounts to about 1–5 nm, so that an angular dependence of $E(\psi)$ close to expression (19) can be expected in the experiments using trilayers grown by layer epitaxy. As the quality of the interface decreases (in our scheme, this corresponds to $\Lambda \longrightarrow 1/2$, $\rho\zeta \longrightarrow \infty$, the probability of rather wide regions appearing in the AFM layer with N = const becomes negligibly small and the interlayer coupling acquires the traditional form of Eq. (21).

Let us briefly consider the dependence of the interlayer exchange on the AFM layer thickness and the temperature. Demokritov *et al.* [20] treated experimental data obtained using Kerr magnetometry and the Brillouin light scattering for symmetric Fe(100 Å)/Cr(0–20 Å)/Fe(100 Å) structures with relatively thick FM layers within the framework of a bilinear–biquadratic exchange model. Using qualitative estimates obtained under the assumptions that the magnetic structure of the chromium layer is weakly distorted and that iron layers are homogeneously magnetized, the authors of [20] justified this model and obtained an angular dependence of the interlayer exchange coupling energy that coincides substantially with that given by formula (21) calculated above in the limit of $\rho\zeta \ll 1$ and $\Lambda \approx 1/2$. The experiments also showed that the biquadratic interaction parameter decreases in inverse proportion to the Cr spacer thickness and linearly decreases with increasing temperature in the interval from 77 to 473 K. This behavior of quantity

$J_2(L, T)$ is quite consistent with the estimate $J_2(L, T) \sim (c_2(T)L)^{-1}$ predicted in Section 4 (for the temperature variation of the coefficient $c_2(T)$, see, e.g., [10]).

In the structures with thin (not exceeding 20 Å) iron layers and highly perfect Fe/Cr interfaces (to which the model of almost isolated monosteps is applicable), the state with a 180° FM domain wall described by formulas (29)–(30) seems to be energetically most favorable. This very state was directly observed in the well-known experiments using SEM in combination with polarization analysis for the unique structures grown in the optimum regime on a thick iron layer (whisker) [1]. The surface of a whisker is almost perfectly smooth ($\rho \approx 1\ \mu m^{-1}$). Using a structure with a wedge-shaped chromium spacer, it is possible to obtain (within the same sample) a sequence of very wide ($l_o = l_e = R = 10\ \mu m$) and almost perfectly smooth plane terraces of regular shapes covering the AFM spacer fragments with different numbers N of chromium monolayers.

In accordance with the results obtained in Section 5 for the interlayer exchange according to a mechanism of the 180° FM domain walls, the magnetization of a thin Fe layer is oriented noncollinearly relative to the whisker magnetization only in a narrow region (on the order of ζ). Pierce *et al.* [1] pointed out that the noncollinear coupling between FM layers exhibited a sharp drop at a chromium spacer thickness close to 24 monolayers (i.e., under conditions of the interlayer exchange coupling phase slip). Our analysis suggests that this fact is related to a complex rearrangement of the spatial structure of SDWs in the chromium layer of variable layer thickness $L = 2l$, which was recently considered in [21]. This rearrangement naturally leads to a change in the SDW amplitude at the Fe/Cr interface, thus determining, according to formula (26), the scale $\zeta \sim \Delta(l)^{-1/2}$ of the magnetization orientation fluctuations in the iron layer. It should be noted in conclusion that an $E(\psi)$ dependence resembling relation (19) for $\Lambda \equiv 1/2$ and $\zeta\rho \equiv 0$ was obtained in [22] by using a completely different microscopic model.

In our opinion, the effective exchange coupling in most experiments (both with Fe/Cr/Fe(001) trilayers and [Fe/Cr](001) superlattices) is related to variations of the SDW vector orientation in the chromium layer as described in Sections 3 and 4. The situation with the appearance of 180° FM domain walls considered in Section 5 is the exception rather than the rule. A rough qualitative estimate of the region of model parameters corresponding to the states with inhomogeneous magnetization of the iron layers is provided by the relation $d\gamma < \delta L$, where the quantity $\delta = c_2 \nu_F^2 \Delta^2/2$ can be naturally treated as the exchange stiffness of a thin chromium layer.

ACKNOWLEDGMENTS

The authors are grateful to Yu.V. Kopaev, Head of the Workshop at the Lebedev Institute of Physics; to participants of this workshop for fruitful discussion of the results of this study; and to N.M. Kreines and D.I. Kholin for detailed elucidation of the experimental situation.

This study was supported in part by the Russian Foundation for Basic Research, project no. 01-02-16175.

REFERENCES

1. D. T. Pierce, J. Unguris, R. J. Celotta, and M. D. Stiles, J. Magn. Magn. Mater. **200**, 290 (1999).
2. H. Zabel, J. Phys.: Condens. Matter **11**, 9303 (1999).
3. R. S. Fishman, J. Phys.: Condens. Matter **13**, R235 (2001).
4. C. M. Schmidt, D. E. Burgler, D. M. Schaller, *et al.*, Phys. Rev. B **60**, 4158 (1999).
5. D. T. Pierce, J. A. Stroscio, J. Unguris, and R. J. Celotta, Phys. Rev. B **49**, 14564 (1994).
6. B. Heinrich, J. F. Cohran, T. Monchesky, and R. Urban, Phys. Rev. B **59**, 14520 (1999).
7. M. Avignon, V. Men'shov, and V. Tugushev, Europhys. Lett. **56**, 132 (2001).
8. V. N. Men'shov and V. V. Tugushev, Zh. Éksp. Teor. Fiz. **120**, 899 (2001) [JETP **93**, 786 (2001)].
9. V. N. Men'shov and V. V. Tugushev, Zh. Éksp. Teor. Fiz. **122**, 1044 (2002) [JETP **95**, 901 (2002)].
10. V. V. Tugushev, in *Electronic Phase Transitions*, Ed. by W. Hanke and Yu. V. Kopaev (North-Holland, Amsterdam, 1992), Modern Problems in Condensed Matter Sciences, Vol. 32, p. 239.
11. N. I. Kulikov and V. V. Tugushev, Usp. Fiz. Nauk **144**, 643 (1984) [Sov. Phys. Usp. **27**, 954 (1984)].
12. E. E. Fullerton, C. H. Sowers, and S. D. Bader, Phys. Rev. B **56**, 5468 (1997).
13. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun, 2nd ed. (Dover, New York, 1971; Nauka, Moscow, 1979).
14. R. Rajaraman, *Solitons and Instantons: an Introduction to Solitons and Instantons in Quantum Field Theory* (North-Holland, Amsterdam, 1982; Mir, Moscow, 1985).
15. J. C. Slonczewski, J. Magn. Magn. Mater. **150**, 13 (1995).
16. R. S. Fishman, Phys. Rev. Lett. **81**, 4979 (1998).
17. E. J. Escorcia-Aparicio, Hyuk J. Choi, W. L. Ling, *et al.*, Phys. Rev. Lett. **81**, 2144 (1998).
18. H. Hopster, Phys. Rev. Lett. **83**, 1227 (1999).
19. A. Schreyer, J. F. Anker, Th. Zeidler, *et al.*, Phys. Rev. B **52**, 16066 (1995).
20. S. O. Demokritov, A. B. Drovosekov, N. M. Kreĭnes, *et al.*, Zh. Éksp. Teor. Fiz. **122**, 1233 (2002) [JETP **95**, 1062 (2002)].
21. V. N. Men'shov and V. V. Tugushev, Fiz. Tverd. Tela (St. Petersburg) **44**, 1650 (2002) [Phys. Solid State **44**, 1727 (2002)].

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 1, 2004, pp. 135–143.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 1, 2004, pp. 150–159.

SOLIDS
Electronic Properties

The Energy Band Structure and Optical Spectra of $FeBO_3$ Calculated with Allowance for Strong Electron Correlations

S. G. Ovchinnikov* and V. N. Zabluda

Kirensky Institute of Physics, Siberian Division, Russian Academy of Sciences, Krasnoyarsk, 660036 Russia
**e-mail: sgo@iph.krasn.ru*

Received May 16, 2003; in final form, September 19, 2003

Abstract—A model of the energy band structure of iron borate ($FeBO_3$) is proposed that combines a one-electron description of the *sp* states of boron and oxygen with a many-electron description of the *d* states of iron. The Green functions of *d* electrons are calculated using the exact Lehmann spectral representation. The energies of the *d*-type quasiparticles are calculated using terms of the d^4, d^5, and d^6 electron configurations. The optical absorption spectrum of $FeBO_3$ is determined by local excitons and by the electron excitations with charge transfer. The latter excitations control the nature of the dielectric gap in $FeBO_3$ crystals. The model parameters are determined from a comparison to the exciton energies. The density of single-particle states in $FeBO_3$ is calculated. The main bands in the calculated optical absorption spectrum agree well with experimental data for energies up to 3 eV. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Iron borate $FeBO_3$ is one of a few magnets combining transparency in the visible spectral range with spontaneous magnetization at room temperature. This is a weak ferromagnet with nearly antiparallel spin sublattices of Fe^{3+} ions in the (111) base plane at temperatures below the Néel temperature $T_N = 348$ K [1]. The $FeBO_3$ crystals possess a calcite structure belonging to the space group $R\bar{3}c$ (D_{3d}^6) [1, 2], in which Fe^{3+} ions are surrounded by an oxygen octahedron of an almost cubic symmetry; bond lengths: Fe–O, 2.028 Å and Fe–Fe, 3.601 Å; O–Fe–O bond angles, 91.82° and 88.18° [3]. Under normal ambient conditions, $FeBO_3$ is an insulator with a fundamental absorption edge at $E_g^{(0)} = 2.9$ eV [4]. Despite many years of research, there has been permanent interest in studying the properties of $FeBO_3$ crystals. Recent investigations revealed a structural phase transition in $FeBO_3$ [5], a collapse of the magnetic moment of Fe^{3+} ions under pressure [6], peculiarities in the concentration dependence of the magnetic and optical properties of some solid solutions of the $V_xFe_{1-x}BO_3$ system [7], and the light-induced breakage of the magnetic order under conditions of pulsed optical pumping [8].

At the same time, relationships between the observed properties and the electron structure of $FeBO_3$ have not yet been established even on a qualitative level. There are difficulties in application of the standard band theory to $FeBO_3$, which are related to strong electron correlations involving the *d* electrons of iron. Indeed, a one-electron approach to the d^5 electron configuration of Fe^{3+} ion leads to a partly filled band and the metallic state. In the Hubbard model with strong electron correlations, whereby $U \gg W$ (U is the Coulomb interaction parameter and W is the width of a half-filled *d* band), we obtain an antiferromagnetic state of the Mott–Hubbard dielectric. However, in $FeBO_3$ (as well as in many other real substances), a simple pattern based on the Hubbard model is complicated by the presence of a large number of *d*(*f*) orbitals.

This paper proposes a many-electron model taking into account all *d* orbitals and strong electron correlations involving *d* electrons. Within the framework of this model, the density of single-particle states of *d* electrons contains contributions due to local quasiparticles with energies

$$\Omega_{ij} = E_i(d^{n+1}) - E_j(d^n),$$

where $E_i(d^n)$ denotes the *i*th term of the d^n configuration. In the case of $FeBO_3$, the energies of both high-spin and various low-spin terms of Fe^{2+}, Fe^{3+}, and Fe^{4+} ions become significant. An analogous approach was employed in the analysis of magnetism in *d* metals [9, 10] and layered cuprates [11]. The model parameters are determined from a comparison to the energies of exciton peaks in the optical absorption spectrum. The calculated density of single-particle states, $N(E)$, is compared to the experimental absorption spectrum in a broad range of energies $E \leq 3$ eV.

1063-7761/04/9801-0135$26.00

The paper is organized as follows. Section 2 describes the proposed many-electron multiband model of $FeBO_3$. Section 3 is devoted to calculation of the local Green functions of *d* electrons, which is compared to the exact Lehmann spectral representation. Section 4 considers the experimental absorption spectrum of $FeBO_3$ measured in a broad energy range. Section 5 compares the calculated density of states $N(E)$ to the experimental absorption spectra. Finally, in Section 6 we will discuss the temperature dependence of these spectra.

2. A MANY-ELECTRON MULTIBAND MODEL OF THE ELECTRON STRUCTURE OF $FeBO_3$

The *ab initio* one-electron energy band calculations performed for $FeBO_3$ using the density functional method in the local spin density approximation [12] and the generalized gradient approximation [13], together with the calculation of molecular orbitals of a FeB_6O_6 cluster [7], revealed the following electron structure of $FeBO_3$. The empty conduction band ε_c consists predominantly of the *s* and *p* states of boron. The top of the valence band ε_v is formed mostly by the *s* and *p* states of oxygen. The energy gap E_{g0} between valence and conduction bands in the antiferromagnetic phase amounts to 2.5 eV [12], which is quite close to the fundamental absorption edge ($E_{g0} = 2.9$ eV). The band of *d* electrons occurs at the top of the valence band, and the crystal field parameter is $\Delta \approx 1$ eV [12]. The degree of hybridization of the *d* electrons of iron with the *s* and *p* electrons of oxygen is very small [7, 12], much smaller as compared to the case of 3*d* metal oxides. This is related to a very strong hybridization inside the BO_3 group, where the $(BO_3)^{3-}$ ion does in fact exist and the electron orbitals of oxygen are closed to boron (which accounts for the small *p–d* hybridization). This circumstance significantly simplifies the many-electron model, for which the d^n ($n = 4, 5, 6$) terms of iron in the crystal field can be calculated, rather than the terms of a metal–oxygen complex (as in copper oxides [11]).

The intraatomic part of the Hamiltonian for *d* electrons can be written as

$$H_{\text{at}} = \sum_{\lambda,\sigma}\left(\varepsilon_\lambda n_{\lambda\sigma} + \frac{U_\lambda}{2} n_{\lambda\sigma} n_{\lambda\bar{\sigma}}\right) + \sum_{\substack{\lambda,\lambda' \\ (\lambda\neq\lambda')}} \sum_{\sigma,\sigma'} (V_{\lambda\lambda'} n_{\lambda\sigma} n_{\lambda'\sigma'} - J_{\lambda\lambda'} a^\dagger_{\lambda\sigma} a_{\lambda\sigma'} a^\dagger_{\lambda'\sigma'} a_{\lambda'\sigma}), \quad (1)$$

where $n_{\lambda\sigma} = a^\dagger_{\lambda\sigma} a_{\lambda\sigma}$, $a_{\lambda\sigma}$ is the operator of *d* electron creation on one of the five orbitals λ with the spin projection σ ($\bar{\sigma} = -\sigma$). The first term in (1) describes the atomic *d* levels in the crystal field. A small uniaxial component of the crystal field is ignored and it is assumed that

$$\varepsilon(t_{2g}) = \varepsilon_d - 2\Delta/5, \quad \varepsilon(e_g) = \varepsilon_d + 3\Delta/5.$$

The other terms in Hamiltonian (1) represent the Coulomb intraorbital (U_λ) and interorbital ($V_{\lambda\lambda'}$) repulsion and the Hund exchange $J_{\lambda\lambda'}$. For the sake of simplicity, we neglect the orbital dependence of the Coulomb matrix elements, assuming that the three parameters (*U*, *V*, and *J*) are related by the well-known condition $U = 2V + J$.

The kinetic energy of *d* electrons, as determined by interatomic hopping, is described by the Hamiltonian

$$H_t = \sum_{i,j,\sigma} \sum_{\lambda,\lambda'} t_{ij}^{\lambda\lambda'} a^\dagger_{i\lambda\sigma} a_{j\lambda'\sigma} + \text{H.c.}, \quad (2)$$

where $t_{ij}^{\lambda\lambda'}$ is the matrix element of hopping between *i*th and *j*th lattice sites. The main matrix element corresponds to the hops between nearest neighbors: $t \sim t_{pd}^2/|\varepsilon_p - \varepsilon_d|$. However, in view of the weak *p–d* hybridization between Fe and O atoms, this element is also small, $t \ll U$, which accounts for the strong electron correlation effects. Thus, the model parameters are the two Coulomb integrals, *U* and *J*, the crystal field magnitude Δ, the position of the one-electron *d* level relative to the top of the valence band ε_v ($\delta = \varepsilon_d - \varepsilon_v$), and the hopping integral *t*. The parameters will be determined by comparison with the experimental optical and photoemission spectra (see Section 4).

The Fe^{3+} ion has a d^5 configuration that can occur in various spin and orbital terms. The considerations below will also imply the knowledge of the terms of d^4 (Fe^{4+}) and d^6 (Fe^{2+}) configurations for description of the hole and electron creation in the many-electron system. The energies of terms in each of these d^n configurations are expressed via the Racah parameters *A*, *B*, and *C* [14, 15]. The *B*, *C* and Δ values for the terms of Fe^{3+} ion were determined in [16]: $B = 680$ cm^{-1}, $C = 3160$ cm^{-1}, and $\Delta = 12700$ cm^{-1}.

With neglect of a small uniaxial component of the crystal field, three t_{2g} levels and two e_g levels are degenerate. For the d^5 configuration, the ground state $^6A_1$ (with $S^z = +5/2$) is described by the wave function

$$|d^5, S = 5/2, S^z = 5/2\rangle = t^+_{1\uparrow} t^+_{2\uparrow} t^+_{3\uparrow} e^+_{1\uparrow} e^+_{2\uparrow} |0\rangle, \quad E_{5/2}(d^5) = 5\varepsilon_d + 10V - 10J, \quad (3)$$

where $t^+_{\lambda\sigma}$ ($\lambda = 1, 2, 3$) and $e^+_{\lambda\sigma}$ ($\lambda = 1, 2$) are the operators of creation of t_{2g} and e_g electrons, respectively, in one of the orbital states λ with the spin projection σ; and $|0\rangle$ is the vacuum state for *d* electrons. The lowest

excited term $^4T_1$ has a nonzero orbital moment and the spin $S = 3/2$.

The other excited terms with $S = 3/2$ can be written in a similar manner. For example, the term $^4A_1$ has a configuration of $t_{2g}^{3\uparrow}e_g^{\uparrow}e_g^{\downarrow}$ with an energy of

$$E''_{3/2}(d^5) = 5\varepsilon_d + 10V - 6J.$$

The low-spin excited term $^2T_2$ with a configuration of $t_{2g}^{3\uparrow}t_{2g}^{2\downarrow}$ has an energy of

$$E_{1/2}(d^5) = 5\varepsilon_d - 2\Delta + 2U + 8V - 4J.$$

Let us also write the ground and lower excited terms of the d^4 and d^6 configurations. For d^4, the main term ^5E for $S = 2$, $S^z = 2$ has a configuration of $t_{2g}^{3\uparrow}e_g^{\uparrow}$ with an energy of

$$E_2(d^4) = 4\varepsilon_d - 3\Delta/5 + 6V - 6J.$$

In what follows, we will also consider the term $^3T_1$ with $S = 1$ and an energy of

$$E_1(d^4) = 4\varepsilon_d - 8\Delta/5 + U + 5V - 3J$$

and the term $^1E(t_{2g}^{2\uparrow}t_{2g}^{2\downarrow})$ with $S = 0$ and the energy

$$E_0(d^4) = 4\varepsilon_d - 8\Delta/5 + 2U + 4V - 2J.$$

For d^6, the main term $^5T_2$ has a configuration of $t_{2g}^{3\uparrow}t_{2g}^{\downarrow}e_g^{2\uparrow}$ with an energy of

$$E_2(d^6) = 6\varepsilon_d - 2\Delta/5 + U + 14V - 10J,$$

the excited spin triplet $^3T_1$ has an energy of $E_1(d^6)$, and the spin singlet $^1A_1(t_{2g}^{3\uparrow}t_{2g}^{\downarrow})$ has an energy of $E_0(d^6)$. For the given values of B and Δ for each term, the corresponding energies relative to the lowest term of each configuration can be numerically determined using the Tanabe–Sugano diagrams [14, 15] (see Fig. 1 below). Note that a half-occupied d^5 configuration should possess the electron–hole symmetry. This symmetry is revealed when the one-electron energies are measured from the chemical potential level (see Section 5).

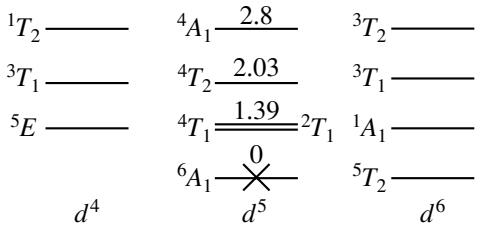

Fig. 1. A diagram of terms for Fe^{4+}, Fe^{3+}, and Fe^{2+} ions in $FeBO_3$. The cross indicates the occupied lowest sublevel of the term $6A_1$ of Fe^{3+} ion at $T = 0$. Figures at the levels indicate the energies (in eV) relative to the lowest sublevel.

3. ONE-PARTICLE GREEN FUNCTION OF *d* ELECTRONS

For establishing a relationship between many-electron terms and the spectrum of one-particle excitations determining the density of single-particle states, $N(E)$, we use an approach based on the generalization of Hubbard's ideas. Since the hops between atoms are small, the exact one-particle Green function $G_{k\sigma}(\omega) = \langle\langle a_{k\sigma} a_{k\sigma}^{\dagger} \rangle\rangle$ in the zero-order approximation with respect to t reduces to the local function $G^{(0)}(\omega)$. A consistent method for calculating the Green functions at $t/U \ll 1$ is provided by the representation of Hubbard's X operators constructed on the eigenstates of H_{at}. In our case, these are the aforementioned terms of d^n configurations. For the initial Hubbard model without orbital degeneracy, the corresponding perturbation theory has been developed in [17, 18] and the case of arbitrary degeneracy was considered in [19].

The structure of the Green functions of *d* electrons is revealed by the exact Lehmann spectral representation [20], in which electrons are described as superpositions of various quasiparticles. According to this, for $T = 0$,

$$G_{\sigma}(k, \omega) = \sum_m \left(\frac{A_m(k, \omega)}{\omega - \Omega_m^+} + \frac{B_m(k, \omega)}{\omega - \Omega_m^-} \right), \qquad (4)$$

where the quasiparticle energies are

$$\begin{aligned} \Omega_m^+ &= E_m(N+1) - E_0(N) - \mu, \\ \Omega_m^- &= E_0(N) - E_m(N-1) - \mu, \end{aligned} \qquad (5)$$

and their spectral weights are determined by the matrix elements

$$\begin{aligned} A_m(k, \omega) &= |\langle 0, N | a_{k\sigma} | m, N+1 \rangle|^2, \\ B_m(k, \omega) &= |\langle m, N-1 | a_{k\sigma} | 0, N \rangle|^2. \end{aligned} \qquad (6)$$

Here, $|m, N\rangle$ denotes the mth many-electron eigenstate of a system with N electrons,

$$H|m, N\rangle = E_m|m, N\rangle,$$

so that index m is essentially the band index numerating quasiparticles possessing the spin 1/2, the charge e (as

seen in the matrix elements), the energy Ω_m^+ (Ω_m^-), and the spectral weight A_m (B_m).

At a finite temperature, Lehmann's representation can be written, for example, for the retarded Green function (see [21, 22]),

$$G_\sigma^R(k, \omega) = \sum_{m,n} W_n \frac{A_{mn}(k, \omega)}{\omega - \Omega_{mn}^+ + i0} \times [1 + \exp(-\Omega_{mn}^+/T)], \quad (7)$$

where $\Omega_{mn}^+ = E_m(N+1) - E_n(N) - \mu$ and W_n is the statistical weight of state $|n\rangle$ determined by the Gibbs distribution with the thermodynamic potential Ω:

$$W_n = \exp(\Omega - E_n + \mu N)/T.$$

At $T \neq 0$, both the ground state $|0, N\rangle$ and excited states $|n, N\rangle$ are populated. In this case, quasiparticles are denoted by two indices, m and n, and are considered as excitations in a many-electron system, whereby electron added to the N-electron system in the state $|n, N\rangle$ induces a transition to the final $(N+1)$-electron state $|m, N+1\rangle$.

In Lehmann's representation, $|m, N\rangle$ is the unknown state of the whole crystal. As will be shown below, the same structure is inherent in the local Green function $G^{(0)}$ according to the generalized tight binding method [19]. This function is determined by the local many-electron terms $|m, N\rangle$ obtained in Section 2. In the case of $FeBO_3$, significant contributions result from the terms with $N = 4$, 5, and 6. Denoting $|m, N\rangle \equiv |p\rangle$, we define Hubbard's X operator at site f as

$$X_f^{pq} = |p\rangle\langle q| = |mN\rangle\langle m'N'|. \quad (8)$$

In standard writing, X operators appear with cumbersome notation indicating the initial and final states. In order to simplify this notation, we will use the idea of Zaitsev [17], according to which a pair of indices (p, q) is replaced by the so-called root vector $(p, q) \longleftrightarrow \boldsymbol{\alpha}(p, q) \equiv \boldsymbol{\alpha}$. Since the set of these vectors is denumerable, we introduce the numeration $\boldsymbol{\alpha} \longleftrightarrow \boldsymbol{\alpha}_n$ and then indicate only the number n of the root vector:

$$X_f^{pq} \longleftrightarrow X_f^{\boldsymbol{\alpha}(p,q)} \longleftrightarrow X_f^{\boldsymbol{\alpha}_n} \longleftrightarrow X_f^n.$$

This essentially implies that we construct a table of the correspondence between pairs (p, q), vectors $\boldsymbol{\alpha}_n$, and indices n necessary for explicitly calculating the commutation relations. Let us define vectors $\boldsymbol{\alpha}$ so as to correspond to the process of electron annihilation, $N_q - N_p = +1$. Then, the operators of electron creation (annihilation) in state $|f\lambda\sigma\rangle$ can be written in the X representation as

$$a_{f\lambda\sigma} = \sum_n \gamma_{\lambda\sigma}(n) X_f^n,$$
$$a_{f\lambda\sigma}^\dagger = \sum_n \gamma_{\lambda\sigma}^*(n) (X_f^n)^\dagger, \quad (9)$$

$$\gamma_{\lambda\sigma}(n) = \langle p|a_{f\lambda\sigma}|q\rangle = \langle mN|a_{f\lambda\sigma}|m', N+1\rangle. \quad (10)$$

Since the Hamiltonian H_{at} in the representation of the Hubbard operators is diagonal, the local Green functions of d electrons are immediately calculated as

$$G_{\lambda\lambda'\sigma}^{(0)}(k, \omega) = \sum_n \gamma_{\lambda\sigma}(n)\gamma_{\lambda'\sigma}^*(n) \frac{F(n)}{\omega - \Omega_n + i0}, \quad (11)$$

where $\Omega_n = E_{m'}(N+1) - E_m(N)$ is the quasiparticle energy and $F(n) = \langle X_f^{pp}\rangle + \langle X_f^{qq}\rangle$ is the occupation factor. Evidently, the Green functions (11) realize Lehmann's representation inside the unit cell but, in contrast to noncomputable energies and matrix elements in such a representation, all quantities entering into expression (11) can be calculated via the local characteristics of terms. Here, index n numerates quasiparticles with a charge e, spin 1/2, energy Ω_n, and spectral weight $A_{\lambda\lambda'\sigma}(n) = \gamma_{\lambda\sigma}(n)\gamma_{\lambda'\sigma}^*(n)F(n)$. By virtue of completion of the basis set of many-electron states $|p\rangle$, the total spectral weight is the same as that of free electrons.

In the diagram technique developed for X operators [17–19], the series of perturbation theory are constructed for the matrix Green function,

$$D_{nn'}(k, \omega) = \langle\langle X_k^n | (X_k^{n'})^\dagger \rangle\rangle_\omega,$$

rather than for the electron Green function related to the former in the X representation (9) as

$$G_{\lambda\lambda',\sigma}(k, \omega) = \sum_{n,n'} \gamma_{\lambda\sigma}(n)\gamma_{\lambda'\sigma}^*(n') D_{nn'}(k, \omega).$$

It is possible to write a generalized Dyson equation for the Green function $\hat{D}$, in which the perturbation renormalizes both the mass operator and the spectral weight. In the simplest Hartree–Fock approximation, the mass operator is determined as the Fourier transform of the hopping integral $t_{nn'}(k)$. As a result, the dispersion of quasiparticles is described by the following equation:

$$\det\|\delta_{nn'}(\omega - \Omega_n)/F(n) - t_{nn'}(k)\| = 0. \quad (12)$$

There is an obvious analogy between Eq. (12) and the dispersion equation obtained in the one-electron tight binding method: the structures of these expressions are identical. However, there are important distinctions as well: first, the local energies Ω_n include (unlike the one-electron energies $\varepsilon_{\lambda\sigma}$) intracell Coulomb interactions; second, the band index n of a quasiparticle is determined by a pair of indices of the initial and final states (differing from the band index λ of free electrons); third, the band structure of quasiparticles depends (via the occupation factors $F(n)$) on the density of electrons, temperature, and external fields; and fourth, a one-electron rigid band model cannot be developed for quasiparticles.

For determining the occupation numbers and the factors $F(n)$, it is necessary to solve an equation for the chemical potential. In the X representation, this equation can be written as

$$N_e = \sum_{f,m,N} N\langle X_f^{mN,mN}\rangle, \tag{13}$$

where $\langle X_f^{mN,mN}\rangle$ is the occupation number for the mth term of d^N configuration at the f site. Each term of d^N contributes N electrons to their total number N_e. A solution of this equation for $FeBO_3$ at $T = 0$ appears as

$$\langle X_f^{mN,mN}\rangle = 0$$

for all m and $N \neq 5$, and as

$$\langle X_f^{+5/2,+5/2}\rangle = 1$$

for $N = 5$. For the other d^5 configurations, the occupation numbers are zero. We take into account that, for $S = 5/2$ in a magnetically ordered phase, the term $E_{5/2}(d^5)$ is split with respect to the spin projection and only the lowest sublevel is occupied in each sublattice (+5/2 and –5/2 for sublattices A and B, respectively). Of course, there are zero-point quantum spin fluctuations leading to small population of the sublevels adjacent to $S = 5/2$ ($S^z = 3/2$); this small effect is considered below (see Section 5).

Interatomic hopping in the antiferromagnetic phase is suppressed by the spin–polaron effect [23]. For the hops between nearest neighbors, the effective hopping integral is determined by the product of occupation factors for the two sites belonging to different sublattices (A and B) [24]. For the lowest Hubbard band, the effective hopping Hamiltonian t_v differs from the one-electron integral t,

$$t_v^2 = t^2(\langle X_A^{+5/2,+5/2}\rangle + \langle X_A^{+2,+2}\rangle) \times (\langle X_B^{+5/2,+5/2}\rangle + \langle X_B^{+2,+2}\rangle), \tag{14}$$

where $\langle X_A^{pp}\rangle$ and $\langle X_B^{pp}\rangle$ are the occupation numbers of state $|p\rangle$ in the sublattices A and B, respectively; $|+5/2\rangle$ and $|+2\rangle$ are the spin sublevels of the terms $E_{5/2}(d^5)$ and $E_2(d^4)$ split with respect to the spin projection in the internal molecular field. For the sublattice A, the level $|+5/2\rangle$ is occupied (being the lowest sublevel), while for the sublattice B (where the lowest level is $|-5/2\rangle$) the level $|+5/2\rangle$ at $T = 0$ is unoccupied. Therefore, for $FeBO_3$ at $T = 0$, $\langle X_B^{+5/2,+5/2}\rangle = 0$ and, hence, the occupation numbers of all d^4 and d^6 sublevels are also zero and the widths of the Hubbard bands are close to zero too.

As a result, it is the poles of the local Green function (11) that determine single-particle contributions of the d-type to $N(E)$. Figure 1 shows the lowest levels of the d^4, d^5, and d^6 configurations (the cross indicates the occupied lowest sublevel of the term $^6A_1$ of the Fe^{3+} ion). Nonzero occupation factors are inherent in the transitions $^6A_1 \longrightarrow d^4$ (hole creation) and $^6A_1 \longrightarrow d^6$ (electron creation), but the matrix elements $\gamma_{\lambda\sigma}(n)$ given by formula (10) are nonzero only when the difference between the spins of terms $|p\rangle$ and $|q\rangle$ is 1/2. In the case under consideration, this implies that nonzero spectral weight and nonzero contribution to the density of states $N(E)$ will be only due to transitions between the lowest terms of all configurations:

$$\Omega_v = E_{5/2}(d^5) - E_2(d^4),$$
$$\Omega_c = E_2(d^6) - E_{5/2}(d^5). \tag{15}$$

The energy levels Ω_v and Ω_c, or the energy band $\Omega_v(k)$ and $\Omega_c(k)$ appearing with allowance for the weak interatomic hopping are analogs of the lower and upper Hubbard subbands. In addition, it is of interest to consider the quasiparticles for which the matrix element (10) differs from zero, while the spectral weight in the ground state is zero because of zero occupation numbers: such states are referred to as virtual. The virtual states can acquire nonzero weights upon a change in the electron configuration (e.g., in CuO_2 layers after hole doping [12]) or upon optical pumping of excited levels. For $FeBO_3$, an example of such a virtual d state is offered by a quasiparticle with an energy of

$$\Omega_v' = E_{3/2}(d^5) - E_2(d^4). \tag{16}$$

For comparison with experiment, it is necessary to determine the model parameters as discussed below. Previously, the optical absorption was studied separately in various spectral intervals. For this reason, the next section is devoted to the experimental absorption spectrum of $FeBO_3$ measured in a broad energy range, $E \leq 3$ eV, covering the entire bandgap width E_{g0}.

4. OPTICAL ABSORPTION SPECTRUM OF $FeBO_3$ IN A BROAD ENERGY RANGE

Previously [25–27], the optical absorption and the magneto-optical Faraday effect in iron borate were

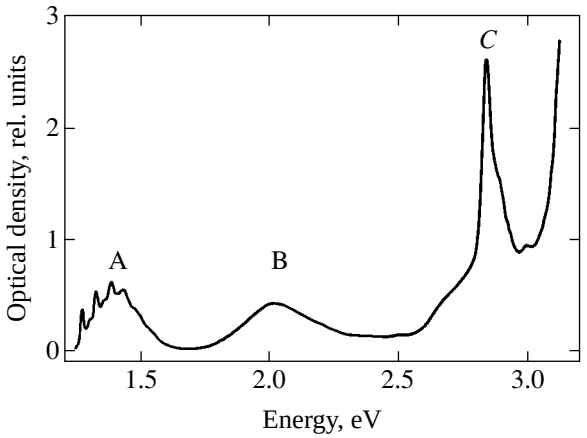

Fig. 2. The optical absorption spectrum of a $FeBO_3$ single crystal measured at 83 K.

studied in the visible and near-infrared (1.0–2.3 eV) spectral range. The results of analogous measurements in the region of strong absorption (2.6–3.3 eV) were reported in [4]. For the sake of convenience, we present the absorption spectrum of $FeBO_3$ measured in a broad energy range.

$FeBO_3$ single crystals were grown by V.V. Rudenko by spontaneous crystallization from solution melt. The crystals had the shape of thin hexahedral plates of a greenish color. The thicknesses of plates selected for the optical measurements were about 80 μm for the former spectral interval and 20 μm for the latter, the sample area in both cases being about 2 mm^2. Orientation of the plates corresponded to the easy magnetization plane, with the hard axis (coinciding with the optical axis of the crystal) being normal to the plate surface. Thus, by applying a small external field parallel to the plane of the crystal, it was possible to readily change the magnetic moment direction in the plane. The optical absorption spectra were measured using an automated spectrometer in a temperature range from 80 to 300 K.

The combination of a high Néel temperature (T_N = 348 K) and transparency in the visible spectral range allows us to perform a detailed comparison of the optical absorption and magneto-optical effects in the transmission mode in the region of three absorption bands wit minimum energies. Our measurements revealed the same three groups of absorption bands (A, B, and C, Fig. 2) as those reported in [4, 25–27]. These bands can be interpreted within the framework of the proposed many-electron model as described below. The main difference of our interpretation consists in that, in addition to the *d–d* transitions ${}^6A_{1g}({}^6S) \longrightarrow {}^4T_{1g}({}^4G)$ (group A), ${}^6A_{1g}({}^6S) \longrightarrow {}^4T_{2g}({}^4G)$ (group B), and ${}^6A_{1g}({}^6S) \longrightarrow {}^4A_{1g}$, ${}^4E_g({}^4G)$ (group C), the C band contains contributions due to the *p–d* transitions with charge transfer. We use the data of Fig. 2 and the X-ray photoelectron spectrum [12] for determining the Coulomb parameters of the model.

Proceeding from expressions for the *d–d* exciton energies, the parameters of A–C bands, and the additional peaks observed in the Kerr effect in terms of the Racah parameters, it was found [28] that B = 680 cm^{-1} and $\Delta \approx 12700$ cm^{-1} for $FeBO_3$; from the same data, we readily obtain C = 3160 cm^{-1}. Note that these values of B and C are somewhat lower than the analogous parameters of the free Fe^{3+} ion, but the ratio C/B = 4.65 is typical. The crystal field parameter Δ = 1.57 eV is greater as compared to the result (Δ = 1 eV) of the band calculations [12]. Using the known values of B and C, we determine the positions of the lowest excited terms of d^5 configurations with spins 3/2 and 5/2 (see Fig. 1) relative to the ground term ${}^6A_1$ from the Tanabe–Sugano diagrams [14]. The lowest terms of d^4 and d^6 configurations are also schematically depicted in Fig. 1. Their quantitative characteristics are not presented here because, generally speaking, each d^n configuration has its own level (depending on the chemical potential) from which the energies are measured. Moreover, even determination of the positions of excited terms relative to the lowest term for d^4 and d^6 configuration require knowledge of the corresponding B and Δ values.

Although the energies of these terms will not be required below, we present here for reference the corresponding energies determined from the Tanabe–Sugano diagrams assuming that the B and Δ for F^{4+}, Fe^{3+}, and Fe^{2+} are the same (in eV):

$$Fe^{4+}: E({}^5E) = 0, \quad E({}^3T_1) = 0.59, \quad E({}^1T_2) = 1.60,$$

$$Fe^{2+}: E({}^5T_2) = 0, \quad E({}^1A_1) = 0.17, \quad E({}^3T_1) = 0.76,$$

$$E({}^3T_2) = 1.18.$$

At the same time, the difference $E(d^{n+1})$–$E(d^n)$ of the energies of these terms has the meaning of energy increment per added electron. A peak at this energy is present on the density of single-particle states $N(E)$. In particular, for the lowest and highest Hubbard subbands, we obtain

$$\begin{aligned} \Omega_v &= \varepsilon_d + 3\Delta/5 + 4V - 4J, \\ \Omega_c &= \varepsilon_d - 2\Delta/5 + U + 4V. \end{aligned} \tag{17}$$

In $FeBO_3$ at T = 0, the level Ω_v is filled, while the level Ω_c is empty. This implies that the level Ω_v determines the *d*-type peak in the experimental X-ray photoelectron spectra or the X-ray absorption spectra. Indeed, such a peak was observed in the X-ray photoelectron spectra at a binding energy of 1.4 eV [12]. Measuring the energies of single-particle states from the top of the valence band ε_v, we set Ω_v–ε_v = –1.4 eV.

As can be seen from the optical absorption spectra, the intensity of peak C is much greater than those of peaks A and B. According to the commonly accepted interpretation of this fact, peak C is formed not only by *d–d* exciton (${}^6A_1 \longrightarrow {}^4A_1$), but makes a contribution due

to the $p^6d^5 \longrightarrow p^5d^6$ transition with charge transfer. The latter transition, reflecting the formation of a hole at the top of the valence band and the filling of level Ω_c, has the energy Ω_c–ε_v = 2.8 eV. Using the optical data, we can also determine the Hund exchange parameter J. The ground term $^6A_1$ and the excited term $^4A_1$ of the d^5 configuration possess the energies (independent of the crystal field) indicated in Section 2. The difference in these energies, determining the exciton energy for band C (22600 cm^{-1}), is $E(^4A_1)$–$E(^6A_1)$ = $4J$, from which it follows that J = 5650 cm^{-1} = 0.70 eV. This value of the Hund exchange is typical of 3d elements.

5. THE DENSITY OF SINGLE-PARTICLE STATES IN $FeBO_3$

A scheme of the density of states obtained for the proposed model is depicted in Fig. 3. The diagram shows empty s and p conduction bands with the bottom of the band ε_c, filled valence s and p bands with the top of the band ε_v, and the bandgap $\varepsilon_c - \varepsilon_v = E_{g0}$ = 2.9 eV. Thin solid lines (with neglect of the electron dispersion and damping, described by delta functions) show the energies of local d quasiparticles. With allowance for the spin–polaron suppression of interatomic d–d hopping in the magnetically ordered phase (14), the diagram shows only the local d maxima. A fluctuational contribution to the formation of narrow d bands certainly exists, being estimated as $t_v^2 \sim t^2 n_0$, where n_0 is the concentration of zero-point quantum fluctuations [24]. For a three-dimensional isotropic antiferromagnet, the typical value of $S - \langle S^z \rangle \approx 0.078$ [29] yields n_0 = 0.03 and $t_v \approx 0.035$ eV, with the corresponding bandwidth of $2zt_v \approx 0.42$ eV.

The upper filled d band Ω_v (15) is situated below the top of the valence band, while the lower empty d band Ω_c is below the bottom of the conduction band (inside the bandgap). Thus, the dielectric gap is determined by the excitations with charge transfer, $p^6d^5 \longrightarrow p^5d^6$, from the top of the valence band to the conduction Ω_c (charge transfer gap in terms of Zaanen *et al.* [30]. Note that the energy of transitions between lower and upper Hubbard bands,

$$\Omega_c - \Omega_v = E_2(d^6) + E_2(d^4) - 2E_{5/2}(d^5), \qquad (18)$$

can be considered as the effective Coulomb repulsion energy U_{eff}. In the Hubbard model $U_{eff} = U$, but in our case $U_{eff} \neq U$ (because of the orbital effects): $U_{eff} = U + 4J - \Delta$. This parameter (in comparison to the d-band width) determines the character of strong electron correlations in the system studied. The experimental values of Ω_v and Ω_c presented in Section 4 yield U_{eff} = 4.2 eV. For J = 0.7 eV and Δ = 1.57 eV, we obtain U = 2.97 eV and $V = (U–J)/2$ = 1.15 eV.

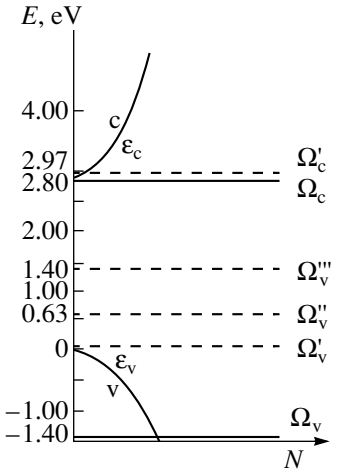

Fig. 3. Schematic diagram of the density of states in a magnetically ordered phase of $FeBO_3$. The Fermi level is situated above the top of the valence band ε_v.

A solution of Eq. (13) for the filled d^5 configuration is the chemical potential occurring between the empty level Ω_c and the filled level Ω_v:

$$\mu = \frac{1}{2}(\Omega_c + \Omega_v) = \varepsilon_d + \frac{U}{2} + \frac{\Delta}{10} + 4V - 2J.$$

Measuring energies relative to the chemical potential clearly reveals the electron–hole symmetry of the system:

$$\Omega_c - \mu = \frac{U}{2} - \frac{\Delta}{2} + 2J = -(\Omega_v - \mu).$$

The spectral weights of states Ω_v and Ω_c (with allowance for spin) is unity, rather than two as in the case of free electrons. In the diagram of Fig. 3, primed symbols indicate the virtual levels Ω'_v (16), $\Omega''_v = E(^4T_2) - E_2(d^4)$, and $\Omega'''_v = E(^4A_1) - E_2(d^4)$. The spectral weight of these levels in the ground state is zero, while the energies of transitions Ω'_v–Ω_v, Ω''_v–Ω_v, and Ω'''_v–Ω_v coincide with the exciton energies ε_A, ε_B, and ε_C. Under the conditions of optical pumping of the terms $^4T_1$, $^4T_2$, and $^4A_1$, their populations are no longer zero and the spectral weights of the virtual levels Ω'_v, Ω''_v, Ω'''_v are proportional to the concentration of optically excited Fe^{3+} ions. Thus, the exciton transitions inside the same d^n configuration can be represented by a virtual level in the one-electron density of states, and the $\Omega_v \longrightarrow \Omega'_v$ transition corresponds to the appearance of a hole in the band Ω_v and electron in the band Ω'.

Let us consider interpretation of the optical absorption spectrum within the framework of the proposed model. Since the exciton band A was used for determining the model parameters, the coincidence of theoreti-

cal and experimental energies for this band is trivial. At low temperatures, band *A* exhibits splitting into components A_1–A_4. Line A_1 is interpreted as a magnon satellite of the pure exciton line, and bands A_2–A_4, as magnon repetitions of the exciton–magnon line A_1 [25]. Indeed, at low temperatures, the spin levels $E_S(d^n)$ are split by the internal molecular field $I\langle S^z\rangle$ with respect to the spin projection M_z:

$$E_S(d^n, M_z) = E_S(d^n) - I\langle S^z\rangle M_z. \quad (19)$$

At $T = 0$, only the sublevel $M_z = +5/2$ of the term $^6A_1$ is occupied, so that transitions to the lower sublevel $M_z = +5/2$ of the $^4T_1$ term require the participation of a magnon.

As for the peak B, this absorption band corresponds to the exciton with $\varepsilon_B = E(^4T_2)$–$E(^6A_1)$. The band C contains contributions due to exciton $\varepsilon_C = E(^4A_1)$–$E(^6A_1)$ and due to transitions from the top of the valence band to the bands Ω_c (excitation with charge transfer).

6. TEMPERATURE DEPENDENCE OF THE INTENSITY OF BAND A

The temperature dependence of the band structure of local quasiparticles is revealed by general formula (7) showing temperature blurring of the distribution function. However, magnetic materials exhibit a stronger dependence due to interrelated electron and magnetic subsystems. All the absorption lines A_1–A_4 shift by 40 cm^{-1} toward lower energies when the temperature increases in the range from 30 to 200 K [28]. For the A_1 component, relation (19) yields

$$\Delta E(T) = E_{5/2}(d^5, +5/2) - E_{3/2}(d^5, +3/2)$$
$$= \Delta E(0) + I\langle S^z\rangle.$$

As the temperature *T* grows, the value of $\langle S^z\rangle$ decreases so that ΔE shifts toward smaller energies. The results of measurements of the sublattice magnetization $\langle S^z\rangle(T)$ [31] show that

$$\frac{\langle S^z\rangle(30\text{ K}) - \langle S^z\rangle(200\text{ K})}{\langle S^z\rangle(30\text{ K})} \approx \frac{1}{8}.$$

Using this estimate and the shift of exciton A_1,

$$\Delta E(30\text{ K}) - \Delta E(200\text{ K}) \approx \frac{5}{2}I \times \frac{1}{8},$$

it is possible to evaluate the Fourier transform of the interatomic exchange integral for $q = 0$ as $I \approx 0.015$ eV. This value determines the Néel temperature and, in the simplest variant of the mean field approximation,

$$T_N = IS(S+1)/3.$$

This yields $T_N = 317$ K, which is quite close to the experimental value of $T_N = 348$ K. We can also relate the magnetic and electron parameters by assuming that $I = Jz$, where *J* is the indirect exchange interaction between neighboring Fe^{3+} ions. Estimating this quantity as

$$J \sim 2t_v^2/U_{\text{eff}},$$

we obtain $t_v \approx 0.05$ eV. In Section 4, the fluctuation contribution was estimated as $t_v \approx 0.035$ eV. Therefore, the electron, magnetic, and optical properties of $FeBO_3$ in the proposed model exhibit a sufficiently good mutual agreement.

7. CONCLUSIONS

A question can arise as to how correct are the results of one-electron energy band calculations [12, 13] and can these results be applied to a system such as $FeBO_3$ with electron–spin correlations. Indeed, calculations [12] performed in the approximation of the local spin density functional ignore the correlation effects. As a result, the Fermi level falls within a partly occupied *d* band that implies the metallic state. Calculations [13] performed in the generalized gradient approximation take into account nonlocal corrections to the density functional, although it is not clear whether this approach adequately describes the regime of strong electron correlations. Nevertheless, the antiferromagnetic phase exhibits a dielectric state [13]. The calculation of pressure-induced changes in the magnetic state also rather well reproduces the magnetic and structural phase transitions observed recently [5, 6].

We believe that the results of band calculations in the local density functional approximation can be used as the initial information that should be supplemented by corrections for the transition from one-electron description of *d* electrons to local quasiparticles–excitations between d^n and d^{n+1} terms. There are no reasons for not believing the results of band calculations for the *s* and *p* states of boron and oxygen. The bandgap width $E_g^{(0)}$ is close to the experimental value, the crystal field Δ is 1.5 times the value according to the band theory, and the *d* band width in this theory is significantly overestimated.

On the other hand, it is by no means possible to use the level positions and occupation statistics obtained for the one-electron *d* band. The strong electron correlations not only split the *d* band into Hubbard's subbands, but (even more importantly) change the statistics of quasiparticles of the *d* type. As was demonstrated above, this gives rise to very unusual virtual states with the spectral weights determined by the nonstoichiometry or the incident light intensity.

Recently, [32], we interpreted the phase transition under pressure in $FeBO_3$ within the framework of the same model as being due to the intersection of the levels of terms $^6A_1$ and $^4T_2$ caused by increasing crystal field Δ. The model parameters in [32] were partly deter-

mined using the results of band calculations [12] and partly based on the optical data. Subsequently, it was established that this approach can lead to ambiguous results, since the theoretical and experimental values of the same quantity are not independent. In this study, we have used only experimental data for determining the model parameters. As a result, the values of $U \approx 3$ eV and $J \approx 0.7$ eV have proved to be much greater than those obtained in [32]. However, the conclusions [32] concerning the nature of the phase transition in $FeBO_3$ under pressure remain fully valid.

To summarize, we have constructed a many-electron model of the band structure of $FeBO_3$ taking into account both the one-electron *s* and *p* states of boron and oxygen and many-electron terms of Fe^{2+}, Fe^{3+}, and Fe^{4+} ions formed under the conditions of strong intraatomic *d–d* correlations. The density of one-electron states exhibits a set of narrow peaks related to local quasiparticles of the *d* type on the background of valence and conduction bands. Each quasiparticle corresponds to an electron with charge *e*, spin 1/2, and a reduced spectral weight. Only the sum of the spectral weights of all quasiparticles gives the one-electron spectral weight. Using this approach, it is possible to identify, with good fit to experiment, the main features of the absorption spectrum of $FeBO_3$ related both to excitons and the electron excitations with charge transfer. The parameters of electron and magnetic structures are also well consistent.

ACKNOWLEDGMENTS

The authors are grateful to A.V. Malakhovskiĭ, R.V. Pisarev, D.I. Khomskiĭ, and I.S. Édel'man for fruitful discussions of results and critical remarks.

This study was supported by the Russian Foundation for Basic Research (project no. 03-02-16286) and by the "Strongly Correlated Electrons" Program of the Department of Physical Sciences of the Russian Academy of Sciences.

REFERENCES

1. R. Wolff, A. J. Kurtzig, and R. C. Lecraw, J. Appl. Phys. **41**, 1218 (1970).
2. I. Bernal, C. W. Struck, and J. G. White, Acta Crystallogr. **16**, 849 (1963).
3. R. Diehl, Solid State Commun. **17**, 743 (1975).
4. I. S. Édel'man, A. V. Malakhovskiĭ, T. I. Vasil'eva, and V. N. Seleznev, Fiz. Tverd. Tela (Leningrad) **14**, 2810 (1972) [Sov. Phys. Solid State **14**, 2442 (1972)].
5. A. G. Gavriliuk, I. A. Trojan, R. Boehler, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 25 (2002) [JETP Lett. **75**, 23 (2002)].
6. V. A. Sarkisyan, I. A. Troyan, M. S. Lyubutin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 788 (2002) [JETP Lett. **76**, 664 (2002)].
7. N. B. Ivanova, V. V. Rudenko, A. D. Balaev, *et al.*, Zh. Éksp. Teor. Fiz. **121**, 354 (2002) [JETP **94**, 299 (2002)].
8. A. V. Kimel, R. V. Pisarev, J. Hohlfeld, and Th. Rasing, Phys. Rev. Lett. **89**, 287401 (2002).
9. R. O. Zaĭtsev, Pis'ma Zh. Éksp. Teor. Fiz. **65**, 881 (1997) [JETP Lett. **65**, 925 (1997)].
10. R. O. Zaĭtsev, Pis'ma Zh. Éksp. Teor. Fiz. **72**, 109 (2000) [JETP Lett. **72**, 77 (2000)].
11. V. A. Gavrichkov, S. G. Ovchinnikov, A. A. Borisov, and E. G. Goryachev, Zh. Éksp. Teor. Fiz. **118**, 422 (2000) [JETP **91**, 369 (2000)].
12. A. V. Postnikov, St. Bartkowski, M. Neumann, *et al.*, Phys. Rev. B **50**, 14849 (1994).
13. K. Parlinski, Eur. Phys. J. B **27**, 283 (2002).
14. Y. Tanabe and S. Sugano, J. Phys. Soc. Jpn. **9**, 766 (1954).
15. D. T. Sviridov and Yu. F. Smirnov, *Theory of Optical Spectra of Transition-Metal Ions* (Nauka, Moscow, 1977).
16. I. S. Édel'man and A. V. Malakhovskiĭ, Fiz. Tverd. Tela (Leningrad) **15**, 3084 (1973) [Sov. Phys. Solid State **15**, 2056 (1973)].
17. R. O. Zaĭtsev, Zh. Éksp. Teor. Fiz. **70**, 1100 (1976) [Sov. Phys. JETP **43**, 574 (1976)].
18. Yu. A. Izyumov, M. I. Katsnel'son, and Yu. N. Skryabin, *Magnetism of Collectivized Electrons* (Fizmatgiz, Moscow, 1994).
19. V. V. Val'kov and S. G. Ovchinnikov, *Quasi-Particles in Strongly Correlated Systems* (Sib. Otd. Ross. Akad. Nauk, Novosibirsk, 2001).
20. H. Lehmann, Nuovo Cimento **11**, 342 (1954).
21. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics* (Fizmatgiz, Moscow, 1962; Prentice Hall, Englewood Cliffs, N.J., 1963).
22. E. M. Lifshitz and L. P. Pitaevskiĭ, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics* (Nauka, Moscow, 1978; Pergamon, New York, 1980), Part 2.
23. É. L. Nagaev, *Physics of Magnetic Semiconductors* (Nauka, Moscow, 1979).
24. S. G. Ovchinnikov, Zh. Éksp. Teor. Fiz. **107**, 796 (1995) [JETP **80**, 451 (1995)].
25. B. Andlauer, O. F. Schirmer, and J. Schneider, Solid State Commun. **13**, 1655 (1973).
26. A. J. Kurtzig, R. Wolff, R. C. Le Graw, and J. W. Nielsen, Appl. Phys. Lett. **14**, 350 (1969).
27. K. Egashira, T. Manabe, and H. Katsiraki, J. Phys. Soc. Jpn. **31**, 602 (1971).
28. A. V. Malakhovskii and I. S. Edelman, Phys. Stat. Sol. **B74**, K145 (1976).
29. P. W. Anderson, Phys. Rev. **86**, 694 (1952).
30. J. Zaanen, G. A. Sawatzky, and J. W. Allen, Phys. Rev. Lett. **55**, 418 (1985).
31. A. M. Kadomtseva, R. Z. Levitin, Yu. F. Popov, *et al.*, Fiz. Tverd. Tela (Leningrad) **14**, 214 (1972) [Sov. Phys. Solid State **14**, 172 (1972)].
32. S. G. Ovchinnikov, Pis'ma Zh. Éksp. Teor. Fiz. **77**, 808 (2003)[JETP Lett. **77**, 676 (2003)].

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 1, 2004, pp. 14–23.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 1, 2004, pp. 19–29.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Exchange Dipole Interaction in a Multilevel Cooperative System of Atoms

E. V. Orlenko* and B. G. Matisov

St. Petersburg State Technical University, St. Petersburg, 195251 Russia

**e-mail: quark@citadel.stu.neva.ru*

Received March 10, 2003

Abstract—Atomic systems with three or more equidistant energy levels, in which a cascade process is possible, are considered. Hamiltonians obtained for such systems are analogous to Heisenberg Hamiltonians, but for systems with integral spins. For Dicke states in multilevel systems, quantum-mechanical mean values of the energy of a cooperative system are calculated taking into account weak interactions between atoms. The type of emission preceding superradiant avalanche emission of the system is analyzed. It can be expected that a coherent state may be formed, in which collective processes affect one another not only via population of the intermediate common layer, but also via phasing of pure quantum states. The single superfluorescence pulse that can be formed in this case is not a simple superposition of two or more pulses of sequential superradiant transitions in two-level systems. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known that simple quantum systems (ensembles of two-level atoms interacting via a radiation field, electrostatic field, dipole–dipole interaction, etc.) exhibit a number of light-induced phase transitions [1]. Analysis of such nonequilibrium phase transitions has shown that a deep-rooted analogy exists with second-order nonequilibrium phase transitions that emerge in a spin system when there is interaction with a constant (Coulomb exchange interaction) that exceeds the energy of thermal motion, leading to spin disorientation. Spontaneous spin coorientation arising in the system in this case is manifested on a macroscopic scale in residual magnetization. Phase transitions are also induced in quantum-optical systems by an interaction leading to the establishment of a definite order in the orientation of so-called energy spins (or isospins) [2, 3] observed in the case of Dicke superradiance [4]. The correlation of optical dipole moments of individual atoms observed in this case leads to the formation of a macroscopic dipole moment, which is proportional to the number of emitters. Consequently, the superradiant intensity turns out to be proportional to the square of this number, while the superfluorescence time is inversely proportional to this number. Moreover, as was shown in [1] from analysis of the polariton generation state in an open medium of two-level atoms interacting via the Stokes field during Raman scattering, the emergence of the superradiance regime is substantially one possible phase transition.

Superfluorescence belongs to the class of coherent optical phenomena. The concept of coherence in this case pertains to the emitting system rather than to the electromagnetic field. The reason for the emergence and evolution of coherence is assumed to be associated with the total radiation field of atoms that affects the state of each atom. Consequently, the mechanism of interaction via the reradiated field is regarded as the most universal type of interaction in such systems. However, an atomic ensemble in the medium of two-level atoms with constant dipole moments can also be collectivized due to the static dipole–dipole interaction. In crystals, an additional interaction via phonons also takes place.

The analogy with equilibrium second-order phase transitions occurring in magnetic systems is so much more deep-rooted that the Hamiltonian describing the behavior of two-level atoms in the radiation field and taking into account the interaction between atoms is analogous to the Heisenberg Hamiltonian for spin systems. Attempts at reducing the Hamiltonian directly to the Heisenberg Hamiltonian [5] as in the case of spin systems were repeatedly made in quantum optics.

At the same time, the group-theoretical approach [6, 7], which in fact generalizes the Dicke theory to the case of multilevel emitters, was developed and effectively applied even in early works on superradiance [6, 7]. It was noted that it is important in the Dicke theory to choose the stationary states of the unperturbed Hamiltonian in the form of irreducible representations of the SU_2 group in the energy space of a system. In the case of a multilevel system, the basis of irreducible representation of the SU_n group (n is the number of energy levels) plays a similar role.

The electric dipole interaction plays an important role in the stabilization of superradiant states even when atoms possess only dipole moments of transitions. The effect of the Coulomb interaction on super-

fluorescence of a system of two-level atoms was studied in [8] using a semiclassical approach. It was shown that the Coulomb interaction induces coherent transport of excitation in the system of atoms, which leads to approximate spatial homogeneity of inversion in a chain of atoms. Thus, the statement was formulated and proved consistently that the Coulomb dipole–dipole interaction not only preserves (contrary to the prevailing opinion) the superradiant state, but even facilitates the stabilization of this state. It was also proved that the Coulomb interaction must be taken into account in all systems with a small Fresnel number, since the time τ_c of "exchange" of excitations in such systems is much shorter than the superradiance time τ_R. An attempt at including the Coulomb interaction for a small number of atoms was also made in earlier publications [9], where it was proved that the electrostatic interaction in the semiclassical approximation does not affect the superfluorescence dynamics, but leads to phase modulation. However, a more detailed analysis of superfluorescence dynamics taking into account the Coulomb interaction [8] revealed that the fluorescence as a function of time exhibits clearly manifested oscillations, which could be attributed to the propagation of wave-type excitations in the system.

Mechanisms of phasing and the role of the dipole–dipole interatomic interaction in cooperative systems were studied in [10], where superfluorescence effects in various physical systems, including small (Dicke collection of atoms) and extended (coherent waves in a magnetized plasma) systems, were studied from a unified point of view.

The Hamiltonian of a system of two-level atoms, which takes into account the Coulomb dipole interaction as well as interactions via the reradiated field and is similar to the Heisenberg Hamiltonian in the theory of magnetism, was derived from the first principles in [11]. The Hamiltonian obtained in this way is used for studying wave excitations of the system, which are similar to spin waves in ferro- or antiferromagnets. It was shown that it is these excitations which lead to the characteristic temperature dependence of the superfluorescence intensity. In addition, the critical temperature at which superfluorescence disappears and a second-order phase transition occurs in the system was calculated.

The case of three-level systems was studied in detail in [6] on the basis of the group-theoretical method. Namely, complete classification of coherent states is given and properties of spontaneous radiation are analyzed.

In a system of three-level atoms, the following modes leading to qualitatively different dynamics are possible.

1. Cascade superradiance [12], when the 3–2 and 2–1 transitions are allowed and the 3–1 transition is forbidden.

2. Two-frequency superradiance (Λ scheme) [13], when the 3–2 and 3–1 transitions are allowed, but the 2–1 transition is forbidden.

3. Superradiance with a common lower level (V scheme).

In the superradiance limit, when the collective superfluorescence times are much shorter that the time of noncorrelated decay, the equations of pure superradiance generalizing the equation for an aggregate of two-level atoms were derived in [6] and the solutions are classified depending on the initial occupancy of levels; the cases listed above were actually derived in [6].

Cascade superfluorescence was considered in [12] in the semiclassical approximation. It was shown that, under the condition $\tau_1 < \tau_2$ (τ_1 is the time of the 3–2 collective decay and τ_2 is the same of the 2–1 transition), superradiance pulses for the upper and lower transitions do not overlap and the system decay can be described in the two-level model approximation. For $\tau_1 > \tau_2$, the pulses emitted in the first and second transitions overlap and it is assumed that these pulses mutually affect the decay kinetics only via the population of the common level. The same mutual effect of collective processes on two transitions is also typical of other decay modes in three-level systems. Equidistant levels and semiclassical states of the Glauber type for the multilevel Dicke problem are considered in [7].

In the above-mentioned publications, the main mechanism determining the system collectivization is the interatomic interaction via the reradiated field; the same interaction also determines the structure of the Hamiltonian of the system. It should be noted, however, that the constant of this interaction is small [1] in comparison with the equilibrium temperature of the atomic system, with the equilibrium thermal radiation of the system, and, finally, with the intensity of the pumping field which plays the role of the parameter being varied in radiating systems and is an analog of temperature. Tracing the adopted analogy with spin systems, we can recollect that the constant of the direct magnetic-dipole interaction is also infinitesimal as compared to the characteristic constants of the ferromagnetic system, although the operator of this interaction explicitly contains the dot product of the spin operators for pairwise interacting particles. However, it is not this product that determines the collective effect of coorientation of spins in the system. It was proved in [11] that the structure of the Dicke Hamiltonian for a system of two-level atoms might indeed resemble the Heisenberg Hamiltonian, while the interaction constant is of the same exchange nature as the Heisenberg parameter. In addition, this parameter is no longer small and may be as large as several T's in accordance with the estimated given in [11], where T is the gas temperature in energy units. This interaction leads to splitting of energy levels of the system of two-level atoms, which are degenerate in the principal cooperative number.

Here, we consider a special case of a system with three or more equidistant energy levels, in which a cascade process can occur. For such systems, Hamiltonians similar to Heisenberg Hamiltonians [11], but for systems of integral spins were obtained. In contrast to two-level atoms, for which Dicke's states are the eigenstates for the corresponding Hamiltonian, these states for multilevel atoms are not the eigenstates for the Hamiltonian taking into account the dipole interaction. For these states, quantum-mechanical mean energy values for cooperative systems are calculated taking into account weak interatomic interaction. Analysis of the radiation emitted by the system prior to superradiant avalanche emission shows that we can expect the stabilization of a coherent state in which collective processes mutually affect one another not just via the population of the intermediate common level, but also via phase relations between pure quantum states. In this situation, a single superfluorescence pulse can be formed, which is not a simple superposition of two or more pulses of sequential superradiant transitions in two-level systems. It turns out, in addition, that three-level systems must exhibit a delay preceding the final emission of coherent radiation, which is analogous to that observed in a system of two-level atoms [11]. Such a delay should not be observed for four- or five-level systems, and the process of spontaneous emission smoothly transforms into a superradiant state.

2. HAMILTONIAN OF A COOPERATIVE SYSTEM ($j = 1$)

Let us consider a three-level cascade diagram in which the energy levels form a sequence

$$E_1 < E_2 < E_3.$$

The state of each atom can be described by a spinor χ [4], where

$$\chi(1) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \chi(2) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \chi(3) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

indicate that an atom is in an energy state with $E = E_1$, E_2, and E_3, respectively, or, in other notation,

$$\chi(1) = |j = 1, j_z = -1\rangle, \quad \chi(2) = |j = 1, j_z = 0\rangle,$$

$$\chi(3) = |j = 1, j_z = 1\rangle,$$

where j and j_z denote the value of the so-called isoenergy spin and its component. Then the state of a system of two noninteracting atoms can be written in the form of the simple product of spinors χ_I and χ_{II}, corresponding to atoms I and II:

$$\chi_{I, II} = \chi_I(i)\chi_{II}(k),$$

where indices i and k assume values of 1, 2, and 3.

If, analogously to [4], we introduce the operator

$$\hat{j}_z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

the Hamiltonian of the noninteracting system can be written in the form

$$\hat{H}_0 = \varepsilon(\hat{j}_{Iz} + \hat{j}_{IIz}),$$
$$\varepsilon = E_3 - E_2 = E_2 - E_1.$$

In this case, we have

$$\hat{H}_0\chi_{I, II} = \varepsilon(\hat{j}_{Iz} + \hat{j}_{IIz})\chi_{I, II}. \quad (1)$$

Some energy states of a system of two interacting atoms, e.g., such that the first atom is in a state with energy E_1 and the other is in a state with energy E_2 (in this case, the energy of the system is $E = E_1 + E_2$) or, similarly, for a state with $E = E_1 + E_3$, may correspond to the functions

$$\chi_{I, II} = \chi_I(1)\chi_{II}(2), \quad \chi'_{I, II} = \chi_I(2)\chi_{II}(1),$$
$$\chi_{I, II} = \chi_I(2)\chi_{II}(3), \quad \chi'_{I, II} = \chi_I(3)\chi_{II}(2),$$
$$\chi_{I, II} = \chi_I(1)\chi_{II}(3), \quad \chi'_{I, II} = \chi_I(3)\chi_{II}(1).$$

Thus, the system is degenerate. In addition, functions $\chi_{I, II}$ and $\chi'_{I, II}$ are mutually orthogonal.

We will take into account the interaction of atoms whose mechanism may be any of those listed above in accordance with the experimental situation. We will describe this interaction symbolically with the help of operator $\hat{V}_{I, II}$. The constant of any of the interaction types listed above is smaller than the energy difference ε so that conventional perturbation theory with degeneracy is applicable. The correction to the total energy of the system for this perturbation is well known [7]:

$$\varepsilon^{(1)} = K \pm A, \quad (2)$$

where,

$$K = \langle\chi_{I, II}|\hat{V}_{I, II}|\chi_{I, II}\rangle, \quad A = \langle\chi'_{I, II}|\hat{V}_{I, II}|\chi_{I, II}\rangle$$

in the case under consideration. In this case, the isospin functions of the symmetric and antisymmetric forms will be regular wave functions in the zeroth approximation. The symmetric regular wave functions corresponding to the upper sign in formula (2) describe the

states with total isospin J of a pair of atom and with its component J_z:

$$\begin{aligned}
|J = 2, J_z = 2\rangle &= |1, 1\rangle_1|1, 1\rangle_2 \\
|J = 2, J_z = 1\rangle &= \frac{1}{\sqrt{2}}(|1, 1\rangle_1|1, 0\rangle_2 + |1, 0\rangle_1|1, 1\rangle_2), \\
|J = 2, J_z = 0\rangle &= |1, 0\rangle_1|1, 0\rangle_2, \\
|J = 2, J_z = -1\rangle &= \frac{1}{\sqrt{2}}(|1, -1\rangle_1|1, 0\rangle_2 + |1, 0\rangle_1|1, -1\rangle_2), \\
|J = 2, J_z = -2\rangle &= |1, -1\rangle|1, -1\rangle, \\
|J = 0, J_z = 0\rangle &= \frac{1}{\sqrt{3}}(|1, 1\rangle_1|1, -1\rangle_2 \\
&\quad + |1, -1\rangle_1|1, 1\rangle_2 - 2|1, 0\rangle_1|1, 0\rangle_2).
\end{aligned} \tag{3a}$$

The regular antisymmetric wave functions corresponding to the lower sign have the form

$$\begin{aligned}
|J = 1, J_z = 1\rangle &= \frac{1}{\sqrt{2}}(|1, 1\rangle_1|1, 0\rangle_2 - |1, 0\rangle_1|1, 1\rangle_2), \\
|J = 1, J_z = 0\rangle &= \frac{1}{\sqrt{2}}(|1, 1\rangle_1|1, -1\rangle_2 - |1, -1\rangle_1|1, 1\rangle_2), \\
|J = 1, J_z = -1\rangle &= \frac{1}{\sqrt{2}}(|1, -1\rangle_1|1, 0\rangle_2 - |1, 0\rangle_1|1, -1\rangle_2).
\end{aligned} \tag{3b}$$

It can be seen from relation (2) that, if the interaction is taken into account, antisymmetric states are more advantageous for $A > 0$ and, conversely, symmetric states are more advantageous for $A < 0$. The energy states of a diatomic system with $E = 2E_1$ and $E = 2E_3$ are initially nondegenerate; for such states, only one (namely, symmetric) state is realized. The state with $E = 2E_2$ is doubly degenerate; on account of the interaction, this state splits into two states (symmetric state $|J = 0, J_z = 0\rangle$ and antisymmetric state $|J = 1, J_z = 0\rangle$) corresponding to different energy sublevels.

We introduce the operator

$$\hat{P}_{1,1} = \hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2 + (\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2)^2 - 1,$$

where $\hat{j}_{1x}$, $\hat{j}_{1y}$, $\hat{j}_{1z}$, $\hat{j}_{2x}$, $\hat{j}_{2y}$, $\hat{j}_{2z}$ are equivalent to the matrices of the operator of the x, y, and z components of the momentum, which is equal to unity. We can verify that the eigenvalue of the operator

$$\hat{J}^2 = (\hat{\mathbf{j}}_1 + \hat{\mathbf{j}}_2)^2$$

for a symmetric state corresponds to the total isospin J of a system of two noninteracting atoms, which is equal to 2 or 0. For an antisymmetric state, the total isospin J of the system is equal to 1. Similarly, it can easily be seen that the eigenvalues of operator $\hat{P}_{1,1}$ are equal to +1 for symmetric states and to –1 for antisymmetric states. Then, the operator describing the interaction of the two atoms and explicitly taking into account their isospin states can be written in the form

$$\hat{h}_{\text{int}} = K + A\hat{P}_{1,1}$$

or

$$\hat{h}_{\text{int}} = K + A(\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2 + (\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2)^2 - 1). \tag{4}$$

If we now consider a system of N atoms and take into account pair interactions only, the corresponding Hamiltonian can be written in the form

$$\begin{aligned}
\hat{H}_{\text{int}} &= (K - A)\frac{N(N-1)}{2} \\
&\quad + A\sum_{k<l}[(\hat{\mathbf{j}}_k \cdot \hat{\mathbf{j}}_l)^2 + \hat{\mathbf{j}}_k \cdot \hat{\mathbf{j}}_l],
\end{aligned} \tag{5}$$

where k and l are the numbers of the atoms. Then the total Hamiltonian describing the system of pairwise interacting three-level atoms has the form

$$\hat{H} = \varepsilon\sum_k \hat{J}_{kz} + A\sum_{k<l}[(\hat{\mathbf{j}}_k \cdot \hat{\mathbf{j}}_l)^2 + \hat{\mathbf{j}}_k \cdot \hat{\mathbf{j}}_l] + NE_2. \tag{6}$$

It should be noted here that, in contrast to a two-level system, the Dicke states [4] characterized by the value of the total isospin of the system (or the so-called the cooperative quantum number R) and the isospin component along the z axis (polarization number m) are not the eigenstates of operator (6) due to the presence of the nonlinear term in this operator.

3. DICKE STATES

We will calculate the energy values for a multiparticle isospin system. The second sum appearing in formula (5) can be expressed in terms of operator $\hat{R}^2$ of the squared total isospin of the system:

$$\sum_{k<l}\hat{\mathbf{j}}_k \cdot \hat{\mathbf{j}}_l = \frac{1}{2}\left(\hat{R}^2 - \sum_k \hat{\mathbf{j}}_k^2\right). \tag{7}$$

The quadratic term can be written in terms of operator $\hat{J}^2$ of the squared pair spin and calculated taking

into account quantum averaging over states with definite values of the pair isospin,

$$\sum_{k<l}(\hat{\mathbf{j}}_k\cdot\hat{\mathbf{j}}_l)^2 = \frac{N(N-1)}{2} \times\sum_J\left\{(\hat{J}^2-2\hat{j}^2)^2\left(\sum_{J_z,j_{z_1}j_{z_2}}\langle J,J_z|j_1j_{z_1}j_2j_{z_2}\rangle^2\right)\right\}, \quad (8)$$

where $\langle J, J_z|j_1j_{z_1}j_2j_{z_2}\rangle^2$ are the Clebsch–Gordon coefficients.

The eigenvalue of operator (7) acting on state $|R, m\rangle$ with total isospin R and its projection m has the form

$$\overline{\sum_{k<l}\hat{\mathbf{j}}_k\cdot\hat{\mathbf{j}}_l} = \frac{1}{2}(R(R+1)-Nj(j+1)). \quad (9)$$

The averaged value of operator (8) has the form

$$\overline{\sum_{k<l}(\hat{\mathbf{j}}_k\cdot\hat{\mathbf{j}}_l)^2} = \frac{4}{3}\frac{R(R-1)}{2}. \quad (10)$$

Then the energy of interaction corresponding to formula (5) is given by

$$E_{\text{int}} = (K-A)\frac{N(N-1)}{2} + A\left[\frac{2R(R-1)}{3}+\frac{1}{2}(R(R+1)-Nj(j+1))\right]. \quad (11)$$

Considering that $j = 1$, for the energy of interaction, we obtain

$$E_{\text{int}} = K\frac{N(N-1)}{2}-A\frac{N(N+1)}{2}+\frac{A}{6}R(7R-1). \quad (12)$$

The total energy of a cooperative system has the form

$$E_{\text{tot}} = \varepsilon m + NE_2 + K\frac{N(N-1)}{2} - A\frac{N(N+1)}{2}+\frac{A}{6}R(7R-1). \quad (13)$$

The states with a definite polarization number m are not degenerate any longer since the degeneracy in the total isospin (or cooperative quantum number of the system) is removed in view of the interaction described by the second term in formula (5). Dicke states $|R, m\rangle$ with different values of R (R can assume values from N to $|m|$), but with the same m now correspond to different energy sublevels as in the case of two-level systems [11].

It can easily be seen that, for $A > 0$, the state corresponding to the minimal energy of a sublevel is the state with the minimal possible value of $R = |m|$ for a given m; i.e., $R = 0$ in a state with $m = 0$. The state

$$|R=0, m=0\rangle = \prod_{k<l}\frac{1}{\sqrt{3}}(|j_k=1, j_{zk}=1\rangle_k|j_l=1, j_{zl}=-1\rangle_l + |j_k=1, j_{zk}=-1\rangle_k|j_l=1, j_{zl}=1\rangle_l - 2|j_k=1, j_{zk}=0\rangle|j_l=1, j_{zl}=0\rangle) \quad (14)$$

is completely symmetric.

According to Dicke [4], the radiation intensity of the system is defined as

$$I = I_0(R+m)(R-m+1),$$

where I_0 is the intensity of radiation emitted by an atom [4, 13]. This expression clearly shows that coherent radiation can be emitted if R is large and $|m|$ is small. In state (14), the system does not emit at all; the probability of a transition from this state is equal to zero.

Let us consider in greater detail the evolution of the system until the instant when its coherent decay begins. For a transition of the system from state $|m\rangle$ to state $|m-1\rangle$, the probability of the system being on one of sublevels degenerate in R is finite. This probability can be calculated with the help of the relation specially derived in [7] for a degenerate system:

$$w_{R-1} = \left|\frac{1}{V_{RR}-V_{R-1R-1}}\sum_r\frac{V_{R-1}V_{rR}}{\varepsilon(m-(m-1))}\right|^2 = \left|\left(K\frac{N(N-1)}{2}-A\frac{N(N+1)}{2}+\frac{A}{6}R(7R-1)\right)\times\left(K\frac{N(N-1)}{2}-A\frac{N(N+1)}{2}+\frac{A}{6}(R-1)(7R-8)\right)\times\left(\varepsilon\frac{A}{6}(14R-8)\right)^{-1}\right|^2.$$

In this case, the system tends to be on a sublevel with the minimal possible R (namely, with $R = m$). This leads to a decrease in the radiation intensity at each step in accordance with the law

$$I\propto I_0(R+m) = 2mI_0.$$

From all possible states $|R, m\rangle$ with $m = 0$, it is precisely the state with $R = 0$ that is most advantageous from the energy point of view. However, it was noted by Dicke [4] that the system does not emit in this state, being in a sort of "frozen" state in a shallow well. The system can perform a transition from a sublevel with $R = 0$ to a

sublevel with $R = N$ only as a result of interaction with the radiation field, which induces avalanche emission of the entire cooperative system [4, 11]. This effect of gradual "freezing" of the system can explain the existing delay preceding the emission of a superradiant pulse. A transition to this state is always accompanied by the emission of incoherent radiation with a decreasing intensity.

Thus, the state preceding a superradiant pulse is the state of the system in which half the atoms are distributed uniformly between the upper level $E = E_3$ and the lower level $E = E_1$, while the other half populate the middle level $E = E_2$.

4. HAMILTONIAN OF A FIVE-LEVEL COOPERATIVE SYSTEM ($j = 2$)

A cooperative system of five-level atoms can be described as a system of particles with an energy spin (isospin) of $j = 2$. In the case of a pair of such atoms, the symmetry of energy levels is as follows: even states with isospins $J = 4, 2, 0$ of the pair of atoms are symmetric, while odd states with isospins $J = 3, 1$ of the pair of atoms are antisymmetric. We will write the complete system of eigenvalues of the dot product operator $\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2$ of atomic isospins:

$$\begin{aligned} J = 4,&\quad \overline{\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2} = 4 \text{ (symmetric)},\\ J = 3,&\quad \overline{\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2} = 0 \text{ (antisymmetric)},\\ J = 2,&\quad \overline{\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2} = -3 \text{ (symmetric)},\\ J = 1,&\quad \overline{\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2} = -5 \text{ (antisymmetric)},\\ J = 0,&\quad \overline{\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2} = -6 \text{ (symmetric)}. \end{aligned} \tag{15}$$

Let us now compose an operator $\hat{P}_{2,2}$ taking into account the symmetry properties of the spin function of a pair of atoms in such a way that the action of this operator on a symmetric state results in an eigenvalue of (+1), while its action on an antisymmetric state results in an eigenvalue of (–1). This operator has the form

$$\begin{aligned} \hat{P}_{2,2} = \frac{1}{36}(\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2)^4 + \frac{1}{6}(\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2)^3\\ -\frac{13}{36}(\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2)^2 - \frac{5}{2}(\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2) - 1. \end{aligned} \tag{16}$$

If we now replace $\pm A$ in expression (2) by this operator, we obtain the following expression for the Hamiltonian of a system of N five-level pairwise interacting atoms:

$$\begin{aligned} \hat{H}_{\text{int}} = (K - A)\frac{N(N-1)}{2} + \frac{A}{6}\\ \times \sum_{i<j}\left\{\frac{1}{6}(\hat{\mathbf{j}}_i \cdot \hat{\mathbf{j}}_j)^4 + (\hat{\mathbf{j}}_i \cdot \hat{\mathbf{j}}_j)^3\right.\\ \left. -\frac{13}{6}(\hat{\mathbf{j}}_i \cdot \hat{\mathbf{j}}_j)^2 - 15(\hat{\mathbf{j}}_i \cdot \hat{\mathbf{j}}_j) - 6\right\}. \end{aligned} \tag{17}$$

5. GROUND STATE OF A SYSTEM WITH $j = 2$

We will calculate the values of the operators appearing in formula (17) and averaged over Dicke states taking into account the Clebsch–Gordon coefficients:

$$\begin{aligned} \overline{\sum_{i<j}(\hat{\mathbf{j}}_i \cdot \hat{\mathbf{j}}_j)^2} = 6R(R-1),\\ \overline{\sum_{i<j}\hat{\mathbf{j}}_i \cdot \hat{\mathbf{j}}_j} = \frac{1}{2}(R(R+1) - 6N),\\ \overline{\sum_{i<j}(\hat{\mathbf{j}}_i \cdot \hat{\mathbf{j}}_j)^3} = -3R(R-1),\\ \overline{\sum_{i<j}(\hat{\mathbf{j}}_i \cdot \hat{\mathbf{j}}_j)^4} = \frac{588}{5}R(R-1). \end{aligned} \tag{18}$$

Then the energy of interaction has the form

$$E_{\text{int}} = (K - A)\frac{N(N-1)}{2} - \frac{A}{2}\left\{\frac{7R}{10}(13R + 37) - 15N\right\}.$$

The total energy of a cooperative system with $j = 2$ has the form

$$\begin{aligned} E_{\text{tot}} = \varepsilon m + (K - A)\frac{N(N-1)}{2}\\ -\frac{A}{2}\left\{\frac{7R}{10}(13R + 37) - 15N\right\}. \end{aligned} \tag{19}$$

It can be seen from this relation that the ground state for $A > 0$ is a state with the maximal cooperative number $R = N$ of the system. Thus, immediately after the pumping to the uppermost state (in m), the system always passes to symmetric states only, remaining on sublevels with the maximal cooperative number, since it is these states that correspond to the lowest sublevels for a

given m. The system can emit radiation in such a state, the radiation intensity increasing until the system attains a state with small values of $|m|$ and with large values of R. In this case, the intensity becomes proportional to the squared number of atoms and emission becomes cooperative.

6. COOPERATIVE SYSTEM OF FOUR-LEVEL ATOMS

A cooperative system of four-level atoms can be described as a system of particles with an isospin of $j = 3/2$. Possible symmetric and antisymmetric pair states of such a system will be listed below. As in the previous cases, all symmetric states correspond to the correction to the pair interaction energy of atoms, which is given by

$$\varepsilon^{(1)} = K + A,$$

while all antisymmetric combinations correspond to the correction given by

$$\varepsilon^{(1)} = K - A.$$

Thus, we can introduce an operator that automatically takes into account the symmetry of states and the corresponding sign of the energy correction:

$$\hat{P}_{3/2,3/2} = \frac{2}{9}(\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2)^3 + \frac{11}{18}(\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2)^2 - \frac{9}{8}(\hat{\mathbf{j}}_1 \cdot \hat{\mathbf{j}}_2) - \frac{67}{32}. \tag{20}$$

This operator has an eigenvalue equal to +1 for the action on symmetric states and an eigenvalue equal to –1 for the action on antisymmetric states.

Symmetric states have the form

$$|J = 3, J_z = 3\rangle = |3/2, 3/2\rangle|3/2, 3/2\rangle,$$

$$|J = 3, J_z = 2\rangle = \frac{1}{\sqrt{2}}(|3/2, 3/2\rangle|3/2, 1/2\rangle + |3/2, 1/2\rangle|3/2, 3/2\rangle),$$

$$|J = 3, J_z = 1\rangle = (|3/2, 3/2\rangle|3/2, -1/2\rangle + |3/2, -1/2\rangle|3/2, 3/2\rangle)\frac{1}{\sqrt{5}} + |3/2, 1/2\rangle|3/2, 1/2\rangle\sqrt{\frac{3}{5}},$$

$$|J = 3, J_z = 0\rangle = (|3/2, 1/2\rangle|3/2, -1/2\rangle + |3/2, -1/2\rangle|3/2, 1/2\rangle)\frac{3}{2\sqrt{5}} + (|3/2, 3/2\rangle|3/2, -3/2\rangle + |3/2, -3/2\rangle|3/2, 3/2\rangle)\frac{1}{2\sqrt{5}}, \tag{21}$$

$$|J = 1, J_z = 1\rangle = -\sqrt{\frac{2}{5}}|3/2, 1/2\rangle|3/2, 1/2\rangle + \sqrt{\frac{3}{10}}(|3/2, 3/2\rangle|3/2, -1/2\rangle + |3/2, -1/2\rangle|3/2, 3/2\rangle),$$

$$|J = 1, J_z = 0\rangle = \frac{3}{2\sqrt{5}}(|3/2, 3/2\rangle|3/2, -3/2\rangle + |3/2, -3/2\rangle|3/2, 3/2\rangle) - \frac{1}{2\sqrt{5}}(|3/2, 1/2\rangle|3/2, -1/2\rangle + |3/2, -1/2\rangle|3/2, 1/2\rangle),$$

while antisymmetric states have the form

$$|J = 2, J_z = 2\rangle = \frac{1}{\sqrt{2}}(|3/2, 3/2\rangle|3/2, 1/2\rangle - |3/2, 1/2\rangle|3/2, 3/2\rangle),$$

$$|J = 2, J_z = 1\rangle = \frac{1}{\sqrt{2}}(|3/2, 3/2\rangle|3/2, -1/2\rangle - |3/2, -1/2\rangle|3/2, 3/2\rangle),$$

$$|J = 2, J_z = 0\rangle = \frac{1}{2}(|3/2, 1/2\rangle|3/2, -1/2\rangle - |3/2, -1/2\rangle|3/2, 1/2\rangle) + \frac{1}{2}(|3/2, 3/2\rangle|3/2, -3/2\rangle - |3/2, -3/2\rangle|3/2, 3/2\rangle), \tag{22}$$

$$|J = 0, J_z = 0\rangle = \frac{1}{2}(|3/2, 3/2\rangle|3/2, -3/2\rangle - |3/2, -3/2\rangle|3/2, 3/2\rangle) - \frac{1}{2}(|3/2, 1/2\rangle|3/2, -1/2\rangle - |3/2, -1/2\rangle|3/2, 1/2\rangle).$$

If we now replace $\pm A$ in formula (2) by operator (20), we obtain the following expression for the Hamiltonian

of a system of N four-level pairwise interacting atoms:

$$\hat{H}_{\text{int}} = \frac{N(N-1)}{2}K + A\sum_{k<l}\left\{\frac{2}{9}(\hat{\mathbf{j}}_k\cdot\hat{\mathbf{j}}_l)^3 + \frac{11}{18}(\hat{\mathbf{j}}_k\cdot\hat{\mathbf{j}}_l)^2 - \frac{9}{8}(\hat{\mathbf{j}}_k\cdot\hat{\mathbf{j}}_l) - \frac{67}{32}\right\}. \quad (23)$$

Energy values averaged over Dicke states have the form

$$E_{\text{int}} = A\left\{-\frac{2}{9}\cdot\frac{75}{32}\frac{N(N-1)}{2} + \frac{11}{18}\cdot\frac{75}{32}\frac{N(N-1)}{2} - \frac{9}{8}\cdot\frac{1}{2}\left(R(R+1) - \frac{15}{4}N\right) - \left(-\frac{67}{32}\right)\right\} + \frac{N(N-1)}{2}K, \quad (24)$$

$$E_{\text{int}} = m\varepsilon + \frac{N(N-1)}{2}K - A\left\{\frac{N}{384}(227N - 1037) + \frac{9}{16}R(R+1)\right\}.$$

It follows hence that, as in the previous case, sublevels with the maximal cooperative number are more advantageous from the energy point of view for definite values of m. Thus, the system of four-level atoms behaves similarly to a system of atoms with five equidistant energy levels.

7. INTERACTION CONSTANT

It was shown in previous sections that the degeneracy of Dicke states with definite values of m in an atomic system is removed if we take into account the pair dipole–dipole interaction of atoms in this system. The extent of splitting of energy levels with given values of m for different values of cooperative number R is completely determined by the intensity of the exchange and not direct dipole interaction [11]. Before the final emission of a coherent superradiant pulse, systems of two- and three-level atoms are in the state $|R = 0, m = 0\rangle$ of intermediate equilibrium in a potential well with a depth determined by the constant of the exchange dipole interaction. For systems of four- and five-level atoms, sublevels with the maximal number R remain the most advantageous states with respect to energy; consequently, as the system passes to states with lower and lower values of m, the system emits with an increasing intensity

$$I = I_0(R+m)(R-m+1)$$

until it ultimately passes to a state with $m = 0$, after which a coherent superradiant pulse is emitted.

Let us estimate the exchange dipole interaction constant, which determines the depth of the well for an intermediate quasi-equilibrium state preceding the superradiant emission (the system does not emit in this state). For this purpose, we take into account the overlapping of coordinate wave functions Φ of identical interacting atoms. In accordance with the exchange perturbation theory [6], the correction to the interaction energy associated with transposition of atoms is defined by the formula

$$\varepsilon(r) = (\Phi'|\hat{V}(r)|\Phi), \quad (25)$$

where the prime marks the wave function of atoms, which differs from the initial (unprimed) function in the transposition of nuclei, and $\hat{V}(r)$ is the potential energy of interaction of atoms in the dipole approximation, r being the distance between the nuclei.

The computational model similar to that used in [14] is as follows. We describe the interaction of atoms in the form of a potential of the type of the Sutherland potential

$$V(r) = \begin{cases} -\dfrac{C}{r^3}, & r \geq \alpha, \\ \infty, & r < \alpha, \end{cases}$$

where C is a constant defined as the product of the dipole moments of transitions between "operating" levels of a three-, four-, and five-level system and α is the distance at which the repulsive forces operate between the atoms.

Suppose that T is the gas temperature. Then the probability of the distance between neighboring atoms being equal to R is defined by the Boltzmann function

$$W(r) = \frac{\exp(-V(r)/T)}{Z},$$

where Z is the normalization factor, which can be determined from the condition

$$\int_0^\rho W(r)dr = 1$$

(ρ is the mean distance between the atoms). The averaged value of energy can be determined using the modified formula

$$\bar{\varepsilon} = (\Phi'|\hat{V}(r)W(r)|\Phi), \quad (26)$$

which is equivalent to averaging with the density matrix in the coordinate representation. The vector of state of the relative motion of atoms can be written in the form

$$|\Phi) = e^{ikz} + f(\Theta)/r,$$

where $z = r\cos\Theta$ is the coordinate measured along the axis connecting the atoms; $f(\Theta)$ is the scattering amplitude determined by the interatomic interaction, i.e., by potential V; $k = p/\hbar$ is the wave number; and $\mathbf{p}$ is the

momentum of the relative motion of atoms. The vector of state in which the transposition of atoms is taken into account has the form

$$|\Phi') = e^{-ikz} + f(\pi - \Theta)/r.$$

Using these definitions, we obtain the following expression for the required energy [14]:

$$\bar{\varepsilon} = \frac{3b^2T}{k\alpha^6}\frac{1}{r^0}\exp\left(\frac{4}{3}kr_0 - \frac{b}{\alpha^3}\right) \times\left(1 + \frac{2\bar{f}}{r_0} - \frac{1.5}{\alpha}\right) = A; \quad (27)$$

here, A is the interaction constant appearing in formula (2); r_0 is the effective radius of exchange interaction, which is determined by the extent of overlapping of coordinate wave functions; $b = C/T$; and $\bar{f}$ is the scattering amplitude averaged over the region of overlapping of coordinate wave functions.

Let us consider numerical estimates [14] for atomic hydrogen. In the temperature range $T = 4.2 \times 10^{-14}$ erg (≈300 K), constant $b = 5 \times 10^3 c$ (where $c \approx 0.3$) and $k \approx 1$ (the values are given in atomic units). The potential barrier radius in the Sutherland model is usually chosen in the region of the Van der Waals minimum and is equal approximately to $\alpha \approx 5$, while the effective radius $r_0 \approx 10$. For such values of the quantities appearing in formula (27), we obtain $A \approx 3T$. The constant of direct dipole–dipole interaction for the given parameters is $A \approx 2 \times 10^{-14}$ erg, i.e., equal to one-sixth of the above value.

The increase in the constant of interaction of atoms due to exchange effects is a consequence of interference-induced redistribution of the atomic concentration in the gas in such a way that the probability of the "interference" contact of atoms, which are connected in pairs via the dipole interaction, increases significantly.

8. CONCLUSIONS

The cascade scheme of emission of multilevel atoms with equidistant energy states considered here is such in the classical meaning of this term; however, it is treated as a unified system characterized by the quantum isospin number and its component along a conditionally preferred axis rather than as a cascade of consecutive two-level transitions. It is this analogy with a system of particles with states completely characterized by a certain momentum and its component that makes it possible to use standard methods for describing spin systems. The cooperative number and its component, which characterize Dicke states, are precisely the total isospin of the system and its component in this terminology.

It was demonstrated above that the realization of the ground state of atomic superradiant systems is governed by a certain regularity: atoms with isospins $j = 1/2$ and $j = 1$ are mainly condensed to a state with a cooperative number of $R = 0$, followed by a "suspension" of the system, after which superradiation is emitted with a certain delay. On the other hand, the ground state of a system with particles having isospins $j = 3/2$ and $j = 2$ is characterized by the maximal cooperative number, which ensures a smooth transition to the regime of collective coherent emission of a superradiance pulse. This regularity can be formally explained by the structure of the Hamiltonians describing the behavior of a system of atoms with the pair interaction (namely, the power of dot product operators for isospins). For example, this operator appears to the first power in a system of two-level atoms [11]; to the first and second powers in a system of three-level atoms; to the first, second, and third powers in a system of four-level atoms; and from the first to the fourth powers in a system of five-level atoms. In the case of four- and five-level atoms, the terms with even powers of these operators practically compensate one another in the range of variation of the total isospin of a pair. Thus, the cubic term plays a decisive role here, while it is the quadratic term with a typical parabolic well and with a minimum corresponding to $R = 0$ that determines the situation in a system of three-level atoms.

The constants of exchange dipole interaction estimated above make it possible to judge the degree of splitting of collective energy levels and, hence, the extent to which the dipole interaction factor affects the (presence or absence of a) delay preceding a superradiation pulse. For example, the estimates of the dipole exchange interaction constants for alkali atoms belonging to the sodium group give a value of 0.06–0.07 eV, and the above splitting plays a significant role for transitions in a wavelength range from 3 to 9 μm.

Superradiance effects, which occur due to transitions between closely spaced Landau levels in a strongly magnetized plasma [10], open prospects for realizing the idea of a system consisting of a large number of equidistant levels (e.g., in low-dimensional systems of the type of semiconducting films placed in a magnetic field). Such systems can ensure the presence of any number of discrete Landau levels, which are equidistant a priori. It is such systems that can be used for experimental observation of the Dicke effect in multilevel systems in pure form.

ACKNOWLEDGMENTS

This study was supported financially by the Russian Foundation for Basic Research (project no. 02-02-17686), the program "Universities of Russia" (project no. UR.01.01.040), and the Ministry of Education of the Russian Federation (grant no. E02-3.2-287).

REFERENCES

1. A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Phenomena in Optics* (Nauka, Moscow, 1988); S. A. Akhmanov, Yu. E. D'yakov, and A. S. Chirkin, *Introduction to Statistical Radio Physics and Optics* (Nauka, Moscow, 1981); Yu. L. Klimontovich, *The Kinetic Theory of Electromagnetic Processes* (Nauka, Moscow, 1980; Springer, Berlin, 1983).
2. R. Bonifacio, P. Schwendiman, and F. Haake, Phys. Rev. A **4**, 302, 854 (1971).
3. F. De Martini and G. Preparata, Phys. Lett. **48**, 43 (1974); B. Coffey and R. Friedberg, Phys. Rev. A **17**, 1033 (1978).
4. R. H. Dicke, Phys. Rev. **93**, 99 (1954).
5. V. I. Yukalov, Acta Phys. Pol. A **57**, 295 (1980).
6. T. M. Makhviladze and L. A. Shelepin, Zh. Éksp. Teor. Fiz. **62**, 2066 (1972) [Sov. Phys. JETP **35**, 1080 (1972)]; L. A. Shelepin, Zh. Éksp. Teor. Fiz. **54**, 1463 (1968) [Sov. Phys. JETP **27**, 784 (1968)].
7. T. M. Makhviladze and L. A. Shelepin, Preprint No. 145, FIAN (Inst. of Physics, USSR Academy of Sciences, Moscow, 1971); É. Sh. Teplitskiĭ, Teor. Mat. Fiz. **2**, 399 (1970).
8. A. I. Zaĭtsev, V. A. Malyshev, and E. D. Trifonov, Zh. Éksp. Teor. Fiz. **84**, 475 (1983) [Sov. Phys. JETP **57**, 275 (1983)].
9. K. Nakamura and S. J. Washimiga, J. Phys. C **13**, 3483 (1980); H. Steudel, Ann. Phys. (Leipzig) **37**, 57 (1980); C. R. Stroud, J. H. Eberly, W. L. Lama, and L. Mandel, Phys. Rev. A **5**, 1094 (1972).
10. L. I. Men'shikov, Usp. Fiz. Nauk **169**, 113 (1999) [Phys. Usp. **42**, 107 (1999)].
11. E. V. Orlenko and B. G. Matisov, Zh. Éksp. Teor. Fiz. **116**, 1148 (1999) [JETP **89**, 612 (1999)].
12. A. V. Andreev, R. V. Arutyunyan, and Yu. A. Il'inskiĭ, Opt. Spektrosk. **50**, 1050 (1981) [Opt. Spectrosc. **50**, 578 (1981)]; A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, Usp. Fiz. Nauk **131**, 653 (1980) [Sov. Phys. Usp. **23**, 493 (1980)].
13. V. A. Malyshev, I. V. Ryzhov, E. D. Trifonov, and A. I. Zaĭtsev, Opt. Commun. **180**, 59 (2000); F. Haake and R. Reibold, Phys. Lett. A **92**, 29 (1982); Phys. Rev. A **29**, 3208 (1984); Opt. Acta **31**, 107 (1984).
14. E. V. Orlenko and A. A. Rumyantsev, Fiz. Nizk. Temp. **15**, 485 (1989) [Sov. J. Low Temp. Phys. **15**, 272 (1989)]; E. Orlenko, I. Mazets, and B. Matisov, Zh. Tekh. Fiz. **73**, 30 (2003) [Tech. Phys. **48**, 26 (2003)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 98, No. 1, 2004, pp. 144–155.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 1, 2004, pp. 160–172.

SOLIDS
Electronic Properties

The Crystal Field and Exchange Coupling in Iron Group Metal Carbonates

V. F. Meshcheryakov

Moscow Institute of Radio Engineering, Electronics, and Automatics (Technical University), Moscow, 117454 Russia

e-mail: niin@ranet.ru

Received May 27, 2003

Abstract—The wave functions of Co^{2+} and Fe^{2+} ions near the ground state in the $CaCO_3$-type lattice have been calculated from EPR data in the Abragam–Pryce approximation. The orbital angular momentum contributions to the anisotropic and antisymmetric parts of exchange coupling are determined assuming that this interaction between the magnetic ions occurring in nonequivalent positions is isotropic with respect to spin orientations. It is shown that, in the given approximation, the exchange coupling components in the basal plane for such Fe^{2+}–Fe^{2+} and Co^{2+}–Fe^{2+} ion pairs are missing. This fact explains the uniaxial antiferromagnetic ordering in $FeCO_3$ and the presence of a low-lying oscillation branch for Fe^{2+} impurity ions in antiferromagnetic $CoCO_3$. The EPR spectra of exchange-coupled Co^{2+}–Co^{2+}, Fe^{2+}–Fe^{2+}, and Co^{2+}–Fe^{2+} pairs occupying nonequivalent positions have been calculated and their parameters have been numerically estimated. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The phenomenon of weak ferromagnetism in antiferromagnetic compounds has been observed since 1940 and was originally studied by Schultz [1], Bizette [2], and Néel and Pauthenet [3]. The hypothesis that this ferromagnetism is caused by the sloped state of magnetic moments and is related to a specific symmetry of their local environment was originally formulated by Borovik-Romanov and Orlova [4]. Transition metal carbonates of the $MeCO_3$ type with Me = Mn, Co, and Ni were studied by Borovik-Romanov and coworkers [5–20] for many years and became the model objects for the investigation of this phenomenon. A phenomenological description of weak ferromagnetism was developed by Dzyaloshinski [21]. An analysis of the symmetry of these compounds revealed that the presence of a sloped state is determined by the antisymmetric part of exchange coupling. It was established in [21] that the magnitude of this interaction is determined by relativistic corrections and must be significantly smaller than the total exchange value. A microscopic approach developed by Morija [22] allowed the magnitude of the exchange coupling to be calculated, which has proved to be on the order of the $\Delta g/g$ fraction of the superexchange energy. This estimate was obtained under the assumptions that the lowest energy level in the magnetic ion is the orbital singlet and the spin–orbit coupling is small as compared to the splitting of orbital levels by the crystal field. This implies that the magnitude of Δg (the deviation of g from the pure-spin value, caused by an admixture of the orbital angular momentum of excited states), is relatively small—in agreement with the term "weak" ferromagnetism. However, the orbital angular momentum in many cases is by no means suppressed and the antisymmetric exchange (expressed via effective spins) may be no less important than the isotropic exchange: sometimes the former can even exceed the latter in magnitude.

At the present time, the magnetic properties of the compounds $MnCO_3$, $FeCO_3$, $CoCO_3$, and $NiCO_3$ are known in sufficient detail. However, some facts still have to be explained, including the following.

(i) The compound $FeCO_3$ is a uniaxial antiferromagnet with two sublattices, the magnetic moments of which are oriented along the C_3 axis of the crystal [23], whereas the other compounds in this group at low temperatures are easy-plane antiferromagnets with sloped magnetic moments of the sublattices [5, 7, 15].

(ii) The magnitude of the Dzyaloshinski field H_D (characterizing the antisymmetric part of exchange coupling) varies in this group within two orders of magnitude: from H_D = 4 kOe for $MnCO_3$ [5] to H_D = 160 kOe for $NiCO_3$ [15]. For $CoCO_3$ and $NiCO_3$, the antisymmetric coupling component is of the same order of magnitude as the isotropic exchange. At the same time, the H_D values for $CoCO_3$ determined from the static and resonance measurements exhibit an almost twofold difference: the magnetization measurements give H_D = 27 kOe [7], while the resonance experiment yields H_D = 52 kOe [14]. The g values obtained from the electron paramagnetic resonance (EPR) measurements [24, 25] and from the antiferromagnetic resonance (AFMR) [9, 14] data do not coincide either.

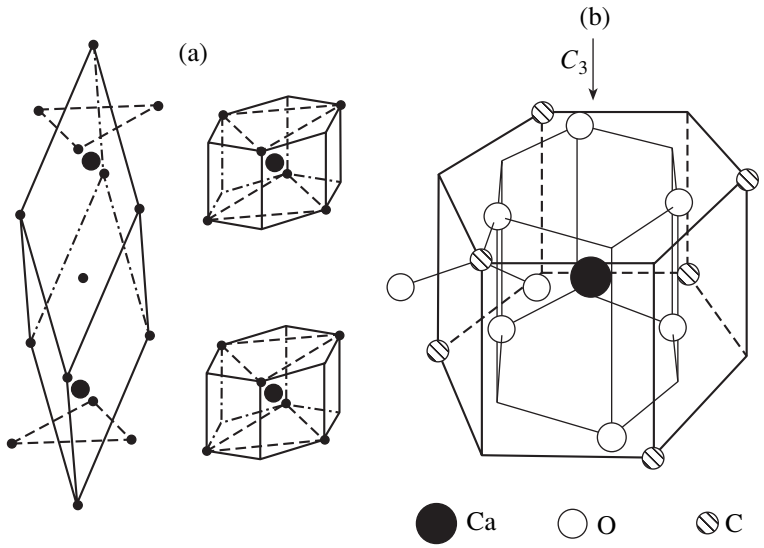

Fig. 1. Schematic diagrams of (a) the crystallographic unit cell and (b) the local environment of a magnetic atom in the $CaCO_3$ (calcite) type lattice.

(iii) The frequency of splitting in the AFMR spectrum of $CoCO_3$, which is related to the impurity of Fe^{2+} ions, corresponds to an exchange field of $H_E \approx 6$ kOe. This value is significantly lower as compared to the exchange fields in $CoCO_3$ and $FeCO_3$.

(iv) The AFMR spectrum of $MnCO_3$ [11] displays a gap related to the hyperfine interaction. No such gap is found in the spectrum of $CoCO_3$, although the values of hyperfine splittings observed for the EPR on Mn^{2+} and Co^{2+} ions in isomorphous crystal lattices are comparable [24–27].

Evidently, these variations in the magnetic properties are related to differences in the microscopic state of magnetic ions. A significantly weaker magnitude of the Dzyaloshinski interaction in $MnCO_3$ as compared to that in $CoCO_3$ and $NiCO_3$ is most probably explained by the absence of the orbital angular momentum contribution to the ground state of the Mn^{2+} ion (for which $L = 0$ and $S = 5/2$). A small contribution from the orbital angular momentum appears due to an admixture of excited states according to the model used by Morija [22]. At the same time, the magnetic moments in the ground states of Co^{2+} and Ni^{2+} ions are determined by the contributions from both orbital angular momentum and spin. These states can be calculated using data for the EPR on these ions. The spatial motion of electrons in the magnetic ions is not only manifested in the magnetic moment of an individual atom, but determines the character of exchange coupling as well. Since the form of this interaction in the general case depends on the particular exchange integrals of certain electron levels in the magnetic ions, the calculation of exchange coupling requires the knowledge of the electron wave functions.

However, knowledge of the wave function of the ground state of a magnetic ion described in terms of the total spin and the orbital angular momentum also allows the contributions to the anisotropic and antisymmetric parts of exchange coupling to be determined. If the exchange coupling between ions, considered as a function of the true spins, is isotropic, these additional contributions appear on passage from description of exchange coupling in terms of the true spin to use of the effective spin variables [28], since the latter variables are employed in description of the EPR spectra. It will be demonstrated below that this transformation influences the positions of levels of the exchange-coupled pairs provided only that (i) the state of magnetic ions is determined by the admixture of the orbital angular momentum and (ii) these pairs of magnetic ions occupy nonequivalent positions in the crystal lattice. Using this approach, it is possible to show that the exchange coupling components in the basal plane for such Fe^{2+}–Fe^{2+} and Co^{2+}–Fe^{2+} pairs are absent. This fact explains the uniaxial antiferromagnetic ordering in $FeCO_3$ and the presence of a low-lying oscillation branch due to Fe^{2+} impurity ions in the $CoCO_3$ lattice.

Figure 1a shows a schematic diagram of the crystallographic unit cell in the $CaCO_3$ (calcite) type lattice. Here, Ca^{2+} ions occupy two nonequivalent positions on the rhombohedron diagonal (C_3 axis). Their local environment shown in Fig. 1b includes six oxygen atoms occurring at the corners of a hexagonal prism. The upper and lower bases of this prism represent equilateral triangles rotated 60° relative to each other. The prisms surrounding the two Ca^{2+} ions are also mutually rotated by 60°.

The EPR spectra of individual ions in isomorphous structures have been studied in sufficient detail for

Table 1. Experimental data obtained from the EPR spectra of Co^{2+} and Fe^{2+} ions in isomorphous $CdCO_3$ and $CaCO_3$ lattices

Sample	f, GHz	$g_{\parallel}$	$g_{\perp}$	$\lvert A\rvert \times 10^4$, cm^{-1}	$\lvert B\rvert \times 10^4$, cm^{-1}
$CdCO_3 + Co^{2+}$	33.87	3.07 ± 0.03	4.96 ± 0.02	37 ± 4	164 ± 10
	21.31	3.06 ± 0.01	4.94 ± 0.01	39 ± 3	154 ± 7
$CaCO_3 + Co^{2+}$	9.2	3.406 ± 0.01	4.817 ± 0.01	53 ± 1	132 ± 3
$CaCO_3 + Fe^{2+}$	9.4	9.85 ± 0.01	<0.2	–	–

Mn^{2+} [26, 27], Co^{2+} [24, 25], Ni^{2+} [15], and Fe^{2+} [29]. At low temperatures, the ground states of Co^{2+} and Fe^{2+} ions is characterized by an effective spin of $S' = 1/2$ and that of Ni^{2+} ions, by a spin of $S' = 1$. Below we will consider only the states with $S' = 1/2$. The estimates of this interaction will be also obtained only for $CoCO_3$ and the impurity of Fe^{2+} in the $CoCO_3$ matrix.

Thus, the results of investigations of the states of paramagnetic ions on a microscopic level are of considerable importance for understanding the nature of exchange coupling. The aim of this study was to calculate, based on the EPR data, the functions describing the ground state of magnetic Co^{2+} and Fe^{2+} ions and to determine the orbital angular momentum contributions to the anisotropic and antisymmetric parts of exchange coupling. Based on these results, the EPR spectra of exchange-coupled Co^{2+}–Co^{2+}, Fe^{2+}–Fe^{2+}, and Co^{2+}–Fe^{2+} ion pairs occupying nonequivalent positions will be calculated and the effect of the antisymmetric contribution on the EPR spectrum will be determined depending on the presence of the anisotropic exchange.

2. WAVE FUNCTION OF THE GROUND STAGE OF Co^{2+} ION IN THE CALCITE-TYPE LATTICE DETERMINED BASED ON THE EPR DATA

The EPR data for individual ions are usually interpreted in terms of the spin Hamiltonian. For the Co^{2+} ions under consideration, this Hamiltonian can be written as

$$\begin{aligned}\hat{H} &= g_{\parallel}\beta H_z\hat{S}'_z + g_{\perp}\beta(H_x\hat{S}'_x + H_y\hat{S}'_y)\\ &+ A\hat{I}_z\hat{S}'_z + B(\hat{I}_x\hat{S}'_x + \hat{I}_y\hat{S}'_y),\end{aligned} \quad (1)$$

where β is the Bohr magneton; $g_{\parallel}$ and $g_{\perp}$ are the longitudinal and transverse g factors, respectively; A and B are hyperfine interaction constants, $\hat{\boldsymbol{I}}$ is the nuclear spin; and S' is the effective spin ($S' = 1/2$). The experimental g factors and hyperfine interaction constants for spin Hamiltonian (1) of some ions are presented in Table 1.

Ground state ^4F of the free Co^{2+} ion has an electron configuration of $3d^7$ and exhibits a sevenfold orbital degeneracy. The next term, ^4P, belonging to the same free ion configuration, is situated 14500 cm^{-1} above the ground state level. In the calcite-type lattice, Co^{2+} ions occur in an octahedral environment. The cubic crystal field component splits the ^4F term into a singlet and two underlying triplets. When the quantization axis is directed along C_3, the wave functions of the lower triplet are as follows [30]:

$$\begin{aligned}|0\rangle &= (2/3)\psi_{3,0} + \sqrt{5/18}(\psi_{3,3} - \psi_{3,-3}),\\ |1\rangle &= \sqrt{5/6}\psi_{3,2} + \sqrt{1/6}\psi_{3,-1},\\ |-1\rangle &= \sqrt{5/6}\psi_{3,-2} - \sqrt{1/6}\psi_{3,1}.\end{aligned} \quad (2)$$

Since the matrix elements of the orbital angular momentum projections in the representation of these functions differ by a constant factor of $\alpha = -3/2$ from analogous elements in the representation of eigenfunctions of the free Co^{2+} ion in the ^4F state with $L = 1$, the former elements can be considered as eigenfunctions of an apparent operator of the angular momentum $l' = 1$. The trigonal component of the crystal field and the spin–orbit coupling further split this triplet into six Kramers doublets, of which three doublets are characterized by the quantum number $m = \pm 1/2$, two doublets possess $m = \pm 3/2$, and one doublet has $m = \pm 5/2$. On the energy scale, these doublets are separated by hundreds of inverse centimeters. The general scheme of these levels is depicted in Fig. 2.

Now let us calculate the distances between levels near the ground state and the wave functions of cobalt ions in the calcite-type lattice using the results of calculations performed by Abragam and Pryce [31] and the published EPR data [24, 25]. In order to use these results for the ground state wave function calculations, it is necessary to reproduce the Abragam–Pryce calculation scheme, which will be also used below for determining the ground state of Fe^{2+}.

The optical absorption measurements for $CoCO_3$ show that the splitting of levels reaches 22000 cm^{-1} [32], which is much greater than the distance to the ^4P term of the free Fe^{2+} ion. For this reason, Abragam and Pryce proceeded from a set of functions of the ^4F and ^4P states that were different from set of functions (2). Assuming

that the set includes the doublet φ_x and φ_y and the singlet φ_z, with the energy separation Δ, and choosing the orbital basis functions in the form of

$$\begin{aligned} |1\rangle &= -\sqrt{1/2}(\varphi_x + i\varphi_y), \\ |0\rangle &= \varphi_z, \\ |-1\rangle &= \sqrt{1/2}(\varphi_x - i\varphi_y), \end{aligned} \tag{3}$$

we have to determine the eigenvalues of the Hamiltonian

$$\hat{W} = \Delta(1 - \hat{l}_z'^2) - \alpha\lambda\hat{l}_z'\hat{S}_z - \alpha'\lambda(\hat{l}_x'\hat{S}_x + \hat{l}_y'\hat{S}_y). \tag{4}$$

Here, the angular momentum operator is written in the form of $-\alpha\hat{\boldsymbol{l}}'$, where $\hat{\boldsymbol{l}}'$ is the apparent orbital angular momentum operator with an eigenvalue $l' = 1$; α and α' are constants equal approximately to 3/2 for Co^{2+}. The deviations from 3/2 are determined by an admixture of the ^4P state and by the influence of the trigonal crystal field component. The first term of Hamiltonian (4) reflects the fact that the levels with $l_z' = \pm 1$ are separated by Δ from the level with $l_z' = 0$. The second term described the spin–orbit coupling (for Co^{2+}, $\lambda = -180$ cm^{-1}).

Since the operator $l_z' + \hat{S}_z$ commutates with Hamiltonian (4), the eigenvalues of this operator ($m = \pm 1/2$, $\pm 3/2$, and $\pm 5/2$) can be used for the classification of levels. If the corresponding functions are written in the form of $|l_z', S_z\rangle$, where $l_z' = 0, \pm 1$ and $S_z = \pm 1/2, \pm 3/2$, the matrix of operator $\hat{W}$ separates into three matrices:

$$\begin{array}{c} \quad |\mp 1, \pm 3/2\rangle \quad |0, \pm 1/2\rangle \quad |\pm 1, \mp 1/2\rangle \\ \begin{array}{r}\langle \mp 1, \pm 3/2| \\ \langle 0, \pm 1/2| \\ \langle \pm 1, \mp 1/2|\end{array} \left| \begin{array}{ccc} (3/2)\alpha\lambda & -\sqrt{3/2}\alpha'\lambda & 0 \\ -\sqrt{3/2}\alpha'\lambda & \Delta & -\sqrt{2}\alpha'\lambda \\ 0 & -\sqrt{2}\alpha'\lambda & (1/2)\alpha\lambda \end{array} \right| \end{array} \tag{5}$$

for $m = \pm 1/2$;

$$\begin{array}{c} \quad |0, \pm 3/2\rangle \quad |\pm 1, \pm 1/2\rangle \\ \begin{array}{r}\langle 0, \pm 3/2| \\ \langle \pm 1, \pm 1/2|\end{array} \left| \begin{array}{cc} \Delta & -\sqrt{3/2}\alpha'\lambda \\ -\sqrt{3/2}\alpha'\lambda & -(1/2)\alpha\lambda \end{array} \right| \end{array} \tag{6}$$

for $m = \pm 3/2$, and

$$\langle \pm 1, \pm 3/2 | -\sqrt{3/2}\alpha\lambda | \quad \text{for} \quad m = \pm 5/2. \tag{7}$$

For the lower doublet corresponding to $m = \pm 1/2$, the wave function can be written as

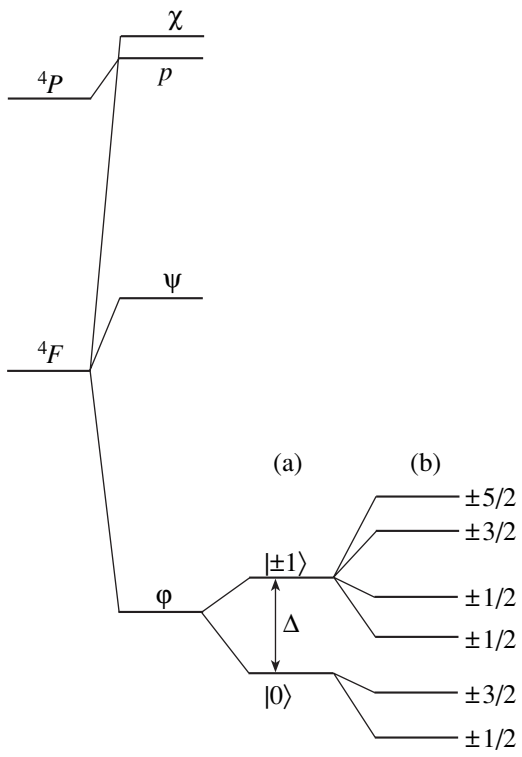

Fig. 2. The general scheme of splitting of the ground state energy levels for Co^{2+} ion in a rhombohedral environment under the action of (a) a trigonal crystal field Δ and (b) spin–orbit coupling.

$$\begin{aligned} |1/2\rangle &= a|-1, 3/2| + b|0, 1/2| + c|1, -1/2\rangle, \\ |-1/2\rangle &= a|1, -3/2| + b|0, -1/2| + c|-1, 1/2\rangle. \end{aligned} \tag{8}$$

Here and below, the first and second terms in brackets indicate the values of projections of the orbital and spin momenta, respectively. Using the normalization condition

$$a^2 + b^2 + c^2 = 1 \tag{9}$$

introducing the parameter x such that

$$E = \frac{1}{2}\alpha\lambda(x + 3) \tag{10}$$

(with the energy E being an eigenvalue of matrix (5)), and taking into account representation (8), we obtain the relations

$$a : b : c = \frac{\sqrt{6}}{x} : -\frac{\alpha}{\alpha'} : \frac{\sqrt{8}}{x+2}, \tag{11}$$

$$\Delta = -\frac{\lambda\alpha'^2}{\alpha}\left(\frac{3}{x} + \frac{4}{x+2}\right) + \frac{\lambda\alpha}{2}(x+3). \tag{12}$$

Table 2. Hyperfine interaction constants calculated for various values of the parameter p for Co^{2+} ions in isomorphous $CdCO_3$ and $CaCO_3$ lattices

Parameters	$CdCO_3 + Co^{2+}$			$CaCO_3 + Co^{2+}$		
p	0.6	1.0	1.4	0.6	1.0	1.4
x	3.53	3.26	3.08	3.02	2.08	2.68
α	1.24	1.56	1.85	1.26	1.55	1.81
$g_{S\parallel}$	2.66	2.60	2.55	2.86	2.79	2.53
$g_{L\parallel}$	0.40	0.46	0.51	0.55	0.62	0.88
$(g_{S\parallel} + g_{L\parallel})_{exp}$		3.06			3.4	
$A \times 10^4$, cm^{-1}	–7	8	22	8	37	60
$A_{exp} \times 10^4$, cm^{-1}		39			53	
$g_{S\perp}$	3.61	3.67	3.70	3.53	3.59	3.61
$g_{L\perp}$	1.34	1.28	1.25	1.28	1.22	1.20
$g_{S\perp} + g_{L\perp}$	4.95	4.95	4.95	4.81	4.81	4.81
$(g_{S\perp} + g_{L\perp})_{exp}$		4.95			4.817	
$B \times 10^4$, cm^{-1}	170	154	146	159	143	138
$B_{exp} \times 10^4$, cm^{-1}		154 ± 7			132 ± 3	
Δ, cm^{-1}	–144	–407	–598	–3	–275	–458

Now let us express the g factor components as functions of parameters α, α', and x. Substituting functions (8) in the expressions

$$g_\parallel = 2\langle 1/2|\hat{L}_z + 2\hat{S}_z|1/2\rangle,$$
$$g_\perp = 2\langle 1/2|\hat{L}_x + 2\hat{S}_x|-1/2\rangle, \quad (13)$$

we arrive at

$$\begin{aligned} g_{S\parallel} &= 4\langle 1/2|\hat{S}_z|1/2\rangle = 6a^2 + 2b^2 - 2c^2, \\ g_{S\perp} &= 4\langle 1/2|\hat{S}_x|-1/2\rangle = 4b^2 + 4\sqrt{3}ac, \\ g_{L\parallel} &= 2\langle 1/2|\hat{L}_z|1/2\rangle = -2\alpha\langle 1/2|\hat{l}'_z|1/2\rangle \\ &= 2\alpha(a^2 - c^2), \\ g_{L\perp} &= 2\langle 1/2|\hat{L}_x|-1/2\rangle = -2\alpha'\langle 1/2|\hat{l}'_z|-1/2\rangle \\ &= -\sqrt{8}\alpha' bc. \end{aligned} \quad (14)$$

Taking into account the normalization condition (9) and using relation (11), we eventually obtain the formulas

$$g_\parallel = 2 + \frac{4(\alpha+2)\left(\frac{3}{x^2} - \frac{4}{(x+2)^2}\right)}{p + \frac{6}{x^2} + \frac{8}{(x+2)^2}},$$
$$g_\perp = 4\frac{p + \frac{2\alpha}{x+2} + \frac{12}{(x+2)x}}{p + \frac{6}{x^2} + \frac{8}{(x+2)^2}}, \quad (15)$$

where $p = (\alpha/\alpha')^2$.

The hyperfine interaction constants A and B in spin Hamiltonian (1) can be expressed via the g value components [31]. These constants are determined by three contributions: from orbital angular momentum (A_L, B_L), spin momentum (A_S, B_S), and unpaired electrons (A_{sk}, B_{sk}). The latter contribution appears as a result of the configuration interaction between $3sd^7 4s$ and $3s^2 4d^7$ states; in the case of Co^{2+} ions, this contribution is usually ignored. Taking this into account, the A and B values can be written in the following form:

$$A = A_L + A_S = Pg_{L\parallel} - (k/2)Pg_{S\parallel},$$
$$B = B_L + B_S = Pg_{L\perp} - (k/2)Pg_{S\perp}, \quad (16)$$

where $P = 2\gamma\beta\beta_N r^{-3}$ (for Co^{2+}, $P = 0.0225$ cm^{-1} and $k = 0.325$).

Based on the results of calculations performed by Abragam and Pryce [31], the distances between levels are calculated as follows. First, we set $p = 0.6$, 1.0, and 1.4 in Eqs. (15) and substitute the $g_\perp$ and $g_\parallel$ values from Table 1 to obtain the corresponding sets of parameters x and α (Table 2). In order to select the optimum p value, we have to compare the hyperfine interaction constants A and B calculated using expressions (16) to the experimental data. To this end, we find from expression (14) the orbital and spin components of the g value and substitute these quantities into expressions (16). The results of these calculations and the experimental data (A_{exp}, B_{exp}) are also presented in Table 2. For both constants, the best fit to the experimental hyperfine interaction constants and g factors is observed for $p = 1.4$. The distributions of energy levels near the ground state can be obtained by solving the secular

Table 3. Energy levels and the corresponding g factors calculated for Co^{2+} ions in $CdCO_3$ lattice

m	Energy, cm^{-1}	$g_\parallel$	$g_\perp$
±1/2	–1018	3.05	4.95
±3/2	–732	5.0	0
±1/2	–381	4.7	1.96
±1/2	131	1.75	1
±3/2	300	0.72	0
±5/2	502	2.28	0

equations based on matrices (5)–(7) for the given parameters p, x and α. The energies of these levels and the corresponding g values are presented in Table 3 and the mutual arrangement of levels is depicted in Fig. 2. The distance to the nearest excited state is about 300 cm^{-1}. The ground state corresponds to a doublet described by the wave function (8) with the coefficients presented in Table 4.

3. THE WAVE FUNCTION OF THE GROUND STATE OF Fe^{2+} ION IN THE CALCITE-TYPE LATTICE

The experimentally observed spectrum of Fe^{2+} ion in the calcite-type lattice is described, in the absence of nuclear spin, by Hamiltonian (1) with the g value components given in Table 1 [29]. The numerical calculation with allowance for all states of the free ion by diagonalization of a 25-order matrix was also reported in [29]. Analytical expressions obtained for the wave functions of Fe^{2+} ion near the ground state and the corresponding g values are derived below based on the spin Hamiltonian method described in Section 2.

The ground state ^5D of the free Fe^{2+} ion has an electron configuration of $3d^6$ and exhibits a fivefold orbital degeneracy. In the calcite-type lattice, where Fe^{2+} ions occur in a cubic crystal field of octahedral symmetry, the ^5D term splits into T_{2g} triplet and an above-lying E_g doublet with a separation on the order of 10^4 cm^{-1}. The axial field of trigonal symmetry splits the triplet into a singlet and the so-called non-Kramers doublet separated by more than 10 cm^{-1}. The mutual arrangement of the doublet and singlet depends on the sign of the axial field: in a calcite-type lattice, the doublet corresponds to a lower level. The spin–orbit coupling splits the orbital triplet into $3 \times 5 = 15$ levels with a separation on the order of 500 cm^{-1}. The general scheme of these levels is depicted in Fig. 3.

The wave functions of the lower triplet for the quantization axis directed along C_3 are as follows [30]:

$$\begin{aligned} &\psi_{2,0}, \\ &\sqrt{2/3}\psi_{2,2} - \sqrt{1/3}\psi_{2,-1}, \\ &\sqrt{2/3}\psi_{2,-2} + \sqrt{1/3}\psi_{2,1}. \end{aligned} \tag{17}$$

According to the spin Hamiltonian method, the orbital triplet T_{2g} is assigned an effective orbital angular momentum $l' = 1$ and the calculation is performed using Hamiltonian (4) with $S = 2$, $\lambda = -100$ cm^{-1}, and $\alpha = 1$. In this case, the matrix of operator $\hat{W}$ separates into four matrices characterized by eigenvalues of operator $\hat{l}'_z + \hat{S}_z$ with $m = 0, \pm1, \pm2$, and ±3:

$$\begin{array}{c|ccc} & |1,-1\rangle & |0,0\rangle & |-1,1\rangle \\ \langle 1,-1| & \alpha\lambda & -\sqrt{3}\alpha'\lambda & 0 \\ \langle 0,0| & -\sqrt{3}\alpha'\lambda & \Delta & -\sqrt{3}\alpha'\lambda \\ \langle -1,1| & 0 & -\sqrt{3}\alpha'\lambda & \alpha\lambda \end{array} \tag{18}$$

for $m = 0$,

$$\begin{array}{c|ccc} & |\pm1,0\rangle & |0,\pm1\rangle & |\mp1,\pm2\rangle \\ \langle \pm1,0| & 0 & -\sqrt{3}\alpha'\lambda & 0 \\ \langle 0,\pm1| & -\sqrt{3}\alpha'\lambda & \Delta & -\sqrt{2}\alpha'\lambda \\ \langle \mp1,\pm2| & 0 & -\sqrt{2}\alpha'\lambda & 2\alpha\lambda \end{array} \tag{19}$$

for $m = \pm1$,

$$\begin{array}{c|cc} & |\pm1,\pm1\rangle & |0,\pm2\rangle \\ \langle \pm1,\pm1| & -\alpha\lambda & -\sqrt{2}\alpha'\lambda \\ \langle 0,\pm2| & -\sqrt{2}\alpha'\lambda & \Delta \end{array} \tag{20}$$

for $m = \pm2$,

$$\langle \pm1,\pm2|-2\alpha\lambda| \quad \text{for} \quad m = \pm3. \tag{21}$$

The lower doublet corresponds to $m = \pm1$. The wave

Table 4. Wave functions of Co^{2+} and Fe^{2+} ions in the vicinity of the ground state in the $CaCO_3$-type lattice

Ion	Wave functions	Coefficients
Co^{2+}	$\|\pm\rangle = a\|\mp1, \pm3/2\rangle + b\|0, \pm1/2\rangle + c\|\pm1, \mp1/2\rangle$	$a = 0.53$, $b = -0.76$, $c = 0.37$
Fe^{2+}	$\|\pm1\rangle = d\|\pm1, 0\rangle + e\|0, \pm1\rangle + f\|\mp1, \pm2\rangle$	$d = 0.17$, $e = 0.22$, $f = 0.96$
	$\|0\rangle = \alpha\|1, 0\rangle + \beta\|0, 1\rangle + \gamma\|-1, 2\rangle$	$\alpha^2 = \gamma^2 = 0.45$, $\beta^2 = 0.10$

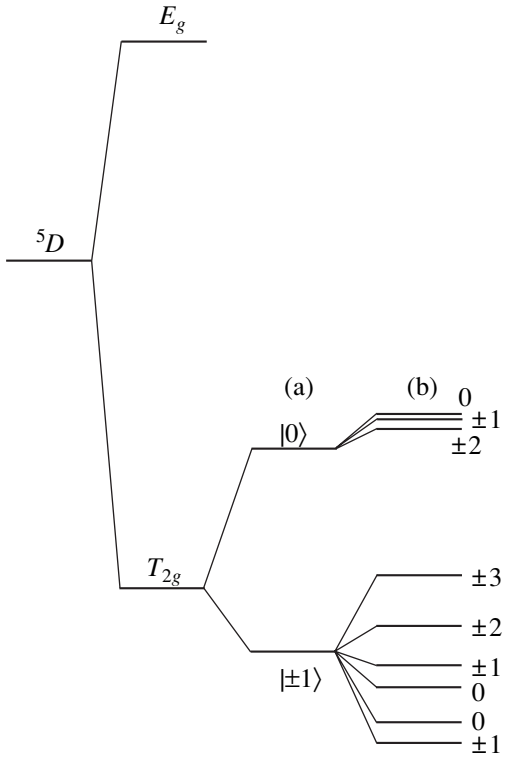

Fig. 3. The general scheme of splitting of the ground state energy levels for Fe^{2+} ion in a rhombohedral environment under the action of (a) a trigonal crystal field and (b) spin–orbit coupling.

function of this state can be represented as

$$\begin{aligned} |+1\rangle &= d|1, 0\rangle + e|0, 1\rangle + f|-1, 2\rangle, \\ |-1\rangle &= d|-1, 0\rangle + e|0, -1\rangle + f|1, -2\rangle. \end{aligned} \tag{22}$$

Using the normalization condition (9), we obtain for matrix (19) the relations

$$d^2 = \frac{3\lambda^2\alpha'^2(E-2\lambda\alpha)^2}{\Sigma}, \quad e^2 = \frac{E^2(E-2\lambda\alpha)^2}{\Sigma}, \quad f^2 = \frac{2\lambda^2\alpha'^2E^2}{\Sigma}, \tag{23}$$

where

Table 5. Energy levels calculated for Fe^{2+} ions in a $CaCO_3$-type lattice

m	Energy, cm^{-1}		
0	–185	–100	609
±1	–233	–43	600
±2	57	567	–
±3	200	–	–

$$\Sigma = 3\lambda^2\alpha'^2(E-2\lambda\alpha)^2 + E^2(E-2\lambda\alpha)^2 + 2\lambda^2\alpha'^2E^2,$$

and E is the energy determined by diagonalization of matrix (19). Substituting functions (22) into the expressions

$$\begin{aligned} g_\parallel &= 2\langle 1|\hat{L}_z + 2\hat{S}_z|1\rangle, \\ g_\perp &= 2\langle 1|\hat{L}_x + 2\hat{S}_x|-1\rangle, \end{aligned} \tag{24}$$

in the approximation employed, we obtain

$$g_\parallel = 4(e^2 + 2f^2) + 2\alpha(f^2 - d^2), \quad g_\perp = 0. \tag{25}$$

Let us estimate the energies of levels and the g value using data [33] on the Mössbauer spectroscopy of Fe^{2+} ions in a $CoCO_3$ matrix. The temperature dependence of the quadrupole splitting showed that $\Delta = 525\ cm^{-1}$. Using Hamiltonian (4) with this value of Δ, $\alpha = \alpha' = 1$, and $\lambda = -100\ cm^{-1}$, we obtain the energies of levels presented in Table 5. For the lowest level with $m = \pm1$, $E = -233\ cm^{-1}$; for the next level with $m = 0$, $E = -185\ cm^{-1}$; thus, the separation of these levels amounts to $48\ cm^{-1}$. The general arrangement of levels is depicted in Fig. 3.

Substituting the energy of the lowest level into formulas (23), we obtain the coefficients of the wave function (22) of the ground state of Fe^{2+} ion (Table 4). Using formula (25), we obtain for the g value in the ground state $g_\parallel = 9.4$. The difference of this result from the experimental value ($g_\parallel = 9.85$) was shown [29] to be due to an admixture of excited states. In addition to the wave functions of the ground state of the Fe^{2+} ion, subsequent analysis will require the function of the adjacent excited state with $m = 0$. The form of this function and the corresponding coefficients determined using matrix (18) are presented in Table 4.

4. THE EPR SPECTRA OF EXCHANGE-COUPLED Co^{2+}–Co^{2+}, Fe^{2+}–Fe^{2+}, AND Co^{2+}–Fe^{2+} ION PAIRS OCCUPYING NONEQUIVALENT CRYSTALLOGRAPHIC POSITIONS

The results of EPR measurements for exchange-coupled ion pairs allow the parameters obtained to be used for the description of magnetically ordered systems.The possibility to measure the EPR spectra of such pairs and the reliability of interpretation of these spectra are limited by the following factors. First, the intensity of signals in the spectrum, which is proportional to the squared concentration of paramagnetic couples, turns out to be two to three orders of magnitude lower than the intensity of signals from individual ions. Second, the presence of a large number of nearest neighbors featuring nonzero exchange coupling (e.g., in Cr_2O_3 there are 11 such neighbors at a distance of 5.73 Å with exchange coupling values ranging from 240 to about $0.5\ cm^{-1}$ [34, 35]) results in that an exper-

imental EPR spectrum contains a continuous series of lines in the vicinity of signals due to individual ions.

For these reasons, the EPR spectra of exchange-coupled pairs are usually observed for the ions with frozen or missing orbital angular momentum. In order to ensure reliable interpretation of the EPR data, the analysis is performed using the results of simultaneous optical absorption, EPR, and electron-nuclear double resonance (ENDOR) measurements. The most reliable EPR and ENDOR data correspond to magnetic fields that significantly differ from the resonance fields of individual ions. An additional useful piece of information is provided by investigations of the angular dependence of the spectra and the temperature dependence of the intensities of signals determined by the population of levels in the exchange-coupled ion pairs.

Let us consider the possibility of observing the resonance absorption due to exchange-coupled Co^{2+}–Co^{2+}, Fe^{2+}–Fe^{2+}, and Co^{2+}–Fe^{2+} ion pairs occurring inside the unit cell depicted in Fig. 1a. In order to determine the energy levels for these pairs, it necessary first to perform transformation of the spin projections to a common coordinate system and second, to derive an expression for the exchange coupling (see Eq. (26) below) using projections of the effective spin $S' = 1/2$, rather than of the true spin, as the variables in Hamiltonian (1).

The exchange coupling Hamiltonian is usually written in the commonly accepted form proposed by Heisenberg,

$$\begin{aligned}\hat{H}_{ex} &= J\hat{\mathbf{S}}_1 \cdot \hat{\mathbf{S}}_2 \\ &= (J/2)(\hat{S}_1^+\hat{S}_2^- + \hat{S}_1^-\hat{S}_2^+) + J\hat{S}_{1z}\hat{S}_{2z},\end{aligned} \tag{26}$$

where $\hat{\mathbf{S}}_1$ and $\hat{\mathbf{S}}_2$ are the true spins of magnetic ions. For $S_1 = S_2 = 1/2$, the eigenvalues of the Hamiltonian are $E_1 = J/4$ for the total spin of $S = 1$ and $E_0 = -(3/4)J$ for $S = 0$. On the passage to a common quantization axis for the ions occupying nonequivalent positions, it is necessary to use a transformation of the spin projections upon rotation of the coordinate axes in the xy plane:

$$S_x = S'_x\cos\varphi - S'_y\sin\varphi, \quad S_y = S'_x\sin\varphi + S'_y\cos\varphi.$$

In the common coordinate system with $\varphi = 60°$, Hamiltonian (26) takes the following form:

$$\begin{aligned}\hat{H}_{ex} &= J_{xy}(\hat{S}_{1x}\hat{S}_{2x} + \hat{S}_{1y}\hat{S}_{2y}) \\ &+ J_z\hat{S}_{1z}\hat{S}_{2z} + D(\hat{S}_{1x}\hat{S}_{2y} - \hat{S}_{2x}\hat{S}_{1y}) \\ &= \frac{1}{2}(J_{xy} + iD)\hat{S}_1^+\hat{S}_2^- + \frac{1}{2}(J_{xy} - iD)\hat{S}_1^-\hat{S}_2^+ + J_z\hat{S}_{1z}\hat{S}_{2z},\end{aligned} \tag{27}$$

where $J_{xy} = J/2$, $D = \sqrt{3}\,J_{xy}$, $J_z = J$; the prime in the spin notation is omitted and $\hat{\mathbf{S}}$ still denotes the total spin. Thus, nonequivalent positions of the paramagnetic ions result in the exchange coupling becoming anisotropic and in the appearance of an antisymmetric term.

Denoting $|1/2\rangle \equiv |+\rangle$ and $|-1/2\rangle \equiv |-\rangle$, we can write the matrix of Hamiltonian (27) in the $|\pm, \pm\rangle \equiv |\pm\rangle|\pm\rangle$ representation as

$$\begin{array}{c|cccc} & |+,+\rangle & |+,-\rangle & |-,+\rangle & |-,-\rangle \\ \langle+,+| & J_z/4 & 0 & 0 & 0 \\ \langle+,-| & 0 & -J_z/4 & (J_{xy}+iD)/2 & 0 \\ \langle-,+| & 0 & (J_{xy}-iD)/2 & -J_z/4 & 0 \\ \langle-,-| & 0 & 0 & 0 & J_z/4 \end{array}, \tag{28}$$

with the eigenvalues for $S = 1$,

$$E_0 = -J_z/4 + J_{xy}, \quad E_{\pm1} = J_z/4, \tag{29}$$

and for $S = 0$,

$$E_0 = -J_z/4 - J_{xy}. \tag{30}$$

For a purely spin state (in the absence of an orbital angular momentum), the levels of Hamiltonian with the matrix (27) are the same as those in the case of isotropic exchange (26). In the presence of a contribution due to the orbital angular momentum, the exchange coupling becomes anisotropic ($J_{xy} \neq J_z/2$) and the form of the Hamiltonian matrix (27) qualitatively changes the system of levels. The anisotropy leads to splitting of the states in the pair with $S = 1$ according to (29), and the D value, as will be shown below, influences the positions of levels in the case when the magnetic field vector is in the plane perpendicular to the C_3 axis.

In order to determine the exchange coupling components in terms of the effective spin $S' = 1/2$, it is necessary to find the matrix of Hamiltonian (27) in the representation of the ground state wave functions (presented in Table 4) and compare this with matrix (28). The resulting exchange constants J_{xy} and J_z determined in this way for Hamiltonian (27) are given in Table 6. As can be seen for Fe^{2+}–Fe^{2+} and Co^{2+}–Fe^{2+} pairs, the exchange coupling in the plane perpendicular to the C_3 axis is absent. Based on these results, it is possible to explain why the magnetic moments in antiferromagnetic $FeCO_3$ are aligned in the triple axis [23] and the impurity mode frequency of Fe^{2+} ion in the antiferromagnetic $FeCO_3$ is very small.

For experimental observation of the EPR spectrum of exchange-coupled pairs, it is necessary to estimate the parameters of the spin Hamiltonian. Substituting the coefficients for Co^{2+} from Table 4, we obtain

$$J_{xy} = 1.68J, \quad J_z = 1.65J, \quad D = 2.91J.$$

The energy levels of a pair are $E_{\pm1} = 0.41J$, $E_0 = 1.27J$ for $S = 1$ and $E_0 = -2.09J$ for $S = 0$. The exchange cou-

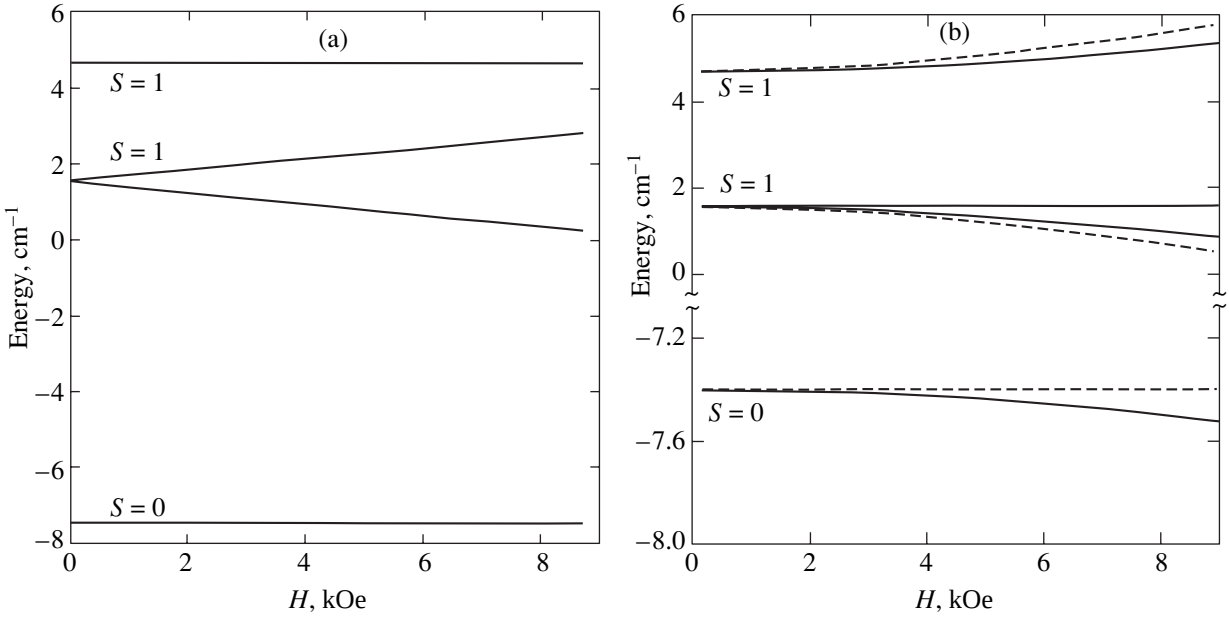

Fig. 4. Plots of the energy levels versus magnetic field strength **H** for exchange-coupled Co^{2+}–Co^{2+} ion pairs: (a) $\mathbf{H} \parallel C_3$; (b) $\mathbf{H} \perp C_3$. Dashed lines show the position of energy levels in the case when magnetic ions occur in equivalent crystallographic positions ($\varphi = 0$).

pling J can be estimated using the Néel temperature T_N of the antiferromagnetic ordering. For $CoCO_3$, T_N = 18.1 K, which corresponds to an exchange coupling energy of about –12 cm^{-1}. Since the exchange field in this compound is oriented in the basal plane and is created by two nearest neighbors, the exchange coupling energy per atom amounts to $J_{xy} \approx 6$ cm^{-1}. Taking into account that $J_{xy} = 2(\sqrt{3}\,ac + b^2)^2 J \approx 6$ cm^{-1}, we obtain $J \approx 3.6$ cm^{-1} and estimate the energy levels as $E_{\pm 1}$ = 1.48 cm^{-1}, $E_0 \approx 4.58$ cm^{-1} for $S = 1$ and $E_0 = -7.52$ cm^{-1} for $S = 0$. As can be seen from these estimates, the population of levels of the upper triplet at 4–10 K will be sufficient to observe the resonance transitions with $\Delta m = \pm 1$ in the region of frequencies corresponding to a wavelength below 3 mm. The identification of lines in the EPR spectrum of this ion pair would be significantly facilitated by data on the frequency and temperature dependences. Using the triplet splitting determined in this way, it is possible to evaluate the exchange coupling.

Figure 4 (solid curves) shows the energy levels of the exchange-coupled Co^{2+}–Co^{2+} pair calculated as functions of the magnetic field strength. For the comparison, dashed lines indicate the positions of levels in the case when magnetic ions occupy the equivalent positions ($\varphi = 0$). As can be seen from these data, the antisymmetric exchange influences the positions of levels only when the magnetic field vector is in the plane perpendicular to the C_3 axis. It should be noted that, in the absence of an anisotropic contribution to exchange coupling, the presence of nonequivalent positions does not affect the energies of levels.

The value of exchange coupling in $CoCO_3$ is much smaller than the separation of the ground and first excited levels of Co^{2+} ion and, hence, the presence of this interaction does not influence the wave function of the ground state. In the case of Fe^{2+} ion, the distance to the excited level amounts to tens of inverse centimeters and the exchange coupling in the basal plane is absent. Therefore, it would be of interest to consider the influence of the excited states on the position of levels for

Table 6. Exchange coupling parameters of Hamiltonian (27) calculated in the Abragam–Pryce approximation

Ion pair	J_{xy}	D	J_z
Co^{2+}–Co^{2+}	$2(\sqrt{3}ac + b^2)^2 J$	$2\sqrt{3}(3ac + b^2)^2 J$	$(3a^2 + b^2 - c^2)^2 J$
Fe^{2+}–Fe^{2+}	0	0	$4(e^2 + 2f^2)^2 J$
Co^{2+}–Fe^{2+}	0	0	$2(3a^2 + b^2 - c^2)(e^2 + 2f^2)\,J$

the exchange-coupled Fe^{2+}–Fe^{2+} and Co^{2+}–Fe^{2+} ion pairs.

Let us find the exchange Hamiltonian with allowance for the excited state. To this end, we proceed from a Hamiltonian in the form

$$\hat{H} = \hat{W} + \hat{H}_{ex}, \tag{31}$$

where $\hat{W}$ is the operator of crystal field (4). For the Fe^{2+}–Fe^{2+} pair, the matrix of Hamiltonian (31) in the representation of the wave functions of the ground and excited states of Fe^{2+} ion separates into two one-dimensional matrices with $m = \pm 2$,

$$\langle \pm 1, \pm 1 | 2E_0 + A^2 J |, \tag{32}$$

two two-dimensional matrices with $m = \pm 1$,

$$\begin{array}{c} \\ \langle +1, 0| \\ \langle 0, +1| \end{array} \begin{vmatrix} |+1, 0\rangle & |0, +1\rangle \\ -ABJ + E_0 + E_1 & D_1 J \\ D_1 J & -ABJ + E_0 + E_1 \end{vmatrix},$$

$$\begin{array}{c} \\ \langle -1, 0| \\ \langle 0, -1| \end{array} \begin{vmatrix} |-1, 0\rangle & |0, -1\rangle \\ ABJ + E_0 + E_1 & D_2 J(1 + i\sqrt{3})/2 \\ D_2 J(1 - i\sqrt{3})/2 & ABJ + E_0 + E_1 \end{vmatrix}, \tag{33}$$

and one three-dimensional matrix with $m = 0$,

$$\begin{array}{c} \\ \langle +1, -1| \\ \langle 0, 0| \\ \langle -1, +1| \end{array} \begin{vmatrix} |+1, -1\rangle & |0, 0\rangle & |-1, +1\rangle \\ -A^2 J + 2E_0 & FJ(1 + i\sqrt{3})/2 & 0 \\ FJ(1 - i\sqrt{3})/2 & B^2 J + 2E_0 & FJ \\ 0 & FJ & -A^2 J + 2E_0 \end{vmatrix}. \tag{34}$$

where

$$A = e^2 + 2f^2, \quad B = \alpha^2 - \gamma^2,$$
$$D_1 = (\sqrt{6}d\alpha + \sqrt{6}e\beta + 2f\gamma)^2,$$
$$D_2 = (\sqrt{6}d\gamma + \sqrt{6}e\beta + 2f\alpha)^2,$$
$$F = \alpha\gamma(6d^2 + 4f^2) + 2\sqrt{6}df(\alpha^2 + \gamma^2) + 6e^2\beta^2 + e\beta(\alpha + \gamma)(6d + 2\sqrt{6}f),$$

E_0 is the energy of the ground state with $m = \pm 1$, and E_1 is the energy of the excited state with $m = 0$. From this it follows that, in the vicinity of the ground state with an energy of $2E_0$, there are four levels with energies of

$$\varepsilon_1 = \varepsilon_2 = 2E_0 + A^2 J = 2E_0 + 0.25J,$$
$$\varepsilon_3 = 2E_0 - A^2 J = 2E_0 - 0.25J,$$
$$\varepsilon_4 = 2E_0 - A^2 J - \frac{(A^2 + B^2)^2 J^2}{8(E_1 - E_0)} - \frac{F^2 J^2}{E_1 - E_0}$$
$$= 2E_0 - 0.25J - \frac{0.06J^2}{42000} - \frac{21J^2}{525}.$$

Let us obtain numerical estimates for the above results. Substituting the values of parameters into the expansions of wave functions, we obtain

$$A = 0.5, \quad B = 0, \quad D_1 = D_2 = 4.61, \quad F = 4.63.$$

As can be seen from these data, the corrections to positions of the levels are significant for the distances from excited levels on the order of 40 cm^{-1}.

Analogous calculations can be performed for the Co^{2+}–Fe^{2+} ion pair. In representation of the functions from Table 4, the matrix for determining the eigenvalues of Hamiltonian (31) is as follows:

$$\begin{array}{c} \\ \langle +, +1| \\ \langle +, -1| \\ \langle -, 0| \\ \langle -, +1| \\ \langle +, 0| \\ \langle -, -1| \end{array} \begin{vmatrix} |+, +1\rangle & |+, -1\rangle & |-, 0\rangle & |-, +1\rangle & |+, 0\rangle & |-, -1\rangle \\ AJ - E & 0 & 0 & 0 & 0 & 0 \\ 0 & -AJ - E & BJ & 0 & 0 & 0 \\ 0 & BJ & DJ - E & 0 & 0 & 0 \\ 0 & 0 & 0 & -AJ - E & CJ & 0 \\ 0 & 0 & 0 & CJ & DJ - E & 0 \\ 0 & 0 & 0 & 0 & 0 & AJ - E \end{vmatrix}, \tag{35}$$

where

$$A = (e^2/2 + f^2)(3a^2 + b^2 - c^2) = 0.31,$$
$$B = (\sqrt{3}ac + b^2)(2\sqrt{6}d\gamma + 2\sqrt{6}e\beta + 4f\alpha) = 1.47,$$
$$C = (\sqrt{3}ac + b^2)(2\sqrt{6}d\alpha + 2\sqrt{6}e\beta + 4f\gamma) = 1.47,$$
$$D = (3a^2 + b^2 - c^2)(\alpha^2 - \gamma^2) = 0.$$

From this it follows that, in the vicinity of the ground state with the energy $2E_0$, the levels are displaced approximately by $1.5J$ and their splitting amounts to $0.2J$.

5. CONCLUSIONS

An analysis of the behavior of exchange-coupled ion pairs in the magnetic field allows the role of the orbital angular momentum in the formation of their energy spectrum to be rationalized. In the general case, the influence of this momentum reduces to rendering the exchange coupling anisotropic. This results in splitting of the state with $S = 1$. An antisymmetric contribution related to the nonequivalent positions of magnetic ions in the pair influences the dependence of their level energies on the magnetic field oriented in the plane perpendicular to the C_3 axis. In the case when the anisotropic contribution to exchange coupling is absent, the antisymmetric contribution does not influence the positions of energy levels of the exchange-coupled pairs. It should be also noted that the positions of levels of the exchange-coupled Fe^{2+}–Fe^{2+} and Co^{2+}–Fe^{2+} ion pairs are significantly affected by the presence of a singlet excited state of the Fe^{2+} ion.

The results of this investigation help us understand the facts pointed out in the Introduction. In the case when Fe^{2+} ions enter the $CaCO_3$-type lattice, the presence of an orbital angular momentum results in the complete absence of the exchange coupling component in the basal plane perpendicular to the C_3 axis. This explains why $FeCO_3$ is a uniaxial antiferromagnet and accounts for a small value of the exchange field at the iron impurity ions in antiferromagnetic $CoCO_3$. If the magnetic ions occupy nonequivalent crystallographic positions, there appears an exchange coupling component that leads to sloped magnetic moments of the sublattices in the case of antiferromagnetic ordering. In $CoCO_3$ and $NiCO_3$, the magnitude of antisymmetric exchange exceeds the isotropic exchange in the basal plane.

The absence of a gap related to the hyperfine splitting in $CoCO_3$ can be understood based on the analysis of data presented in Table 2. As can be seen, the hyperfine interaction constants can vary from negative to positive values in response to small changes in the parameter p characterizing the configuration interaction. This is evidence of high sensitivity of these constants to the form of the function describing the ground state of a magnetic ion. Therefore, the presence of strong exchange fields can significantly influence these values for $CoCO_3$.

To summarize, the results of this investigation are as follows. For Co^{2+} ions in the calcite-type crystal lattice:

(i) The energy levels of the Co^{2+} ion and the corresponding wave functions are determined in the Abragam–Pryce approximation using the experimental g values and hyperfine interaction constants.

(ii) Expressions for Hamiltonians of the exchange coupling and the Dzyaloshinski interaction are found proceeding from the known wave functions of the ground state.

For Fe^{2+} ions in the calcite-type crystal lattice:

(i) The g values, energy levels, and wave functions of Fe^{2+} ion in the vicinity of the ground state are calculated in the Abragam–Pryce approximation.

(ii) The exchange coupling in the basal plane of a calcite-type lattice for the Fe^{2+}–Fe^{2+} and Co^{2+}–Fe^{2+} ion pairs is shown to be absent in the Abragam–Pryce approximation.

(iii) The energy levels of the exchange-coupled Fe^{2+}–Fe^{2+} and Co^{2+}–Fe^{2+} ion pairs are determined with allowance for the excited states.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project no. 02-02-17798.

REFERENCES

1. B. H. Schultz, Physica (Amsterdam) **7**, 413 (1940).
2. H. Bizette, J. Phys. Radium **12**, 161 (1951).
3. L. Néel and R. Pauthenet, C. R. Acad. Sci. (Paris) **234**, 2172 (1952).
4. A. S. Borovik-Romanov and M. P. Orlova, Zh. Éksp. Teor. Fiz. **31**, 579 (1956) [Sov. Phys. JETP **4**, 531 (1957)].
5. A. S. Borovik-Romanov, Zh. Éksp. Teor. Fiz. **36**, 75 (1959) [Sov. Phys. JETP **9**, 54 (1959)].
6. A. S. Borovik-Romanov, Zh. Éksp. Teor. Fiz. **36**, 766 (1959) [Sov. Phys. JETP **9**, 539 (1959)].
7. A. S. Borovik-Romanov and V. I. Ozhogin, Zh. Éksp. Teor. Fiz. **39**, 27 (1960) [Sov. Phys. JETP **12**, 18 (1961)].
8. A. S. Borovik-Romanov, N. M. Kreĭnes, and L. A. Prozorova, Zh. Éksp. Teor. Fiz. **45**, 64 (1963) [Sov. Phys. JETP **18**, 46 (1964)].
9. E. G. Rudashevskiĭ, Zh. Éksp. Teor. Fiz. **46**, 134 (1964) [Sov. Phys. JETP **19**, 96 (1964)].
10. A. S. Borovik-Romanov and E. G. Rudashevskiĭ, Zh. Éksp. Teor. Fiz. **47**, 2095 (1964) [Sov. Phys. JETP **20**, 1407 (1964)].
11. A. S. Borovik-Romanov and V. A. Tulin, Pis'ma Zh. Éksp. Teor. Fiz. **1** (5), 18 (1965) [JETP Lett. **1**, 134 (1965)].
12. V. I. Ozhogin, Zh. Éksp. Teor. Fiz. **48**, 1307 (1965) [Sov. Phys. JETP **21**, 874 (1965)].

13. A. S. Borovik-Romanov and V. F. Meshcheryakov, Zh. Éksp. Teor. Fiz. **53**, 853 (1967) [Sov. Phys. JETP **26**, 519 (1968)].
14. G. D. Bogomolov, Yu. F. Igonin, L. A. Prozorova, and F. S. Rusin, Zh. Éksp. Teor. Fiz. **54**, 1069 (1968) [Sov. Phys. JETP **27**, 572 (1968)].
15. L. A. Prozorova, Zh. Éksp. Teor. Fiz. **57**, 1967 (1969) [Sov. Phys. JETP **30**, 1065 (1970)].
16. A. S. Borovik-Romanov and V. F. Meshcheryakov, Pis'ma Zh. Éksp. Teor. Fiz. **8**, 425 (1968) [JETP Lett. **8**, 262 (1968)].
17. B. S. Dumesh, V. M. Egorov, and V. F. Meshcheryakov, Zh. Éksp. Teor. Fiz. **61**, 320 (1971) [Sov. Phys. JETP **34**, 168 (1972)].
18. A. N. Bazhan, Zh. Éksp. Teor. Fiz. **66**, 1086 (1974) [Sov. Phys. JETP **39**, 531 (1974)].
19. A. N. Bazhan, Zh. Éksp. Teor. Fiz. **67**, 1520 (1974) [Sov. Phys. JETP **40**, 757 (1975)].
20. A. S. Borovik-Romanov, V. G. Zhotikov, and N. M. Kreĭnes, Zh. Éksp. Teor. Fiz. **74**, 2286 (1978) [Sov. Phys. JETP **47**, 1188 (1978)].
21. I. E. Dzyaloshinski, Zh. Éksp. Teor. Fiz. **32**, 1547 (1957) [Sov. Phys. JETP **5**, 1259 (1957)].
22. T. Morija, Phys. Rev. **120**, 91 (1960).
23. I. S. Jacobs, J. Appl. Phys. **34**, 1106 (1963).
24. A. A. Antipin, V. M. Vinokurov, and M. M. Zaripov, Fiz. Tverd. Tela (Leningrad) **6**, 2178 (1964) [Sov. Phys. Solid State **6**, 2737 (1964)].
25. A. S. Borovik-Romanov, N. Yu. Ikornikova, V. F. Meshcheryakov, and E. G. Rudashevskiĭ, Kristallografiya **12**, 488 (1967) [Sov. Phys. Crystallogr. **12**, 417 (1967)].
26. B. N. Grechushnikov and V. F. Koryagin, Fiz. Tverd. Tela (Leningrad) **7**, 3123 (1965) [Sov. Phys. Solid State **7**, 2527 (1965)].
27. Chihiro Kikuchi and L. M. Matarrese, J. Chem. Phys. **33**, 601 (1960).
28. A. Abragam and B. Bleaney, *Electron Paramagnetic Resonance of Transition Ions* (Clarendon Press, Oxford, 1970; Mir, Moscow, 1972), Vol. 1.
29. V. F. Meshcheryakov, B. N. Grechushnikov, and I. N. Kalinkina, Zh. Éksp. Teor. Fiz. **66**, 1870 (1974) [Sov. Phys. JETP **39**, 920 (1974)].
30. S. A. Al'tshuler and B. M. Kozyrev, *Electron Paramagnetic Resonance in Compounds of Transition Elements*, 2nd ed. (Fizmatgiz, Moscow, 1972; Halsted, New York, 1975).
31. A. Abragam and M. H. L. Pryce, Proc. R. Soc. London, Ser. A **206**, 173 (1951).
32. V. V. Eremenko and A. I. Belyaeva, Fiz. Tverd. Tela (Leningrad) **6**, 3646 (1964) [Sov. Phys. Solid State **6**, 2918 (1964)].
33. Hang Nam Ok, Phys. Rev. **181**, 563 (1969).
34. M. J. Berggren, G. F. Imbusch, and P. L. Scott, Phys. Rev. **188**, 187 (1969).
35. L. Rimai, H. Statz, M. J. Weber, *et al.*, Phys. Rev. Lett. **4**, 125 (1960).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 1, 2004, pp. 156–161.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 1, 2004, pp. 173–180.

SOLIDS
Electronic Properties

Carrier Capture and Recombination in CdSe/ZnSe Quantum Dots

V. S. Dneprovskiĭ[a,*], E. A. Zhukov[a,], O. A. Shalygina[a], V. P. Evtikhiev[b], and V. P. Kochereshko[b]**

[a] *Moscow State University, Vorob'evy gory, Moscow, 119992 Russia*
[b] *Ioffe Physicotechnical Institute, Russian Academy of Sciences, St. Petersburg, 194021 Russia*
**e-mail: scon281@phys.msu.su*
***e-mail: zhukov@qwires.phys.msu.su*
Received April 11, 2003

Abstract—Time-resolved photoluminescence (PL) spectra of self-assembled CdSe/ZnSe quantum dots (QDs) are measured with a view to identifying the QD-size dependence of carrier capture and recombination in a single QD. The PL is excited by optical absorption in the ZnSe barrier layers under weak and strong irradiation with femto- and nanosecond laser pulses, respectively. In the case of weak excitation, the PL dynamics observed in a QD and the barrier layers are attributed to (i) fast carrier diffusion in the barrier layers, (ii) intense capture of carriers by the QD, (iii) fast carrier relaxation to the QD ground state, and (iv) dependence of the carrier capture and recombination times on the QD size. In the case of strong excitation, PL spectra are measured for different levels of excitation intensity and PL intensity is examined as a function of excitation intensity. It is established that (i) an increase in excitation intensity has a stronger effect on a high-frequency part of the spectrum and (ii) the intensity characteristic is essentially nonlinear. These findings are explained by state filling and/or decrease in carrier capture rate as the QD becomes increasingly full. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Recent years have seen considerable interest in II–VI semiconductor quantum dots (QDs). Aside from their unusual physical properties, QDs of this type suggest rich possibilities for creating high-speed, low-power optoelectronic devices, such as light-emitting diodes and lasers for the blue–green spectral range [1–9].

This paper presents a photoluminescence (PL) spectroscopy study of linear and nonlinear optical properties of self-assembled CdSe/ZnSe QDs grown by molecular-beam epitaxy (MBE). The goal is to identify the QD-size dependence of carrier capture and recombination in a single QD. The study consists of two parts, in which PL spectra are obtained after weak and strong excitation by femto- and nanosecond laser pulses, respectively.

Due to size quantization, the optical-transition energies of a QD vary with its size, which results in strong inhomogeneous broadening of the PL spectrum. The properties of a single QD have been examined by near-field optical microscopy [1] and spatially resolved techniques [10, 11]. In the latter case, the researchers employed a nanoaperture [10] or fabricated a reasonable number of QD islands by lithography [11].

A different approach was followed in this study. In the first part, we measured time-resolved inhomogeneously broadened PL spectra of a whole array of QDs at 20 K and analyzed the PL dynamics in different spectral regions. The dependence of carrier capture and recombination on QD size was thus determined for a single QD without using spatially resolved techniques. Note that if the excitation is weak enough, QDs luminesce by recombination of electron–hole pairs[1] in the ground state and the homogeneous PL spectral broadening of a single QD is within 1 meV at liquid-helium temperature [12]. For a single CdSe/ZnSe QD, the full width at half maximum (FWHM) measured in [11] was about 0.07 meV at 5 K.

In the second part, we investigated how the PL spectrum varies with excitation intensity and hence with the carrier density in a single QD. A possible mechanism of this variation is analyzed.

2. MATERIALS, METHODS, RESULTS, AND DISCUSSION

Arrays of CdSe/ZnSe QDs were fabricated with a two-chamber MBE machine in the Stranski–Krastanow growth mode on a GaAs(001) substrate misoriented by 6° toward [010]. A GaAs buffer layer was first formed

[1] Although electron–hole pairs bound by Coulomb attraction are commonly referred to as excitons when the discrete excited states of QDs are described, they are not true excitons, i.e., quasiparticles capable of dissociating into free electrons and holes [13, 14]. The Coulomb interaction energy in a QD is approximately $e^2/\varepsilon d$, where d is the mean QD size and ε is permittivity. This energy, though greater than that characteristic of two- and one-dimensional systems, is still much less than the level separation of free electrons and holes.

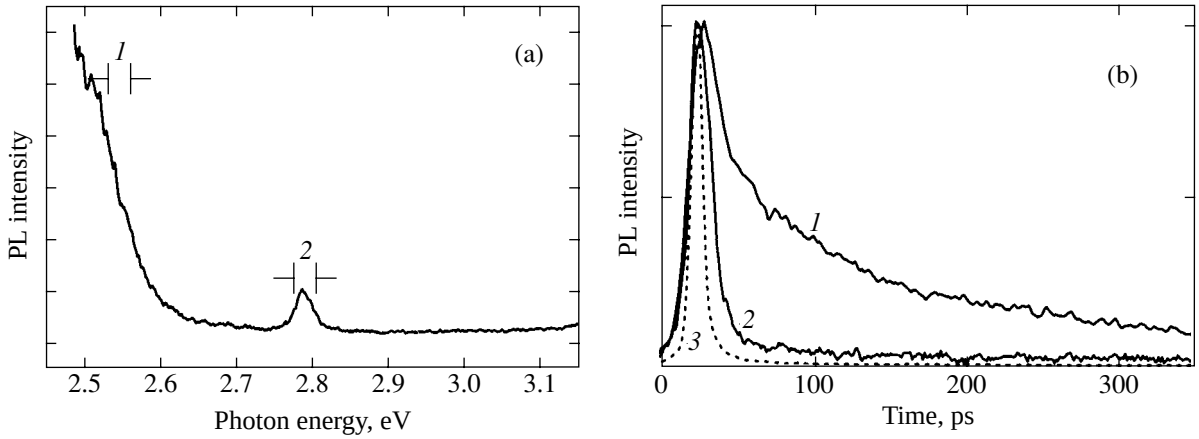

Fig. 1. Weak excitation: (a) a PL spectrum of the specimens at 20 K; (b) time plots of PL and excitation intensity. Spectral regions *1* and *2* in panel (a) correspond to QDs and barrier layers, respectively; curves *1–3* in panel (b), to QDs (spectral region *1*), barrier layers (spectral region *2*), and the excitation pulse, respectively.

on the substrate, and then a trilayered structure was made in which 2.8 monolayers of CdSe were sandwiched between ZnSe layers 40 nm-thick. The three layers were successively grown at 260, 320, and 280°C, respectively. We also used this method to produce specimens in which ten QD layers of CdSe were separated by 12-nm-thick ZnSe barrier layers.

The CdSe thickness exceeded 2.3 monolayers, the point of transition to three-dimensional growth [9, 15–17]. The density of QDs in the layer was about 10^{11} cm^{-2}. The use of a misoriented substrate under the growth conditions indicated above prevented the formation of two-dimensional CdSe islands and ensured step-flow growth (direct incorporation of adatoms into surface steps). When an exactly oriented GaAs(001) substrate is used, transition from two- to three-dimensional CdSe growth does not occur under typical conditions, which may be explained by insufficiently fast diffusion due to low growth temperature [18–21].

Electron and atomic-force microscopy measurements have shown that CdSe/ZnSe QDs created by a similar method and having a similar PL spectrum (peaking at 2.26 to 2.38 eV) tend to be spherical caps about 20 nm in diameter and between 2 and 3 nm in height [11].

In the present study, the linear and nonlinear optical properties of CdSe/ZnSe QDs were measured mainly by laser-based spectroscopic methods, including a time-resolved one.

2.1 Weak-Excitation Photoluminescence

In the case of weak excitation, we explored the dynamics of PL by measuring time-resolved PL spectra at 20 K. PL was excited by the focused second-harmonic emission ($h\nu$ = 3.1 eV) of a Ti:sapphire laser under pulsed conditions (with a pulse width of about 100 fs, a fluence of 0.2 μJ/cm^2, and a repetition rate of 82 MHz), pumped by an Ar laser. In this setting, optical absorption and carrier generation took place mainly in the ZnSe barrier layers.[2] The PL spectra were recorded by crossing a lock-in streak camera (Hamamatsu C1587) and a polychromator. The measuring system provided a time resolution better than 10 ps.

Figure 1 shows a PL spectrum of the specimens and time plots of PL intensity for the QDs and barrier layers over specific spectral regions. An intensity–time plot for the excitation pulse is also included in order to determine the time resolution of the measuring system.

Figures 2a and 2b display time-resolved PL spectra of the QDs and time plots of PL intensity, respectively. The PL spectra shown in Fig. 2a were measured over different periods after excitation. The time plots shown in Fig. 2b correspond to the spectral regions specified in Fig. 2a.

In Fig. 1a, the PL spectrum consists of a strong (2.3–2.5 eV) and a weak band (2.8 eV), which appear to correspond to the QDs and barrier layers, respectively. Note that the former is blue-shifted by more than 0.5 eV relative to the 1.8-eV energy gap of bulk CdSe as a result of size quantization. The large FWHM of the strong band (about 80 meV) can be attributed to the inhomogeneous broadening due to variability of the QD size. The estimated spread of QD size is less than 10%.

It was also established that (i) the PL decay time of the barrier layers is about 7 ps, which is smaller than the recombination time for bulk ZnSe; (ii) PL is much stronger in the QDs as compared to the barrier layers; and (iii) the PL rise time is as short as 8–24 ps, depending on the QD size (see below for details). These find-

[2] Excitation by this method results in intense PL of the QDs. When an Ar laser is used as an excitation source (with a photon energy less than the ZnSe energy gap), a much lower PL intensity is achieved, because the pumping radiation is poorly absorbed by QD layers.

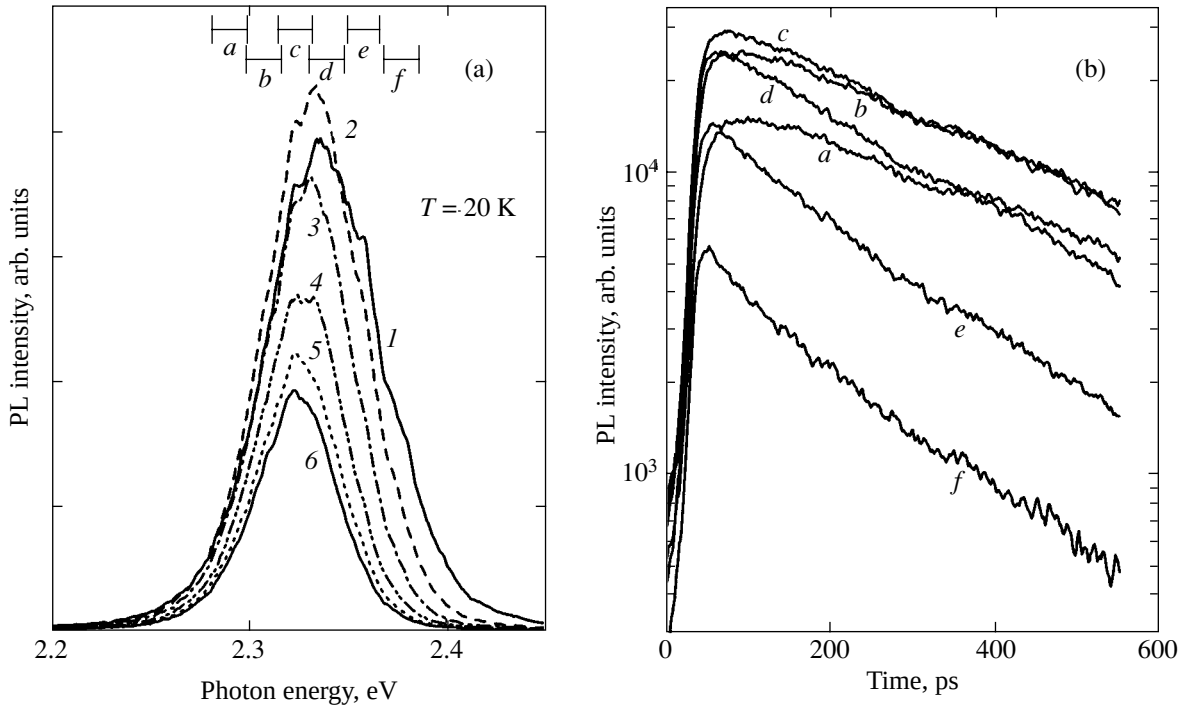

Fig. 2. Weak excitation: (a) time-resolved PL spectra of the QDs measured at 20 K; (b) time plots of PL intensity for the spectral regions specified in panel (a). The PL spectra of Fig. 2a correspond to the periods (*1*) 0–23, (*2*) 23–47, (*3*) 48–72, (*4*) 87–110, (*5*) 122–145, and (*6*) 145–169 ps after the end of excitation.

ings imply (i) fast carrier diffusion in the barrier layers, (ii) intense carrier capture by the QDs, and (iii) fast carrier relaxation to the ground electron–hole state of the QDs. Thus, there is no phonon bottleneck in the specimens under study [22].

Figure 2a demonstrates that PL decays more rapidly in a high-frequency region (the PL peak is red-shifted with time). Note also that the rise and fall times of PL intensity are longer for lower frequency regions (Fig. 2b).

The observed PL dynamics can be explained by variation of QD size. Recombination of electron–hole pairs in a QD can be radiative or radiationless. The rate of radiationless recombination increases with decreasing QD size because of the growing influence of the CdSe/ZnSe interface. This accounts for the changes in the shape of PL spectrum illustrated by Fig. 2a. The role played by nonlinear processes in these changes must be insignificant in view of the relatively low intensity of second-harmonic emission of the Ti:sapphire laser.

The time plots shown in Fig. 2b can be interpreted by invoking a three-level model [23]. To do this, one should introduce two time constants: τ_Σ and τ. The former is the sum of the carrier diffusion time for the barrier layers, the carrier capture time for the QD, and the time of relaxation to the ground electron–hole state in the QD. The time constant τ is the recombination time of an electron–hole pair (excitonic lifetime). The exciton density n as a function of time is governed by the equation

$$\frac{dn}{dt} = \frac{1}{\tau_\Sigma}(n_0\gamma e^{-t/\tau_\Sigma}) - \frac{n}{\tau}, \tag{1}$$

subject to the initial condition $n = 0$ at $t = 0$. Here, n_0 is the excited-carrier density and γ characterizes the efficiency of carrier migration into the QD.

It follows that the PL intensity I evolves as

$$I \propto n = \frac{e^{-t/\tau} - e^{-t/\tau_\Sigma}}{\tau - \tau_\Sigma}\gamma n_0 \tau. \tag{2}$$

The table lists the results obtained by applying the above approach Eq. (2) to the PL time plots shown in Fig. 2b and taking into account the time resolution of the measuring system. It shows how these parameters vary with QD size, which decreases with increasing frequency.

For each specific spectral region, the PL decay can be accurately represented as $e^{-(t/\tau)^\beta}$, with $\beta < 1$ given in the table. The steeper decrease in the recombination time for smaller QDs can be attributed to an increasing contribution of radiationless recombination. The use of a stretched exponential appears to provide an adequate approximation of the dependence of τ on QD size even over a narrow size range.

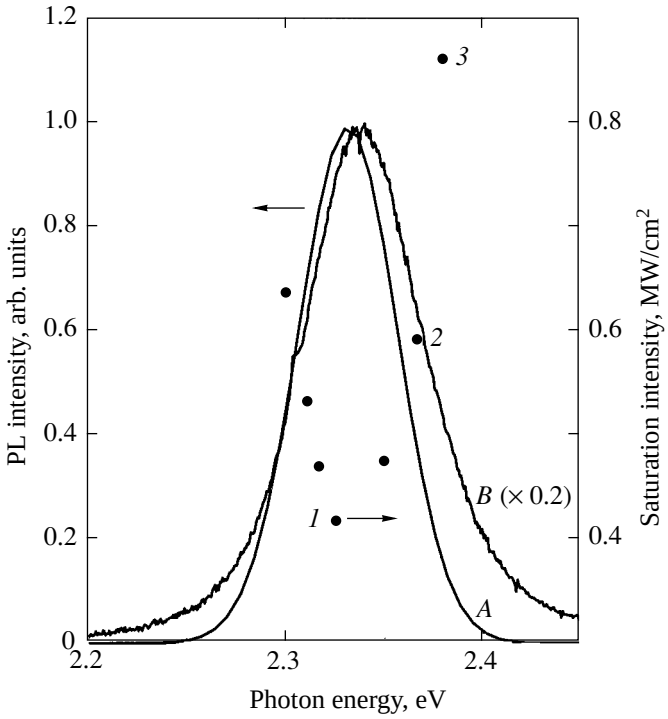

Fig. 3. Strong excitation: PL spectra (80 K) of the QDs for the excitation intensities (*A*) 0.057 and (*B*) 1.5 MW/cm^2. Closed circles represent saturation excitation intensities for different spectral regions.

Note also that τ_Σ increases with QD size. This may be explained by the increase in its capture component. It was shown in [24, 25] that the capture of electrons from a barrier layer into a quantum well may cause negative and positive charging of the quantum well and an adjacent barrier region, respectively. The growing Coulomb barrier inhibits the increase in carrier density in the quantum well. The decline of the carrier capture rate that ensues as the quantum well becomes increasingly full (see [26]) should be more pronounced in larger QDs because of a longer carrier lifetime.

At the same time, the dependence of τ_Σ on the QD size cannot be attributed to its relaxation component. One would rather expect it to be larger for smaller QDs, where the QD levels are less closely spaced, provided that resonant transitions are impossible (the level separation is not equal to the optical-phonon energy [27]). Furthermore, the time of relaxation to the ground state for II–VI QDs was found to be less than 1 ps; i.e., there is no phonon bottleneck [9, 22, 28].

The diffusion component of τ_Σ must be insignificant since the ZnSe barrier thickness is small in the specimens examined in this study.

2.2 Strong-Excitation Photoluminescence

In the case of strong excitation, we investigated how PL varied with excitation intensity. The PL was excited by the third-harmonic emission ($h\nu$ = 3.45 eV) of a Q-switched Nd:$YAlO_4$ laser with a pulse width of 14 ns.

Figure 3 shows the PL spectra of the specimens recorded at 80 K for two levels of excitation intensity. Note that an increase in excitation has a stronger effect on a high-frequency part of the spectrum (the peak is blue-shifted). Figure 4 demonstrates that the PL intensity measured for the high-frequency regions indicated in Fig. 3, which correspond to different QD sizes, is an essentially nonlinear function of excitation intensity.

These results can be explained by analyzing the contributions of the processes that can take place in QDs at high densities of electron–hole pairs: state filling [29, 30], the Stark effect [31], and temperature-dependent change in transition energy. The curves of Fig. 4 should be attributed to the filling of the QD ground level subject to the Pauli exclusion principle. The measured recombination time in the QDs is much smaller than the

Table

Spectral region	τ_Σ, ps	τ, ps	β
a	23 ± 2	410 ± 20	1
b	16 ± 1	418 ± 7	1
c	13 ± 1	357 ± 6	1
d	12 ± 1	244 ± 7	0.9 ± 0.1
e	11 ± 1	144 ± 8	0.7 ± 0.1
f	8 ± 1	112 ± 8	0.6 ± 0.1

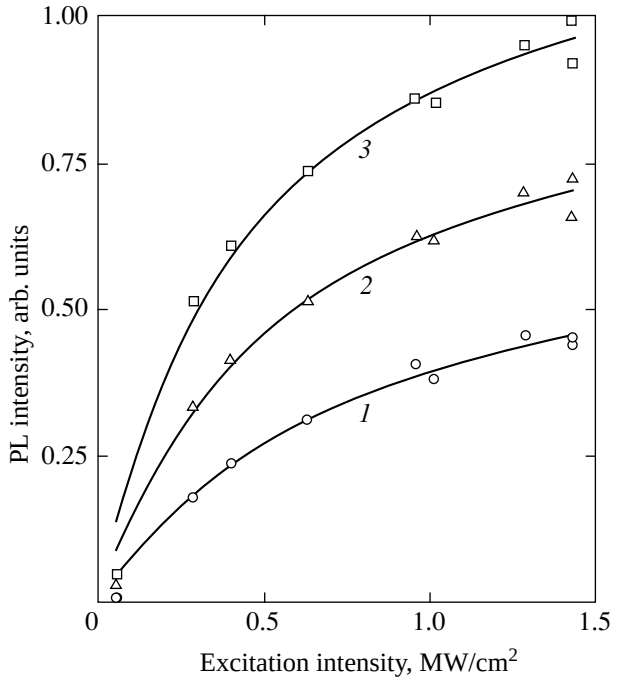

Fig. 4. Strong excitation: PL intensity vs. excitation intensity for spectral regions *1–3* of Fig. 3.

excitation pulse width: $\tau \ll \tau_p$. Accordingly, the steady-state density of the carriers injected into a single QD is

$$n_{QD} \approx \chi \frac{(1-R)(1-e^{-\alpha L})W\tau}{\tau_p Sh\gamma N}, \quad (3)$$

where χ characterizes the carrier-injection efficiency, R is the specimen's reflectance, α is the specimen's absorption coefficient, L is the total thickness of the barrier layers, W is the excitation-pulse energy, S is the area of excitation, and N is the QD density. Since measurements of the specimen's transmittance have shown that $n_{QD} > 10$, the steady-state carrier density is sufficient to ensure level filling in a single QD.

The nonlinear intensity characteristic was approximated by the saturation model

$$I \approx \frac{n}{1 + n/n_s}, \quad (4)$$

where I is the PL intensity, n is the QD carrier density as a function of excitation intensity, and n_s is its saturation level [30]. Furthermore, $n_s \approx P_s$, where P_s is the saturation excitation intensity. The PL intensity at $P = P_s$ is half as high as that reached in the case of a linear intensity characteristic.

At high frequencies, P_s was found to increase with frequency from 0.4 to 0.9 MW/cm^2 (Fig. 3). Both this effect and the blue shift can be attributed to shorter recombination times in smaller QDs.

The behavior of P_s at low frequencies (Fig. 3) may be associated with a considerable red shift in the frequency corresponding to the ground-state transition, due to the Stark effect and a temperature-dependent change in transition energy. Accordingly, the increase in PL intensity (Fig. 3) should result from the contribution of neighboring smaller QDs of higher density. The variation of P_s may also be associated with the decrease in capture rate with increasing QD size (see Section 2.1). This effect is more pronounced for large QDs because they are characterized by longer carrier lifetimes.

3. CONCLUSIONS

PL spectra of self-assembled CdSe/ZnSe QDs were measured by generating carriers in the ZnSe barrier layers by means of weak and strong pulsed laser radiation on time scales of femto- and nanoseconds, respectively.

In the case of weak excitation, time-resolved PL spectra were obtained. It is found that both rise and fall times of PL intensity are shorter for higher frequency regions. This finding is attributed to shorter carrier capture times and the recombination times of electron–hole pairs (excitons) corresponding to smaller QDs.

In the case of strong excitation, PL spectra were measured for different levels of excitation intensity and PL intensity was examined as a function of excitation intensity. The change in PL spectrum and the nonlinear nature of the intensity characteristic were explained by (i) state filling, (ii) a red shift in the ground-state-transition frequency due to the Stark effect and a temperature-dependent change in transition energy, and (iii) decrease in capture rate as the QD becomes increasingly full.

Thus, the use of time-resolved laser PL spectroscopy made it possible to determine the QD-size dependence of carrier capture and recombination in a single QD.

ACKNOWLEDGMENTS

This study was supported, in part, by the Russian Foundation for Basic Research, project nos. 02-02-16167 and 03-02-16762, and under the programs "Physics of Solid-State Nanostructures" and "Integration."

REFERENCES

1. D. Bimberg, M. Grundmann, and N. Ledentsov, *Quantum Dot Heterostructures* (Wiley, Chichester, 1999).
2. Yu. V. Vandyshev, V. S. Dneprovskiĭ, V. I. Klimov, and D. K. Okorokov, Pis'ma Zh. Éksp. Teor. Fiz. **54**, 441 (1991) [JETP Lett. **54**, 442 (1991)].
3. V. S. Dneprovskiĭ, V. I. Klimov, D. K. Okorokov, and Yu. V. Vandyshev, Solid State Commun. **81**, 227 (1992).
4. N. N. Ledentsov, I. L. Krestnikov, M. V. Maximov, *et al.*, Appl. Phys. Lett. **69**, 1343 (1996).
5. S. Nakamura, R. Rinamura, H. Umeya, *et al.*, Electron. Lett. **34**, 2435 (1998).

6. M. Kobayashi, S. Nakamura, K. Wakao, and A. Yoshikawa, J. Vac. Sci. Technol. B **16**, 1316 (1998).
7. M. Kobayashi, S. Nakamura, K. Kitamura, *et al.*, J. Vac. Sci. Technol. B **17**, 2005 (1999).
8. N. Matsumura, H. Endo, and J. Saraie, Phys. Status Solidi B **229**, 1039 (2002).
9. D. Schikora, S. Schwedhelm, I. Kudryashov, *et al.*, J. Cryst. Growth **214**/**215**, 699 (2000).
10. L. M. Robinson, H. Rho, J. C. Kim, *et al.*, Phys. Rev. Lett. **83**, 2797 (1999).
11. A. Hundt, T. Flissikowski, M. Lowisch, *et al.*, Phys. Status Solidi B **224**, 159 (2001).
12. V. Turck, S. Rodt, R. Heitz, *et al.*, Phys. Status Solidi B **224**, 217 (2001).
13. L. Brus, IEEE J. Quantum Electron. **22**, 1909 (1986).
14. L. V. Keldysh, Phys. Status Solidi A **164**, 3 (1997).
15. M. Strassburg, Th. Deniozou, A. Hoffmann, *et al.*, J. Cryst. Growth **214**/**215**, 756 (2000).
16. D. Litvinov, A. Rosenauer, D. Gerthsen, and H. Preis, Phys. Status Solidi B **229**, 523 (2002).
17. V. G. Litvinov, V. I. Kozlovsky, and Yu. G. Sadofyev, Phys. Status Solidi B **229**, 513 (2002).
18. L. M. Robinson, H. Rho, H. E. Jackson, *et al.*, Phys. Status Solidi B **221**, 55 (2000).
19. D. Litvinov, D. Gerthsen, A. Rosenauer, *et al.*, Phys. Status Solidi B **224**, 147 (2001).
20. A. Reznitsky, A. Klochikhin, S. Permogorov, *et al.*, in *Proceedings of 10th International Symposium on Nanostructures: Physics and Technology, St. Petersburg, Russia* (2002), p. 611.
21. H. Kirmse, R. Schneider, M. Rabe, and F. Henneberger, Appl. Phys. Lett. **72**, 1329 (1998).
22. A. L. Efros, V. A. Kharchenko, and M. Rosen, Solid State Commun. **93**, 281 (1995).
23. B. Ohnesorge, M. Albrecht, J. Oshinowo, *et al.*, Phys. Rev. B **54**, 11532 (1996).
24. V. N. Abakumov, V. I. Perel', and I. N. Yassievich, *Nonradiative Recombination in Semiconductors* (Peterb. Inst. Yad. Fiz. Ross. Akad. Nauk, St. Petersburg, 1997).
25. S. A. Solov'ev, I. N. Yassievich, and V. M. Chistyakov, Fiz. Tekh. Poluprovodn. (St. Petersburg) **29**, 1264 (1995) [Semiconductors **29**, 654 (1995)].
26. K. Schmalz, I. N. Yassievich, H. Rucker, *et al.*, Phys. Rev. B **50**, 14287 (1994).
27. P. W. M. Blom, C. Smit, J. E. M. Havercort, and J. H. Wolter, Phys. Rev. B **47**, 2072 (1993).
28. R. Heitz, M. Veit, N. N. Ledentsov, *et al.*, Phys. Rev. B **56**, 10435 (1997).
29. Yu. V. Vandyshev, V. S. Dneprovskiĭ, and V. I. Klimov, Zh. Éksp. Teor. Fiz. **101**, 270 (1992) [Sov. Phys. JETP **74**, 144 (1992)].
30. *Nonlinear Photonics*, Ed. by H. M. Gibbs, G. Khitrova, and N. Peyghambarian (Springer, Berlin, 1990).
31. D. J. Norris, A. Sacra, C. B. Murray, and M. G. Bawendi, Phys. Rev. Lett. **72**, 2612 (1994).

Translated by A. Sharshakov

Journal of Experimental and Theoretical Physics, Vol. 98, No. 1, 2004, pp. 162–169.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 1, 2004, pp. 181–188.

SOLIDS
Electronic Properties

Conductivity of a Periodic Two-Component System of Rhombic Type

Yu. N. Ovchinnikov[a,b]

[a] *Max-Planck-Institut für Physik komplexer Systeme, 01187, Dresden, Germany*
[b] *Landau Institute for Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
e-mail: ovc@itp.ac.ru

Received July 10, 2003

Abstract—The conductivity and the distribution of electric field, current, and charge density in a periodic two-component system composed of rhombs with an arbitrary vertex angle of 2α are investigated. The effective conductivity of such a medium is represented by a tensor with components $\sigma_{\text{eff}}^{11}(\alpha)$ and $\sigma_{\text{eff}}^{22}(\alpha)$ in the principal axes that satisfy the Dykhne relation $\sigma_{\text{eff}}^{11}(\alpha)\,\sigma_{\text{eff}}^{22}(\alpha) = \sigma_1\sigma_2$, where σ_1, σ_2 are the isotropic conductivities of media 1 and 2. In addition, the relation $\sigma_{\text{eff}}^{22}(\alpha) = \sigma_{\text{eff}}^{11}(\pi/2 - \alpha)$ is satisfied. The principal axes are directed along the diagonals of the rhombs. It is shown that there are three lines in the rectangle $0 < \alpha \le \pi/2$, $-1 < Z < 1$ ($Z = (\sigma_1 - \sigma_2)/(\sigma_1 + \sigma_2)$) on which the charge density is expressed in terms elliptic functions. An explicit expression is obtained for all physical quantities on these lines. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A periodic two-component system composed of rhombs with a vertex angle of 2α seems to be the simplest in which the effective conductivity is not isotropic, whereby the Dykhne relation for the components of the conductivity tensor with components $\sigma_{\text{eff}}^{11}(\alpha)$ and $\sigma_{\text{eff}}^{22}(\alpha)$ in the principal axes [1–4],

$$\sigma_{\text{eff}}^{11}(\alpha)\sigma_{\text{eff}}^{22}(\alpha) = \sigma_1\sigma_2, \tag{1}$$

is insufficient for the complete reconstruction of the tensor of effective conductivity. In (1), the quantities σ_1, σ_2 are the isotropic conductivities of the media. The components $\sigma_{\text{eff}}^{11}(\alpha)$ and $\sigma_{\text{eff}}^{22}(\alpha)$ satisfy the following obvious relation:

$$\sigma_{\text{eff}}^{22}(\alpha) = \sigma_{\text{eff}}^{11}(\pi/2 - \alpha). \tag{2}$$

Below, we will show that even the first terms of perturbation theory in the parameter

$$Z = \frac{\sigma_1 - \sigma_2}{\sigma_1 + \sigma_2} \tag{3}$$

for the components of the tensor of effective conductivity nontrivially depend on the angle α. Therefore, it is hardly possible that one can obtain exact algebraic formulas for the tensor components $\sigma_{\text{eff}}^{11}(\alpha)$ and $\sigma_{\text{eff}}^{22}(\alpha)$. Nevertheless, in the domain of parameters ($0 < \alpha < \pi/2$, $-1 < Z < 1$), there exist three lines on which one can obtain a closed-form expression for the charge density, concentrated on the boundaries of domains (1) and (2), in terms of elliptic functions. In the domain $Z \longrightarrow 1$, $\alpha \longrightarrow 0$, the charge density and other physical quantities essentially depend on the ratio of small parameters $(1 - Z)/\alpha$.

2. CHARGE DENSITY IN A PERIODIC SYSTEM COMPOSED OF RHOMBS WITH A VERTEX ANGLE OF 2α

Choose a coordinate system (x, y) with the origin at a vertex of a certain rhomb and the axes directed along the diagonals of the rhombs (see Fig. 1). In this case, the vectors of a unit cell can be chosen as

$$2a\sin\alpha(1, 0); \quad 2a\cos\alpha(0, 1), \tag{4}$$

where a is the length of the side of the rhomb. The continuity of current on the boundaries of rhombs determines the jump of electric field E,

$$\sigma_1\mathbf{E}_{\mathbf{n}}^{(1)} = \sigma_2\mathbf{E}_{\mathbf{n}}^{(2)}, \tag{5}$$

where $\mathbf{n}$ is a normal vector to the side of the rhomb and $E_n^{(1)}$, $E_n^{(2)}$ is a projection of electric field onto the normal vector. The Maxwell equation $\text{div}\mathbf{E} = 4\pi\rho$, where

ρ is the charge density, yields the second equation for the electric field jump

$$E_{\mathbf{n}}^{(1)} - E_{\mathbf{n}}^{(2)} = 4\pi\rho. \tag{6}$$

The relation between the scalar potential φ and the charge density is given by

$$\varphi = E_0 y - 4\pi \int d^2 r_1 G(r - r_1)\rho(r_1), \tag{7}$$

where G is the Green function of the Laplace operator on the plane,

$$G = \frac{1}{2\pi}\ln|r - r_1|. \tag{8}$$

The electric field $\mathbf{E}_0$ is directed along axis y; this allows us to find the component $\sigma_{\text{eff}}^{22}(\alpha)$ of the conductivity tensor in the principal axes. There are obvious relations between the components $\sigma_{\text{eff}}^{11}(\alpha)$ and $\sigma_{\text{eff}}^{22}(\alpha)$ of the conductivity tensor in the principal axes,

$$\begin{aligned}\sigma_{\text{eff}}^{11}(\alpha) &= \sigma_{\text{eff}}^{22}\left(\frac{\pi}{2} - \alpha\right),\\ \sigma_{\text{eff}}^{12}(\alpha) &= \sigma_{\text{eff}}^{21}(\alpha) = 0.\end{aligned} \tag{9}$$

Charges are localized only along the sides of the rhombs. Due to the translation symmetry and two symmetry planes, there exist only two independent functions ρ_1, ρ_2 that determine the charge density on all the sides of the rhombs. As a result, taking into account formulas (4), we obtain the following expression for the scalar potential φ:

$$\begin{aligned}\varphi = E_0 y - 2a &\times \int_0^1 dt \sum_{K,L} \{[\ln((x - at\sin\alpha - 2aK\sin\alpha)^2\\ &+ (y - at\cos\alpha - 2aL\cos\alpha)^2)^{1/2}\\ &+ \ln((x + at\sin\alpha - 2aK\sin\alpha)^2\\ &+ (y - at\cos\alpha - 2aL\cos\alpha)^2)^{1/2}]\rho_1(t)\\ &+ [\ln((x - at\sin\alpha - 2aK\sin\alpha)^2\\ &+ (y + at\cos\alpha - 2aL\cos\alpha)^2)^{1/2}\\ &+ \ln((x + at\sin\alpha - 2a\sin\alpha)^2\\ &+ (y + at\cos\alpha - 2aL\cos\alpha)^2)^{1/2}]\rho_2(t)\}.\end{aligned} \tag{10}$$

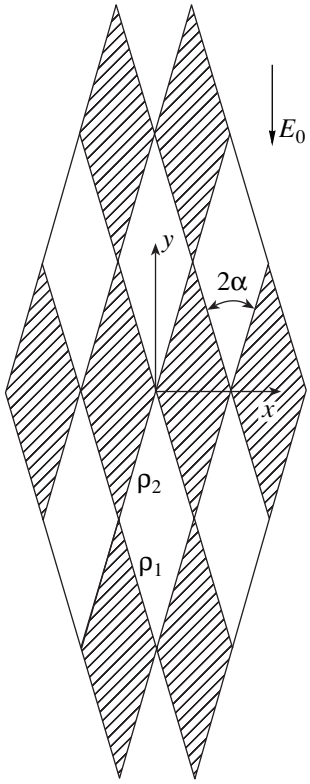

Fig. 1. Periodic two-component system of rhombic type.

Boundary conditions (5) and (6) and expression (10) for the scalar potential φ yield two equations for ρ_1, ρ_2. A simple verification shows that the functions ρ_1, ρ_2 satisfy the simple relation

$$\rho_2(t) = -\rho_1(t). \tag{11}$$

As a result, we have only one equation for $\rho_1(t)$:

$$\begin{aligned}\frac{Z}{2\pi}\Bigg\{E_0\sin\alpha + 2\sin(2\alpha)\int_0^1 dt\rho_1(t)\\ \times \sum_{K,L}\Bigg[\frac{t + K + L}{(t + t' + 2K)^2\sin^2\alpha + (t - t' + 2L)^2\cos^2\alpha}\\ + \frac{K + L}{(t - t' - 2K)^2\sin^2\alpha + (t - t' + 2L)^2\cos^2\alpha}\\ - \frac{K + L}{(t + t' + 2K)^2\sin^2\alpha + (t + t' - 2L)^2\cos^2\alpha}\\ + \frac{t - K - L}{(t - t' - 2K)^2\sin^2\alpha + (t + t' - 2L)^2\cos^2\alpha}\Bigg]\Bigg\} = \rho_1(t').\end{aligned} \tag{12}$$

For $\alpha = \pi/4$, formula (12) reduces to the expression obtained in [3].

In (12), we can perform one summation. As a result, we reduce Eq. (12) to

$$\frac{Z}{2\pi}\Bigg\{E_0\sin\alpha + \pi\int_0^1 dt\rho_1(t)$$

$$\times\sum_K\Bigg[\frac{\sinh(\pi\tan\alpha(t+t'+2K))}{\cosh(\pi\tan\alpha(t+t'+2K))-\cos(\pi(t-t'))}$$

$$+\frac{\sinh(\pi\cot\alpha(t-t'+2K))}{\cosh(\pi\cot\alpha(t-t'+2K))-\cos(\pi(t+t'))}$$

$$+\frac{\sinh(\pi\cot\alpha(t+t'+2K))}{\cosh(\pi\cot\alpha(t+t'+2K))-\cos(\pi(t-t'))}$$

$$+\frac{\sinh(\pi\tan\alpha(t-t'+2K))}{\cosh(\pi\tan\alpha(t-t'+2K))-\cos(\pi(t+t'))} \quad (13)$$

$$+\frac{\sinh(\pi\cot\alpha(t-t'-2K))}{\cosh(\pi\cot\alpha(t-t'-2K))-\cos(\pi(t-t'))}$$

$$-\frac{\sinh(\pi\tan\alpha(t-t'+2K))}{\cosh(\pi\tan\alpha(t-t'+2K))-\cos(\pi(t-t'))}$$

$$+\frac{\sinh(\pi\cot\alpha(t+t'-2K))}{\cosh(\pi\cot\alpha(t+t'-2K))-\cos(\pi(t+t'))}$$

$$-\frac{\sinh(\pi\tan\alpha(t+t'+2K))}{\cosh(\pi\tan\alpha(t+t'+2K))-\cos(\pi(t+t'))}\Bigg]\Bigg\} = \rho_1(t'),$$

where

$$Z = \frac{\sigma_1-\sigma_2}{\sigma_1+\sigma_2}. \quad (14)$$

The function $\rho_1(t)$ is a doubly periodic analytic function of the complex variable t and has two branching points in each parallelogram of periods. Let us determine the behavior of $\rho_1(t)$ in the neighborhood of the branching points $t = \{0, 1\}$. The asymptotics of the function $\rho_1(t)$ as $t \longrightarrow 0$ is proportional to a power of t:

$$\rho_1(t) \propto A/t^{2\nu}. \quad (15)$$

The coefficient ν can be obtained from the following equation derived from Eq. (13):

$$Z\sin(2\alpha)\int_0^\infty dt t^{1-2\nu}$$

$$\times\Bigg[\frac{1}{t^2+1-2t\cos(2\alpha)}+\frac{1}{t^2+1-2t\cos(2\alpha)}\Bigg] = \pi. \quad (16)$$

Calculating the integrals in (16), we obtain

$$\int_0^\infty\frac{dtt^{1-2\nu}}{t^2+1+2t\cos(2\alpha)} = \frac{\pi}{\sin(2\alpha)}\frac{\sin(2\alpha(1-2\nu))}{\sin(2\pi\nu)},$$

$$\int_0^\infty\frac{dtt^{1-2\nu}}{t^2+1-2t\cos(2\alpha)} = \frac{\pi}{\sin(2\alpha)}\frac{\sin(2\pi\nu+2\alpha(1-2\nu))}{\sin(2\pi\nu)}. \quad (17)$$

Then, the equation for the parameter ν follows from Eqs. (16) and (17):

$$Z\frac{\sin(\pi\nu+2\alpha(1-2\nu))}{\sin(\pi\nu)} = 1. \quad (18)$$

For $\alpha = \pi/4$, Eq. (18) coincides with the result of [3]. In the neighborhood of the point $t = 1$, the function $\rho_1(t)$ can be represented as

$$\rho_1(t) \approx \frac{B}{(1-t)^{2\mu}}. \quad (19)$$

Parameter μ satisfies the equation

$$Z\frac{\sin(\pi\mu+2\alpha(1-2\mu))}{\sin(\pi\mu)} = -1. \quad (20)$$

Equations (18) and (20) were obtained in [5].

The two periods (τ, τ') can be determined by Eq. (13) and are given by

$$\tau = \frac{2}{1+i\tan\alpha}, \quad \tau' = \frac{2i\tan\alpha}{1+i\tan\alpha}. \quad (21)$$

The charge density $\rho_1(t)$ can be represented as

$$\rho_1(t) = A(\phi(t))^\nu + A^*((\phi(t^*))^*)^\nu, \quad (22)$$

where $\phi(t)$ is a doubly periodic function.

Notice the following important property of the kernel in Eq. (13): the integral with respect to external variable t' does not depend on angle α and, hence, coincides with the expression obtained in [3]. Integrating Eq. (13) with respect to external variable t', we obtain

$$\frac{Z}{2\pi}\Bigg\{E_0\sin\alpha + 2\pi\int_0^1 dt(1-2t)\rho_1(t)\Bigg\} = \int_0^1 dt\rho_1(t). \quad (23)$$

The component $\sigma_{\text{eff}}^{22}(t)$ of the conductivity tensor is expressed in terms of the charge density $\rho_1(t)$ by the formula

$$\sigma_{\text{eff}}^{22}(\alpha) = \frac{4\pi}{E_0 \sin\alpha}\frac{\sigma_1\sigma_2}{\sigma_1-\sigma_2}\int_0^1 dt\rho_1(t). \quad (24)$$

3. PERTURBATION THEORY WITH RESPECT TO PARAMETER Z

Perturbation theory in Eqs. (12) and (13) is trivial. Setting

$$\rho_1(t) = E_0\tilde{\rho}(t)\sin\alpha, \quad (25)$$

we obtain the following expression up to the second-order terms in Z:

$$\tilde{\rho}(t') = \frac{Z}{2\pi} + \frac{Z^2}{2\pi^2}\sin(2\alpha)$$

$$\times\int_0^1 dt\sum_{K,L}\left[\frac{t+K+L}{(t+t'+2K)^2\sin^2\alpha+(t-t'+2L)^2\cos^2\alpha}\right.$$

$$+\frac{K+L}{(t-t'-2K)^2\sin^2\alpha+(t-t'+2L)^2\cos^2\alpha} \quad (26)$$

$$-\frac{K+L}{(t+t'+2K)^2\sin^2\alpha+(t+t'-2L)^2\cos^2\alpha}$$

$$\left.+\frac{t-K-L}{(t-t'-2K)^2\sin^2\alpha+(t+t'-2L)^2\cos^2\alpha}\right]+\ldots.$$

In third-order perturbation theory, we obtain the following expression for the conductivity $\sigma_{\text{eff}}^{22}(\alpha)$ from (23), (24), and (26):

$$\sigma_{\text{eff}}^{22}(\alpha) = \frac{2\sigma_1\sigma_2}{\sigma_1+\sigma_2}\left\{1+\frac{Z^2}{2}\int_0^1 dt(1-2t')\right.$$

$$\times\int_0^1 dt\sum_{k=-\infty}^{\infty}\left[\frac{\sinh(\pi\tan\alpha(t+t'+2K))}{\cosh(\pi\tan\alpha(t+t'+2K))-\cos(\pi(t-t'))}\right.$$

$$+\frac{\sinh(\pi\cot\alpha(t-t'+2K))}{\cosh(\pi\cot\alpha(t-t'+2K))-\cos(\pi(t+t'))}$$

$$+\frac{\sinh(\pi\cot\alpha(t+t'-2K))}{\cosh(\pi\cot\alpha(t+t'-2K))-\cos(\pi(t-t'))}$$

$$+\frac{\sinh(\pi\tan\alpha(t-t'-2K))}{\cosh(\pi\tan\alpha(t-t'-2K))-\cos(\pi(t+t'))} \quad (27)$$

$$-\frac{\sinh(\pi\tan\alpha(t-t'-2K))}{\cosh(\pi\tan\alpha(t-t'-2K))-\cos(\pi(t-t'))}$$

$$+\frac{\sinh(\pi\cot\alpha(t-t'+2K))}{\cosh(\pi\cot\alpha(t-t'+2K))-\cos(\pi(t-t'))}$$

$$+\frac{\sinh(\pi\cot\alpha(t+t'-2K))}{\cosh(\pi\cot\alpha(t+t'-2K))-\cos(\pi(t+t'))}$$

$$\left.\left.-\frac{\sinh(\pi\tan\alpha(t+t'+2K))}{\cosh(\pi\tan\alpha(t+t'+2K))-\cos(\pi(t+t'))}\right]\right.$$

It follows from (27) that the conductivity $\sigma_{\text{eff}}^{22}(\alpha)$ can be represented as

$$\sigma_{\text{eff}}^{22}(\alpha) = \frac{2\sigma_1\sigma_2}{\sigma_1+\sigma_2}\left\{1+\frac{Z^2}{2}(F_1(\alpha)+F_2(\alpha))\right\}, \quad (28)$$

where

$$F_1(\alpha) = F_1\left(\frac{\pi}{2}-\alpha\right), \quad F_2(\alpha) = -F_2\left(\frac{\pi}{2}-\alpha\right). \quad (29)$$

It follows from the Dykhne theory that the function $F_1(\alpha)$ is independent of α and equals unity:

$$F_1(\alpha) \equiv 1. \quad (30)$$

From (27) and (28), we derive the function

$$F_2(\alpha) = \int_0^1 dt'(1-2t')$$

$$\times\int_0^1 dt\sum_{K=-\infty}^{\infty}\left[\frac{\sinh(\pi\cot\alpha(t-t'+2K))}{\cosh(\pi\cot\alpha(t-t'+2K))-\cos(\pi(t-t'))}\right.$$

$$+\frac{\sinh(\pi\cot\alpha(t+t'-2K))}{\cosh(\pi\cot\alpha(t+t'-2K))-\cos(\pi(t+t'))} \quad (31)$$

$$-\frac{\sinh(\pi\tan\alpha(t-t'-2K))}{\cosh(\pi\tan\alpha(t-t'-2K))-\cos(\pi(t-t'))}$$

$$\left.-\frac{\sinh(\pi\tan\alpha(t+t'+2K))}{\cosh(\pi\tan\alpha(t+t'+2K))-\cos(\pi(t+t'))}\right].$$

Integrating by parts the right-hand side of Eq. (31),

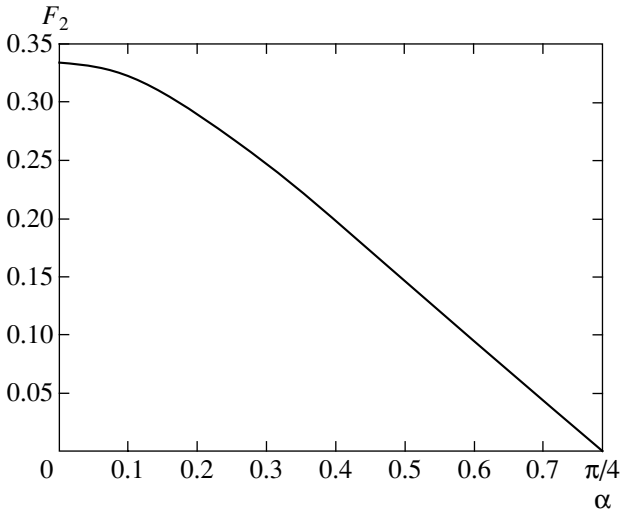

Fig. 2.

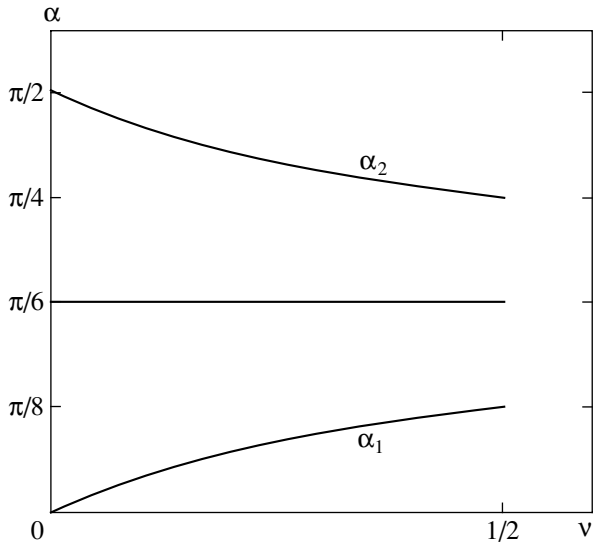

Fig. 3.

we obtain

$$F_2(\alpha) = 4\int_0^1 dt'(t' - (t')^2) \times \sum_{K=-\infty}^{\infty}\left[\frac{\sinh(\pi\cot\alpha(t'+2K))}{\cosh(\pi\cot\alpha(t'+2K)) - \cos(\pi t')} - \frac{\sinh(\pi\tan\alpha(t'+2K))}{\cosh(\pi\tan\alpha(t'+2K)) - \cos(\pi t')}\right]. \quad (32)$$

In the limit as $\alpha \longrightarrow 0$, we obtain

$$F_2(\alpha)_{\alpha\to 0} = 4\int_0^1 dt(t - t^2)(1-t) = \frac{1}{3}. \quad (33)$$

The function $F_2(\alpha)$ is a nontrivial function of α. The results of the numerical calculation of the function $F_2(\alpha)$ are given in Fig. 2.

4. EXACT SOLUTIONS OF THE EQUATION FOR ρ_1

There are three lines in the rectangle $\{0 < \alpha < \pi/2, 0 < \nu < 1/2\}$ on which the function $\phi(t)$ is elliptic. One of these is the straight line $\alpha = \pi/4$. In this case, $\tau = (\tau')$, $\mu = -\nu$, and $\phi = B\wp^\nu$, where $\wp$ is the Weierstrass elliptic function [3] and B is a numerical factor. There are two lines on which

$$\mu = -1 - \nu, \quad (34)$$

and, hence, ϕ is an elliptic function. On these lines, the function $\alpha_1(\nu)$, $\alpha_2(\nu)$ is defined by formulas (18), (20), and (34), which lead to the equation

$$\frac{\sin(\pi\nu + 2\alpha(1-2\nu))}{\sin(\pi\nu)} + \frac{\sin(\pi\mu + 2\alpha(1-2\mu))}{\sin(\pi\mu)} = 0. \quad (35)$$

Solutions to this equation are given by

$$\alpha_1(\nu) = \frac{\pi\nu}{2(1+2\nu)}, \quad \alpha_2 = \frac{\pi}{2} - \alpha_1 = \frac{\pi(1+\nu)}{2(1+2\nu)}. \quad (36)$$

All three of these lines are shown in Fig. 3. On lines (36), the elliptic function ϕ can be expressed in terms of function θ_1 [6]:

$$\theta_1(z) = i\sum_{K=-\infty}^{\infty}(-1)^K\exp\left(-\pi\tan\alpha\left(K - \frac{1}{2}\right)^2\right) \times \exp(i(2K-1)\pi z). \quad (37)$$

It follows from (37) that

$$\theta_1\left(\frac{z+\tau}{\tau}\right) = e^{-i\pi}\theta_1\left(\frac{z}{\tau}\right), \quad \theta_1\left(\frac{z+\tau'}{\tau}\right) = -\exp\left(\pi\tan\alpha - \frac{2i\pi z}{\tau}\right)\theta_1\left(\frac{z}{\tau}\right). \quad (38)$$

Formulas (38) make it possible to express the elliptic function ϕ with periods $\{\tau, \tau'\}$ as

$$\phi = \left(\frac{\theta_1\left(\frac{1-z}{\tau}\right)}{e^{i\pi z/\tau}\theta_1\left(\frac{z}{\tau}\right)}\right)^2. \quad (39)$$

The function $\phi(z)$ satisfies the relation

$$\phi(z)\phi(1-z) = e^{\pi\tan\alpha - i\pi}. \tag{40}$$

We also present the following useful relations:

$$e^{i\pi z/\tau}\theta_1\left(\frac{z}{\tau}\right) = ie^{-\pi\tan\alpha/4} \times \sum_{K=-\infty}^{\infty}(-1)^K\exp[-\pi\tan\alpha(K^2 - K(1-z)) + i\pi Kz], \tag{41}$$

$$\phi(1-z) = \left(\frac{2\sum_{K=1}^{\infty}(-1)^{K+1}\exp\left[-\pi\tan\alpha\left(K-\frac{1}{2}\right)^2\right]\sin\left(\pi z\left(K-\frac{1}{2}\right)\frac{e^{i\alpha}}{\cos\alpha}\right)}{ie^{(-(\pi/4)\tan\alpha)}\left[1+2\sum_{K=1}^{\infty}e^{-\pi\tan\alpha K^2}\cosh\left(\pi Kz\frac{e^{-i(\pi/2-\alpha)}}{\cos\alpha}\right)\right]}\right)^2.$$

The function $\phi(z)$ satisfies the differential equation

$$(\phi')^2 + C_1\phi^3 + C_2\phi^2 - C_1e^{\pi\tan\alpha}\phi = 0, \tag{42}$$

where

$$C_1 = \left(\frac{4\pi B_0 e^{-i(\pi/2-\alpha)}}{B_1\cos\alpha}e^{-(\pi/4)\tan\alpha}\right)^2, \tag{43}$$

$$C_1 = \frac{12\pi^2}{\cos^2\alpha}e^{2i\alpha}\left(\frac{B_2}{3B_0} - \frac{2B_3}{B_1}\right),$$

$$B_0 = \sum_{K=1}^{\infty}(-1)^{K+1}\left(K-\frac{1}{2}\right) \times \exp\left(-\pi\tan\alpha\left(K-\frac{1}{2}\right)^2\right),$$

$$B_1 = 1 + 2\sum_{K=1}^{\infty}e^{-\pi\tan\alpha K^2},$$

$$B_3 = \sum_{K=1}^{\infty}K^2e^{-\pi\tan\alpha K^2},$$

$$B_2 = \sum_{K=1}^{\infty}(-1)^{K+1}\left(K-\frac{1}{2}\right)^3 \times \exp\left(-\pi\tan\alpha\left(K-\frac{1}{2}\right)^2\right).$$

Formulas (22), (34), and (41) imply that, in the interval (0, 1) of the real axis, the charge density can be represented as

$$\rho_1(t) = \{A\phi_{(t)}^{\nu} + A^*(\phi_{(t)}^*)^{\nu}\}E_0\sin\alpha, \tag{44}$$

where

$$A = |A|e^{i(\pi/2 + \nu(\pi-2\alpha))}. \tag{45}$$

The coefficient $|A|$ is given by Eq. (23):

$$\frac{Z}{2\pi}\left\{1 + 2\pi i|A|\int_0^1 dt(1-2t) \times\left[(\phi(t)e^{i(\pi-2\alpha)})^{\nu} - ((\phi(t)e^{i(\pi-2\alpha)})^{\nu})^*\right]\right\} \tag{46}$$

$$= i|A|\int_0^1 dt\left[(\phi(t)e^{i(\pi-2\alpha)})^{\nu} - ((\phi(t)e^{i(\pi-2\alpha)})^{\nu})^*\right].$$

The component $\sigma_{\text{eff}}^{22}(\alpha)$ is defined by (24):

$$\sigma_{\text{eff}}^{22}(\alpha) = 4\pi i|A|\frac{\sigma_1\sigma_2}{\sigma_1-\sigma_2} \times\int_0^1 dt\left[(\phi(t)e^{i(\pi-2\alpha)})^{\nu} - ((\phi(t)e^{i(\pi-2\alpha)})^{\nu})^*\right]. \tag{47}$$

Consider the limit case as $\alpha \longrightarrow \pi/2$. From (41), we obtain

$$\phi(1-z) = \exp[\pi z(\tan\alpha - i)], \quad \phi(z) = \exp[\pi(1-z)(\tan\alpha - i)]. \tag{48}$$

The parameters ν, A, and Z are given by formulas (18), (36), and (45):

$$\nu = 1 - \frac{2\alpha}{\pi}, \quad A = |A|\exp\left(i\left(\frac{\pi}{2} + \pi\nu^2\right)\right), \quad Z = \frac{1}{2}, \quad \nu\tan\alpha = \frac{2}{\pi}. \tag{49}$$

Using (48) and (49), we can reduce the expression for the charge density to

$$\rho_1(t) = E_0 \sin\alpha\{2\pi\nu(1-t)e^{2(1-t)}\},$$
$$i\{(\phi(t)e^{i(\pi-2\alpha)})^\nu - ((\phi(t)e^{i(\pi-2\alpha)})^\nu)^*\} \quad (50)$$
$$= 2\pi\nu(1-t)e^{2(1-t)}.$$

The coefficient $|A|$ is defined by (46) and (50):

$$2\pi\nu|A| = \frac{Z}{2\pi}\left\{\int_0^1 dtt\left(\frac{3}{2}-t\right)e^{2t}\right\}^{-1} = \frac{4Z}{\pi(5+e^2)}. \quad (51)$$

Using (47), (50), and (51), we obtain the following expression for the component $\sigma_{\text{eff}}^{22}(\alpha)$ of the conductivity tensor as $\alpha \longrightarrow \pi/2$:

$$\sigma_{\text{eff}}^{22}(\alpha) = \frac{4\sigma_1\sigma_2}{\sigma_1+\sigma_2}\frac{e^2+1}{e^2+5} \approx 0.67713\sigma_1. \quad (52)$$

The component $\sigma_{\text{eff}}^{11}(\alpha)$ is reconstructed by the Dykhne relation

$$\sigma_{\text{eff}}^{11}(\alpha) = 0.49227\sigma_1, \quad \alpha \longrightarrow \pi/2. \quad (53)$$

Note that $\alpha = 0$ is a singular point. In particular, $F_2(\alpha)$ (see Eq. (31)) is a periodic odd function of α with period π such that $F_2(\alpha \longrightarrow +0) = 1/3$. Hence, an expansion of this function in trigonometric series contains infinitely many harmonics.

5. CONDUCTIVITY IN THE DOMAIN $Z \longrightarrow 1$

Consider the range of values of the parameter Z that are close to unity:

$$1 - Z \ll 1. \quad (54)$$

Suppose also that angle α satisfies the conditions

$$\alpha \gg 1 - Z, \quad \frac{\pi}{2} - \alpha \gg 1 - Z. \quad (55)$$

Under conditions (54) and (55), parameter ν is close to 1/2:

$$\nu = \frac{1}{2} - \delta, \quad \delta \ll 1. \quad (56)$$

From Eq. (18), we determine

$$\delta = \left(\frac{2(1-Z)}{\pi^2[1-(1-4\alpha/\pi)^2]}\right)^{1/2}. \quad (57)$$

Under conditions (56), the charge density ρ_1 in the zero approximation is given by (15):

$$\rho_1 \approx E_0 \sin\alpha \frac{\tilde{C}(\alpha)}{t^{1-2\delta}}, \quad (58)$$

where $\tilde{C}(\alpha)$ is a value of order unity, provided that

$$\delta \gg 1 - Z. \quad (59)$$

The component $\sigma_{\text{eff}}^{22}(\alpha)$ is defined by (24) and, under conditions (56) and (59), is equal to

$$\sigma_{\text{eff}}^{22}(\alpha) = 2\pi\frac{\sigma_1\sigma_2}{\sigma_1-\sigma_2}\frac{\tilde{C}(\alpha)}{\delta}. \quad (60)$$

When $\alpha = \pi/4$, we have $\tilde{C}(\pi(4)) = 1/\pi$. Thus, for $Z \longrightarrow 1$ and a wide range of angles α satisfying condition (55), the components $\sigma_{\text{eff}}^{22}(\alpha)$ and $\sigma_{\text{eff}}^{11}(\alpha)$ of the tensor of effective conductivity differ only by a numerical factor of order unity.

In the domain where $\alpha \longrightarrow 0$ and $Z \longrightarrow 1$, we determine both parameters μ and ν from Eqs. (18) and (20):

$$\mu = -\frac{\alpha}{\pi}, \quad 1 - 2\nu = \frac{1-Z}{2\alpha}\tan(\pi\nu). \quad (61)$$

Equation (61) implies that the parameter ν is a function of the ratio of two small parameters $\{\alpha, 1 - Z\}$ with the range of variation (0, 1/2). In this domain, the charge density experiences a strong variation and the ratio of the components $\sigma_{\text{eff}}^{22}(\alpha)$ and $\sigma_{\text{eff}}^{11}(\alpha)$ is also an essential function of the parameter $\alpha/(1 - Z)$.

6. CONCLUSIONS

We have investigated the conductivity and the charge and current distributions in a periodic two-component system composed of rhombs with an arbitrary vertex angle of 2α. Investigation of special cases has shown that one can hardly expect simple algebraic expressions for the components of the tensor of effective conductivity [7]. We have shown that, in the plane of parameters (α, ν), there are three lines on which an explicit expression is obtained for the charge density and the tensor of effective conductivity in terms of elliptic functions. At a point of general position on the plane $\{\alpha, \nu\}$, the charge density is a doubly periodic analytic function of parameter t with two branching points in the parallelogram of periods.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research and the Ministry of Education of the Russian Federation.

REFERENCES

1. A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1971)].
2. V. G. Marikhin, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 391 (2000) [JETP Lett. **71**, 271 (2000)].
3. Yu. N. Ovchinnikov and A. M. Dyugaev, Zh. Éksp. Teor. Fiz. **117**, 1013 (2000) [JETP **90**, 881 (2000)].
4. Yu. N. Ovchinnikov and I. A. Luk'yanchuk, Zh. Éksp. Teor. Fiz. **121**, 239 (2002) [JETP **94**, 203 (2002)].
5. L. G. Fel and I. V. Kaganov, J. Phys. A **36**, 5349 (2003).
6. M. A. Lavrent'ev and B. V. Shabat, *Methods of the Theory of Functions of a Complex Variable* (Gostekhizdat, Moscow, 1958; Springer, Berlin, 1959).
7. I. M. Khalatnikov and A. Yu. Kamenshchik, Zh. Éksp. Teor. Fiz. **118**, 1456 (2000) [JETP **91**, 1261 (2000)].

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 98, No. 1, 2004, pp. 24–30.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 1, 2004, pp. 30–37.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Broadening of a Two-Photon Resonance in a Zinc Atom in Collisions with CO_2, CO, and NO Molecules

P. A. Bokhan, Dm. E. Zakrevskiĭ, and N. V. Fateev*

Institute of Physics of Semiconductors, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia

**e-mail: fateev@isp.nsc.ru*

Received June 2, 2003

Abstract—The broadening of a two-photon resonance is studied experimentally at the $4s\ ^1S_0$–$6s\ ^3S_1$ transition in a zinc atom upon absorption of two waves with a small detuning from an intermediate state in collisions with CO_2, CO, and NO molecules. The measured absolute values of broadening cross sections greatly exceed gas-kinetic cross sections and are $(9.4 \pm 2.4, 6.5 \pm 1.6, \text{and } 3.9 \pm 1.0) \times 10^{-14}$ cm^2 for CO_2, CO, and NO, respectively. Anomalously large rate constants and cross sections obtained in experiments are explained by the efficient resonance quenching of the excited states of zinc atoms in collisions with molecules accompanied by transfer of the energy of excited atoms to vibrational-rotational degrees of freedom of molecules. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Collisions of electronically excited atoms with molecules have been extensively studied over a long period of time (see, for example, [1, 2]). These collisions are accompanied by a variety of processes such as quenching of excited states, ionization, and chemical reactions. The investigation of these processes is important for the understanding of the collision kinetics, the physics of plasmas and lasers, laser isotope separation, etc. The most popular method for measuring the absolute values of the rate constants of such processes is the detection of the time dependence of the intensity of fluorescence from excited levels as a function of the concentration of another gas. An alternative method is the detection of the widths of absorption lines of atoms. This method also gives information on spectral shifts of absorption lines caused by various physical processes. The first studies in this field were performed by classical absorption methods [3]. To exceed the Doppler broadening, the experiments were performed at high gas pressures. Later, the authors of [4] proposed a method based on the simultaneous absorption of two photons by an atom with the same energies and oppositely directed momenta. In this case, the shape of the absorption line is determined only by homogeneous broadening. The absence of Doppler broadening allows the experiments to be performed at comparatively low pressures (a few orders of magnitude lower than in the one-photon method), and the spectral resolution is determined only by a radiation source.

The broadening and shift of two-photon absorption resonances have been experimentally studied to date mainly for the Rydberg atoms of alkali metals [5–7]. This is explained by the fact that, to study two-photon absorption for other elements, an intense tunable UV radiation source is required.

In this work, we studied experimentally for the first time the broadening of the two-photon resonance at the $4s\ ^1S_0$–$6s\ ^3S_1$ transition in a zinc atom colliding with CO_2, CO, and NO molecules. This transition is of interest because it is involved in the process of excitation of zinc atoms upon photochemical separation of zinc isotopes [8].

2. EXPERIMENTAL

Experiments were performed on a setup for laser separation of zinc isotopes [8]. The scheme of the setup for excitation of zinc atoms is shown in Fig. 1. The atoms were excited in an interaction chamber representing a horizontally positioned quartz cylinder of diameter 3 cm and length 100 cm. Granulated zinc with residual impurities of other elements less than 10^{-4}% was placed at the bottom of the cylinder. The interaction region was heated with an external furnace with automatic temperature control, which was measured with a chromel–copel thermocouple with an accuracy of 0.5°C. After careful degassing in vacuum upon pumping by a magnetic-discharge pump, the gas under study was pumped through the interaction region at a rate of 0.5 l/s. The gas was pumped by a backing pump and its pressure was measured with a McLeod gauge. We used in our experiments CO_2 with residual impuri-

ties of O_2 and H_2O in amounts less than 1%, CO with impurities of O_2 and formic acid less than 0.5%, and NO with impurities of NO_2 and O_2 less than 0.5%. Control experiments were performed with helium and argon. The working temperature of the chamber was 275°C, and the calculated concentration of zinc atoms in the interaction region was 7×10^{12} at./cm^3.

Zinc atoms were excited to the $6s\,^3S_1$ state by two counterpropagating beams at the wavelengths $\lambda_1 = 0.307$ μm and $\lambda_2 = 0.303$ μm from two lasers described in [9] (see Fig. 2). A cw dye laser pumped by an argon laser was used as a master oscillator. Narrowband laser radiation with a linewidth lower than 5 MHz was amplified in a three-stage scheme of amplifiers pumped by a pulsed Cu laser. The amplified radiation was doubled in a nonlinear BBO crystal. The average power of each of the laser beams in the interaction chamber was approximately 1 W, the pulse duration was 10 ns, the linewidth was no more than 40 MHz, the pulse repetition rate was 12 kHz, and the beam diameter was 1 cm. The counterpropagating pulses were combined using a delay line in such a way that their power maxima in time coincided in the center of the region of interaction with zinc vapors. Both waves had linear and mutually orthogonal polarizations because the two-photon transition under study is forbidden for parallel polarizations [10]. The beam frequencies were detuned by 9 GHz from the exact resonance with the $4p\,^3P_1^0$ level. This circumstance and the closeness of the energies of the two counterpropagating photons reduces the Doppler broadening, thereby making it possible to resolve the isotopic structure. After excitation of zinc atoms to the upper $6s\,^3S_1$ state, this state decays via direct and cascade optical transitions to long-lived $4p\,^3P_j^0$ levels with a lifetime of 10 μs. Two-photon absorption was detected by the luminescence signal at the wavelengths 1.3 and 0.48 μm with a germanium photodiode *9* (Fig. 1) and a selective voltmeter *11* at a 130-Hz frequency of modulation of one of the beams with the help of a mechanical chopper. This detection technique provided a better signal-to-noise ratio than direct detection of the UV radiation absorption with photodiode *10* and selective voltmeter *12*. The output signals of the selective voltmeter were detected by a computer by continuously varying the wavelength of one of the beams, the wavelength of the other beam being constant. Each of the experimental points corresponded to a measurement time of 10 ms (averaging was performed over about one hundred laser pulses). The wavelength of tunable radiation was detected simultaneously with a wavemeter. The absolute error of the UV radiation frequency measurement was 400 MHz, and the relative error during the time of measurements of 1 min was smaller than 20 MHz.

3. SHAPE OF THE TWO-PHOTON RESONANCE LINE IN A ZINC ATOM

To measure the width of the two-photon $4s\,^1S_0$–$6s\,^3S_1$ line resonance line (we will denote the states by *g* and *f*, respectively), it is necessary to determine the shape of the absorption line for counterpropagating waves with unequal frequencies and a small detuning δ of radiation frequencies from the intermediate $4p\,^3P_1^0$ state, which we will denote by *n*. For the fixed, ν_1, and tunable, ν_2, frequencies of exciting radiation and the frequencies ν_{fn} and ν_{ng} of transitions in the atom, the value of δ is determined by the condition

$$\delta = \nu_1 - \nu_{ng} \approx \nu_{fn} - \nu_2 > 0. \quad (1)$$

The requirement of the resonance leads to the condition

$$\nu_1 + \nu_2 + (\nu_2 - \nu_1)\frac{v}{c} = \nu_{fg}, \quad (2)$$

where v is the projection of the velocity of an atom on the direction of radiation propagation. According to [11], the rate of the two-photon $g \longrightarrow f$ transition in the field of two counterpropagating waves is determined by the expression

$$W_{gf}(\nu_2) = \frac{\Omega_1^2\Omega_2^2}{4\delta^2} \times \gamma\left\{\left[\nu_0 + (\nu_1 - \nu_2)\frac{v}{c} - \nu_1 - \nu_2 - \Delta\right]^2 + \left(\frac{\gamma}{2}\right)^2\right\}^{-1}, \quad (3)$$

where Ω_1 and Ω_2 are the Rabi frequencies for the one-photon $g \longrightarrow n$ and $n \longrightarrow f$ transitions, γ is the line-

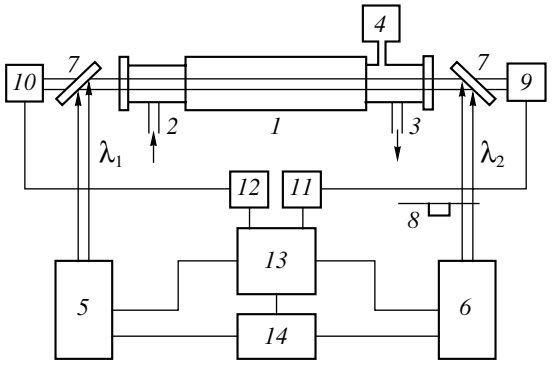

Fig. 1. Scheme of the experimental setup: (*1*) interaction chamber; (*2*) gas inlet; (*3*) gas evacuation; (*4*) pressure gauge; (*5*, *6*) laser systems; (*7*) mirrors; (*8*) mechanical chopper; (*9*, *10*) photodetectors; (*11*, *12*) selective voltmeters; (*13*) computer; (*14*) wavemeter.

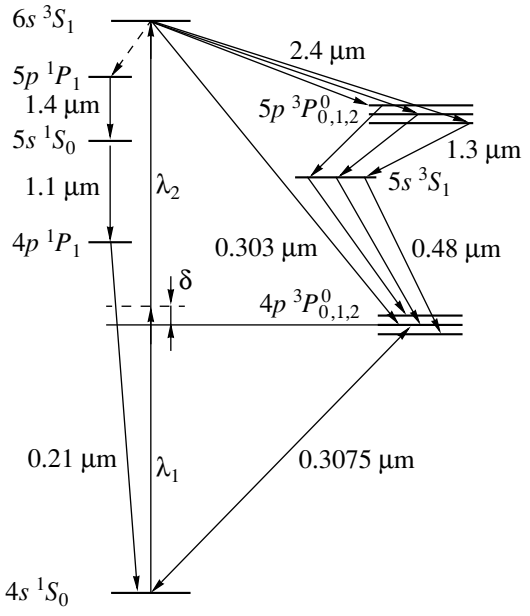

Fig. 2. Energy level diagram of a zinc atom.

width of the two-photon transition, and Δ is the resonance frequency shift. The linewidth γ is determined by the radiative and collision broadenings:

$$\gamma = \gamma_0 + kn/\pi, \tag{4}$$

where k is the rate constant of collision broadening, n is the concentration of molecules, and γ_0 is the transition linewidth in the absence of collisions. The averaging of expression (3) over the Maxwell distribution of atoms over velocities gives a profile with a diminished Doppler width $(\nu_1 - \nu_2)u/c$, where u is the most probable velocity of atoms in gas. This residual Doppler width of the two-photon resonance is equal to 16 MHz, which is lower than other widths (widths of the spontaneous and stimulated transitions from the upper level and laser linewidths), which are independent of the atom velocity. As a result, we obtain the dependence of the probability of two-photon absorption on frequency

$$W_{gf}(\nu_2) = \frac{\Omega_1^2\Omega_2^2}{4\delta^2}\gamma\left\{[\nu_0 - \nu_1 - \nu_2 - \Delta]^2 + \left(\frac{\gamma}{2}\right)^2\right\}^{-1}. \tag{5}$$

The profile of the absorption line proves to be shifted with respect to the center by the value Δ, which is determined by the total contribution of the collision shift Δ_{col} and the field shift caused by the field at the frequency ν_2:

$$\Delta = \Delta_{col} + \Omega_2^2/4\delta. \tag{6}$$

The oscillator strengths of the working transitions in a zinc atom $f_{ng} \approx 1.4 \times 10^{-4}$ and $f_{fn} \approx 1.6 \times 10^{-2}$ [12] differ by two orders of magnitude. At equal intensities of exciting radiation, the field shift at the *f–n* transition leads to an increase in the transition frequency approximately by 200 MHz for a pulse power of I = 10 kW/cm^2 and is comparable to the experimental width of the resonance.

The number of atoms excited to the 6s $^3S_1$ state by a 10-ns pulse is

$$N_f(\nu_2) = \frac{\Omega_1^2\Omega_2^2 N_1 V\gamma\tau}{4\delta^2[(\Delta\nu)^2 + (\gamma/2)^2]}, \tag{7}$$
$$\Delta\nu = \nu_0 - \nu_1 - \nu_2 - \Delta,$$

where N_1 is the concentration of atoms in the ground state and V is the volume of the region of interaction between laser radiation and atoms. The absorbed radiation power is determined by the number of excited atoms and monotonically decreases with increasing γ. Estimates made from (7) show that approximately 10% of the atoms are excited per pulse in the interaction region. This agrees with the calculation of the number of photons absorbed when 1% of a radiation is absorbed. Excited atoms can relax over several channels, with a total lifetime equal to $1/2\pi\gamma$. In this case, a detected luminescence signal is described by the expression

$$S(\nu_2) = CN_f(\gamma_1 + k_1 n)/\gamma,$$

where γ_1 is the decay rate of the level in the absence of quenching collisions with molecules and C is a constant. Therefore, a signal detected in experiments is described by the expression

$$S(\nu_2) = \frac{C\Omega_1^2\Omega_2^2 N_1 V\tau(\gamma_1 + k_1 n)}{4\delta^2[(\Delta\nu)^2 + (\gamma/2)^2]}. \tag{8}$$

Note that the value of $S(\nu_2)$ increases with increasing relaxation rate of the f level due to the population of low-lying levels at which relaxation occurs and from which spontaneous or stimulated radiation is detected. Figure 3 shows the experimental dependence of the luminescence intensity of the ^{64}Zn isotope on frequency ν_2 (curve *1*) obtained at temperature T = 548 K, pressure P_{CO_2} = 1 Torr, and δ = 9 GHz. Unlike (8), the two-photon absorption band is asymmetric, which we observed for all the spectra studied. The asymmetric shape of the absorption band is caused by fluctuations of radiation intensity from pulse to pulse (up to 10%). Each of the measured points was obtained by averaging the signal over more than 100 pulses. In this case, the averaging of $S(\nu_2)$ over the interval of variation in the radiation intensity gives, in the first approximation, the

following expression for the experimentally detected signal:

$$S_{\exp}(\nu_2) = \frac{A}{(\Delta\nu - a)^2 + (\gamma/2)^2} \times\left(1 + \frac{B\Delta\nu}{(\Delta\nu - a)^2 + (\gamma/2)^2}\right), \quad (9)$$

where the parameters A, B, and a, which are independent of ν_2, are selected by fitting the experimental data. Figure 3 shows the dependences (curves *2* and *3*) corresponding to the approximation of experimental data by expression (9) and Lorentzian profile (8) by the least-squares method. In the case of approximation by expression (9), which describes an asymmetric contour, the error of measurements of γ is somewhat lower than that for the Lorentzian contour, while the absolute value is larger. The processing of several experimental curves by both these methods gives values of γ which differ from each other no more than by 15%, which is within the accuracy of other measurements. In this study, experimental data were approximated by Lorentzian profiles, because processing based on (9) would be too cumbersome.

Two-photon absorption of two counterpropagating waves, which were slightly detuned from an intermediate state, was first observed in [13] at the 3*S*–4*D* transition in a sodium atom; however, the lineshape was not studied. The two-photon absorption line observed upon pulsed excitation can have a broad symmetric pedestal due to the short duration of radiation (a few nanoseconds) and a high pulse power (a few tens of MW/cm^2). This was experimentally demonstrated for a helium atom at the 2*S*–3*S* transition [14]. In our case, such a pedestal was not observed because of a significantly lower pulse power.

4. RESULTS

Experiments with different gases were performed at the same temperature $T = 548$ K and $\delta = 9$ GHz. Figure 4 shows the luminescence spectra obtained at different CO_2 pressures in the interaction chamber. Because natural zinc consists mainly of even isotopes, 64 (48.6%), 66 (27.9%), and 68 (18.8%), they were detected experimentally. The isotope shifts at this two-photon transition were $\Delta\nu_{66,64} = 860$ MHz and $\Delta\nu_{68,66} = 930$ MHz. These values differ somewhat from the values $\Delta\nu_{66,64} = 689$ MHz and $\Delta\nu_{68,66} = 677$ MHz obtained earlier at the one-photon $4s\ ^1S_0$–$4p\ ^3P_1$ transition [15]. One can see from Fig. 4 that the spectrum broadens with increasing gas pressure in the interaction region. The dashed curve in Fig. 4 shows the approximation of the spectrum by three Lorentzian profiles for $P_{CO_2} = 7$ Torr. Analysis of the extrapolation obtained in this way showed that width γ was the same within 5% for the three isotopes for all absorption spectra detected.

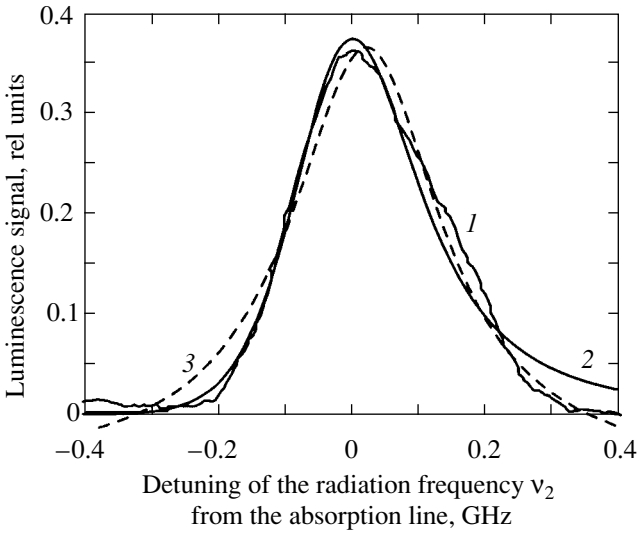

Fig. 3. Two-photon absorption spectrum of the ^{64}Zn isotope. Curve *1* is experiment at $T = 548$ K, $P_{CO_2} = 1$ Torr, and $\delta = 9$ GHz; curves *2* and *3* are the approximations of the experimental spectrum by the least-squares method by expressions (9) and (8), respectively.

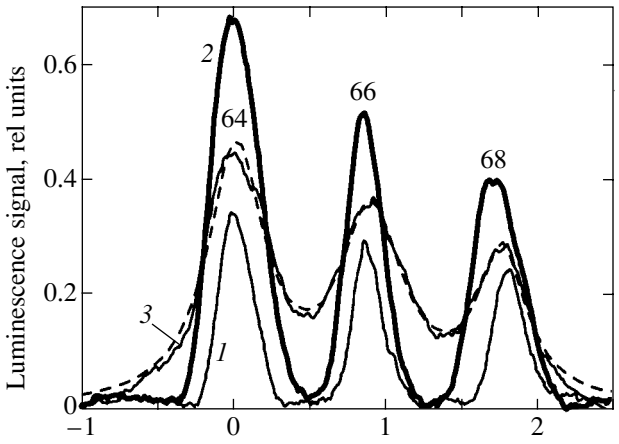

Fig. 4. Two-photon absorption spectra at the $4s\ ^1S_0$–$6s\ ^3S_1$ transition in a zinc atom at $P_{CO_2} = 0.3$ (*1*), 1 (*2*), and 7 Torr (*3*); The dashed curve is the approximation of curve *3* by Lorentzian profiles. The isotope numbers are shown over the peaks.

Figure 5 shows the two-photon absorption linewidths measured as functions of the CO_2, CO, and NO gas pressures in the interaction region. These dependences are well approximated by straight lines. We determined from the slopes of the straight lines the broadening of the two-photon resonance in a zinc atom

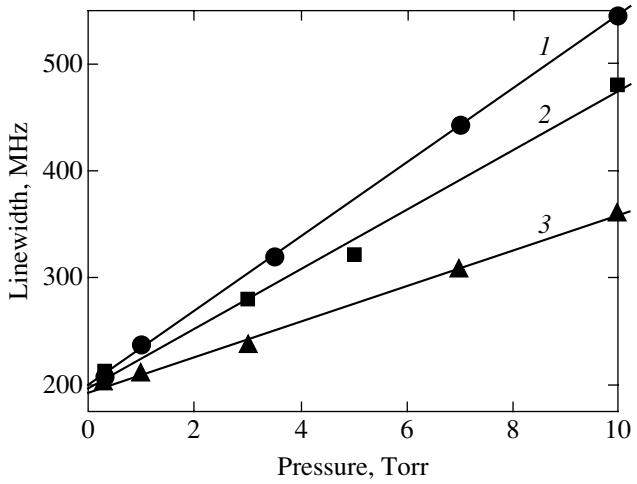

Fig. 5. Dependence of the two-photon absorption linewidth on the gas pressure: (*1*) CO_2; (*2*) CO; (*3*) NO.

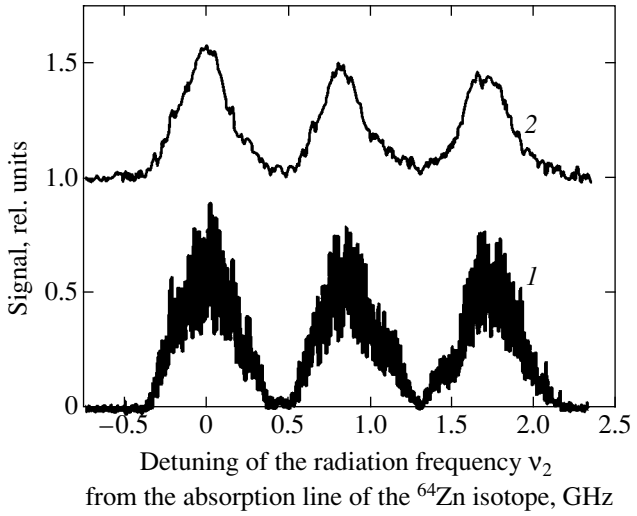

Fig. 6. Two-photon absorption spectra of a zinc atom at a pressure of P_{CO_2} = 3.5 Torr measured (*1*) from the absorbed UV radiation power and (*2*) from the luminescence signal.

in the presence of different gases. The results are presented in the table. The absolute values of the rate constant of the broadening process were determined by expression (4). The collision broadening cross sections were calculated by the expression $\sigma = k/\bar{v}$, where $\bar{v} = \sqrt{\bar{v}_1^2 + \bar{v}_2^2}$ is the average value of the relative velocity of two different particles in a gas [16]. The approximation by a straight line in Fig. 5 in the region of zero pressure gives the initial broadening γ_0 = 200 MHz. This value is determined by the Stark broadening and the decay of the $6s\,^3S_1$ level for a time much shorter than the spontaneous decay time, equal to 30 ns [17]. This is caused by superradiance to lower states due to large cross sections for optical transitions $\sigma_{6s-5p} \approx 4 \times 10^{-11}$ cm^2 and $\sigma_{6s-4p} \approx 3 \times 10^{-13}$ cm^2. In this case, the condition for the appearance of superradiance over length L is readily fulfilled ($\sigma N_f L > 10$), which strongly shortens the upper-level lifetime.

Experimental results

Molecule	Broadening γ/P, MHz/Torr	Rate constant k, 10^{-9} cm^3/s	Cross section σ, 10^{-14} cm^2
CO_2	35 ± 9	6.2 ± 1.6	9.4 ± 2.4
CO	28 ± 7	5.0 ± 1.3	6.5 ± 1.6
NO	16 ± 4	2.9 ± 0.7	3.9 ± 1.0

Figure 6 shows the two-photon absorption spectra obtained by direct detection of the absorbed average radiation power and from the luminescence signal. The approximation of these spectra by Lorentzian profiles at different pressures gives coincident linewidths. However, the signal-to-noise ratio in the case of direct detection of absorption is substantially lower than that under recording luminescence, as illustrated in Fig. 6. The fraction of absorbed radiation power is usually a few percent. No pressure shifts of the absorption bands were observed within the absolute accuracy of measurement of the radiation frequency, equal to 400 MHz under any experimental conditions studied.

The amplitude of the absorbed power resonance monotonically decreases with increasing molecular pressure in the interaction region due to an increase in the transition linewidth. A luminescence signal detected at the resonance center for the ^{64}Zn isotope behaves in a more complicated way, as shown in Fig. 7. For CO_2 and CO molecules, this signal first increases, achieving a maximum, and then decreases with increasing pressure. This experimental fact is described by expression (8) and is caused by an increase in the rate of decay of the excited level to the intermediate $5p\,^3P^0_{0.1.2}$ state, from which emission is recorded with a measuring system. This makes it possible to measure the rate constants independently by a different method. The processing of the experimental dependences of the luminescence intensity by expression (8) using the least-squares method gives broadening values coincident with those presented in the table. This confirms the validity of the rate constants and cross sections obtained from the two-photon resonance broadening. The maximum of curve *3* for NO molecules is located at higher pressures due to a lower value of the two-photon resonance broadening. Note in conclusion that the two-photon resonance broadening in collisions with inert gas atoms at pressures of up to 5 Torr does not exceed the initial value of 200 MHz.

5. DISCUSSION

The cross sections for collisions of different electronically excited atoms with molecules are usually on the order of 10^{-15} cm^2 [1, 2]. As follows from the results presented in the table, the experimental cross sections for collisions of excited zinc atoms ($6s\ ^3S_1$) with CO_2, CO, and NO molecules are much greater and substantially exceed gas-kinetic cross sections. Such anomalously high cross sections can be explained by the resonance process, in which the excitation energy is transferred to the vibrational state:

$$\mathrm{Zn}(6s\ ^3S_1) + \mathrm{M} \longrightarrow \mathrm{Zn}(5p\ ^3P_j^0) + \mathrm{M}^* + \Delta E. \quad (10)$$

It is known [18] that the dependence of the probability of such processes on the energy gap ΔE between the initial and final states has a distinct resonance nature. An almost exact resonance corresponds to rather large cross sections on the order of 10^{-13} cm^2. Moderate cross sections on the order of 10^{-16} cm^2 correspond to detuning from the resonance by a few kT, and when the detuning exceeds a few tenths of an electronvolt, the cross section becomes negligibly small. Figure 8 illustrates the correspondence between the transition energy in a zinc atom and vibrational transition energies in CO_2, CO, and NO molecules. The energy level diagram of a zinc atom is taken from [19], the vibrational-rotational energies of CO_2 from [20], and those for NO and CO from [21, 22]. It follows from Fig. 8 that the values of ΔE for all the three molecules studied in the paper are lower than the kinetic energy kT of colliding particles. Such a situation was studied in [23], where the cross section $\sigma \approx 3 \times 10^{-14}$ cm^2 was obtained for collision resonance processes of energy transfer from Na(ns) (n = 5–11) atoms to CH_4 and CD_4 molecules. In such processes, the fundamental molecular vibrations are excited. The cross section for collisions of zinc atoms with CO_2 molecules is three times larger than the above value, although the molecules are excited into less intense (by three to four orders of magnitude) overtone vibrations. The excited zinc atoms also can decay at the $6s\ ^3S_1$–$5p\ ^1P_1$ transition due to collisions with molecules, emitting then luminescence at 1.4 and 1.1 μm (see Fig. 2). The $6s\ ^3S_1$–$5p\ ^1P_1$ transition energy is in resonance with intense vibrational-rotational ν_3 absorption bands of CO_2 molecules ($\Delta E = 2 \times 10^{-2}$ eV) and the 0–1 transition in CO ($\Delta E = 4 \times 10^{-2}$ eV), these value of ΔE being smaller than the relative kinetic energy of colliding particles ($kT = 5 \times 10^{-2}$ eV). For NO molecules, $\Delta E = 8 \times 10^{-2}$ eV. The $6s\ ^3S_1$–$5p\ ^1P_1$ transition, being an intersystem crossing transition, is weakly allowed and is shown in Fig. 2 by the dashed arrow. However, the collision transfer of the energy of this transition to the vibrational energy of molecules can be rather efficient.

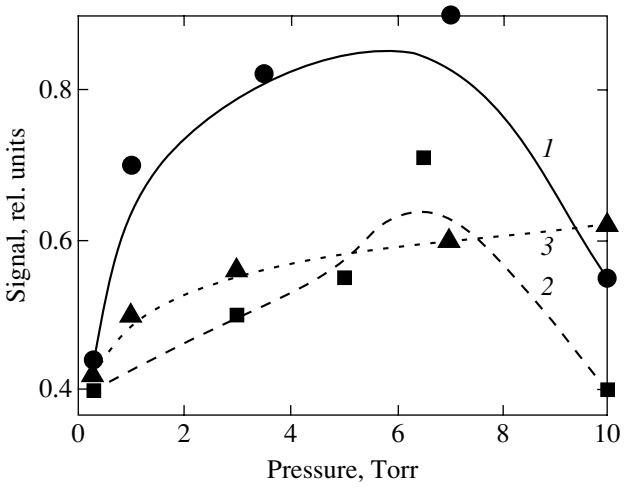

Fig. 7. Dependence of the luminescence signal at the absorption line center of the ^{64}Zn isotope on the molecular gas pressure in the interaction region: (*1*) CO_2; (*2*) CO; (*3*) NO.

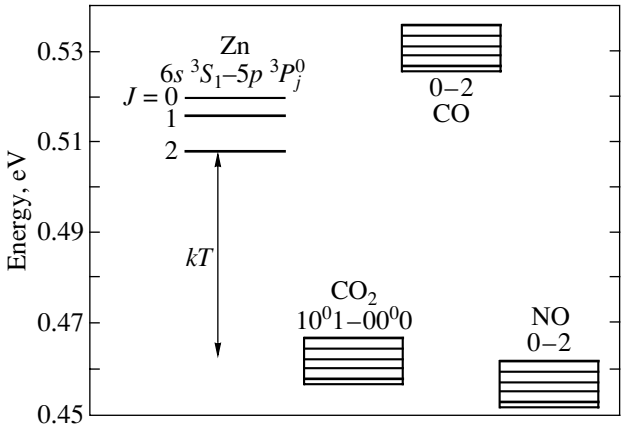

Fig. 8. Energy level diagram of a zinc atom and the transition energies of molecules studied in the paper. The arrow shows the average energy of particles in gas.

6. CONCLUSIONS

We have studied the decay of the upper $6s\ ^3S_1$ state of a zinc atom caused by collisions with CO_2, CO, and NO molecules by the method of two-photon laser spectroscopy in counterpropagating waves with close frequencies. Unlike the popular time-resolved method of detection of luminescence at a fixed transition, the study of the broadening of the absorption line gives complete information on the decay rate upon collisions with other particles. The absolute cross sections measured in the paper greatly exceed gas-kinetic cross sec-

tions. This can be explained by the resonance electronic energy transfer to the vibrational degrees of freedom.

Such processes were earlier experimentally studied only one-electron atoms of alkali metals.

ACKNOWLEDGMENTS

The authors thank A.Yu. Stepanov for his great help in constructing the experimental setup and V.A. Kim for his help in the experiments.

REFERENCES

1. H. Okabe, *Photochemistry of Small Molecules* (Wiley, New York, 1978; Mir, Moscow, 1981).
2. B. M. Smirnov, *Excited Atoms* (Énergoizdat, Moscow, 1982).
3. M. A. Mazing and P. D. Serapinas, Zh. Éksp. Teor. Fiz. **60**, 541 (1971) [Sov. Phys. JETP **33**, 294 (1971)].
4. L. S. Vasilenko, V. P. Chebotaev, and A. V. Shishaev, Pis'ma Zh. Éksp. Teor. Fiz. **12**, 161 (1970) [JETP Lett. **12**, 113 (1970)].
5. K. H. Webe and K. Niiemax, Z. Phys. A **307**, 13 (1982).
6. D. C. Thompson, E. Weinberger, G.-X. Xu, *et al.*, Phys. Rev. A **35**, 690 (1987).
7. D. C. Tompson, E. Kammermayer, B. P. Stoicheeff, *et al.*, Phys. Rev. A **36**, 2134 (1987).
8. P. A. Bokhan, V. V. Buchanov, D. É. Zakrevskiĭ, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 705 (2000) [JETP Lett. **71**, 483 (2000)].
9. P. A. Bokhan, D. É. Zakrevskiĭ, S. A. Kochubeĭ, *et al.*, Kvantovaya Élektron. (Moscow) **31**, 132 (2001).
10. P. A. Bokhan, V. V. Buchanov, D. É. Zakrevskiĭ, *et al.*, Kvantovaya Élektron. (Moscow) **32**, 570 (2002).
11. V. S. Letokhov and V. P. Chebotaev, *Nonlinear Laser Spectroscopy*, 2nd ed. (Nauka, Moscow, 1975; Springer, Berlin, 1977).
12. *Physical Quantities: A Handbook*, Ed. by I. S. Grigor'ev and E. Z. Meĭlikhov (Énergoatomizdat, Moscow, 1991).
13. J. E. Bjorkholm and P. F. Liao, Phys. Rev. Lett. **33**, 128 (1974).
14. N. V. Vitanov, B. W. Shore, L. Yatsenko, *et al.*, Opt. Commun. **199**, 117 (2001).
15. P. Campbell, J. Billowes, and I. S. Grant, J. Phys. B **30**, 2351 (1997).
16. E. W. McDaniel, *Collision Phenomena in Ionized Gases* (Wiley, New York, 1964; Mir, Moscow, 1967).
17. A. L. Osherovich, Ya. F. Verolaĭnen, and V. I. Privalov, Opt. Spektrosk. **46**, 1092 (1979) [Opt. Spectrosc. **46**, 617 (1979)].
18. H. S. W. Massey and E. H. S. Burhop, *Electronic and Ionic Impact Phenomena* (Clarendon Press, Oxford, 1952; Inostrannaya Literatura, Moscow, 1958).
19. N. M. Brown and S. G. Tilfold, J. Opt. Soc. Am. **65**, 1404 (1975).
20. G. Herzberg, *Infrared and Raman Spectra of Polyatomic Molecules* (Van Nostrand, New York, 1945; Inostrannaya Literatura, Moscow, 1949).
21. E. F. Horn and F. P. Dickey, J. Chem. Phys. **41**, 1614 (1964).
22. K. S. Krasnov, V. S. Timoshinin, T. G. Danilova, and S. V. Khandozhko, *Molecular Constants of Inorganic Compounds*, Ed. by K. S. Krasnov (Khimiya, Leningrad, 1979), p. 15.
23. T. F. Gallagher, G. A. Ruff, and K. A. Safinya, Phys. Rev. A **22**, 843 (1980).

Translated by M. Sapozhnikov

Journal of Experimental and Theoretical Physics, Vol. 98, No. 1, 2004, pp. 31–38.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 1, 2004, pp. 38–47.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Two-Photon Interference in the Presence of Absorption

S. P. Kulik[a,*], G. A. Maslennikov[a], S. P. Merkulova[b], A. N. Penin[a], L. K. Radchenko[a], and V. N. Krasheninnikov[b]

[a]*Moscow State University, Vorob'evy gory, Moscow, 119992 Russia*
[b]*Institute of Spectroscopy, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia*
**e-mail: postmast@qopt.phys.msu.su*

Received July 17, 2003

Abstract—Experiments on two-photon interference are discussed in the case when there is absorption of all the modes participating in the process of spontaneous parametric down-conversion (SPDC) of light. The objects of investigation are 10- to 80-Å-thick ultrathin gold films deposited on fused-silica substrates. Conditions are determined under which the effect of absorption of the signal and pump waves on the interference pattern is small. It is shown that, under these conditions, the visibility of the interference pattern and the shape of the frequency–angular spectrum at the signal frequency are determined by the optical parameters of the medium at the idler frequency, which belongs to the near-infrared region. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

For over two decades, two-photon interference [1–3] has attracted the attention of physicists. This interest is primarily associated with the interpretation of a number of experiments based on two-photon interference, as well as with the fact that two-photon (nonclassical) states of light are relatively easy to obtain. The most efficient source of such states is the spontaneous parametric down-conversion (SPDC) of light. Recently, various aspects of two-photon interference have been investigated in the context of the physics of quantum information: various fields of quantum information, such as generation of entangled states, quantum cryptography, quantum teleportation, etc. [4], intensively use the accumulated experience in the preparation, transformation, and measurement of two-photon light. This experience seems to be useful for developing quantum communication devices that employ nonclassical states of light as information carriers. At the same time, another property of two-photon interference, which may be useful in spectroscopy, has not received due attention. The point is that the interference pattern itself bears information about the properties of the medium (or several media) in which the generation and transformation of two-photon light occurs. Hence, one can solve the inverse problem; namely, one can recover the properties of the scattering and/or transforming medium from the interferograms of two-photon interference, as it is done in Raman or polariton spectroscopy. The latter method is the limiting case of the SPDC when the frequency of one of the waves falls within the range of lattice oscillations of a nonlinear crystal [5]. In this sense, the interferometry of spontaneous parametric down-conversion is a generalization of the method of polariton spectroscopy to nonlinear media, where $\chi^{(2)} = 0$.

The line shape of two-photon SPDC in a separate layer was considered in [6, 7] when the interference phenomena associated with the reflection and absorption of all the waves participating in the process were taken into account. The spectroscopic aspects of two-photon interference were discussed in the literature in the context of nonlinear diffraction [8, 9]. In [10], a method of diagnosing quasiregular domain structures by the frequency–angular spectra of SPDC was considered. In [11–14], the authors analyzed the capabilities of two-photon interferometry as a method that makes it possible to evaluate the optical parameters of substances placed in a nonlinear interferometer.

In the present paper, we discuss the application of the method of two-photon interference to the study of thin metal films deposited on fused-silica substrates.

2. TWO-PHOTON INTERFERENCE IN THE MACH–ZEHNDER SCHEME

2.1. Mach–Zehnder Nonlinear Interferometer with Several Layers

Consider a system of $n + 2 = x$ plane layers (Fig. 1). The first and the last layers have a nonzero value of quadratic susceptibility $\chi^{(2)} \equiv \chi$, while in the intermediate layers, this parameter is equal to zero and the layers differ only in their permittivity ε_q, where q is the layer number. In the literature, such a system was called the Mach–Zehnder nonlinear interferometer (MZNI) [15]. This term reflects the fact that a laser beam propagating across the layers induces nonlinear polarization in the first and the last layers due to χ; these layers are analogous to beam-splitters that divide/mix spatial–frequency modes. The optical fields with new frequencies generated by the nonlinear process have different phase delays while propagating across intermediate layers

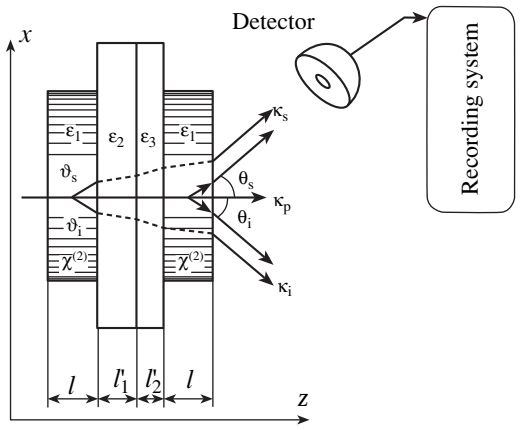

Fig. 1. Scheme of the Mach–Zehnder nonlinear interferometer with two nonlinear crystals.

($q = 1, ..., n - 1$) and, hence, may interfere at the output of the system if they are coherent.[1] The phase delays of the field components are determined by the dispersion of dielectric permittivity $\varepsilon_l(\omega)$. The oscillating behavior of the intensity of the generated fields as a function of a certain parameter of the system, for example, the optical thickness of intermediate (linear) layers, has been a subject of study by the method of nonlinear interferometry [12, 16]. In our case, a biphoton field is generated in the first and the second nonlinear crystals during the SPDC [17]. Recall that, during the SPDC, a photon (p) of the laser pump spontaneously decays into a pair of photons, the so-called signal (s) and idler (i) photons. In the stationary case, the photon frequencies are related by the energy conservation law,

$$\omega_p = \omega_s + \omega_i, \tag{1}$$

while the propagation direction of the generated wave is determined by the dispersion law of the medium,

$$k = \frac{\omega}{c}\sqrt{\varepsilon(\omega)} \tag{2}$$

via the phase-matching condition

$$\mathbf{k}_p + \mathbf{k}_s + \mathbf{k}_i + \mathbf{\Delta}. \tag{3}$$

Here, $\mathbf{k}_j$ (j = p, s, i) are the wave vectors and $\mathbf{\Delta}$ is the wave mismatch associated with the dimensions of nonlinear crystals. Since the layers are assumed to be infinite in the transverse direction, we have $\Delta_\perp \longrightarrow 0$, where $\Delta_\perp = k_p^x - k_s^x - k_i^x$. When the pump wave propagates along the z axis, we have $k_p = k_p^z$. As a result, we obtain a strict relation between the scattering directions of the signal and idler photons: $k_s^x = -k_i^x$ or $k_s \sin\vartheta_s = -k_i \sin\vartheta_i$, where ϑ_s and ϑ_i are the angles between the z axis and the scattering directions of the signal and idler photons in the crystals, respectively. In the low-absorption approximation, which is valid in the transparency regions of the crystals, the dispersion relation (2) contains the real parts of the wave vector $k_j \equiv 2\pi n_j/\omega_j$ (j = s, p, i) and the dielectric permittivity.

Formally, the intensity of the SPDC as a function of frequency and the scattering angle (the line shape) is proportional to the squared modulus of the sum of amplitudes of biphoton fields emitted from different macroscopic regions [18]. In our case, there are two such regions, which are nonlinear crystals in the MZNI:

$$I_s(\omega_s, \theta_s) \propto \left|\sum_{m=1}^{2} f_m\right|^2 = |f_1|^2 + |f_2|^2 + 2\mathrm{Re}(f_1 f_2^*). \tag{4}$$

The amplitudes f_m determine the wave function of the biphoton filed. They depend on the intensity of the laser pump field, the quadratic susceptibility of a crystal, the frequencies of interacting fields, etc. The third term in (4) describes the periodic modulation of the scattering intensity as a function of the relative phase of the two amplitudes. In stationary experimental conditions, two-photon interference, or interference of biphotons in the second order in the field, manifests itself as alternating maxima and minima in the intensity of the frequency–angular spectra of the SPDC; the relative phase in (4) depends on the variation of the direction and/or frequency of observation [11].

If the crystals are transparent at all three frequencies ω_p, ω_s, and ω_i, then the intensity of the observable (signal) wave as a function of frequency and scattering angle is given by

$$I_s(\theta_s, \omega_s) \propto \left[\frac{\sin(\Delta l/2)}{\Delta l/2}\cos\left\{\frac{1}{2}\left(\Delta L + \sum_{q=1}^{n}\Delta_q' l_q'\right)\right\}\right]^2, \tag{5}$$

where $\Delta' = k_p' - k_s' - k_i'$ is the z component of the wave mismatch in the intermediate media. The first coefficient in (5) describes the frequency–angular line shape of the spontaneous parametric down-conversion in a plane nonlinear layer [17]. The second coefficient is due to the interference between the signal fields generated in the extreme layers; this coefficient is responsible for the modulation of the line shape due to the contributions of the phase mismatches $\delta' = \Delta' l'$ in the intermediate materials. The explicit form of the wave mismatches δ and δ' as functions of the parameters observed in the experiment (the wavelength λ_s and the angle θ_s) is given in [12].

[1] The coherence of the components of interfering fields is guaranteed by common laser pumping, which is assumed to be classical and fixed.

Formula (5) is a particular case of the expression obtained in [19] for the MZNI containing several transparent linear media. This formula is obtained when one takes into account the contribution of the dispersion of all intermediate layers (without taking into account reflections) to the propagators of the signal and idler modes, as well as of the pump mode. Note that, in the real interferometer schemes used in the experiments in [12–14], the observed line shape of down-conversion is given precisely by (5) because there always exist air gaps between nonlinear crystals and the dispersive substance, so that the whole scheme actually consists of five, rather than three, layers.

2.2. Taking into Account Losses at the Idler Frequency

From the viewpoint of interpreting the experimental results, the following case is of interest. In the scheme shown in Fig. 1, one detects a signal wave, while the idler (nonobservable) modes experience losses due to, for example, absorption or reflection. This case was first considered by Mandel and colleagues in [20]. In the scheme suggested, they succeeded in spatially separating the signal and idler modes generated in different nonlinear crystals. In this configuration, it is convenient to control the transmission in the idler mode by inserting filters with different optical densities. The effect observed in this case was called induced coherence because the visibility of the interference pattern observed in a signal mode depends on the transmission coefficient of the filter inserted into the idler mode. In [11–14], it was pointed out that the induced coherence can find application in spectroscopy when the idler modes fall within the infrared region of the spectrum, while the signal mode is detected in the visible region.

The physical scheme considered in the present paper (Fig. 1) does not essentially differ from the scheme proposed in [20]. However, from the experimental point of view, the MZNI scheme considered here is more convenient because the optical-path difference between the signal and idler modes is maintained constant automatically [11]. In this scheme, the line shape of the signal wave for a finite amplitude of the transmission coefficient τ_i of a certain intermediate layer at idler frequencies is determined by the following expression, which takes losses into account:

$$I_s(\omega_s, \theta_s) \propto \frac{1}{2}\left[\frac{\sin(\Delta l/2)}{\Delta l/2}\right]^2 \times\left\{1 + |\tau_i|\cos\left(\Delta l + \sum_{q=1}^{p}\Delta_q' l_q'\right)\right\}. \quad (6)$$

The transmission coefficient τ_i relates the annihilation operators of photons in idler modes after the first and second nonlinear crystals [21]. Note that the idler field generated in the first crystal has a characteristic spectral luminosity on the order of $10^{-8} \ll 1$ photons per mode and does not influence the SPDC process in the second crystal (a spontaneous regime); this yields the relation

$$\hat{a}_{i2}(\omega_i) = \tau_i \hat{a}_{i1}(\omega_i) + r_i^{\text{eff}} \hat{a}_{\text{vac}}(\omega_i). \quad (7)$$

The second term in (7) is attributed to the unitarity of the transformation: the operator $\hat{a}_{\text{vac}}$ describes the vacuum field that is admixed to the idler mode with the weight r_i^{eff} for $|\tau_i|^2 < 1$; in this case,

$$|\tau_i|^2 + |r_i^{\text{eff}}|^2 = 1.$$

The coefficient r_i^{eff} describes losses due to reflection and absorption.

According to the scheme shown in Fig. 1, for identical nonlinear crystals ($\varepsilon_1 \equiv \varepsilon_4$ and $\chi_1 \equiv \chi_4$), the idler modes $\mathbf{k}_{i1}$ and $\mathbf{k}_{i2}$ are degenerate; therefore, formally, formula (7) makes it possible to take losses into account. A detailed analysis of (6) in the multimode case was carried out in [21].

Formula (6) implies that the visibility of interference pattern, which is defined in a standard way [22] by

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \quad (8)$$

falls to zero as $\tau_i \longrightarrow 0$; this fact was pointed out in [20]. For instance, if

$$|\tau_i|^2 \propto \exp\{-\alpha_i l'\},$$

where α_i is the Bouguer absorption coefficient at idler frequency, attenuation of the idler mode due to absorption in the intermediate medium deteriorates the visibility of the interference pattern. Formula (6) displays an essential property of two-photon interference: losses in the idler mode do not change the integral intensity of the SPDC but only affect its shape. This property underlies the two-photon interferometric method for estimating the absorption coefficient of nonlinear crystals in the near-infrared region [14, 23].

2.3. Taking into Account Losses at Pump and Signal Frequencies

If the linear layers between two nonlinear crystals introduce losses in the *p* and *s* modes, then, on the one hand, the pump amplitude in the second crystal decreases,

$$E_p^{(2)} = \tau_p E_p^{(1)};$$

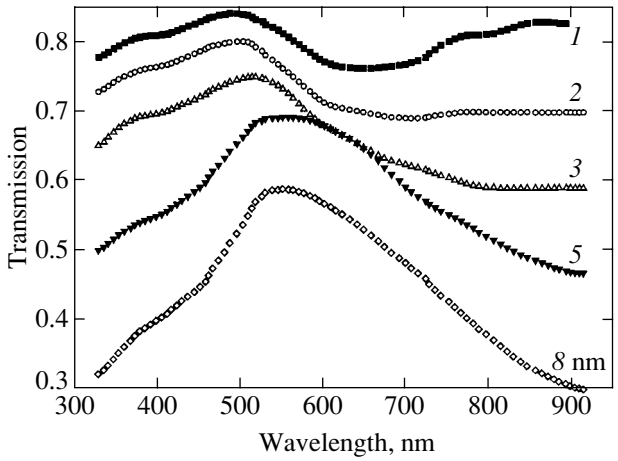

Fig. 2. Transmission coefficient of samples as a function of wavelength.

therefore, the amplitude of the signal wave generated in the second crystal is given by

$$f_2 \propto \chi E_p^{(2)} = \chi \tau_p E_p^{(1)}.$$

On the other hand, due to the losses of the signal wave generated in the first crystal while passing through intermediate layers, this amplitude decreases at the input of the second crystal:

$$f_1|_{\text{right}} \xrightarrow{\text{losses}} f_1|_{\text{left}} = \tau_s f_1 \propto \tau_s \chi E_p^{(1)},$$

where $f_1|_{\text{right}}$ is the amplitude on the right boundary of the first crystal and $f_1|_{\text{left}}$ is the amplitude on the left boundary of the second crystal. Thus, amplitudes that differ not only in phase but also in absolute value contribute to the interference of biphoton fields. This may deteriorate the visibility of two-photon interference. The difference in amplitude is the more conspicuous, the greater the frequency dispersion of the transmission coefficient $\tau_{s,p}(\omega)$, or the greater the difference of the parameter τ_s/τ_p from unity. Using (4), we can estimate the visibility of interference pattern in two-photon interference for different amplitudes:

$$V \sim \frac{2|\tau_s/\tau_p|}{|\tau_s/\tau_p|^2 + 1}. \quad (9)$$

It follows from (9) that, even for $|\tau_s/\tau_p| = 2$, the visibility of interference is still sufficiently high: $V \approx 80\%$. Thus, we can assert that losses in the pump modes and in the signal (observable) mode do not affect the visibility of two-photon interference while the parameter $|\tau_s/\tau_p|$ does not differ too much from unity. For instance, $V \geq 90\%$ if

$$0.6 \leq |\tau_s/\tau_p| \leq 1.6. \quad (10)$$

Hence, under condition (10), the observable deterioration of the visibility of two-photon interference is mainly attributed to the losses in the idler (nonobservable) mode. This fact serves as a basis for the analysis of experimental data obtained in the present work. Under condition (10), formula (6) gives a relation between the visibility of two-photon interference, which can be measured experimentally, and the losses at the idler frequency:

$$V = |\tau_i|. \quad (11a)$$

If we take into account that $|\tau_i|^2 \propto \exp\{-\alpha_i l_2'\}$, then the absorption coefficient at the idler frequency proves to be logarithmically related to V:

$$\alpha_i = -\frac{2\ln V}{l_2'}. \quad (11b)$$

Thus, measuring the visibility of two-photon interference, one can directly evaluate the absorption coefficient of a substance (in the infrared region) placed between nonlinear crystals.

Note that the losses in the *s* and *p* modes do not lead to a decrease in the integral intensity of the signal of parametric down-conversion. Therefore, it would be interesting to verify experimentally which of the two factors proves to be dominant as losses increase in all the modes *s*, *p*, and *i*: the disappearance of interference under a still appreciable total intensity of the signal of parametric down-conversion or the total disappearance of the signal.

3. EXPERIMENT

3.1. Description of Samples

We used 0.2-mm-thick polished fused-silica plates with an area of 15×15 mm^2 as the substrates. A gold film was deposited on these substrates by the cathode sputtering method. We investigated films with integral thicknesses of 10, 20, 30, 50, and 80 Å in the working area. Due to the small thickness, the films did not continuously cover the substrate; they were characterized by a cluster structure when the thickness was less than 30 Å and by a porous structure for greater thickness.

The transmission coefficient of the samples was measured by an M400 spectrophotometer as a function of the wavelength. The results of these measurements are shown in Fig. 2.

3.2. Experimental Setup and Measurement Technique

In the experiment, we measured two-dimensional frequency–angular spectra of spontaneous parametric down-conversion of light emitted from the MZNI. As the nonlinear media, we used lithium niobate crystals doped with magnesium oxide, $LiNbO_3$: MgO (5%). Between these nonlinear media, we placed gold films of various thicknesses deposited on fused-silica substrates. The thickness of crystals was 440 μm, and the

thickness of the fused-silica substrates was 200 μm. An argon laser operating at the 488-nm line with an output power of 1 W and a beam diameter of 2 mm^2 served as the pump source; the scattered field was collimated by an objective lens and focused onto the input slit of an ISP-51 spectrograph.

To obtain panoramic spectra, we used a photographic technique that is the standard one for SPDC spectroscopy [24]. Quantitative information about the line shape of down-conversion was obtained by two methods. In the first method, the angular distribution of intensity at several signal wavelengths was recovered after taking into account the nonlinear dependence of the blackening of photographic film as a function of light intensity. In the second method, the line shape was recorded directly by the angular scanning of two-dimensional spectra at a fixed wavelength. A Hamamatsu R5600U photomultiplier tube was placed in the focal plane of the spectrographic camera and could move along two coordinates (the angle θ_s and the wavelength λ_s). The output pulses of the photomultiplier tube were amplified, subjected to amplitude discrimination, and fed to a counter. The entire electronic part of the receiving system was assembled in the CAMAC standard.

Typical photographs of the spectra and the corresponding angular intensity distributions obtained during scanning are shown in Figs. 3 and 4.

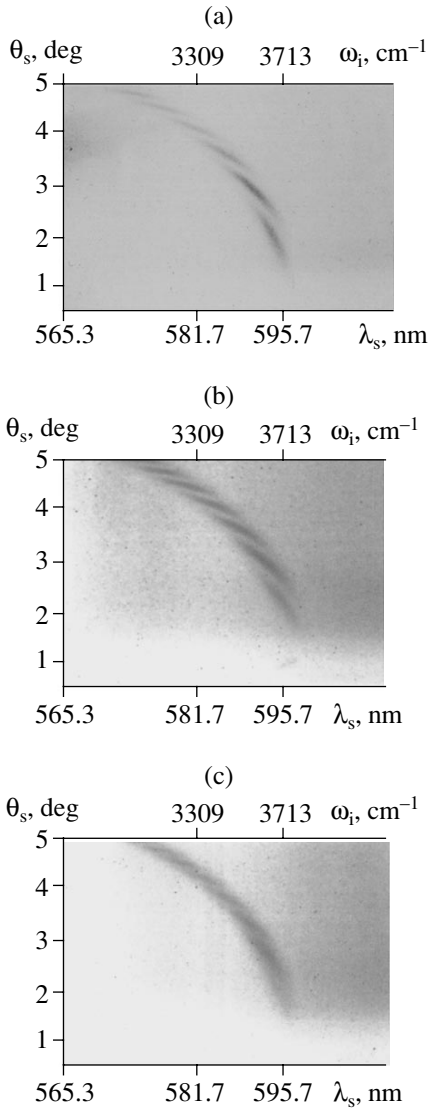

Fig. 3. Photographs of angular-frequency spectra of the SPDC obtained in the system nonlinear crystal–fused-silica substrate–gold film–nonlinear crystal for films of various thicknesses; (a) in the absence of a film, (b) for a film thickness of 10 Å, and (c) for a film thickness of 50 Å.

4. DISCUSSION

The diagrams shown in Fig. 2 allow one to estimate the parameter $|\tau_s/\tau_p|$ introduced in the preceding section. For the signal wavelengths 5685, 5707, and 5731 Å (the corresponding idler frequencies, defined by (1), are equal to 2902, 2970, and 3043 cm^{-1}, respectively) at which the spectra were processed, the values of the parameter $|\tau_s/\tau_p|$ are given in the table. One can see that this parameter satisfies condition (10). Thus, at these wavelengths, the visibility of two-photon interference is mainly determined by losses at idler frequencies in the infrared region.

The visibility of two-photon interference as a function of losses was experimentally investigated when gold films of various thicknesses were placed into the MZNI. Figures 3 and 4 show that interference phenomena in two-photon light virtually disappear for a film thickness of 50 Å. A test photograph of an 80-Å-thick sample shows that there is no angular–frequency modulation. At the same time, we certainly observed an SPDC signal even in films 100 Å in thickness. Hence, we can conclude that the method of two-photon interference is more sensitive to the losses τ_i at idler frequencies than to the losses τ_p and τ_s at pump and signal frequencies, which result in a decrease in the integral intensity of the spectra.

The transmission coefficient at a frequency of 2970 cm^{-1} determined by formulas (6) and (11a) from the angular scans of interferograms is shown in Fig. 5 as a function of film thickness. We did not observe an appreciable difference in the behavior of these functions in the range of idler frequencies from 2900 to 3040 cm^{-1}. This fact suggests that the absorption coefficient of the films shows weak dispersion in this spectral domain. A similar conclusion can be drawn from a visual analysis of the frequency–angular spectra (Fig. 3) that were obtained in a wider range (2400–3700 cm^{-1}): in these spectrograms, one cannot distinguish regions where the visibility is appreciably varied.

According to (11), the visibility of the interference pattern depends on the transmission coefficient at the idler frequency. In this range, the optical properties of metals are primarily determined by free electrons. The

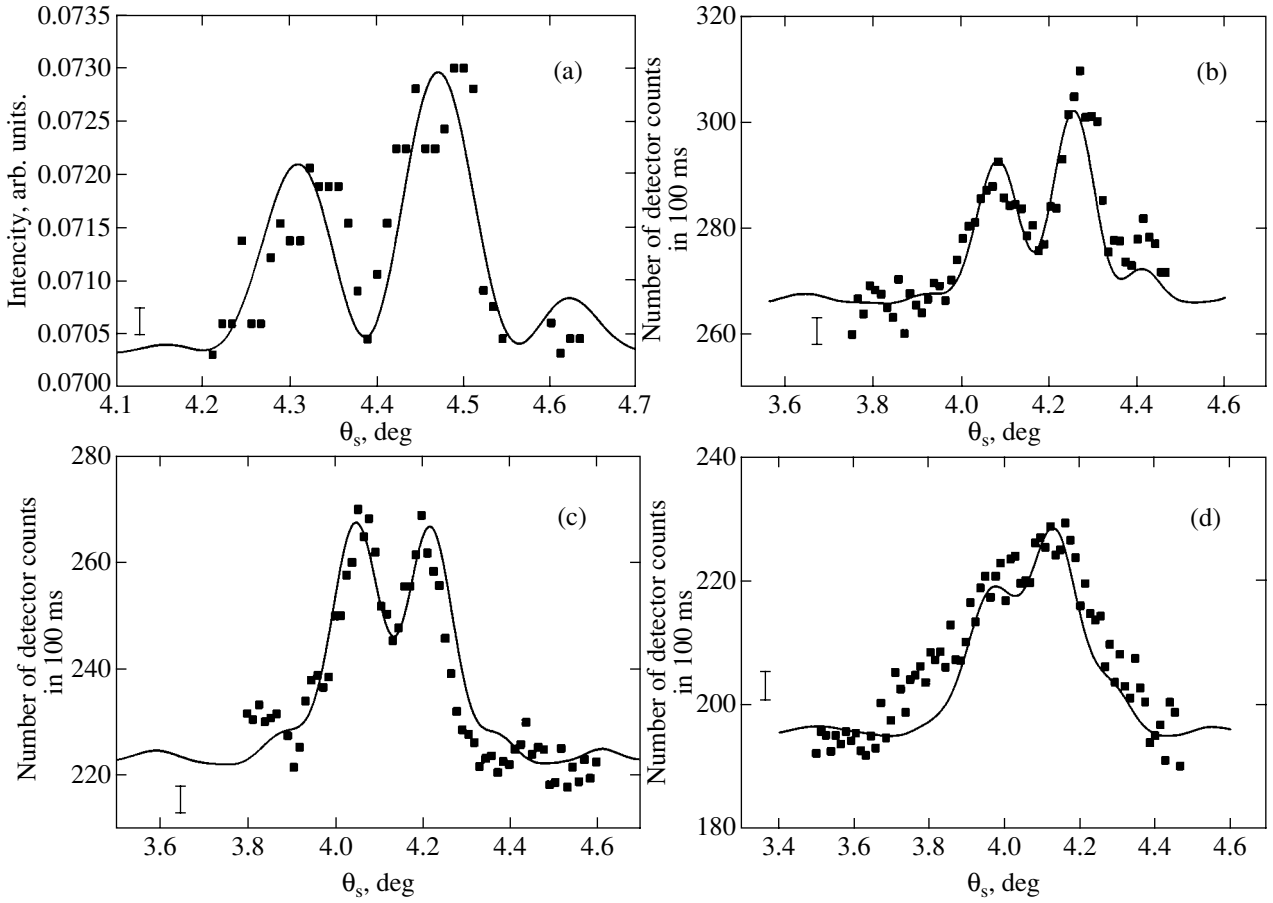

Fig. 4. Angular distribution of the SPDC intensity for films of various thicknesses; (a) in the absence of a film, (b) for a film thickness of 10 Å, (c) for a film thickness of 30 Å, and (d) for a film thickness of 50 Å. Distribution (a) is obtained by recovering the intensity from the blackening level of a film by Photoshop 6.0 software for λ_s = 5684 Å. Distributions (b)–(d) are obtained by photoelectric recording of signals for λ_s = 5707 Å.

relation between the dielectric permittivity and the basic optical constants is given by the following formulas [22]:

$$\mathrm{Re}\varepsilon = n^2(1-\kappa^2), \tag{12}$$

$$\mathrm{Im}\varepsilon = \frac{4\pi\sigma}{\omega} = 2n^2\kappa, \tag{13}$$

where n is the real part of the refractive index $\hat{n} = n(1 + i\kappa)$ and κ is the extinction coefficient. From (12) and (13), we can derive the absorption coefficient,

$$\alpha_i = \frac{2\omega_i}{c} n\kappa = \frac{4\pi}{\lambda_{0i}}\kappa, \tag{14}$$

where λ_{0i} is the idler wavelength in vacuum. Approximating the experimental curve (Fig. 5) by an exponential function (solid curve), we obtain $\alpha_i \approx 2.8 \times 10^6$ cm^{-1} at a frequency of $\omega_i \approx 3000$ cm^{-1} (or $\lambda_{0i} \approx 3.3$ μm). Unfortunately, we had no information about the optical con-

Table

No.	Sample thickness, Å	$\lvert\tau_s\rvert^2$			$\lvert\tau_p\rvert^2$	$\lvert\tau_s/\tau_p\rvert$		
		5685 Å	5707 Å	5731 Å	4880 Å	5685 Å	5707 Å	5731 Å
1	10	0.79	0.79	0.79	0.84	0.974	0.974	0.974
2	20	0.75	0.74	0.74	0.80	0.968	0.962	0.962
3	30	0.71	0.71	0.70	0.74	0.98	0.98	0.973
4	50	0.69	0.69	0.69	0.64	1.038	1.038	1.038
5	80	0.58	0.58	0.58	0.5	1.077	1.077	1.077

stants of the films in the infrared region. Therefore, we used the data given in [25]. Substituting the values of the refractive index $n \approx 0.8$ and the extinction coefficient $\kappa \approx 20$ into (13), we obtain $\alpha_i \approx 6.1 \times 10^5$ cm^{-1}, which is about five times less than the value obtained in our experiments.

The difference between the experimental and calculated values of the absorption coefficient, which is substantially greater than the measurement error, can be attributed to several factors.

First, the optical constants of gold presented in [25] refer to thick films. It is well known that the optical constants in ultrathin films, of thickness less than 10 nm, are significantly affected by the film structure: an ultrathin film is considered as a set of separate islands. For instance, it was pointed out in [26] that the cluster structure leads to a significant increase in absorption. A giant increase in the absorption of infrared radiation in metal particles was pointed out, for example, in [27]. The optical properties of fractal clusters, in particular, an anomalous behavior of susceptibility, was considered in [28]. Recently, the optical properties of metal island films near the percolation threshold have been intensively discussed in the literature [29].

Second, the measurement technique for the absorption coefficient, based on two-photon interference, may give results that substantially differ from those obtained earlier. The point is that, in conventional methods for investigating metal films, one measures the reflection and transmission coefficients of "free" waves, i.e., waves with nonzero mean occupation of modes incident to a sample from free space. The dispersion relations for the transmission coefficient of films in the visible region (Fig. 2) have been obtained precisely in this way. However, in the method of two-photon interference, the optical parameters at a nonobservable (idler) frequency depend on fluctuating vacuum fields with zero mean occupation number of modes. There are examples in the literature where a comparison of these two methods shows a significant discrepancy precisely when measuring the absorption coefficient: the absorption measured by the SPDC spectroscopy (field fluctuations) [30] proves to be about an order of magnitude greater than that obtained by four-wave coherent scattering by polaritons (excitation of polaritons by biharmonic pumping) [31] or by direct measurement of infrared transmission [32]. The physical nature of this discrepancy has not yet been revealed, and one may suppose that experiments on two-photon interference in the Mach–Zehnder scheme will provide an answer to this question.[2]

Note that the approach considered in the present paper does not allow us to make any conclusions about

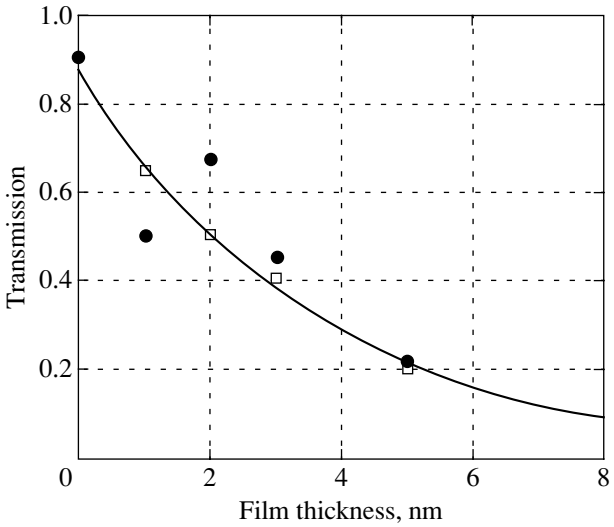

Fig. 5. Amplitude transmission coefficient at a frequency of 2970 cm^{-1} as a function of film thickness, obtained from angular scans of interferograms. The solid line represents an exponential approximation of the experimental data; circles, photographically acquired data; squares, data acquired by the photomultiplier.

the dispersion of the real part of the refractive index of films. This is associated with the small thickness of the films. As was shown in [11, 13], the distinct features of the dispersion of the refractive index that arise near resonance frequencies must be accompanied by the variation of the curvature of interference orders or with the appearance of crooks in the spectra of two-photon interference. This is associated with the fact that, by definition, the wave mismatch $\delta' = \Delta' l_2'$ of the film involves the real parts of the wave vectors

$$\mathrm{Re}(\hat{k}) = \frac{\omega}{c}\mathrm{Re}(\hat{n}) = \frac{\omega}{c}n.$$

The strong dispersion of n leads to an increase in the wave mismatch Δ', which leads to a variation in the phase of the interference pattern versus frequency/angle. The value of n of gold ranges from 0.2 to 1.2 as the wavelength varies from 0.3 to 1 μm. In the near-infrared region, when $\lambda_i \approx 3$ μm, the value of n proves to be on the order of unity and weakly increases with wavelength [24, 34, 35]. However, since the film thickness is about tens of angstroms and $l_2' \ll l, l_1'$, the corresponding mismatch δ_2' is small, and its contribution to the line shape (6) is negligible compared with the contributions of the mismatches in the nonlinear crystals, $\delta = \Delta l$, and in the quartz substrate, $\delta_1' = \Delta' l_1'$.

5. CONCLUSIONS

We have discussed experiments on two-photon interference in the presence of absorption in all the

[2] Note that recent publications (see, for example, [33]) on the spectroscopic applications of the so-called frequency-entangled photon pairs do not answer the question posed since these works deal with the contribution of the direct transmission of a real idler wave to the distribution of photocount coincidences.

modes that participate in the spontaneous parametric down-conversion of light. The objects of study were ultrathin gold films 10, 20, 30, 50, and 80 Å in thickness deposited on a fused-silica substrate.

We have determined conditions under which the effect of the absorption for the signal and pump wave on the interference pattern is negligible. We have shown that, under these conditions, the visibility of the interference pattern and the form of the frequency–angular spectrum on the signal wave are determined by the optical parameters of the medium at an idler frequency, which belongs to the near-infrared region.

The results obtained have allowed us to reveal a number of features of the effect of absorption on the spontaneous parametric down-conversion of light. By the method of two-photon interferometry, we have measured the absorption coefficient of gold films in the neighborhood of $\lambda \approx 3.3$ μm; the value of this coefficient, $\alpha \approx 2.8 \times 10^6$ cm^{-1}, is greater than the appropriate value for thick films by a factor of five.

ACKNOWLEDGMENTS

We are grateful to A.A. Nikulin and A.N. Rubtsov for fruitful discussions of the experimental results.

This work was supported in part by the Russian Foundation for Basic Research (project nos. 02-02-16843, 03-02-16444, and 01-02-17893), by project no. 15 of Special Federal Research Program "Investigation and Development of Priority Directions in Science and Technology," and by INTAS (project no. 2122-01).

REFERENCES

1. D. N. Klyshko, Laser Phys. **2**, 997 (1992).
2. L. Mandel, Rev. Mod. Phys. **71**, S274 (1999).
3. A. Zeilinger, Rev. Mod. Phys. **71**, S288 (1999).
4. *The Physics of Quantum Information: Quantum Cryptography, Quantum Teleportation, Quantum Computation*, Ed. by D. Bouwmeester, A. K. Ekert, and A. Zeilinger (Springer, Berlin, 2000; Postmarket, Moscow, 2002).
5. Yu. N. Polivanov, Usp. Fiz. Nauk **126**, 185 (1978) [Sov. Phys. Usp. **21**, 805 (1978)].
6. G. Kh. Kitaeva, D. N. Klyshko, and I. V. Tubin, Kvantovaya Élektron. (Moscow) **9**, 561 (1982).
7. A. A. Malygin and A. V. Sergienko, Available from VINITI, No. 5294-84 (1984), p. 21.
8. G. Kh. Kitaeva, S. P. Kulik, and A. N. Penin, Zh. Éksp. Teor. Fiz. **90**, 1051 (1986) [Sov. Phys. JETP **63**, 613 (1986)].
9. G. Kh. Kitaeva, S. P. Kulik, and A. N. Penin, Fiz. Tverd. Tela (St. Petersburg) **34**, 3440 (1992) [Sov. Phys. Solid State **34**, 1841 (1992)].
10. G. Kh. Kitaeva, Doctoral Dissertation in Physics and Mathematics (Moscow State Univ., Moscow, 2002).
11. A. V. Burlakov, S. P. Kulik, A. N. Penin, and M. V. Chekhova, Zh. Éksp. Teor. Fiz. **113**, 1991 (1998) [JETP **86**, 1090 (1998)].
12. D. Yu. Korystov, S. P. Kulik, and A. N. Penin, Kvantovaya Élektron. (Moscow) **30**, 921 (2000).
13. D. Yu. Korystov, S. P. Kulik, and A. N. Penin, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 248 (2001) [JETP Lett. **73**, 214 (2001)].
14. A. V. Burlakov, M. V. Chekhova, O. A. Karabutova, *et al.*, Laser Phys. **12**, 825 (2001).
15. A. V. Burlakov, M. V. Chekhova, D. N. Klyshko, *et al.*, Phys. Rev. A **56**, 3214 (1997).
16. R. Stolle, G. Marovsky, E. Schwartzberg, and G. Berkovie, Appl. Phys. B **63**, 491 (1996).
17. D. N. Klyshko, *Photons and Nonlinear Optics* (Nauka, Moscow, 1980), p. 256.
18. A. V. Belinsky and D. N. Klyshko, Laser Phys. **4**, 663 (1994).
19. D. N. Klyshko, Zh. Éksp. Teor. Fiz. **104**, 2676 (1993) [JETP **77**, 222 (1993)].
20. L. J. Wang, X. Y. Zou, and L. Mandel, Phys. Rev. Lett. **67**, 318 (1991).
21. L. J. Wang, X. Y. Zou, and L. Mandel, Phys. Rev. A **44**, 4614 (1991).
22. M. Born and E. Wolf, *Principles of Optics*, 4th ed. (Pergamon Press, Oxford, 1969; Nauka, Moscow, 1970).
23. A. V. Burlakov, Yu. B. Mamaeva, A. N. Penin, and M. V. Chekhova, Zh. Éksp. Teor. Fiz. **120**, 67 (2001) [JETP **93**, 55 (2001)].
24. D. N. Klyshko, A. N. Penin, and B. F. Polkovnikov, Pis'ma Zh. Éksp. Teor. Fiz. **11**, 11 (1970) [JETP Lett. **11**, 5 (1970)].
25. *Physikalisch–Chemische Tabellen* (Springer, Berlin, 1929), Vol. 165B.
26. R. P. Devaty and A. J. Sievers, Phys. Rev. Lett. **52**, 1344 (1984).
27. A. V. Plyukin, A. K. Sarychev, and A. M. Dykne, Phys. Rev. B **59**, 1685 (1999).
28. V. M. Shalaev and M. I. Shtokman, Zh. Éksp. Teor. Fiz. **92**, 509 (1987) [Sov. Phys. JETP **65**, 287 (1987)].
29. A. K. Sarychev, V. A. Shubin, and V. M. Shalaev, Phys. Rev. E **59**, 7239 (1999).
30. A. V. Burlakov, M. V. Chekhova, S. P. Kulik, and A. N. Penin, Opt. Commun. **165**, 39 (1999).
31. G. Kh. Kitaeva, K. A. Kuznetsov, A. A. Mikhailovsky, and A. N. Penin, J. Raman Spectrosc. **31**, 767 (2000).
32. K. A. Kuznetsov, Candidate's Dissertation in Physics and Mathematics (Moscow State Univ., Moscow, 2002).
33. A. Yabushta and T. Kobayashi, quant-ph/0306154.
34. M.-L. Theye, Phys. Rev. B **2**, 3060 (1970).
35. P. B. Jonson and R. W. Christy, Phys. Rev. B **6**, 4370 (1972).

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 98, No. 1, 2004, pp. 39–52.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 1, 2004, pp. 48–62.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Hyperfine Ground-State Structure of Muonic Hydrogen

A. P. Martynenko[a,*] and R. N. Faustov[b,]**

[a]*Samara State University, Samara, 443011 Russia*
[b]*Scientific Council on Cybernetics, Russian Academy of Sciences, Moscow, 117333 Russia*
**e-mail: mart@info.ssu.samara.ru*
***e-mail: faustov@theory.sinp.msu.ru*
Received August 7, 2003

Abstract—Corrections of orders α^5 and α^6 to the superfine ground-state structure of the muonic hydrogen atom were calculated. The calculations took into account the effects of the structure of the nucleus on one- and two-loop Feynman amplitudes with the help of the electromagnetic form factors of the proton and the modification of the superfine part of the Breit potential caused by the electronic polarization of the vacuum. The total splitting of the 1*S* state is 182.725 meV; this value can be used as a reliable estimate in conducting a corresponding experiment with an accuracy of 30 ppm. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Precision measurements of the energy spectra of the simplest atomic systems (muonium, positronium, hydrogen atom, muonic hydrogen, etc.) are of great importance for high-accuracy verification of the Standard Model and the theory of bound states. Two-particle bound states are an effective tool for refining the values of fundamental physical constants (the fine structure constant, the ratio between the masses of the muon and electron, the Rydberg constant, the proton charge radius, etc.) used for creating standards of units [1]. It can be claimed that a necessary supplement to the use of large accelerators for penetrating deep into particles and to the search for new interactions is the observation of fine effects in low-energy physics, including bound states in quantum electrodynamics. These effects can be used to extract such details of interaction behavior at small distances that can only be observed at very high energies [2].

Like electronic hydrogen, muonic hydrogen (μp) is the simplest atomic system, whose energy spectrum is to a substantial extent determined by strong interaction effects. These effects are primarily related to two proton form factors (electric G_E and magnetic G_M), which describe the charge and magnetic moment distributions. For the Lamb shift, the major contribution to the energy spectrum of order $(Z\alpha)^4$ is determined by the differential characteristic of these distributions, namely, the proton charge radius r_p. For this reason, a comparison of the experimental Lamb shift value and its theoretical estimate obtained with corrections high-order in α can be used to more precisely determine the r_p value. For instance, measurements of the $2P$–$2S$ Lamb shift in μp with an accuracy of 30 ppm would allow us to obtain the proton charge radius with an accuracy one order of magnitude higher than that obtainable by the other methods [3]. When calculating corrections for the structure of the nucleus to the superfine splitting of energy levels (see [4–11]), we must know the electromagnetic form factors of the proton themselves. The most recent experimental measurements of G_E and G_M were performed in Mainz 20 years ago [12].

Another important strong interaction contribution to the energy spectrum of the hydrogen atom is related to the polarizability of the proton [13–16]. This contribution arises already in one-loop muon (electron)–proton interaction amplitudes, when, for instance, various baryon resonances can be created in the intermediate state as a result of virtual Compton scattering by the proton. Precise calculations of this effect can be performed using experimental data and theoretical constructions of the polarization structural functions of the nucleon. The proton structure and polarizability effects introduce major theoretical uncertainty into the equations for various energy levels, primarily into the superfine splitting of the ground state of the hydrogen atom, which has the form

$$\Delta E^{HFS}_{\rm theor} = E^F(1 + \delta^{QED} + \delta^{\rm str} + \delta^{\rm pol} + \delta^{HVP}),$$
$$E^F = \frac{8}{3}\alpha^4\frac{m_P m_1^2 m_2^2}{(m_1+m_2)^3}, \quad (1)$$

where μ_P is the magnetic moment of the proton in nuclear magnetons, m_1 is the mass of the muon (electron), m_2 is the mass of the proton, δ^{QED} is the quantum electrodynamic contribution, δ^{HVP} is the contribution of the hadronic vacuum polarization, and the corrections $\delta^{\rm str}$ and $\delta^{\rm pol}$ are the contributions of strong interactions related to the structure of the nucleus and its polarizability. Equation (1) is valid for both muonic and elec-

tronic hydrogen, but the particular forms of corrections are substantially different for these two species. The superfine splitting of the ground state of the electronic hydrogen atom was measured to a high accuracy many years ago [17],

$$\Delta\nu_{\rm exp}^{HFS}(ep) = 1420405.7517667(9)\ \text{kHz}. \qquad (2)$$

The present-day discrepancy between theory and experiment (without taking proton polarizability into account), which can be represented by the ratio [18]

$$\frac{\Delta E_{\rm theor}^{HFS}(ep) - \Delta E_{\rm exp}^{HFS}(ep)}{E^F(ep)} = -4.5(1.1)\times 10^{-6}, \qquad (3)$$

contains one of the principal errors caused by the inaccuracy of determining the proton form factors. The main part of the one-loop correction for the structure of the proton is given by the following equation (the Zemach correction) [4]:

$$\Delta E_Z = E^F\frac{2\mu\alpha}{\pi^2}\int\frac{d\mathbf{p}}{(\mathbf{p}^2+W^2)^2} \times\left[\frac{G_E(-\mathbf{p}^2)G_M(-\mathbf{p}^2)}{\mu_P}-1\right] = E^F(-2\mu\alpha)R_p, \qquad (4)$$

where

$$W = \alpha\mu, \quad \mu = m_1m_2/(m_1+m_2)$$

is the reduced mass of two particles and R_p is the Zemach radius. In the coordinate representation, the Zemach correction (4) is determined by the convolution of the distribution density of the proton magnetic moment $\rho_M(r)$ and the distribution density of the electric charge $\rho_E(r)$. The Zemach radius, which is the integral characteristic of proton structure effects on the superfine splitting of energy levels, can be treated as a new proton parameter, which coincides for the electronic and muonic hydrogen atoms to within $O(\alpha)$. The Zemach contribution for muonic hydrogen is

$$\Delta E_Z = -1.362 \pm 0.068\ \text{meV}, \qquad (5)$$

where the error, estimated at 5%, includes the uncertainty in proton form factor measurements [12]. Therefore, along with measurements for electronic hydrogen, designing an experiment in which the superfine splitting of the ground state of muonic hydrogen can be measured at the same accuracy of 30 ppm as with Lamb shift measurements would provide new information on possible values of the $\delta^{\rm str}$ and $\delta^{\rm pol}$ contributions [19].

Performing such an experiment requires corrections of various orders to be calculated equally accurately. Although calculations of the superfine structure of the hydrogen atom have been conducted over many years and have already reached an accuracy of 10^{-8} for quantum electrodynamic contributions δ^{QED} [20, 18], their results cannot be directly applied to muonic hydrogen by merely replacing the mass of the electron by the mass of the muon. The main reason for this is corrections for the structure of the nucleus (see above). Indeed, with muonic hydrogen, the main region of intermediate loop momentum integrals is on the order of the mass of the muon. It follows that high-accuracy theoretical calculations of such amplitudes can only be performed by directly integrating them taking into account experimental data on the electromagnetic form factors of the proton.

Various contributions to the energy levels of muonic atoms were studied long ago [21]. For this reason, a new, more complete analysis of all possible corrections to the superfine structure of μp at an accuracy level of 30 ppm is required. The principal corrections of order α^5 to the superfine structure of the $2S$ state of μp were studied in [22]. These corrections are very important for determining the $2P$–$2S$ Lamb shift from experiment. In this work, we calculate various contributions of orders α^5 and α^6 to the superfine structure of muonic hydrogen that are determined by the electromagnetic and strong interaction effects. Our purpose was to numerically determine the superfine structure of muonic hydrogen with the accuracy specified above and obtain a reliable reference value for performing the corresponding experiment. Some problems in designing an experiment for measuring the superfine structure of the μp atom were discussed in [23].

2. VACUUM POLARIZATION EFFECTS IN ONE-PHOTON INTERACTION

Our calculations of various energy levels of hydrogen-like atoms are performed within the framework of the quasi-potential approach, in which a bound state of two particles is described by the Schrödinger-type equation [24]

$$[G^f]^{-1}\psi_M \equiv \left(\frac{b^2}{2\mu_R}-\frac{\mathbf{p}^2}{2\mu_R}\right)\psi_M(\mathbf{p}) = \int\frac{d\mathbf{q}}{(2\pi)^3}V(\mathbf{p},\mathbf{q},M)\psi_M(\mathbf{q}), \qquad (6)$$

where

$$b^2 = E_1^2 - m_1^2 = E_2^2 - m_2^2,$$

$\mu_R = E_1E_2/M$ is the relativistic reduced mass, $M = E_1 + E_2$ is the mass of the bound state, and $E_{1,2}$ are the energies of free particles in the center-of-mass frame. The quasi-potential in (6) is constructed in quantum electrodynamics by perturbation theory with the use of the two-particle scattering amplitude T projected onto the

positive-frequency states outside the mass surface at zero relative energies of the particles,

$$V = V^{(1)} + V^{(2)} + V^{(3)} + \ldots,$$
$$T = T^{(1)} + T^{(2)} + T^{(3)} + \ldots, \quad (7)$$

$$V^{(1)} = T^{(1)},$$
$$V^{(2)} = T^{(2)} - T^{(1)} \times G^f \times T^{(1)}, \ldots. \quad (8)$$

The initial approximation to the quasi-potential $V(\mathbf{p}, \mathbf{q}, M)$ for a bound system was selected in the form of the usual Coulomb potential,

$$V(\mathbf{p}, \mathbf{q}, M) = V^C(\mathbf{p} - \mathbf{q}) + \Delta V(\mathbf{p}, \mathbf{q}, M).$$

The increase in the lepton mass in muonic hydrogen compared with its electronic counterpart decreases the radius of the Bohr orbit in μp. As a result, the Compton wave length of the electron and the radius of the Bohr orbit become commensurate [1],

$$\frac{\hbar^2}{\mu e^2} : \frac{\hbar}{m_e c} = 0.737384$$

(m_e is the mass of the electron and μ is the reduced mass in the μp atom). This substantially enhances the role played by vacuum polarization effects in the energy spectrum of the μp atom [25]. Corrections for vacuum polarization in one-photon interaction are shown in Fig. 1.

To determine the contribution of diagram *a* in Fig. 1 (electronic vacuum polarization) to the particle interaction operator, we must perform the following substitution in the photon propagator [25]:

$$\frac{1}{k^2} \longrightarrow \frac{\alpha}{\pi}\int_0^1 dv \frac{v^2(1 - v^2/3)}{k^2(1 - v^2) - 4m_e^2}. \quad (9)$$

If

$$-k^2 = \mathbf{k}^2 \sim \mu_e^2 (Z\alpha)^2 \sim m_e^2 (Z\alpha)^2$$

(electronic hydrogen, μ_e is the reduced mass of two particles in the hydrogen atom), then, ignoring the first term in the denominator in the right-hand side of (6), we obtain

$$-\frac{\alpha}{15\pi m_e^2}.$$

However, if

$$\mathbf{k}^2 \sim \mu^2 (Z\alpha)^2 \sim m_1^2 (Z\alpha)^2$$

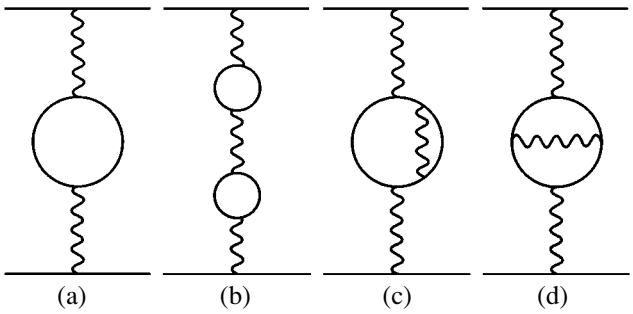

Fig. 1. One- and two-loop vacuum polarization effects in one-photon interaction.

(muonic hydrogen, m_1 is the mass of the muon), then $\mu\alpha$ and m_e are values of one order and we cannot expand the denominator in (9) in α. With muonic hydrogen, we must construct the superfine part of the potential in the 1γ approximation using the exact equation (9). Further, when we consider values of orders α^5 and α^6, we take into account that the appearance of the mass of the electron $m_e \sim \mu\alpha$ in the denominator of the amplitude effectively decreases the order of the contribution in α by one. It is well known that, in the one-photon approximation, the quasi-potential of the superfine interaction between the muon and proton has the form [26]

$$V_{1\gamma}^{HFS}(\mathbf{k}) = \frac{4\pi Z\alpha}{m_1 m_2}\frac{1 + \kappa}{4}\frac{1}{\mathbf{k}^2} \times [(\boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2)\mathbf{k}^2 - (\boldsymbol{\sigma}_1 \cdot \mathbf{k})(\boldsymbol{\sigma}_2 \cdot \mathbf{k})]. \quad (10)$$

For *S* states, (10) reduces to

$$V_{1\gamma}^{HFS}(\mathbf{k}) = \frac{8\pi Z\alpha}{3m_1 m_2}\frac{\boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2}{4}(1 + \kappa), \quad (11)$$

where $\kappa = 1.792847337(29)$ is the anomalous magnetic moment of the proton. Averaging potential (11) over Coulomb wave functions yields the main contribution of order $(Z\alpha)^4$ to the superfine splitting of the $1S$ state of the μp atom, or the Fermi energy

$$E^F = \frac{8}{3}(Z\alpha)^4 \frac{\mu^3}{m_1 m_2}(1 + \kappa) = 182.443 \text{ meV}. \quad (12)$$

The modification of the Coulomb potential

$$V^C(\mathbf{k}) = -Ze^2/\mathbf{k}^2$$

caused by vacuum polarization is determined taking

into account (9) [22],

$$V^C_{VP}(\mathbf{k}) = -4\pi Z\alpha\frac{\alpha}{\pi}\int_1^\infty \frac{\sqrt{\xi^2-1}}{3\xi^4}\frac{(2\xi^2+1)}{\mathbf{k}^2+4m_e^2\xi^2}d\xi. \quad (13)$$

The Fourier transform of (13) gives the corresponding operator in the coordinate representation,

$$V^{(C)}_{VP}(r) = \frac{\alpha}{3\pi}\int_1^\infty d\xi\frac{\sqrt{\xi^2-1}(2\xi^2+1)}{\xi^4} \times\left(-\frac{Z\alpha}{r}\exp(-2m_e\xi r)\right). \quad (14)$$

Similarly, we can calculate the contribution of the electronic polarization of the vacuum to the superfine part of the 1γ potential for the S states in the momentum and coordinate representations, respectively,

$$V^{HFS}_{1\gamma,\,VP}(\mathbf{k}) = \frac{4\pi Z\alpha}{m_1m_2}\frac{(1+\kappa)}{4}\frac{2}{3}(\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2)\mathbf{k}^2\frac{\alpha}{\pi} \times\int_1^\infty\frac{\sqrt{\xi^2-1}(2\xi^2+1)}{3\xi^4(\mathbf{k}^2+4m_e^2\xi^2)}d\xi, \quad (15)$$

$$V^{HFS}_{1\gamma,\,VP}(r) = \frac{8Z\alpha(1+\kappa)}{3m_1m_2}\frac{(\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2)}{4}\frac{\alpha}{\pi} \times\int_1^\infty\frac{\sqrt{\xi^2-1}(2\xi^2+1)}{3\xi^4}d\xi \times\left[\pi\delta(\mathbf{r})-\frac{m_e^2\xi^2}{r}\exp(-2m_e\xi r)\right]. \quad (16)$$

The last equation can be used to calculate the correction for the electronic polarization of the vacuum to the superfine structure of the μp atom of order α^5. Bearing in mind that the wave function of the $1S$ state of μp has the form

$$\psi_{100}(r) = \frac{W^{3/2}}{\sqrt{\pi}}e^{-Wr}, \quad W = \mu Z\alpha, \quad (17)$$

let us write this correction as

$$\Delta E^{HFS}_{1\gamma,\,VP} = \frac{8\mu^3(Z\alpha)^4(1+\kappa)}{3m_1m_2}\frac{\alpha}{\pi}\frac{m_e^3}{3W^3} \times\int_{m_e/W}^\infty\frac{\sqrt{(W^2/m_e^2)\xi^2-1}}{\xi^4}\left(2\frac{W^2}{m_e^2}\xi^2+1\right)d\xi \times\left[1-\int_0^\infty\exp\left(-\frac{rm_e(\xi+1)}{W\xi}\right)r\,dr\right] = 0.398\ \text{meV}. \quad (18)$$

The contribution of the muonic vacuum polarization can be obtained from (16), in which m_e should be replaced by m_1. This correction is of order α^6 for the reason specified in discussing (9). Its value is

$$\Delta E^{HFS}_{1\gamma,\,MVP} = 0.004\ \text{meV}. \quad (19)$$

Contributions of the same order (α^6) arise from the diagrams of two-loop electronic polarization of the vacuum shown in Figs. 1b, 1c, and 1d. The interparticle interaction potential corresponding to the amplitude with two sequential electronic loops can be obtained by applying replacement (9) two times in the photon propagator. In the coordinate representation, this propagator has the form

$$V^{HFS}_{1\gamma,\,VP-VP}(r) = \frac{8\pi Z\alpha(1+\kappa)}{3m_1m_2}\frac{(\boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2)}{4}\left(\frac{\alpha}{\pi}\right)^2 \times\int_1^\infty\frac{\sqrt{\xi^2-1}(2\xi^2+1)}{3\xi^4}d\xi\int_1^\infty\frac{\sqrt{\eta^2-1}(2\eta^2+1)}{3\eta^4}d\eta \times\left[\delta(\mathbf{r})-\frac{m_e^2}{\pi r(\eta^2-\xi^2)} \times(\eta^4\exp(-2m_e\eta r)-\xi^4\exp(-2m_e\xi r))\right] \quad (20)$$

and gives the following result for the energy spectrum:

$$\Delta E^{HFS}_{1\gamma,\,VP-VP} = 0.001\ \text{meV}. \quad (21)$$

The contributions of diagrams 1c and 1d (Fig. 1), which are determined by the second-order polarization operator, can be calculated after the following replacement in the photon propagator [27]:

$$\frac{1}{k^2}\longrightarrow\left(\frac{\alpha}{\pi}\right)^2\int_0^1\frac{f(v)}{4m_e^2+k^2(1-v^2)}dv = \left(\frac{\alpha}{\pi}\right)^2\frac{2}{3}\int_0^1 dv\frac{v}{4m_e^2+k^2(1-v^2)} \times\left\{(3-v^2)(1+v^2)\left[\mathrm{Li}_2\left(-\frac{1-v}{1+v}\right)+2\mathrm{Li}_2\left(\frac{1-v}{1+v}\right)+\frac{3}{2}\ln\frac{1+v}{1-v}\ln\frac{1+v}{2}-\ln\frac{1+v}{1-v}\ln v\right]\right. \quad (22)$$

$$+\left[\frac{11}{16}(3-v^2)(1+v^2)+\frac{v^2}{4}\right]\ln\frac{1+v}{1-v}$$

$$+\left[\frac{3}{2}v(3-v^2)\ln\frac{1-v^2}{4}-2v(3-v^2)\ln v\right]$$

$$+\frac{3}{8}v(5-3v^2)\Bigg\}.$$

The contribution value can then conveniently be calculated in the coordinate representation with reduction of the interparticle interaction potential to the form

$$\Delta V^{HFS}_{1\gamma,\,2\text{-loop }VP}(r) = \frac{8\pi Z\alpha(1+\kappa)}{3m_1m_2}\left(\frac{\alpha}{\pi}\right)^2\int_0^1\frac{f(v)dv}{(1-v^2)} \times\left[\delta(\mathbf{r})-\frac{m_e^2}{\pi r(1-v^2)}\exp\left(-\frac{2m_er}{\sqrt{1-v^2}}\right)\right]. \quad (23)$$

Operator (23) gives the following contribution to the superfine structure of the μp atom:

$$\Delta E^{HFS}_{1\gamma,\,2\text{-loop }VP} = 0.002\ \text{meV}. \quad (24)$$

Note that, as we determine contributions to the energy spectrum numerically, the corresponding results are given with an accuracy of 0.001 meV.

3. SECOND-ORDER PERTURBATION THEORY

Second-order perturbation theory corrections to the energy spectrum are determined by the reduced Coulomb Green function [28], whose partial expansion is written as

$$\tilde G_1(\mathbf{r},\mathbf{r}') = \sum_{l,m}\tilde g_{nl}(r,r')Y_{lm}(\mathbf{n})Y^*_{lm}(\mathbf{n}'). \quad (25)$$

The $\tilde g_{nl}(r,r')$ radial function was obtained in [28] in the form of the Sturm expansion in Laguerre polynomials. For the $1S$ state, this function is written as

$$\tilde g_{10}(r,r') = -4\mu^2Z\alpha\left(\sum_{m=2}^{\infty}\frac{L^1_{m-1}(x)L^1_{m-1}(x')}{m(m-1)} +\frac{5}{2}-\frac{x}{2}-\frac{x'}{2}\right)\exp\left(-\frac{x+x'}{2}\right), \quad (26)$$

where $x = 2\mu Z\alpha r$ and L^m_n are the usual Laguerre polynomials defined as

$$L^m_n(x) = \frac{e^xx^{-m}}{n!}\left(\frac{d}{dx}\right)^n(e^{-x}x^{n+m}). \quad (27)$$

As several quasi-potential terms contain $\delta(\mathbf{r})$, we must know $\tilde G_1(\mathbf{r},0)$. The corresponding equation for the reduced Coulomb Green function was obtained in [29] using the Hostler representation for the Coulomb Green function and subtracting the pole term. This gave

$$\tilde G_{1S}(\mathbf{r},0) = \frac{Z\alpha\mu^2}{4\pi}\frac{2e^{-x/2}}{x} \times[2x(\ln x+C)+x^2-5x-2], \quad (28)$$

where $C = 0.5772\ldots$ is the Euler constant. The main contribution of order α^5 is determined in second-order perturbation theory by the equation

$$\Delta E^{HFS}_{1\ SOPT} = \sum_{n=2}^{\infty}\frac{\langle\psi^c_1|V^C_{VP}|\psi^c_n\rangle\langle\psi^c_n|V^{HFS}_{1\gamma}|\psi^c_1\rangle}{E^c_1-E^c_n}, \quad (29)$$

in which $\Delta V^{HFS}_{1\gamma}\sim\delta(\mathbf{r})$. Using (14) and (28), (29) can be rewritten as

$$\Delta E^{HFS}_{1\ SOPT} = -E^F\frac{2\alpha}{3\pi}\int_1^{\infty}d\xi\frac{\sqrt{\xi^2-1}}{\xi^2}\left(1+\frac{1}{2\xi^2}\right) \times\int_0^{\infty}dx\exp\left(-x\left(1+\frac{m_e\xi}{W}\right)\right) \times[2x(\ln x+C)+x^2-5x-2] = 0.795\ \text{meV}. \quad (30)$$

The contribution of order α^6 in second-order perturbation theory, which is determined by vacuum polarization, can be obtained from (29) after the replacement

$$\Delta V^{HFS}_{1\gamma}\longrightarrow\Delta V^{HFS}_{1\gamma\ VP}.$$

We used the explicit equations for the wave function $\psi^c_1(\mathbf{r})$, reduced Coulomb Green function (25), and (28) to obtain this correction in the form

$$\Delta E^{HFS}_{2\ SOPT} = -E^F\alpha^2\frac{m_e^2}{W^2}\frac{8}{9\pi^2} \times\int_1^{\infty}d\xi\left(1+\frac{1}{2\xi^2}\right)\frac{\sqrt{\xi^2-1}}{\xi^2} \times\int_1^{\infty}d\eta\left(1+\frac{1}{2\eta^2}\right)\frac{\sqrt{\eta^2-1}}{\eta^2}H\left(\xi,\eta,\frac{m_e}{W}\right), \quad (31)$$

$$H\left(\xi,\eta,\frac{m_e}{W}\right) = \frac{1}{\left(1+\frac{m_e\xi}{W}\right)^2}\frac{\eta^2}{\left(1+\frac{m_e\eta}{W}\right)^2}$$

$$\times\left[\frac{1}{\frac{W}{m_e\xi}+\frac{W}{m_e\eta}+\frac{W^2}{m_e^2\xi\eta}}-\ln\frac{\left(\frac{W}{m_e\xi}+\frac{W}{m_e\eta}+\frac{W^2}{m_e^2\xi\eta}\right)}{\left(1+\frac{W}{m_e\xi}\right)+\left(1+\frac{W}{m_e\eta}\right)}\right]$$

$$+\eta^2\left[\frac{5}{2\left(1+\frac{m_e\xi}{W}\right)^2\left(1+\frac{m_e\eta}{W}\right)^2}\right. \tag{32}$$

$$\left.-\frac{1}{\left(1+\frac{m_e\xi}{W}\right)^2\left(1+\frac{m_e\eta}{W}\right)^3}-\frac{1}{\left(1+\frac{m_e\xi}{W}\right)^3\left(1+\frac{m_e\eta}{W}\right)^2}\right]$$

$$+\frac{W^2}{m_e^2}\left[\frac{1}{\left(1+\frac{m_e\xi}{W}\right)^2}\left(1-\ln\left(1+\frac{m_e\xi}{W}\right)\right)\right.$$

$$\left.-\frac{5}{2\left(1+\frac{m_e\xi}{W}\right)^2}+\frac{1}{\left(1+\frac{m_e\xi}{W}\right)^3}-\frac{1}{\left(1+\frac{m_e\xi}{W}\right)}\right].$$

The corresponding contribution was estimated at

$$\Delta E^{HFS}_{2\ SOPT} = 0.002\ \text{meV}. \tag{33}$$

Second-order perturbation theory also gives other relativistic corrections of order $(Z\alpha)^6$, including recoil effects, which were studied in [30, 31, 32]. Their values are listed in the table.

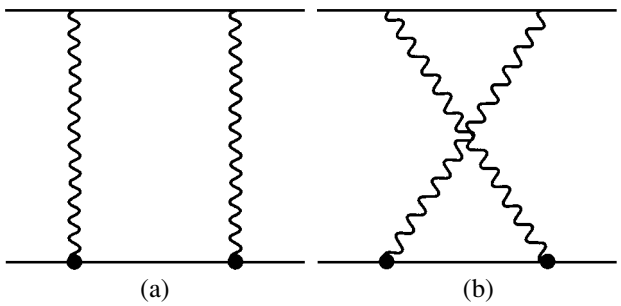

Fig. 2. Proton structure corrections of order $(Z\alpha)^5$. The solid circle is the proton vertex operator.

4. PROTON STRUCTURE AND VACUUM POLARIZATION EFFECTS

Corrections for the structure of the proton make a large relative contribution to the energy spectrum of muonic hydrogen compared with its electronic counterpart, because the mass ratio between the electron and muon is $m_e/m_1 = 4.83633210(15)\times10^{-3}$ [1]. These corrections to the superfine structure of μp are determined by the one-loop diagrams shown in Fig. 2.

In order to construct the quasi-potential that corresponds to these two diagrams, let us write the protonic tensor in the form

$$M^{(p)}_{\mu\nu} = \bar u(q_2)\left[\gamma_\mu F_1+\frac{i}{2m_2}\sigma_{\mu\omega}k^\omega F_2\right]$$
$$\times\frac{\hat p_2-\hat k+m_2}{(p_2-k)^2-m_2^2+i0}\left[\gamma_\nu F_1-\frac{i}{2m_2}\sigma_{\nu\lambda}k^\lambda F_2\right]u(p_2), \tag{34}$$

where p_2 and q_2 are the 4-momenta of the proton in the initial and final states. Sequentially projecting the muon and proton onto the $^1S_0$ and $^3S_1$ states with the use of the projection operators

$$\hat\pi(^1S_0) = [u(p_2)\bar v(p_1)]_{S=0} = \frac{(1+\gamma^0)}{2\sqrt{2}}\gamma_5,$$
$$\hat\pi(^3S_1) = [u(p_2)\bar v(p_1)]_{S=1} = \frac{(1+\gamma^0)}{2\sqrt{2}}\hat\epsilon \tag{35}$$

(ϵ^μ is the polarization vector of the state with spin $S=1$) and ignoring the momenta of the relative motion of the particles in the initial and final states, we obtain the following contribution to the superfine structure:

$$\Delta E^{HFS}_{\rm str} = E^F\frac{Z\alpha m_1m_2}{8\pi n^3(1+\kappa)}\delta_{l0}$$
$$\times\int\frac{id^4k}{\pi^2(k^2)^2}\left[\frac{16k^6k_0^2}{m_2^2}F_2^2+\frac{32k^8}{m_2^2}F_2^2-64k^2k_0^4F_2^2\right.$$
$$+16k^4k_0^2F_1^2+128k^4k_0^2F_1F_2+64k^4k_0^2F_2^2+32k^6F_1^2 \tag{36}$$
$$\left.+64k^6F_1F_2\right]\frac{1}{(k^4-4m_1^2k_0^2)(k^4-4m_2^2k_0^2)}.$$

Let us pass to the integration over the four-dimensional Euclidean space in (36),

$$\int d^4k = 4\pi\int k^3dk\int\sin^2\phi\,d\phi,\quad k_0 = k\cos\phi. \tag{37}$$

After the analytic integration in angle ϕ, we can repre-

Corrections of order α^5 and α^6 to the hyperfine structure (HFS) of the ground state of the muonic hydrogen atom

Contribution to the HFS of the μp atom	Contribution, meV	Refs.
Fermi energy E^F	182.443	[18], (12)
Correction for the anomalous magnetic moment of the muon $a_\mu E^F$ of order α^5, α^6	0.213	[18]
Relativistic correction $(3/2)(Z\alpha)^2 E^F$ of order α^6	0.015	[43]
Relativistic and radiative corrections for recoil taking into account κ of the nucleus of order α^6	0.014	[30]
Contribution of one-loop electronic polarization of the vacuum to 1γ interaction of order α^5	0.398	(18)
Contribution of one-loop muonic polarization of the vacuum to 1γ interaction of order α^6	0.004	(19)
Second-order perturbation theory corrections determined by the polarization of the vacuum of orders α^5 and α^6	0.797	(30) + (33)
Correction for the structure of the nucleus of order α^5	–1.215	[22], (40)
Correction for the structure of the nucleus of order α^6	–0.014	[8]
Contribution of the electronic polarization of the vacuum + corrections for the structure of the nucleus of order α^6	–0.021	(43)
Contribution of the two-loop electronic polarization of the vacuum to 1γ interaction of order α^6	0.003	(21) + (24)
Correction for the intrinsic muon energy + corrections for the structure of the nucleus of order α^6	0.008	(50)
Vertex corrections + corrections for the structure of the nucleus of order α^6	–0.014	(61)
Jellyfish diagram correction + corrections for the structure of the nucleus of order α^6	0.004	(66)
Correction for the hadronic polarization of the vacuum of order α^6	0.004	(45)
Correction for the polarizability of the proton of order α^5	0.084	[16]
Contribution of weak interaction	0.002	[36]
Total correction	182.725 ± 0.062	

sent (36) in the form of a one-dimensional integral in loop momentum k,

$$\Delta E^{HFS}_{\rm str} = -E^F \frac{Z\alpha}{8\pi n^3(1+\kappa)}\delta_{l0}\int_0^\infty \frac{dk}{k}V(k),$$

$$V(k) = \frac{2F_2^2k^2}{m_1m_2} + \frac{\mu}{(m_1-m_2)k(k+\sqrt{4m_1^2+k^2})}$$

$$\times\left[-128F_1^2m_1^2 - 128F_1F_2m_1^2 + 16F_1^2k^2 + 64F_1F_2k^2 + 16F_2^2k^2 + \frac{32F_2^2m_1^2k^2}{m_2^2} + \frac{4F_2^2k^4}{m_1^2} - \frac{4F_2^2k^4}{m_2^2}\right] \quad (38)$$

$$+ \frac{\mu}{(m_1-m_2)k(k+\sqrt{4m_2^2+k^2})}[128F_1^2m_2^2 + 128F_1F_2m_2^2 - 16F_1^2k^2 - 64F_1F_2k^2 - 48F_2^2k^2].$$

To remove infrared divergence in (38), we must take into account the contribution of the iterative term of quasi-potential (8) to the superfine structure of the μp atom,

$$\Delta E^{HFS}_{\rm iter,\,str} = -\langle V_{1\gamma}\times G^f\times V_{1\gamma}\rangle^{HFS}_{\rm str} = -\frac{64}{3}\frac{\mu^4(Z\alpha)^5(1+\kappa)}{m_1m_2\pi n^3}\int_0^\infty\frac{dk}{k^2}, \quad (39)$$

where angle brackets denote averaging the interaction operator over the Coulomb wave function of the ground

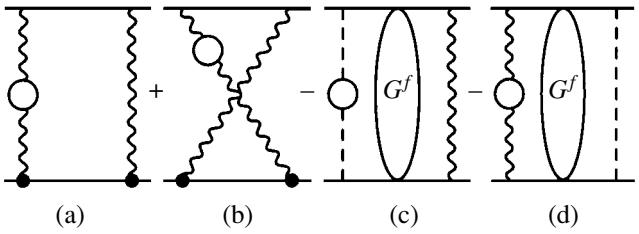

Fig. 3. Proton structure and vacuum polarization corrections of order $\alpha(Z\alpha)^5$. The dashed line denotes the Coulomb photon.

state, and index HFS is indicative of the separation of the superfine part in the iterative term of quasi-potential (8). The total contribution of (38) and (39) coincides with the result obtained in [22]. The integration in (38) and (39) was performed using the parametrization for the electromagnetic form factors of the proton obtained by analyzing elastic lepton–nucleon scattering [12]. The correction for the proton structure of order $(Z\alpha^5)$ was found to be

$$\Delta E^{HFS}_{\rm str} + \Delta E^{HFS}_{\rm iter,\,str} = -1.215\ \text{meV}. \tag{40}$$

The proton structure effects must be taken into account also in the amplitudes of a higher order in α shown in Fig. 3.

The contribution of diagrams 3a and 3b (Fig. 3) to the potential can be determined as with the amplitudes shown in Fig. 2. Replacement (9) then should be made in the propagator of one of the exchange photons. The corresponding correction to the superfine splitting of the energy level takes the form

$$\Delta E^{HFS}_{\rm str,\,VP} = -E^F\frac{Z\alpha}{8\pi(1+\kappa)n^3}2\frac{\alpha}{\pi} \times\int_0^1\frac{v^2(1-v^2/3)dv}{k^2(1-v^2)+4m_e^2}\int_0^\infty dk V_{VP}(k), \tag{41}$$

where the $V_{VP}(k)$ potential only differs from $V(k)$ in (38) by the additional multiplier k^2. Although the integral in (41) is finite, the amplitude terms of the quasi-potential in Figs. 3a and 3b should be augmented by two iterative terms shown in Figs. 3c and 3d. The first term $\langle V^c\times G^f\times\Delta V^{HFS}_{VP}\rangle$, which is of order $\alpha(Z\alpha)^4$, should be subtracted, because the 2γ amplitudes 3a and 3b (Fig. 3) reproduce the contribution of the lower order. The second term $\langle V^c_{VP}\times G^f\times V^{HFS}_{1\gamma}\rangle$, also of order $\alpha(Z\alpha)^4$, is similar in structure to equation (29) of second-order perturbation theory. The contributions of the specified iterative terms to the superfine structure of the μp atom coincide,

$$\begin{aligned}\Delta E^{HFS}_{\rm iter,\,VP+str} &= -2\langle V^C\times G^F\times\Delta V^{HFS}_{VP}\rangle^{HFS}\\ &= -2\langle V^c_{VP}\times G^f\times\Delta V^{HFS}_{\gamma}\rangle^{HFS}\\ &= -E^F\frac{4(Z\alpha)\mu\alpha}{m_e\pi^2}\int_0^\infty dk\int_0^1\frac{v^2(1-v^2/3)dv}{k^2(1-v^2)+1},\end{aligned} \tag{42}$$

where we additionally reduced the integration momentum k to the dimensionless form with the help of the electron mass m_e. The effect of the proton structure and vacuum polarization on the 2γ exchange amplitudes in the superfine structure of the μp atom then amounts to

$$\Delta E^{HFS}_{VP,\,\rm str} + 2\Delta E^{HFS}_{\rm iter,\,VP+str} = -0.021\ \text{meV}. \tag{43}$$

The contribution of the hadronic polarization of the vacuum to the superfine structure of the ground state of μp was studied in [33]. Here, we represent it in a different form using (39) and (42),

$$\Delta E^{HFS}_{HVP} = -E^F\frac{\alpha(Z\alpha)}{4\pi^2(1+\kappa)} \times\int_{4m_\pi^2}^\infty\frac{\rho(s)ds}{k^2+s}\int_0^\infty dk V_{VP}(k). \tag{44}$$

Partitioning the entire range of the integration in s into the intervals within which the cross section of the e^+e^- annihilation into hadrons

$$\rho(s) = \sigma^h(e^+e^-\longrightarrow \text{hadrons})/3s\sigma_{\mu\mu}$$

is known from experiment [34], we can perform the integration in (44). Contribution (44) coincides with the result obtained in [33],

$$\Delta E^{HFS}_{HVP} = 0.004\ \text{meV}. \tag{45}$$

5. PROTON STRUCTURE EFFECTS AND ENERGETIC PROPER AND VERTEX CORRECTIONS OF ORDER $\alpha(Z\alpha)^5$

There exist several other important contributions of order α^6 that are determined by the diagrams shown in Figs. 4 and 5. The radiative corrections of these amplitudes of order $\alpha(Z\alpha)^5$, including recoil effects, were studied earlier for both the Lamb shift and the superfine structure of hydrogen-like atoms [18, 35, 36]. The Fried–Yennie calibration [37–39], in which the amplitudes on the mass surface do not contain infrared divergences, can conveniently be used for radiative photons. The infrared finiteness of Feynman diagrams in this

calibration allows the standard subtraction procedure to be performed on the mass surface without introducing the mass of the photon. Consider radiative corrections that are determined by the energetic embeddings proper into the muon line. The renormalized mass operator in the Fried–Yennie calibration is [18]

$$\Sigma^R(p) = \frac{\alpha}{\pi}(\hat{p}-m)^2\int_0^1 dx\frac{-3\hat{p}x}{m_1^2x+(m_1^2-p^2)(1-x)}. \quad (46)$$

Performing embeddings (46) in the lepton tensor present in two-photon exchange interactions and using projection operators (35), we can construct the superfine part of the quasi-potential for the diagrams shown in Fig. 4. As previously, the proton–photon interaction vertex should be determined by the electric and magnetic form factors, because the characteristic loop momenta are on the order of the muon mass. The convolution of the proton and lepton tensors in Lorentzian indices and calculation of the traces of Dirac γ matrices were performed using the Form system [40]. Going over to the Euclidean space of variable k allows the correction to the superfine structure of the μp atom to be written as

$$\Delta E^{HFS}_{2\gamma,\,SE} = \frac{(Z\alpha)^5\mu^3}{\pi^2 n^3}\delta_{l0}\frac{\alpha}{\pi}\int_0^1 x\,dx\int_0^\infty k\,dk \times\int_0^\pi \sin^2\phi\,d\phi\,V_{SE}(k,\phi,x), \quad (47)$$

$$V_{SE}(k,\phi,x) = \frac{1}{(k^2+4m_2^2\cos^2\phi)[(xm_1^2+\bar{x}k^2)^2+4m_1^2\bar{x}^2k^2\cos^2\phi]} \times\left\{-\frac{4m_1^2}{m_2^2}k^2F_2^2(x+6\bar{x})\cos^2\phi-\frac{8m_1^2}{m_2^2}k^2xF_2^2 + 16m_1^2F_2\cos^4\phi(4F_1\bar{x}-F_2x-2F_2\bar{x}) + 16m_1^2\cos^2\phi(F_1^2x+6F_1^2\bar{x}+2F_1F_2x+8F_1F_2\bar{x} + F_2^2x+2F_2^2\bar{x})+32m_1^2xF_1(F_1+F_2)-\frac{4k^4}{m_2^2}F_2^2\bar{x}\cos^2\phi - \frac{8k^4}{m_2^2}F_2^2\bar{x}-16k^2F_2^2\bar{x}\cos^4\phi + 16k^2\bar{x}\cos^2\phi(F_1^2+4F_1F_2+F_2^2) + 32k^2F_1\bar{x}(F_1+F_2)\right\}. \quad (48)$$

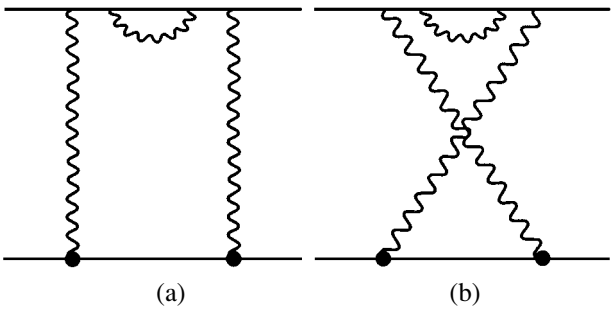

Fig. 4. Proton structure and intrinsic muon energy effects of order $\alpha(Z\alpha)^5$.

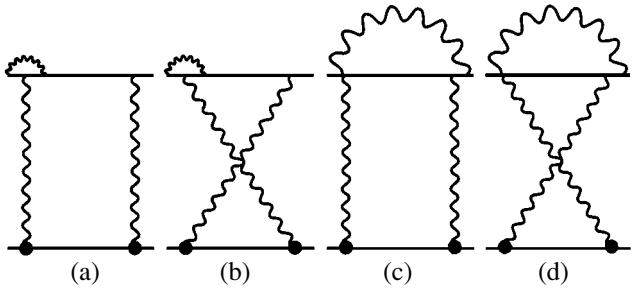

Fig. 5. Proton structure effects and vertex muon corrections of order $\alpha(Z\alpha)^5$.

After the integration in ϕ in (47), we obtain the representation of contribution (47) in the form of the integrals in parameter x and momentum k for use in numerical calculations,

$$\Delta E^{HFS}_{2\gamma,\,SE} = E^F\frac{m_1m_2\alpha(Z\alpha)}{\pi^2(1\kappa)n^3}\delta_{l0}\int_0^1 x\,dx\int_0^\infty dk \times\left\{\left[-\frac{8F_2^2k^2}{m_2^2}+32F_1(F_1+F_2)\right]\frac{1}{h_1(k,x)} + \left[-\frac{k^3F_2^2}{m_2^4}-\frac{6m_1^2k^3F_2^2\bar{x}}{m_2^4(xm_1^2+\bar{x}k^2)}+\frac{4k}{m_2^2}(F_1^2+4F_1F_2+F_2^2)\right] \times\left(\frac{1}{h_2(k,x)}-\frac{k}{h_1(k,x)}\right) + \left[\frac{2km_1^2}{m_2^2}F_2(2F_1+F_2)\bar{x}-\frac{kF_2^2}{m_2^2}(xm_1^2+\bar{x}k^2)\right] \times\left[\frac{2}{h_2^2(k,x)}-\frac{k^2}{m_2^2(xm_1^2+\bar{x}k^2)}\left(\frac{1}{h_2(k,x)}-\frac{k}{h_1(k,x)}\right)\right]\right\}, \quad (49)$$

$$h_1(k,x) = k\sqrt{4m_1^2\bar{x}^2k^2 + (xm_1^2 + \bar{x}k^2)^2} + (xm_1^2 + \bar{x}k^2)\sqrt{4m_2^2 + k^2},$$

$$h_2(k,x) = \sqrt{4m_1^2\bar{x}^2k^2 + (xm_1^2 + \bar{x}k^2)^2} + (xm_1^2 + \bar{x}k^2).$$

Contribution (49) calculated with the use of the F_1 and F_2 form factors from [12] was

$$\Delta E^{HFS}_{2\gamma,\,SE} = 0.008\ \text{meV}. \tag{50}$$

Consider calculations of vertex corrections. The renormalized equation for the one-particle vertex operator in the Fried–Yennie calibration was obtained in [41] ($p^2 = m_1^2$),

$$\Lambda^R_\mu(p, p-k) = \frac{\alpha}{4\pi}\int_0^1 dx\int_0^1 dz\left[\frac{F^{(1)}_\mu}{\Delta} + \frac{F^{(2)}_\mu}{\Delta^2}\right], \tag{51}$$

where

$$\Delta = m_1^2x + 2pk(1-x)z - k^2z(1-xz),$$

and the $F^{(1)}_\mu$ and $F^{(2)}_\mu$ functions were defined in [41]. In accordance with (51), the lepton tensor is separated into two terms,

$$M^{(l)(1)}_{\mu\nu} = \frac{\bar{v}(p_1)F^{(1)}_\nu(-\hat{p}_1 - \hat{k} + m_1)\gamma_\mu v(q_1)(k^2 - 2k^0m_1)[m_1^2x - k^2z(1-xz) + 2m_1k^2\bar{x}^2]}{(k^4 - 4k_0^2m_1^2)[(m_1^2x - k^2z(1-xz))^2 - 4m_1^2k_0^2\bar{x}^2z^2]}, \tag{52}$$

$$M^{(l)(2)}_{\mu\nu} = \frac{\bar{v}(p_1)F^{(2)}_\nu(-\hat{p}_1 - \hat{k} + m_1)\gamma_\mu v(q_1)(k^2 - 2k^0m_1)[m_1^2x - k^2z(1-xz) + 2m_1k^2\bar{x}^2]^2}{(k^4 - 4k_0^2m_1^2)[(m_1^2x - k^2z(1-xz))^2 - 4m_1^2k_0^2\bar{x}^2z^2]^2}. \tag{53}$$

For simplicity, we consider the main contribution in m_1/m_2 only, which allows us to represent the first part of the vertex correction of order $F^{(1)}_\mu$ and the second part of order $F^{(2)}_\mu$ in the form

$$\Delta E^{HFS}_{2\gamma,\,\text{vert1}} = -E^F\left(\frac{\alpha}{\pi}\right)^2\frac{8m_1m_2}{(1+\kappa)\pi n^3}\int_0^1 dx\int_0^1 dz\int_0^\pi \sin^2\phi\,d\phi\int_0^\infty k\,dk \times \frac{V_1(x,k,\phi)[F_1(F_1+F_2) - (1+\kappa)]}{(k^2 + 4m_1^2\cos^2\phi)(k^2 + 4m_2^2\cos^2\phi)[[m_1^2x + k^2z(1-zx)]^2 + 4m_1^2k^2\cos^2\phi\,\bar{x}^2z^2]}, \tag{54}$$

$$V_1(x,k,\phi) = -2m_1^4x^2(1-x) + k^2m_1^2(6x^3z^2 - 8x^2z^2 - 3x^2z + 8xz - 3x) + k^4(4x^3z^4 - 6x^2z^4 - 5x^2z^3 + 12xz^3 - 2xz^2 - 6z^2 + 3z), \tag{55}$$

$$\Delta E^{HFS}_{2\gamma,\,\text{vert 2}} = -E^F\left(\frac{\alpha}{\pi}\right)^2\frac{32m_1^3m_2}{(1+\kappa)\pi n^3}\int_0^1 x(1-x)dx\int_0^1 dz\int_0^\pi \sin^2\phi\,d\phi\int_0^\infty k^3\,dk \times \frac{V_2(x,k,\phi)F_1(F_1+F_2)}{(k^2 + 4m_1^2\cos^2\phi)(k^2 + 4m_2^2\cos^2\phi)[[m_1^2x + k^2z(1-zx)]^2 + 4m_1^2k^2\cos^2\phi\,\bar{x}^2z^2]^2}, \tag{56}$$

$$V_2(x,k,\phi) = m_1^4x^2z(2z-1) - k^2m_1^2xz^2(4xz^2 - 2xz - 4z + 2) + k^4z^3(2x^2z^3 - x^2z^2 - 4xz^2 + 2xz + 2z - 1). \tag{57}$$

The analytic integration in ϕ in (54) and (56) and the subtraction of the iterative contribution

$$\Delta E^{HFS}_{\text{iter, }2\gamma\text{ vert}} = \langle V_{1\gamma}\times G^f\times V_{1\gamma}\rangle^{HFS}_{\text{vert}} = F^F\left(\frac{\alpha}{\pi}\right)^2\int_0^1 dz\int_0^1 dx\int_0^\infty dk\frac{4\mu}{k^2} \tag{58}$$

from the sum of (54) and (56) (we must take one photon, and the second photon must contain the superfine part of the potential with the magnetic form factor value at zero) yields the following equation for the vertex correction, which is determined by the diagrams shown in Figs. 5a and 5b:

$$\Delta E^{HFS}_{2\gamma,\text{ vert}} = -E^F\left(\frac{\alpha}{\pi}\right)^2\int_0^1 dx\int_0^1 dz\int_0^\infty dk$$

$$\times\left\{\frac{F_1(F_1+F_2)}{8k(1+\kappa)m_1^3m_2\bar{x}^2z^2}[-2m_1^4x^2\bar{x}\right.$$

$$+k^2m_1^2x(6x^2z^2-8xz^2-3xz+8z-3)$$

$$\left.+k^4z(4x^3z^3-6x^2z^3-5x^2z^2+12xz^2-2xz-6z-3)\right]$$

$$\times\left[-\frac{\sqrt{1+b^2}}{b(a^2-b^2)(b^2-c^2)}+\frac{\sqrt{1+a^2}}{a(a^2-b^2)(a^2-c^2)}\right.$$

$$\left.+\frac{\sqrt{1+c^2}}{c(b^2-c^2)(a^2-c^2)}\right]+\frac{F_1(F_1+F_2)x}{2(1+\kappa)m_1^3m_2k\bar{x}^3z^4} \tag{59}$$

$$\times[m_1^4x^2z(2z-1)-2k^2m_1^2xz^2(2xz^2-xz-2z+1)$$

$$+k^4z^3(2x^2z^3-x^2z^2-4xz^2+2xz+2z-1)]$$

$$\times\left[-\frac{\sqrt{1+b^2}}{b(a^2-b^2)(b^2-c^2)^2}+\frac{\sqrt{1+a^2}}{a(a^2-b^2)(a^2-c^2)}\right.$$

$$+\frac{1}{2c\sqrt{1+c^2}(b^2-c^2)(a^2-c^2)}-\frac{\sqrt{1+c^2}}{2c^3(b^2-c^2)(a^2-c^2)}$$

$$\left.+\frac{\sqrt{1+c^2}}{c(b^2-c^2)(a^2-c^2)}+\frac{\sqrt{1+c^2}}{c(b^2-c^2)(a^2-c^2)^2}\right]+\frac{4\mu}{k^2}\Bigg\},$$

$$a^2=\frac{k^2}{4m_1^2},\quad b^2=\frac{k^2}{4m_2^2},$$
$$c^2=\frac{[m_1^2x+k^2z(1-xz)]^2}{4m_1^2k^2\bar{x}^2z^2}. \tag{60}$$

Vertex corrections (59) were found to be

$$\Delta E^{HFS}_{2\gamma,\text{ vert}} = -0.014\text{ meV}. \tag{61}$$

Another vertex-type diagram with one enveloping photon and two exchange photons is a "jellyfish"-type diagram. Its contribution to the energy spectrum is of order $\alpha(Z\alpha)^5$. In the region of small loop momenta, this diagram gives a finite result in the Fried–Yennie calibration. The equation for the lepton tensor present in amplitudes 5c and 5d (Fig. 5) obtained in [36] was

$$L^{(\mu)}_{\mu\nu} = \frac{\alpha}{4\pi}\int_0^1 xdx\int_0^1(1-z)dz\sum_{n=1}^3\frac{M^{(n)}_{\mu\nu}}{\Delta^n}, \tag{62}$$

where Δ had the same form as in (51), and the $M^{(n)}_{\mu\nu}$ tensor functions were reported in [36]. The character of the further transformations of amplitudes 5c and 5d (Fig. 5) in constructing the superfine part of the muon–proton interaction operator is the same as with the other amplitudes shown in Figs. 4 and 5. Omitting the details of these transformations performed using form [40], let us write the contributions to the superfine structure of the μp atom that correspond to the three $M^{(n)}_{\mu\nu}$ functions. In the main order in m_1/m_2, they have the form

$$\Delta E^{HFS}_{1,\text{ jellyfish}} = -\frac{64\alpha(Z\alpha)^5\mu^3\delta_{l0}}{\pi^3n^3}$$
$$\times\int_0^1 xdx\int_0^1(1-z)(1-3xz)$$
$$\times\int_0^\infty kdkF_1(F_1+F_2)\int_0^\pi\frac{\sin^2\phi d\phi}{(k^2+4m_2^2\cos^2\phi)}$$
$$\times\frac{[m_1^2x+k^2z(1-xz)]}{[m_1^2x+k^2z(1-xz)]^2+4m_1^2k^2\cos^2\phi\bar{x}^2z^2}, \tag{63}$$

$$\Delta E^{HFS}_{2,\,\text{jellyfish}} = -\frac{128\alpha(Z\alpha)^5\mu^3\delta_{l0}}{3\pi^3 n^3}\int_0^1 x\,dx\int_0^1(1-z)\,dz\int_0^\infty k\,dk F_1(F_1+F_2)$$
$$\times\int_0^\pi \frac{\sin^2\phi\,d\phi}{(k^2+4m_2^2\cos^2\phi)}\frac{[m_1^2x+k^2z(1-xz)]^2[k^2xz^2(1-xz)+m_1^2(x^2z+2xz-x-3z)]}{\{[m_1^2x+k^2z(1-xz)]^2+4m_1^2k^2\cos^2\phi\bar{x}^2z^2\}^2}, \quad (64)$$

$$\Delta E^{HFS}_{3,\,\text{jellyfish}} = \frac{512\alpha(Z\alpha)^5\mu^3\delta_{l0}}{3\pi^3 n^3}$$
$$\times\int_0^1 x\,dx\int_0^1(1-z)z^2\,dz\int_0^\infty k^3\,dk m_1^2F_1(F_1+F_2)$$
$$\times(x+xz-x^2z-1)\int_0^\pi\frac{\sin^2\phi\,d\phi}{(k^2+4m_2^2\cos^2\phi)} \quad (65)$$
$$\times\frac{[m_1^2x+k^2z(1-xz)]^3}{\{[m_1^2x+k^2z(1-xz)]^2+4m_1^2k^2\cos^2\phi\bar{x}^2z^2\}^3}.$$

The integration in angle ϕ in (63)–(65) can be performed analytically. Without writing down the resulting equations, we will give the eventual numerical result for the contribution of diagrams 5c and 5d (Fig. 5) to the superfine structure of the μp atom,

$$\Delta E^{HFS}_{\text{jellyfish}} = \sum_{n=1}^{3}\Delta E^{HFS}_{n,\,\text{jellyfish}} = 0.004\ \text{meV}. \quad (66)$$

Note that, in the point proton approximation, when the form factors of the nucleus, which enter the amplitudes shown in Figs. 4 and 5, are replaced by their values at zero,

$$F_1(0) = 1, \quad F_2(0) = \kappa,$$

contributions (64)–(66) increase approximately twofold.

6. CONCLUSIONS

We calculated various quantum electrodynamic effects, structure effects, and the effects of the polarization of the proton and the hadronic polarization of the vacuum on the superfine structure of the muonic hydrogen atom. The contributions of orders α^5 and α^6 were considered. It was taken into consideration that the $\mu\alpha/m_e$ ratio was close to one for some corrections; for this reason, we did not increase the order in α in these contributions. The numerical results for the contributions obtained in this work are listed in the table. The table also contains the quantum electrodynamic corrections to the Fermi energy for the superfine structure of the hydrogen-like system related to the anomalous magnetic moment of the muon $a_\mu E^F$ [18] (we used the experimental value for the anomalous magnetic moment of the muon $a_\mu^{\text{exp}} = 11659203(8)\times10^{-10}$ [42]), the Breit relativistic correction of order $(Z\alpha)^6$ [43], the relativistic and radiative effects of the same order with recoil $(Z\alpha)^6 m_1/m_2$ taking into account the anomalous magnetic moment of the proton [30], the correction for the structure of the nucleus of order $(Z\alpha)^5$ [22] (see Eq. (40) above), the correction for the structure of the nucleus of order $(Z\alpha)^6\ln(Z\alpha)^2$ [8], and the contributions of the hadronic polarization of the vacuum [33], proton polarizability [16], and weak interaction via exchange of a Z boson [18].

Consider several points related to the calculations performed above.

(1) For muonic hydrogen, of great importance are the effects of vacuum polarization, which modify the spin-dependent part of the one-photon interaction potential.

(2) All loop amplitudes were calculated taking into account the structure of the proton with the help of electromagnetic form factors. The point proton approximation substantially (approximately twofold) exaggerates the results.

(3) The energetic corrections proper and the vertex corrections of order $\alpha(Z\alpha)^5$ were calculated using the equations for the corresponding lepton factors in the amplitude terms of the quasi-potential obtained by Eides, Grotch, and Shelyuto in the Fried–Yennie calibration. We augmented these equations where necessary by subtracting the iterative terms of the interparticle interaction potential.

The first result for the superfine structure of the ground state of muonic hydrogen given in the table can serve as a reliable estimate in performing the corresponding experiment, the designing of which is currently under way [23]. As has been mentioned above, the correction values were obtained with a 0.001 meV accuracy. The theoretical error caused by the uncertainties in the fundamental parameters (fine structure constant, proton magnetic moment, etc.) entering the Fermi energy is around 10^{-5} meV. The other part of the theoretical error is related to corrections higher in order.

This part can be estimated from the leading contribution of a higher order in α and m_1/m_2,

$$\frac{\alpha}{\pi}(Z\alpha)^2\ln(Z\alpha)^2 \approx 0.0005 \text{ meV}$$

(we used the fine structure constant $\alpha^{-1} = 137.03599976(50)$ [1]).

The total contribution to the superfine structure of muonic hydrogen obtained in this work (see table) is useful to compare with the superfine splitting calculated in the point proton approximation, in which only the electromagnetic form factors of the proton at zero are taken into account,

$$G_E(0) = 1, \quad G_M(0) = \mu$$

(the only exception is the Zemach correction). The superfine structure of the ground state can then be obtained only with accuracy $O((m_1/m_2)\alpha^6)$ in the form [18]

$$\Delta E^{HFS}(QED) = E^F\Bigg\{1 - 2\mu\alpha R_p + \frac{3}{2}(Z\alpha)^2 + a_\mu$$

$$+ \alpha(Z\alpha)\left(\ln 2 - \frac{5}{2}\right) + \frac{1}{1+\kappa}\Bigg[-\frac{3\alpha}{\pi}\frac{m_1m_2}{m_2^2 - m_1^2}\ln\frac{m_2}{m_1}$$

$$+ (Z\alpha)^2\frac{\mu^2}{m_1m_2}\Bigg[\left(2(1+\kappa) + \frac{7\kappa^2}{4}\right)\ln(Z\alpha)^{-1} \qquad (67)$$

$$-\left(8(1+\kappa) - \frac{\kappa(12-11\kappa)}{4}\right)\ln 2 + 3\frac{11}{18} + \frac{\kappa(11+31\kappa)}{36}\Bigg]\Bigg]$$

$$-\frac{2}{3}(Z\alpha)^2\ln(Z\alpha)^{-2}m_1^2r_p^2\Bigg\} = 181.177 \text{ meV}.$$

The significant discrepancy between this value and the value obtained by us, 182.725 meV, is explained by several reasons, including the modification of the Breit potential for muonic hydrogen as a result of the electronic polarization of the vacuum, proton structure effects in constructing two- and three-photon interparticle interaction potentials, and the inclusion of the hadronic polarization of the vacuum and proton polarizability in our calculations. A further improvement of the theoretical result given in the table will primarily be related to corrections for the structure and polarizability of the proton, the theoretical error in these values being approximately 340 ppm. The overwhelming part of this error is determined by corrections for the structure of the nucleus, which are of order $(Z\alpha)^5$ (the Zemach correction). It follows that performing experimental measurements of the superfine splitting of the 1*S* and 2*S* energy levels in muonic hydrogen with an accuracy of 30 ppm would allow us to obtain a more reliable (accurate to 10^{-3}) value for the Zemach radius, which can be used to improve the theoretical result for the superfine structure of the ground state of the hydrogen atom and to more accurately estimate the possible contribution of proton polarizability. Increasing the number of problems related to studying the superfine structure of the energy spectra of hydrogen-like atoms, including the superfine structure of excited energy levels [44] and the superfine structure of the "new" simplest atomic systems, will decrease the uncertainties in the fundamental physical parameters and increase the precision of verifying the Standard Model in low-energy physics.

ACKNOWLEDGMENTS

The authors thank D.D. Bakalov, A.V. Borisov, A.I. Studenikin, A.L. Kataev, V.V. Fil'chenkov, and I.B. Khriplovich for useful advice. This work was financially supported by the program "Universities of Russia—Basic Research" (project no. UR.01.02.016).

REFERENCES

1. P. J. Mohr and B. N. Taylor, Rev. Mod. Phys. **72**, 351 (2000).
2. Yu. L. Sokolov, Usp. Fiz. Nauk **169**, 559 (1999) [Phys. Usp. **42**, 481 (1999)].
3. F. Kottmann, F. Biraben, C. A. N. Conde, *et al.*, in *Proceedings of QED 2000 Second Workshop on Quantum Electrodynamics and Physics of the Vacuum*, Ed. by G. Cantatore (New York, 2001), AIP Conf. Proc., Vol. 564, p. 13; R. Pohl, F. Biraben, C. A. N. Conde, *et al.*, Hyperfine Interact. **127**, 161 (2000).
4. A. C. Zemach, Phys. Rev. **104**, 1771 (1956).
5. F. Guerin, Nuovo Cimento A **50**, 1 (1967).
6. G. M. Zinov'ev, B. V. Struminskiĭ, R. N. Faustov, and V. L. Chernyak, Yad. Fiz. **11**, 1284 (1970) [Sov. J. Nucl. Phys. **11**, 715 (1970)].
7. A. I. Mil'shteĭn, S. S. Petrosyan, and I. B. Khriplovich, Zh. Éksp. Teor. Fiz. **109**, 1146 (1996) [JETP **82**, 616 (1996)]; I. B. Khriplovich and A. I. Milshtein, nucl-th/0304069.
8. S. G. Karshenboim, Phys. Lett. A **225**, 97 (1997).
9. J. L. Friar, Can. J. Phys. **80**, 1337 (2002).
10. A. Pineda, Phys. Rev. C **67**, 025201 (2003).
11. R. N. Faustov and A. P. Martynenko, Phys. Rev. A **67**, 052506 (2003).
12. G. G. Simon, Ch. Schmitt, F. Borkowski, *et al.*, Nucl. Phys. A **333**, 381 (1980).
13. S. D. Drell and J. D. Sullivan, Phys. Rev. **154**, 1477 (1967).
14. J. Bernabeu and T. E. O. Ericson, Z. Phys. A **309**, 213 (1983).
15. A. P. Martynenko and R. N. Faustov, Yad. Fiz. **63**, 915 (2000) [Phys. At. Nucl. **63**, 845 (2000)].
16. E. V. Cherednikova, R. N. Fastov, and A. P. Martynenko, Nucl. Phys. A **703**, 365 (2002).

17. H. Hellwig, R. F. C. Vessot, M. W. Levine, *et al.*, IEEE Trans. Instrum. Meas. **19**, 200 (1970).
18. M. I. Eides, H. Grotch, and V. A. Shelyuto, Phys. Rep. **342**, 62 (2001).
19. D. D. Bakalov, E. Milotti, C. Rizzo, *et al.*, in *Proceedings of the 3rd International Symposium on Weak and Electromagnetic Interactions in Nuclei (WEIN-92)* (1992), p. 656.
20. G. T. Bodwin and D. R. Yennie, Phys. Rev. D **37**, 498 (1988).
21. E. Borie and G. A. Rinker, Rev. Mod. Phys. **54**, 67 (1982).
22. K. Pachucki, Phys. Rev. A **53**, 2092 (1996).
23. A. Adamczak, D. Bakalov, K. Bakalova, *et al.*, Hyperfine Interact. **136**, 1 (2001).
24. A. P. Martynenko and R. N. Faustov, Teor. Mat. Fiz. **64**, 179 (1985).
25. V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 2nd ed. (Nauka, Moscow, 1980; Pergamon Press, Oxford, 1982).
26. R. N. Faustov, Fiz. Élem. Chastits At. Yadra **3**, 238 (1972) [Sov. J. Part. Nucl. **3**, 119 (1972)].
27. M. I. Eides, H. Grotch, and V. A. Shelyuto, Phys. Rev. D **65**, 013003 (2001).
28. S. A. Zapryagaev, N. L. Manakov, and V. G. Pal'chikov, *Theory of Multicharge Ions with One and Two Electrons* (Énergoatomizdat, Moscow, 1985).
29. V. G. Ivanov and S. G. Karshenboĭm, Zh. Éksp. Teor. Fiz. **109**, 1219 (1996) [JETP **82**, 656 (1996)].
30. G. T. Bodwin, D. R. Yennie, and M. A. Gregorio, Rev. Mod. Phys. **57**, 723 (1985).
31. T. Kinoshita and M. Nio, Phys. Rev. D **53**, 4909 (1996).
32. A. P. Martynenko and R. N. Faustov, Zh. Éksp. Teor. Fiz. **115**, 1221 (1999) [JETP **88**, 672 (1999)].
33. A. P. Martynenko and R. N. Faustov, Yad. Fiz. **61**, 534 (1998) [Phys. At. Nucl. **61**, 471 (1998)].
34. R. R. Akhmetshin *et al.*, Phys. Lett. B **476**, 33 (2000); **509**, 217 (2001).
35. M. I. Eides, S. G. Karshenboim, and V. A. Shelyuto, Ann. Phys. (N.Y.) **205**, 231 (1991).
36. M. I. Eides, H. Grotch, and V. A. Shelyuto, Phys. Rev. A **63**, 052509 (2001).
37. A. A. Abrikosov, Zh. Éksp. Teor. Fiz. **30**, 96 (1956) [Sov. Phys. JETP **3**, 71 (1956)].
38. L. P. Gor'kov, Zh. Éksp. Teor. Fiz. **30**, 790 (1956) [Sov. Phys. JETP **3**, 762 (1956)].
39. H. M. Fried and D. R. Yennie, Phys. Rev. **112**, 1391 (1958).
40. J. A. M. Vermaseren, math-ph/0010025 (2000); http://www.nikhef.nl/~form.
41. M. I. Eides and V. A. Shelyuto, Eur. Phys. J. C **21**, 489 (2001).
42. G. W. Bennett, B. Bousquet, H. N. Brown, *et al.*, Phys. Rev. Lett. **89**, 101804 (2002).
43. G. Breit, Phys. Rev. **35**, 1447 (1930).
44. S. G. Karshenboim and V. G. Ivanov, Eur. Phys. J. D **19**, 13 (2002).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 98, No. 1, 2004, pp. 53–61.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 1, 2004, pp. 63–71.

PLASMA, GASES

Dipole–Dipole Interactions between Dust Grains in Plasmas¶

D. D. Tskhakaya[a,b] and P. K. Shukla[c,d,*]

[a]*Institute of Physics, Georgian Academy of Sciences, Tbilisi, 380077 Georgia*
[b]*Institute of Theoretical Physics, University of Innsbruck, A-6020 Innsbruck, Austria*
[c]*Institut für theoretische Physik IV, Fakultät für Physik und Astronomie, Ruhr-Universität Bochum, D-44780 Bochum, Germany*
[d]*Department of Plasma Physics, Umeå University, SE-90187 Umeå, Sweden*
**e-mail: ps@tp4.ruhr-uni-bochum.de*

Received May 5, 2003

Abstract—Complete screening of the negative dust grain charge by a cloud of trapped ions in plasmas is investigated. In the plasma electric field, the compound dust particle "dust grain + ion cloud" acquires a dipole moment due to displacement of the centers of positive and negative charges in the opposite directions within the compound particle. By analogy to the van der Waals attractive interaction potential, the dipole–dipole interactions of the compound dust particle can have an attractive behavior. It is shown that, for the electric field strengths typically observed in experiments, the dipole–dipole attractive force exceeds the shadowing force that is connected with the reciprocal interception of ions by the neighboring dust grains. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A cloud of dust particles in plasmas, confined by the walls (electrodes), is characterized by a self-organizing property that reveals itself as the capability of dust grains to form ordered spatial structures in the vicinity of electrodes [1–12]. The dust grains in a cloud usually have an electric charge of the same sign (negative); according to the general consideration, at large intergrain distances, such a capability of self-organization implies the existence of an attractive force between dust grains having the same polarity. In the past, various mechanisms had been proposed for the dust grain attraction in dusty plasmas. They are as follows:

(i) The attraction of dust grains in the wake potential [13–16]: the ions are focused in the negative potential region of the wake field behind a moving dust grain and provide a possibility for attracting the following negatively charged grain in a linear chain [17, 18].

(ii) The shadowing force [19, 20]: the reciprocal shadowing of a pair of dust grains in a nonstreaming plasma and, as a result, reciprocal interception of ions moving from the outside of the system of the grain pair, leading to a net momentum transfer that pushes the grains to meet each other. This, in effect, represents an attractive force between two dust grains.

(iii) When placed into an external electric field (for instance, in the field of another charged dust particle), the dust grain, considered a conductor, is polarized. The excess of charges with a definite sign on one side leads to an anisotropy of the plasma-particle flows to the dust-particle surface. Even when there is an equality of the ion and electron currents to the dust particle surface, the momentum transferred to the dust particle by the ions incident on the surface considerably exceeds the momentum brought by electrons. Therefore, an additional force exerted by the plasma flow acts on a dust particle in the electric field. This additional force has the same direction as the electric field and can exceed the electrostatic force acting on the dust particle in the electric field [21]. The force is proportional to the electric field strength. If the given grain is placed into the electric field of another grain and the distance between the grains is much larger than the Debye radius (which is the most interesting case for investigating the grain–grain interactions), then the electric field and the force become very small as a result of the Debye shielding.

Furthermore, under the condition of a volume distribution of dust particles, the neighboring dust grains surrounding the given grain from every side can intercept the ions flowing towards the grain, and the attractive forces described in items (ii) and (iii) must be substantially less.

It seems more consistent to relate the creation of the attractive force to the screening of the dust charge by a cloud of trapped ions [22]. Below, we assume that the dust electric charge is completely screened by an ion cloud. Such a possibility is investigated and predicted in [23, 24]. Taking into consideration large distances between the grains instead of bare dust grains, we can operate with the grains "dressed" in the jacket of an ion cloud. The system "grain + ion cloud" is said to be a compound (dust) particle in what follows. In an exter-

¶This article was submitted by the authors in English.

nal electric field, the centers of the negative and positive charges within the compound particle are displaced from each other and the compound particle acquires a dipole moment. The dipole–dipole interactions of the compound dust particle can have an attractive nature by analogy with the van der Waals interaction in solid-state physics.

This paper is devoted to a quantitative analysis of the attractive force acting between the compound dust particles. It is shown below that the attractive force connected with dipole–dipole interactions of compound particles can exceed the shadowing force [19, 20]. Hence, a special feature of the interaction potential of dust particles in plasmas must be the existence of some equilibrium distance between dust grains at which the forces of attraction and repulsion balance each other. The paper is organized as follows. In Section 2, we discuss the theory of ion trapping in the potential well and calculate the induced dipole moment in a self-consistent electric field in plasmas. An expression for the attractive force associated with dipole–dipole interactions is obtained. For typical laboratory conditions, the newly found dipole–dipole attractive force predominates over the shadowing force. Section 3 contains a summary and approximations required for developing the present theory.

2. THEORY

We assume plasma to be collisionless, which means that the ion mean free path is much larger than the plasma Debye length, $\lambda_{mfp} \gg \lambda_D$. In [25], the capture of particles by a nonstationary potential well in a collisionless plasma was proposed. A brief description of this nonstationary capture is given in [26]. The nonstationarity of the potential well means that the height of the walls forming the well increases in time and is saturated at some stationary value. Therefore, initially free particles, passing a distance on the order of magnitude of the extent of the well, can collide with the growing wall. After reflection, a particle can meet an analogous obstacle moving in the opposite direction. At the time a stationary well is established, a definite number of particles is captured by the well. The distribution function of trapped particles can be found from the continuity condition for the distribution function at the limiting level of the trapped-particle energy. At this level, the distribution function of trapped particles must be equal to the distribution function of free particles. In our case, capture of ions by the potential well occurs during the process of dust grain charging. The adiabaticity condition [25, 26] (i.e., the condition that the creation of the well proceeds slowly), which is necessary for analytical description of nonstationary particle trapping, is fulfilled: if the Debye radius λ_D exceeds the dust grain size a, then the characteristic time $\tau \approx \lambda_D/a\omega_{pi}$ of dust grain charging (which is the same as the characteristic time in which the potential well is created) is much larger than the time τ_i necessary for an ion to traverse the width of the well (here, ω_{pi} is the ion plasma frequency). As shown below, the width of the well is on the order of λ_D. The time τ_i can be estimated as follows: the potential of the grain is usually given by $|\varphi_0| \approx T_e/e$, where T_e is the electron temperature and $e > 0$ is the ion charge. For the average velocity of ions in the well, we then have

$$v_i \approx \sqrt{\frac{e|\varphi_0|}{m}} \approx \sqrt{\frac{T_e}{m}}$$

(where m is the ion mass), whence

$$\tau \approx \frac{\lambda_D}{v_i} \approx \frac{1}{\omega_{pi}}.$$

Therefore, the condition of the adiabaticity of ion capturing, $\tau \gg \tau_i$, is fulfilled at $\lambda_D \gg a$.

According to [25, 26], upon the adiabatic creation of the well, the distribution function of trapped particles (ions) is constant and equals the value of the distribution function of free particles (ions) at the limiting energy level of the trapped particles. The physical reason for this result is as follows: in a collisionless plasma, the trapped particles do not leave the well and the probability of finding them in the well is one.

To analyze the dust cloud formation, we consider the motion of ions in the field of a negatively charged dust grain. Dust grains are assumed to be absolutely absorbed and are considered spherical with a radius a much smaller than the Debye radius, $a \ll \lambda_D$ [27–30]. At the spherical symmetry of the grain field, the dependence of the ion effective potential energy on the distance r to the center of force, $r = 0$, is

$$U_{\text{eff}}(r; L) = \frac{L^2}{2mr^2} + U(r), \tag{1}$$

where

$$U(r) = e\varphi(r) = -e|\varphi(r)|$$

is the ion potential energy and $\varphi(r)$ is the electric potential. The angular momentum L is an integral of motion. From the equality

$$\frac{L^2}{m} = r^3\frac{dU(r)}{dr}, \tag{2}$$

we can find the extremum values of $U_{\text{eff}}(r; L)$. The qualitative dependence of r^3dU/dr on r is depicted in Fig. 1. It is due to the specific dependence of the potential energy $U(r)$ on r (see [31, pp. 255–266]). At short distances (for r smaller than the Debye radius λ_D), the potential energy $U(r)$ decreases as $1/r$, i.e., slower than $1/r^2$. For $r \gtrsim \lambda_D$, $U(r)$ decreases exponentially due to

the Debye screening, i.e., faster than $1/r^2$. The behavior of $U(r)$ at large distances ($r \gg \lambda_D$) significantly depends on the conditions at the dust grain surface. If the dust grain surface absorbs electrons and ions, the potential energy $U(r)$ decays as $1/r^2$ at $r \longrightarrow \infty$ (see [31, pp. 140–141; 32, 33]). In Fig. 1, the intersection points of the curve with the dashed horizontal lines indicate the extremum points of $U_{\text{eff}}(r; L)$. The characteristic values L_0, L_p, and L_k are determined as follows.

(1) Far from the grain, $r \gg \lambda_D$, the potential energy $U(r)$ can be written as [31, 32]

$$U(r) = U_\infty\left(\frac{a}{r}\right)^2, \tag{3}$$

where U_∞ is a constant. We then have

$$L_0^2 = m\left(r^3\frac{dU}{dr}\right)_{r\to\infty} = 2ma^2|U_\infty|. \tag{4}$$

At $L \le L_0$, the effective potential energy U_{eff} has only one extremum point, which corresponds to a minimum.

(2) The characteristic angular momentum L_p is determined from the condition that the maximum value of the effective potential energy is equal to the value of the effective potential energy on the grain surface, $r = a$ [33, 34]. L_p and the corresponding point $r_{\max}$ of the maximum of the effective potential energy can be found from the system of equations

$$U_{\text{eff}}(a; L_p) = U_{\text{eff}}(r_{\max}(L_p); L_p), \tag{5}$$

$$\frac{L_p^2}{m} = \left(r^3\frac{dU(r)}{dr}\right)_{r = r_{\max}(L_p)}. \tag{6}$$

Equation (5) yields

$$L_p^2 \approx 2ma^2|U(a)|\frac{1 - |U(r_{\max})|/|U(a)|}{1 - a^2/r_{\max}^2} \approx 2ma^2|U(a)|. \tag{7}$$

Usually, $r_{\max} > \lambda_D$. As the angular momentum L increases, the minimum point of $U_{\text{eff}}(r; L)$ moves away from the center and the distance between the extremum points decreases. We stress that only those trapped ions that have the angular momentum $L < L_p$ can reach the grain surface and be absorbed. At $L > L_p$, the edge of the well is far from the grain surface.

(3) The maximum value of r^3dU/dr is reached at a certain point r_k ($>\lambda_D$), where

$$\left[\frac{\partial}{\partial r}\left(r^3\frac{\partial U(r)}{\partial r}\right)\right]_{r_k} = 0, \tag{8}$$

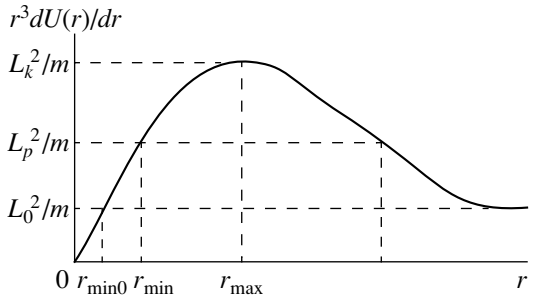

Fig. 1. A qualitative plot of the auxiliary function r^3dU/dr vs. r.

and r_k is always between the maximum and minimum points. The characteristic angular momentum L_k is defined as

$$L_k^2 = m\left(r^3\frac{dU}{dr}\right)_{r = r_k}. \tag{9}$$

At $L = L_k$, the extremum points coincide and the function $U_{\text{eff}}(r; L)$ has an inflection at this point. If $L > L_k$, the function $U_{\text{eff}}(r; L)$ decreases monotonically with increasing r.

(4) At $L > L_0$, Eq. (2) has two roots. For a more detailed description of ion motion, we must also determine the angular momentum L_a at which the small root (corresponding to the minimum of $U_{\text{eff}}(r; L)$) coincides with the radius a of the grain,

$$\frac{L_a^2}{m} = \left(r^3\frac{dU(r)}{dr}\right)_{r = a}. \tag{10}$$

Hence, if the angular momentum is in the range

$$L_0 < L < L_k, \tag{11}$$

then $U_{\text{eff}}(r; L)$ has both maximum and minimum points. The qualitative dependence of $U_{\text{eff}}(r; L)$ on r for various values of the angular momentum is shown in Fig. 2. We can now determine the surface that separates the regions of infinite and finite motion of ions in the velocity space (v_r, v_θ), where v_r and v_θ are the velocity components along and across the radial direction. The standard definition of the angular momentum is

$$L = mv_\theta r. \tag{12}$$

The ions with the angular momentum $L \le L_0$ (or with the velocity component $v_\theta \le (a/r)\sqrt{2|U_\infty|/m}$) and with a negative total energy

$$E(v_r, v_\theta, r) \le 0, \tag{13}$$

$$E(v_r, v_\theta, r) = \frac{mv_r^2}{2} + U_{\text{eff}}(r; L) \tag{14}$$

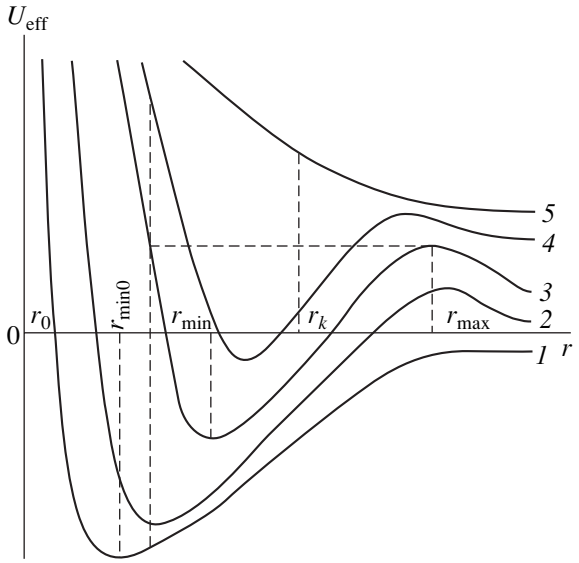

Fig. 2. A qualitative plot of $U_{\text{eff}}(r; L)$ vs. r for different values of the angular momentum: (*1*) $L \le L_0$, (*2*) $L = L_a$, (*3*) $L = L_p$, (*4*) $L_p < L < L_k$, and (5) $L = L_k$.

can be trapped in the potential well. The ions with the angular momentum in the range $L_0 < L < L_k$ can be trapped if their total energy satisfies the condition

$$E(v_r, v_\theta, r) \le U_{\text{eff}}(r_{\max}; L). \tag{15}$$

The ions with the angular momentum larger than L_k ($L > L_k$) are not trapped. The dependence of $|U_\infty|$ on the grain surface potential $|U(a)|$ for equal electron and ion temperatures ($T_e \approx T_i$) is depicted in [31, p. 317] and shows that one always has

$$\frac{|U_\infty|}{|U(a)|} < \frac{1}{3}.$$

With increasing $|U(a)|$, this ratio decreases. A more precise relation between $|U_\infty|$ and $|U(a)|$ can be established from the quasineutrality condition, in the case where the ion and electron densities are roughly equal. This occurs far from the grain ($r \gg \lambda_D$). Under this condition, a calculation quite similar to that given in [34] shows that for a nonisothermal plasma, $T_e \gg T$, and the inequality

$$|U_\infty| < \frac{1}{2}|U(a)| \tag{16}$$

is satisfied for the absorbing grains if

$$\frac{|U(a)|}{T_e} > \frac{1}{2}.$$

The latter relation is usually fulfilled with a large reserve both in laboratory and space plasmas [28]. From (13) and (16), it follows that the zero point r_0 of the effective potential energy $U_{\text{eff}}(r; L)$ at $L \le L_0$ is always close to the center ($r = 0$) in comparison with the grain surface, $r_0 < a$. Indeed, we find from Eq. (1) that $U_{\text{eff}}(r; L_0)$ at $r = a$ is negative (see curve *1* in Fig. 2),

$$U_{\text{eff}}(a; L_0) \le |U_\infty| - |U(a)| < 0, \quad L \le L_0. \tag{17}$$

Consequently, when L is smaller than a certain critical value L_0, the dust grain surface is within the well and the ions falling into the potential well are therefore immediately lost due to absorption onto the dust grain surface. Hence, the formation of trapped ion clouds that can shield the grain electric field is possible only for $L > L_a$.

According to the general theory [25, 26], for the stationary well, the distribution function of the trapped particles f_{itr} is constant and the value of f_{itr} is defined by the value of the distribution function of untrapped particles at the limiting energy, in our case, by the energy level $U_{\text{eff}}(r_{\max}; L)$. Considering the distribution function of untrapped ions as a Maxwell–Boltzmann one, we obtain

$$\begin{aligned} f_{itr} &= n_0(m/2\pi T_i)^{3/2} \\ &\times \exp(-U_{\text{eff}}(r_{\max}(L); L)/T_i). \end{aligned} \tag{18}$$

We emphasize that $r_{\max}$ here depends on the angular momentum L (see Eq. (2)). Because we are interested in distances not very large compared to λ_D, $r \lesssim r_{\max}$ (see also Eq. (21)), we can choose the Debye–Hückel form

$$U(r) = -|U(a)|\frac{a}{r}\exp\left(-\frac{r-a}{\lambda_D}\right) \tag{19}$$

for the potential energy of ion interactions with a dust grain. It should be stressed that the Debye–Hückel law holds even in the nonlinear regime [23]. It is somewhat modified by the ion flow [16]. The latter also produces a wake field, which is not the focus of this paper. The dependence of $U(r)$ on r definitely corresponds to the dependences necessary for the classification of ion motion according to the angular momentum (see Section 1). We note that dependence (3) is valid only for very large distances r, $r \gg \lambda_D$. From (2), (6), and (19), we find the critical value of the angular momentum L_a and the corresponding maximum point $r_{\max}$ of the effective potential energy

$$L_a^2 = ma^2|U(a)|\left(1 + \frac{a}{\lambda_D}\right), \tag{20}$$

as

$$r_{\max}(L_a) = \lambda_D\left\{\ln\left(\frac{\lambda_D}{a}\right) + \ln\left[\frac{r_{\max}(L_a)}{\lambda_D}\left(1+\frac{r_{\max}(L_a)}{\lambda_D}\right)\right]\right\} \quad (21)$$

$$\approx \lambda_D \ln\left\{\frac{\lambda_D}{a}\left[1+\ln\left(\frac{\lambda_D}{a}\right)\right]\ln\left(\frac{\lambda_D}{a}\right)\right\}.$$

Comparing (4) and (20) by means of (16), we find that $L_a > L_0$ and, consequently, the effective potential energy has both a minimum and a maximum for $L \geq L_a$. At $L = L_a$, the minimum point of the effective potential energy coincides with the dust grain radius. For $L < L_a$, the distance of the minimum point of $U_{\text{eff}}(r; L)$ from the center ($r = 0$) is smaller than the grain radius and all trapped ions are absorbed by the dust grain surface. From (2) and (4), it follows that, at $L = L_0$, the point of the minimum of the effective potential energy is given by [9]

$$r_{\min 0} \approx 2a\frac{|U_\infty|}{|U(a)|}. \quad (22)$$

For $|U_\infty| < (1/2)|U(a)|$, we have $r_{\min 0} < a$ (see Fig. 2). This result is physically expected, because the ions with small angular momenta impact the dust grain surface and are absorbed. We can therefore restrict ourselves by considering the ion angular momenta $L \geq L_a$ and distances $r \leq r_{\max}(L_a)$, defined by (21). Obviously, the Debye–Hückel potential (19) is applicable for such distances, and we use it for estimations in what follows. For instance, using the Debye–Hückel shielded potential for the critical distance r_k defined by (8), we obtain $r_k \approx 1.61\lambda_D$.

Different kinds of potential wells that make a contribution to ion trapping can be gathered in two groups. For the angular momentum in the range

$$L_a \leq L < L_p, \quad (23)$$

or

$$ma^2|U(a)|\left(1+\frac{a}{\lambda_D}\right) \leq L^2 < 2ma^2|U(a)|, \quad (24)$$

the distribution function is defined by (18) (with the corresponding $r_{\max}(L)$) and only the ions with an energy

$$E \leq U_{\text{eff}}(a; L) \quad (25)$$

can take part in forming the cloud shielding the grain field. The ions with larger energy $E \geq U_{\text{eff}}(a; L)$ disappear due to absorption on the dust surface. Conditions (25), (12), and (14) allow us to define the limiting value for the velocity component along the radial direction,

$$v_r^2 \leq \left[\frac{L^2}{m^2}\left(\frac{1}{a^2}-\frac{1}{r^2}\right)-\frac{2}{m}(|U(a)|-|U(r)|)\right] \times [\Theta(r-a)-\Theta(r-\bar{r}(L)], \quad (26)$$

where $\Theta(x)$ is the step function ($\Theta(x) = 1$ if $x \geq 0$ and $\Theta(x) = 0$ if $x < 0$). The turning point $\bar{r}(L)$ is the solution of the equation

$$\frac{L^2}{m^2}\left(\frac{1}{a^2}-\frac{1}{\bar{r}^2}\right)-\frac{2}{m}(|U(a)|-|U(\bar{r})|) = 0. \quad (27)$$

It turns out that, for a given L, the distance r can change in the range

$$a \leq r \leq \bar{r}(L) \quad (28)$$

defined by the energy level $U_{\text{eff}}(a)$. According to the general definition, the number density of trapped ions in the angular momentum range (23) is

$$n_1(r) = 2\pi\int v_\theta dv_\theta dv_r f_{itr}, \quad (29)$$

where f_{itr} is defined by (18) and the limits of integration over v_θ and v_r must be chosen according to (12), (23), and (26). Introducing the variable

$$s = L^2/2ma^2|U(a)|,$$

we find

$$n_1(r) = \frac{2}{\sqrt{\pi}}n_0\frac{a^2}{r^2}\left(\frac{|U(a)|}{T_i}\right)^{3/2} \times \int\limits_{\frac{1}{2}\left(1+\frac{a}{\lambda_D}\right)}^{1} ds\exp\left[\frac{|U(a)|}{T_i}\frac{a^2}{r_{\max}^2(s)}s+\frac{|U(r_{\max}(s))|}{T_i}\right] \times\left[s\left(1-\frac{a^2}{\lambda_D^2}\right)-\left(1-\frac{|U(r)|}{|U(a)|}\right)\right]^{1/2} \times [\Theta(r-a)-\Theta(r-\bar{r}(s)], \quad (30)$$

where $r_{\max}(s)$ and $\bar{r}(s)$ are to be found from (2), (19), and (27). In what follows, we assume the grain size to be so small that the inequality

$$\frac{|U(a)|}{T_i}\frac{a^2}{\lambda_D^2} \ll 1 \quad (31)$$

is fulfilled. In accordance with (2) and (19), the exponential function in the integrand of (30) can then be replaced by 1 and we find after integration that

$$n_1(r) = \frac{4}{3\sqrt{\pi}} n_0 \frac{a^2}{r^2}\sqrt{1-\frac{a^2}{r^2}}\left(\frac{|U(a)|}{T_i}\right)^{3/2} \times\left\{1-\left[1-\frac{a}{r}\exp\left(-\frac{r-a}{\lambda_D}\right)\right]\left[1-\frac{a^2}{r^2}\right]^{-1}\right\} \times[\Theta(r-a)-\Theta(r-\bar{r}(1)], \quad (32)$$

where

$$\bar{r}(1) = \lambda_D \ln\left[\frac{\lambda_D}{a}\ln\left(\frac{\lambda_D}{a}\ln\frac{\lambda_D}{a}\right)\right]. \quad (33)$$

Similar calculations can be performed for ions with the angular momentum in the range

$$L_p \le L < L_k, \quad (34)$$

or

$$2ma^2|U(a)| \le L^2 \le 0.419 \times 2ma\lambda_D|U(a)|. \quad (35)$$

In this range, the surface of the grain is outside the well. The energy border of the well is defined by $U_{\text{eff}}(r_{\max}; L)$, and for the limiting value of the velocity component along the radial direction, we have

$$v_r^2 \le \left[\frac{2}{m}(|U(r)| - |U(r_{\max}(L))|) - \frac{L^2}{m^2}\left(\frac{1}{r^2}-\frac{1}{r_{\max}^2(L)}\right)\right] \times[\Theta(r-\hat{r}(L))-\Theta(r-r_{\max}(L))], \quad (36)$$

where $\hat{r}(L)$ $(\neq r_{\max}(L))$ is the solution of the equation

$$\frac{2}{m}(|U(\hat{r})| - |U(r_{\max}(L))|) - \frac{L^2}{m^2}\left(\frac{1}{\hat{r}^2}-\frac{1}{r_{\max}^2(L)}\right) = 0, \quad (37)$$

and $r_{\max}(L)$, which is again the maximum point, also satisfies this equation. The procedure, quite analogous to that used above, gives the following expression for the number density $n_2(r)$ of the trapped ions with the angular momentum in the range (34):

$$n_2(r) = \frac{2}{\sqrt{\pi}} n_0 \frac{a^2}{r^2}\left(\frac{|U(a)|}{T_i}\right)^{3/2} \times \int_1^{0.419\frac{\lambda_D}{a}} ds \exp\left\{-\frac{a^2}{r_{\max}^2(s)}\frac{|U(a)|}{T_i}s + \frac{|U(r_{\max}(s))|}{T_i}\right\}\left[(|U(r)| - |U(r_{\max}(s))|)/|U(a)| - s\left(\frac{a^2}{r^2}-\frac{a^2}{r_{\max}^2(s)}\right)\right]^{1/2} \times[\Theta(r-\hat{r}(s))-\Theta(r-r_{\max}(s))]. \quad (38)$$

Here, $r_{\max}(s)$ is again defined by (2) and (19), and $\hat{r}(s)$ is the root of the expression under the radical (cf. (19) and (37)). From (30) and (38) (and also from the dependence of the wells on the angular momentum described in Sections 2 and 3), it follows that a cloud of trapped ions is localized in a spherical layer restricted by the spheres with the radii a and $\bar{r}(1)$ (the latter is defined by (33)). Integrating the sum of $n_1(r)$ and $n_2(r)$ over the space, we find the total number N of trapped ions, which we assume to be equal to the charge number Z of the grain,

$$N \approx \frac{4}{3}\sqrt{\frac{2}{3}}\left(\frac{4\pi}{3}n_0\lambda_D^3\right)\left\{\frac{a}{\lambda_D}\frac{|U(a)|}{T_i}\right\}^{3/2} = Z. \quad (39)$$

In estimating (39), we used the condition of the smallness of a dust grain, $a \ll \lambda_D$. The possibility of such a compensation of charges was recently predicted in [23, 24]. In our model, therefore, the electric charge of the dust grain is screened by the trapped ion cloud and the interaction of the compound particle (dust grain + ion cloud) at large distances cannot be realized as an interaction of charges. In an external (or induced) electric field, the centers of positive and negative charges within such a compound particle can be shifted and the particles acquire a dipole moment that can lead to dipole–dipole interactions of the compound particles. Below, we find the electric field necessary for shifting the centers of charges over a distance r, with

$$a \ll r \ll \lambda_D, \quad (40)$$

and determine the corresponding induced dipole moment. At shifting distances $r \ll \lambda_D$ (much less than the size of the trapped ion cloud; cf. (33)), we can assume that the form of the dust cloud remains

unchanged under shifting. From (32) and (38), the total ion number density at $r \ll \lambda_D$ is given by

$$n(r) = n_1(r) + n_2(r) = \frac{4}{3\sqrt{\pi}} n_0 \left[\frac{|U(a)|}{T_i}\right]^{3/2} \times \frac{a}{r}\sqrt{\frac{r}{a} - 1}\left(\frac{r}{a} + 1\right)^{-1} \Theta(r - a). \quad (41)$$

For the electric field strength, which is defined as

$$\mathbf{E} = \frac{\mathbf{r}}{r^3} 4\pi \int_a^r dr' r'^2 e n(r'), \quad (42)$$

we then obtain

$$\mathbf{E} = \frac{8}{3\sqrt{\pi}}\left(\frac{4\pi}{3} n_0\right)\left[\frac{a}{r}\frac{|U(a)|}{T_i}\right]^{3/2} e\mathbf{r}. \quad (43)$$

Equation (43) represents the electric field within the cloud of trapped ions generated by these ions. Placing a charged grain at distance r from the center, in order to keep it in equilibrium, one needs to apply an external electric field whose value can be found from Eq. (43). The direction of the external field must be opposite to the displacement of the centers of the positive and negative charges [35–37].

Inequalities (40) and relation (39) give the following restriction on the electric field:

$$\frac{6}{\pi}\frac{e^2 Z^2}{\lambda_D^4} \ll E^2 \ll \frac{6}{\pi}\frac{e^2 Z^2}{\lambda_D^4}\frac{\lambda_D}{a}. \quad (44)$$

According to (39) and (43), the induced dipole moment and the polarizability of the compound particle (dust grain + ion cloud) are given by

$$\mathbf{P} = Ze\mathbf{r} = \alpha(E)\mathbf{E}, \quad (45)$$

where

$$\alpha(E) = \frac{6}{\pi}\left(\frac{Ze}{\lambda_D^2 E}\right)^3 \lambda_D^3. \quad (46)$$

Due to the specific dependence of the electric field within the compound particle at distance r from the center (Eq. (43)), the polarizability reveals a nonlinear behavior. At large distances, the interaction energy between the compound particles in the external electric field can be interpreted as the dipole–dipole interaction

$$V = \frac{1}{R^3}[\mathbf{P}_1 \cdot \mathbf{P}_2 - 3(\mathbf{n} \cdot \mathbf{P}_1) \cdot (\mathbf{n} \cdot \mathbf{P}_2)], \quad (47)$$

where R ($\gg \lambda_D$) is the distance between the dust particles, $\mathbf{n} = \mathbf{R}/R$, and $\mathbf{P}_1$ and $\mathbf{P}_2$ are the dipole moments of the dust particles. Depending on the orientation of the dipole moments, the potential energy can acquire an attractive character. For identical dust particles, the attractive force becomes maximum when the dipole moments are parallel to each other and to $\mathbf{n}$. According to Eqs. (47) and (46), this attractive force is given by

$$\mathbf{F} = \frac{1}{\pi^2}\left(\sqrt{6}\frac{Ze}{\lambda_D E}\right)^6 \frac{E^2}{R^4}\frac{\mathbf{R}}{R}. \quad (48)$$

In [19, 20], the effective attractive force between two isolated dust grains due to their reciprocal shadowing in the plasma has been investigated. According to [29], the value of the shadowing force is

$$F_{\rm sh} = \frac{3}{8}\frac{a^2}{\lambda_D^2}\left(\frac{Ze}{R}\right)^2. \quad (49)$$

Comparison of Eqs. (48) and (49) reveals that the shadowing force is smaller than the force due to dipole–dipole interactions for electric field strengths that are typical in laboratory experiments [38–40]. Indeed, relation $|\mathbf{F}| \gg F_{\rm sh}$ is identical to the inequality

$$E^4 \ll 10 E_0^4, \quad (50)$$

where

$$E_0 = \left(7.6\frac{\lambda_D}{\pi a}\frac{\lambda_D}{R}\right)^{1/2}\frac{Ze}{\lambda_D^2}. \quad (51)$$

For $E = E_0$, conditions (44) and (50) can be satisfied if $\lambda_D^2/aR \gg 1$ ($\lambda_D/R \ll 1$). Taking $a \approx 10^{-4}$ cm, $Z \approx 10^4$, $\lambda_D \approx 1.4 \times 10^{-2}$ cm, and $R \approx 10^{-1}$ cm, we have $E = E_0 \approx$ 50 V/cm. According to Eq. (46), the potential energy and the interacting force decrease as the electric field strength increases. Apparently, the dipole–dipole interaction potential energy, as discussed here, may be responsible for the formation of many-layer structures that have been observed in laboratory experiments [8–11] where the behavior of a dust particle cloud in the plasma discharge was investigated.

3. CONCLUSIONS

We have considered complete shielding of the dust grain charge by the trapped ions in plasmas. In the plasma electric field, a neutral compound particle (dust grain + ion cloud) acquires dipole moments that can lead to their interaction by the potential energy of the

dipole–dipole type. We note that our calculations for dipole–dipole interactions are valid under the following assumptions.

(1) The surface of the dust grain is absolutely absorbing.

(2) For the distances in which we are interested, the spatial dependence of the ion potential energy in the field of a dust grain follows the Debye–Hückel law (see (19)). The latter holds even in the nonlinear regime, as demonstrated in [23]. The ion flow slightly affects the Debye–Hückel potential [16], and, in addition, generates a wake field, which is not the topic of the present paper.

(3) It is assumed that, in a collisionless plasma, ion trapping is the result of adiabatic change of the potential well's shape in time [25, 26]. Therefore, the steady state is reached before dust-neutral interactions take place.

(4) The number density of trapped ions is small compared to the total ion number density. The trapped ions do not take part in the formation of the potential well.

In conclusion, we mention that some aspects of the interaction observed experimentally [8–11], for example, formation of regular equidistant layers of dust grains, can be explained by the theory developed here. Finally, the present dipole–dipole attractive force can be incorporated in molecular dynamics simulation studies of charged dust particle behavior in dusty plasmas.

ACKNOWLEDGMENTS

This research was partially supported by the Deutsche Forschungsgemeinschaft (Bonn) through the Sonderforschungsbereich 591 entitled "Universelles Verhalten gleichgewichtsferner Plasmen: Heizung, Transport und Strukturbildung," and by the European Commission (Brussels) through contract no. HPRN-CT-2000-00140 for carrying out the task of the Human Potential Training Network entitled "Complex Plasmas: The Science of Laboratory Colloidal Plasmas and Mesospheric Charged Aerosols."

REFERENCES

1. H. Ikezi, Phys. Fluids **29**, 1764 (1986).
2. J. H. Chu, J. B. Du, and I. Lin, J. Phys. D **27**, 296 (1994).
3. H. Thomas, G. Morfill, V. Demmel, *et al.*, Phys. Rev. Lett. **72**, 652 (1994).
4. J. H. Chu and I. Lin, Phys. Rev. Lett. **72**, 4009 (1994).
5. Y. Hayashi and K. Tachibana, Jpn. J. Appl. Phys. **33**, L804 (1994).
6. A. Melzer, T. Trottenberg, and A. Piel, Phys. Lett. A **191**, 301 (1994).
7. V. E. Fortov, A. P. Nefedov, O. F. Petrov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 187 (1996) [JETP Lett. **63**, 187 (1996)]; V. E. Fortov, A. P. Nefedov, V. M. Torchinskiĭ, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **64**, 92 (1996) [JETP Lett. **64**, 92 (1996)].
8. K. Takahashi, T. Oishi, K. Shimomai, *et al.*, Jpn. J. Appl. Phys. **37**, 6609 (1998).
9. J. B. Pieper, J. Goree, and R. A. Quinn, Phys. Rev. E **54**, 5636 (1996).
10. U. Mohideen, H. U. Rahman, M. A. Smith, *et al.*, Phys. Rev. Lett. **81**, 349 (1998).
11. Y. Hayashi, Phys. Rev. Lett. **83**, 4764 (1999).
12. V. I. Molotkov, A. P. Nefedov, M. Y. Pustyl'nik, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 152 (2000) [JETP Lett. **71**, 102 (2000)].
13. M. Nambu, S. V. Vladimirov, and P. K. Shukla, Phys. Lett. A **230**, 40 (1995).
14. P. K. Shukla and N. N. Rao, Phys. Plasmas **3**, 1770 (1996).
15. G. Lapenta, Phys. Rev. E **66**, 026409 (2002).
16. A. M. Ignatov, Plasma Phys. Rep. **29**, 296 (2003).
17. K. Takahashi, T. Oishi, K. Shimomai, *et al.*, Phys. Rev. E **58**, 7805 (1998).
18. A. Melzer, V. A. Schweigert, and A. Piel, Phys. Rev. Lett. **83**, 3194 (1999); A. Piel and A. Melzer, Plasma Phys. Control. Fusion **44**, R1 (2002).
19. V. N. Tsytovich, Y. K. Khodataev, and R. Bingham, Comm. Plasma Phys. Control. Fusion **17**, 249 (1996).
20. A. M. Ignatov, Kratk. Soobshch. Fiz. **1–2**, 58 (1995); Plasma Phys. Rep. **22**, 585 (1996).
21. Yu. A. Mankelevich, M. A. Olevanov, and T. V. Rachimova, Zh. Éksp. Teor. Fiz. **121**, 1288 (2002) [JETP **94**, 1106 (2002)].
22. J. Goree, Phys. Rev. Lett. **69**, 277 (1992).
23. A. V. Zobnin, A. P. Nefedov, V. A. Sinel'shchikov, and V. E. Fortov, Zh. Éksp. Teor. Fiz. **118**, 554 (2000) [JETP **91**, 483 (2000)]; V. E. Fortov, A. P. Nefedov, V. I. Molotkov, *et al.*, Phys. Rev. Lett. **87**, 205002 (2001).
24. D. D. Tskhakaya, P. K. Shukla, and F. Subba, Phys. Lett. A **300**, 619 (2002).
25. A. V. Gurevich, Zh. Éksp. Teor. Fiz. **53**, 953 (1967) [Sov. Phys. JETP **26**, 575 (1968)].
26. E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics* (Nauka, Moscow, 1979; Pergamon Press, Oxford, 1981).
27. P. K. Shukla, Phys. Plasmas **8**, 1791 (2001).
28. P. K. Shukla and A. A. Mamun, *Introduction to Dusty Plasma Physics* (Inst. of Physics, Bristol, 2002); P. K. Shukla, *Dust Plasma Interaction in Space* (Nova Sci., New York, 2002).
29. M. Lampe, G. Joyce, G. Ganguli, and V. Gavrishchaka, Phys. Plasmas **7**, 3851 (2000).
30. M. Lampe, V. Gavrishchaka, G. Ganguli, and G. Joyce, Phys. Rev. Lett. **86**, 5278 (2001); M. Lampe, R. Gos-

wami, Z. Sternovsky, *et al.*, Phys. Plasmas **10**, 1500 (2003).

31. Ya. L. Al'pert, A. V. Gurevich, and L. P. Pitaevskiĭ, *Artificial Satellites in a Rarefied Plasma* (Nauka, Moscow, 1964).
32. I. B. Bernstein and I. N. Rabinowitz, Phys. Fluids **2**, 112 (1959).
33. D. D. Tskhakaya, P. K. Shukla, and L. Stenflo, Phys. Plasmas **12**, 5333 (2001).
34. D. D. Tskhakaya, N. L. Tsintsadze, P. K. Shukla, and L. Stenflo, Phys. Scr. **64**, 366 (2001).
35. V. V. Batygin and I. N. Toptygin, *Problems in Electrodynamics* (Fizmatgiz, Moscow, 1962; Academic, London, 1964), problems 81, 82, and 297.
36. A. M. Portis, *Electromagnetic Fields: Sources and Media* (Wiley, New York, 1978), p. 78.
37. H. Mitter, *Elektrodynamik* (Wissenschaftsverlag, Mannheim, 1990), p. 140.
38. A. A. Samarian, O. S. Vaulina, A. P. Nefedov, *et al.*, Phys. Rev. E **64**, 056407 (2001).
39. A. M. Lipaev, V. I. Molotkov, A. P. Nefedov, *et al.*, Zh. Éksp. Teor. Fiz. **112**, 2030 (1997) [JETP **85**, 1110 (1997)].
40. S. Takamura, T. Misawa, N. Ohno, *et al.*, Phys. Plasmas **8**, 1886 (2001).

Journal of Experimental and Theoretical Physics, Vol. 98, No. 1, 2004, pp. 62–92.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 1, 2004, pp. 72–102.

FLUIDS

Light Scattering in Cholesteric Liquid Crystals with a Large Pitch

E. V. Aksenova[a,*], A. Yu. Val'kov[b,], and V. P. Romanov[a,***]**

[a] *St. Petersburg State University (Petrodvorets Branch), Universitetskiĭ pr. 2, Petrodvorets, 198504 Russia*
[b] *St. Petersburg Institute for International Relations, Trade, Economics, and Law, St. Petersburg, 191104 Russia*
**e-mail: aksev@mail.ru*
***e-mail: alexvalk@mail.ru*
****e-mail: V.Romanov@pobox.spbu.ru*
Received June 23, 2003

Abstract—Optical properties of cholesteric liquid crystals with a pitch larger than the wavelength of light are considered. Normal waves of the medium and the Green function of the electromagnetic field are analyzed. A general algorithm based on the application of the Kirchhoff method is proposed for calculating the scattered light intensity in media with a one-dimensional periodic structure. The WKB vector method is used for calculating the spatial correlation function of thermal fluctuations of the director. It is found that the transformation of two fluctuation modes takes place is some regions. The angular and polarization dependences of the intensity of light scattered from fluctuations of the director are calculated. It is found, in particular, that the intensity of scattering is a nonmonotonic function of the size of the system. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, considerable attention has been paid to investigation of liquid crystals (LCs) by various methods. The interest in this problem is provoked by a number of anomalies in physical properties of these systems and by their wide application in graphics display systems (above all, in LC displays).

The light scattering technique is one of the effective methods for studying LC systems. Peculiar optical and structural properties of liquid crystals considerably complicate the description of light scattering. Such properties include strong optical anisotropy, the presence of regular spatial structures, abnormally large fluctuations of the order parameter, and anomalously high optical activity.

In this connection, a number of physical problems arise, which have been studied insufficiently so far. One of such problems is the description of light scattering in media with a smoothly varying periodic structure. Scattering of light is usually considered under the assumption that the medium is spatially homogeneous, or fluctuations and the propagation of waves in inhomogeneous media are described on the basis of small parameters, which make it possible to reduce the solution of the problem to that for a certain effective homogeneous medium. For homogeneous systems, normal waves and the field of a point source (the Green function of the electromagnetic field) are known and the spatial correlation function of thermal fluctuations of permittivity, from which light is scattered, are calculated quite easily. The condition of spatial homogeneity of a system makes it possible to obtain simple expressions for the intensity of scattered light in closed form by passing to the 3D spectrum of Fourier fluctuations.

The problem becomes more complicated when the spatial homogeneity of the medium is violated considerably. This immediately gives rise to several problems such as the description of the structure of the incident field (normal waves in the medium), the calculation of the Green function of the electromagnetic field, and analysis of the correlation function of permittivity fluctuations.

Typical examples are the problems of propagation and scattering of light in media with periodically varying properties (in particular, in media with one-dimensional periodic structures). Such media include cholesteric liquid crystals (CLCs), twisted nematic liquid crystals (NLCs), and certain kinds of smectic liquid crystals (SLCs).

Although this problem has been considered for a long time, it involves significant mathematical difficulties since the problem is reduced to solving a system of differential equations with periodic coefficients, which have no exact solution in the general case. For example, the exact solution of the problem on propagation of electromagnetic waves in CLCs was obtained only for waves propagating along the symmetry axis of the system [1–3]. The formal analytic solution for the problem in the case of oblique incidence [4–6] has the form of an infinite series and is difficult for analysis. For this reason, various approximate and numerical methods are widely used in the optics of layered LCs [7–11]. In this case, main attention is paid to the situation when the wavelength is on the order of the structural period and the

methods developed for X-rays diffraction have been found to be effective [8]. Such an approach remains most popular for CLCs even now [11–15]. This case is characterized by the emergence of forbidden bands. The problems with normal waves and the field of a point source [16–19], as well as the spectrum of thermal vibrations of the director [20–23], have been investigated using this method.

The opposite case, when the wavelength is much smaller than the characteristic size of the LC structure, has been studied insufficiently. However, this problem has become important in recent years in connection with the application of twist cells of weakly twisted NLCs and CLCs with a large pitch in graphics display systems.

It is well known that, when light propagates along the axis of the cholesteric in such systems, the adiabatic regime occurs, when the polarization of waves rotates together with the optical axis [24]. For high intensities of incident light, nonlinear effects (such as the generation of the third harmonic) were observed [25]. In the general case of oblique incidence, it is natural in this case to use the WKB method since the size of inhomogeneities is much larger than the wavelength. It is difficult to directly apply the WKB method for electromagnetic waves since this gives rise of a system of coupled equations [4, 5, 26]. This problem was solved in [27, 28] for electromagnetic waves propagating in locally isotropic media with smooth inhomogeneities. The generalization of the WKB method, proposed in [29] for CLCs with a large pitch, has made it possible to obtain an analytic solution of the problem with oblique incidence of light (in particular, normal waves were determined). On the basis of this method, the field of a point source in such a medium was also obtained [30–32].

In order to describe scattering of light, we must know, in addition to the optical parameters of the system, the correlation function of thermal fluctuations of permittivity. The main contribution to scattering in CLCs comes from fluctuations of the director. The problem of the director thermal noise in CLCs was considered only for fluctuations with characteristic scales on the order of or larger than the structure period ("smecticlike" CLCs) [20–23]. In the opposite case of a "nematiclike" CLCs, fluctuations have not been studied. It should be noted that fluctuations were analyzed for some physical systems with regular inhomogeneities (e.g., in studying the influence of the gravity effect in the vicinity of the liquid–vapor critical point on density fluctuations) [33]. However, the correlation radius of fluctuations in this case is much smaller than the characteristic size of regular inhomogeneities in the system. This problem can be reduced to that for a locally homogeneous medium with parameters smoothly varying from point to point. The system considered here is distinguished by the fact that the correlation length of director fluctuations we are interested in is limited only by the size of the system [34]. For this reason, the local homogeneity approximation is inapplicable in this case.

In this study, we use a vector generalization of the WKB method for calculating the correlation function of director fluctuations in CLCs with a large pitch. A general algorithm is proposed for calculating the intensity of scattered light for layered systems. This enabled us to derive explicit expressions for the angular and polarization dependences of the intensity of single scattering of light in CLCs in the case when the pitch is much larger than the wavelength. The results are represented in a form convenient for comparison with experiment.

The article has the following structure. In Section 2, the general equations describing the elastic energy and fluctuations, as well as the propagation of electromagnetic waves in CLCs, are considered. In Section 3, the general algorithm for calculating the intensity of scattered light in layered media is constructed on the basis of the Kirchhoff method. Section 4 is devoted to determining the normal waves and the Green function of the electromagnetic field in CLCs with a pitch larger than the wavelength of light. In Section 5, the correlation function of director fluctuations in CLCs is calculated. The intensity of scattered light is calculated in Section 6, where various experimental geometries are analyzed. The algorithm of the vector WKB method used for determining the normal waves and the correlation function in CLCs is given in the Appendix.

2. BASIC EQUATIONS

A CLC can be described by free energy in the form [34]

$$F = F_0 + \frac{1}{2}\int d\mathbf{r}[K_{11}(\mathrm{div}\mathbf{n})^2 + K_{22}(\mathbf{n}\,\mathrm{curl}\mathbf{n} + p_0)^2 + K_{33}(\mathbf{n}\times\mathrm{curl}\mathbf{n})^2], \quad (2.1)$$

where F_0 is the energy of a homogeneous system and K_{ll} (l = 1, 2, 3) are the Frank moduli. The unit vector $\mathbf{n} = \mathbf{n}(\mathbf{r})$ of the director characterizes the direction of the local preferred orientation of the longer axes of the molecules. The minimum of energy (2.1) corresponds to a helicoidal equilibrium distribution of the director,

$$\mathbf{n}^0(\mathbf{r}) \equiv \mathbf{n}^0(z) = (\cos\phi, \sin\phi, 0). \quad (2.2)$$

Here, we have introduced the Cartesian system of coordinates with the z axis coinciding with the axis of the cholesteric in the CLCs, $\phi = \phi(z) = p_0 z + \phi_0$, angle ϕ_0 determining the direction of the director on the plane $z = 0$, $p_0 = \pi/d$, where d is the pitch of the cholesteric. Director $\mathbf{n}^0(\mathbf{r})$ in Eq. (2.2) is perpendicular to the z axis and rotates uniformly around this axis.

Optical properties of a cholesteric are determined by permittivity tensor $\hat{\varepsilon}$; in an equilibrium CLC, this tensor has the form [34]

$$\varepsilon^0_{\alpha\beta}(\mathbf{r}) \equiv \varepsilon^0_{\alpha\beta}(z) = \varepsilon_\perp \delta_{\alpha\beta} + \varepsilon_a n^0_\alpha(z) n^0_\beta(z), \quad (2.3)$$

where $\varepsilon_a = \varepsilon_\parallel - \varepsilon_\perp$, $\varepsilon_\parallel$ and $\varepsilon_\perp$ being the permittivities along and across $\mathbf{n}^0$.

The Maxwell equations for a monochromatic wave in such a medium have the form

$$\begin{aligned} \mathrm{curl}\mathbf{E}(\mathbf{r}) &= ik_0 \hat{\mu}(\mathbf{r})\mathbf{H}(\mathbf{r}), \\ \mathrm{curl}\mathbf{H}(\mathbf{r}) &= -ik_0 \hat{\varepsilon}(\mathbf{r})\mathbf{E}(\mathbf{r}), \end{aligned} \quad (2.4)$$

where $\mathbf{E}$ and $\mathbf{H}$ are the vectors of electric and magnetic fields and $k_0 = \omega/c$, ω being the circular frequency and c, the velocity of light in vacuum. Henceforth, we will assume that the medium is nonmagnetic, $\mu_{\alpha\beta}(\mathbf{r}) = \delta_{\alpha\beta}$. Eliminating vector $\mathbf{H}$ from system (2.4), we obtain the wave equation for vector $\mathbf{E}$:

$$(\mathrm{curl}\,\mathrm{curl} - k_0^2 \hat{\varepsilon}(\mathbf{r}))\mathbf{E}(\mathbf{r}) = 0. \quad (2.5)$$

In solving the scattering problem, it is convenient to use the integral representation for wave equation (2.5),

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^0(\mathbf{r}) + k_0^2 \int d\mathbf{r}' \hat{T}^0(\mathbf{r}, \mathbf{r}') \delta\hat{\varepsilon}(\mathbf{r}')\mathbf{E}(\mathbf{r}'), \quad (2.6)$$

where $\delta\hat{\varepsilon}(\mathbf{r}) = \hat{\varepsilon}(\mathbf{r}) - \hat{\varepsilon}^0(\mathbf{r})$ are the fluctuations of the permittivity tensor, and field $\mathbf{E}^0(\mathbf{r})$ and the Green function $\hat{T}^0(\mathbf{r}, \mathbf{r}')$ of the electromagnetic field satisfy the equations

$$(\mathrm{curl}\,\mathrm{curl} - k_0^2 \hat{\varepsilon}(z))\mathbf{E}^0(\mathbf{r}) = 0, \quad (2.7)$$

$$(\mathrm{curl}\,\mathrm{curl} - k_0^2 \hat{\varepsilon}^0(z))\hat{T}^0(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}')\hat{I}. \quad (2.8)$$

Here, $\hat{I}$ is the unit matrix.

Since Eq. (2.7) is homogeneous, field $\mathbf{E}^0(\mathbf{r})$ can be expressed in the form of a linear combination of normal waves of problem (2.7). In order to formulate the problem unambiguously, Eq. (2.8) should be supplemented with the corresponding boundary conditions. In an unbounded medium, such conditions are the radiation conditions [35]. In view of the symmetry of the CLC relative to displacements in the *XY* plane, we have $\hat{T}^0(\mathbf{r}, \mathbf{r}') \equiv \hat{T}^0(\mathbf{r}_\perp - \mathbf{r}'_\perp; z, z')$, where $\mathbf{r}_\perp = (x, y)$.

The second term on the right-hand side of Eq. (2.6) corresponds to scattered field $\mathbf{E}^{(s)}$ generated by incident field $\mathbf{E}^0(\mathbf{r})$. Solving this equation by iterations and confining the analysis to the lowest order in $\delta\hat{\varepsilon}$, we obtain scattered field $\mathbf{E}^{(s)}$ in the single-scattering approximation,

$$\mathbf{E}^{(s)}(\mathbf{r}) = k_0^2 \int d\mathbf{r}' \hat{T}^0(\mathbf{r}_\perp - \mathbf{r}'_\perp; z, z') \delta\hat{\varepsilon}(\mathbf{r}')\mathbf{E}^0(\mathbf{r}'). \quad (2.9)$$

The properties of scattered light are determined by the coherence function

$$\begin{aligned} &\langle E^{(s)}_\alpha(\mathbf{r}_1) E^{(s)*}_\beta(\mathbf{r}_2)\rangle \\ &= k_0^4 \int d\mathbf{r}'_1 d\mathbf{r}'_2 T^0_{\alpha\gamma}(\mathbf{r}_{1\perp} - \mathbf{r}'_{1\perp}; z_1, z'_1) \\ &\times T^{0*}_{\beta\zeta}(\mathbf{r}_{2\perp} - \mathbf{r}'_{2\perp}; z_2, z'_2) \mathcal{G}_{\gamma\nu\zeta\mu}(\mathbf{r}'_1, \mathbf{r}'_2) E^0_\nu(\mathbf{r}'_1) E^{0*}_\mu(\mathbf{r}'_2), \end{aligned} \quad (2.10)$$

where $\mathcal{G}_{\gamma\nu\zeta\mu}(\mathbf{r}'_1, \mathbf{r}'_2) = \langle \delta\varepsilon_{\gamma\nu}(\mathbf{r}'_1) \delta\varepsilon^*_{\zeta\mu}(\mathbf{r}'_2)\rangle$ is the correlation function of permittivity fluctuations, angle brackets $\langle \ldots \rangle$ denote statistical averaging, and the asterisk indicates complex conjugation. In view of the symmetry of the CLC, we have $\hat{\mathcal{G}}(\mathbf{r}'_1, \mathbf{r}'_2) \equiv \hat{\mathcal{G}}(\mathbf{r}'_{1\perp} - \mathbf{r}'_{2\perp}; z'_1, z'_2)$.

Thus, in order to calculate coherence function (2.10), we must know the normal waves determining the form of field $\mathbf{E}^0(\mathbf{r})$, Green function $\hat{T}^0$, and correlation function $\hat{\mathcal{G}}$ for permittivity fluctuations in our system.

The largest contribution to $\delta\hat{\varepsilon}$ in LCs comes from fluctuations of the director [34],

$$\mathbf{n}(\mathbf{r}) = \mathbf{n}^0(z) + \delta\mathbf{n}(\mathbf{r}); \quad (2.11)$$

here, we confine our analysis to these fluctuations only. Under this assumption, not only the equilibrium, but also the fluctuating permittivity tensor has a form analogous to expression (2.3) with the substitution $\mathbf{n}^0(z) \longrightarrow \mathbf{n}(\mathbf{r})$:

$$\varepsilon_{\alpha\beta}(\mathbf{r}) \equiv \varepsilon_\perp \delta_{\alpha\beta} + \varepsilon_a n_\alpha(\mathbf{r}) n_\beta(\mathbf{r}). \quad (2.12)$$

Subtracting expression (2.3) from (2.12), we obtain the relation between fluctuations of the permittivity and the director in the CLC,

$$\delta\varepsilon_{\alpha\beta}(\mathbf{r}) = \varepsilon_a (n^0_a(z)\delta n_\beta(\mathbf{r}) + \delta n_\alpha(\mathbf{r}) n^0_\beta(z)), \quad (2.13)$$

as well as the relation between the corresponding correlation functions,

$$\begin{aligned} \mathcal{G}_{\alpha\beta\gamma\delta}(\mathbf{r}_\perp; z, z') = \varepsilon_a^2 [&n^0_\alpha(z) n^0_\gamma(z') g_{\beta\delta}(\mathbf{r}_\perp; z, z') \\ &+ n^0_\alpha(z) n^0_\delta(z') g_{\beta\gamma}(\mathbf{r}_\perp; z, z') \\ &+ n^0_\beta(z) n^0_\gamma(z') g_{\alpha\delta}(\mathbf{r}_\perp; z, z') \\ &+ n^0_\beta(z) n^0_\delta(z') g_{\alpha\gamma}(\mathbf{r}_\perp; z, z')]. \end{aligned} \quad (2.14)$$

Here,

$$g_{\alpha\beta}(\mathbf{r}_{1\perp} - \mathbf{r}_{2\perp}; z_1, z_2) = \langle \delta n_\alpha(\mathbf{r}_{1\perp}, z_1)\delta n_\beta(\mathbf{r}_{2\perp}, z_2)\rangle \quad (2.15)$$

is the correlation function of director fluctuations.

To calculate the correlation function of director fluctuations in the Gaussian approximation, we can confine ourselves to the contribution to free energy (2.1), which is quadratic in $\delta\mathbf{n}$:

$$\delta F = \frac{1}{2}\int d\mathbf{r}\{K_{11}(\nabla\cdot\delta\mathbf{n})^2 + K_{22}[\mathbf{n}^0\cdot(\nabla\times\delta\mathbf{n})]^2 + K_{33}[(\delta\mathbf{n}\cdot\nabla)\mathbf{n}^0 + (\mathbf{n}^0\cdot\nabla)\delta\mathbf{n}]^2\}. \quad (2.16)$$

In deriving this equation, we took into account the fact that the relations $\mathrm{div}\,\mathbf{n}^0 = 0$ and $\mathrm{curl}\,\mathbf{n}^0 = -p_0\mathbf{n}^0$ are valid for helicoidal structure (2.2). Since $|\mathbf{n}| = |\mathbf{n}^0| = 1$, the condition $\delta\mathbf{n} \perp \mathbf{n}^0$ is satisfied in the main order in δn. Vector $\delta\mathbf{n} = (\delta n_x, \delta n_y, \delta n_z)$ can be parameterized with the help of two quantities. In the case of CLCs, the parameterization [20, 21]

$$\delta n_x(\mathbf{r}) = -u_1(\mathbf{r})\sin\phi(z),\quad \delta n_y(\mathbf{r}) = u_1(\mathbf{r})\cos\phi(z),\quad \delta n_z(\mathbf{r}) = u_2(\mathbf{r}) \quad (2.17)$$

is normally used. Modes u_1 and u_2 determine the fluctuations of the director in the XY plane and along the z axis, respectively (Fig. 1). In vector form, we have

$$\delta\mathbf{n}(\mathbf{r}) = u_1(\mathbf{r})\mathbf{h}^{(1)}(z) + u_2(\mathbf{r})\mathbf{h}^{(2)}, \quad (2.18)$$

where

$$\mathbf{h}^{(1)}(z) = \mathbf{h}^{(2)}\times\mathbf{n}^0(z),\quad \mathbf{h}^{(2)} = \mathbf{e}_z. \quad (2.19)$$

Equation (2.18) leads to the expression for the correlation function of director fluctuations in terms of the correlation matrix of scalar quantities $u_{1,2}$,

$$g_{\alpha\beta}(\mathbf{r}_\perp; z_1, z_2) = \sum_{k,l=1}^{2} G_{kl}(\mathbf{r}_\perp; z_1, z_2)h_\alpha^{(k)}(z_1)h_\beta^{(l)}(z_2), \quad (2.20)$$

where

$$G_{kl}(\mathbf{r}_{1\perp} - \mathbf{r}_{2\perp}; z_1, z_2) \equiv G_{kl}(\mathbf{r}_1, \mathbf{r}_2) = \langle u_k(\mathbf{r}_1)u_l(\mathbf{r}_2)\rangle. \quad (2.21)$$

Substituting relation (2.17) into Eq. (2.16),

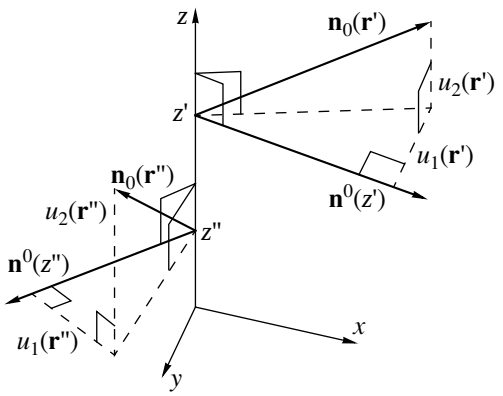

Fig. 1. Modes $u_{1,2}$ of director fluctuations in a CLC.

we obtain

$$\delta F = \frac{1}{2}\int d\mathbf{r}\{K_{11}(-\sin\phi\partial_x u_1 + \cos\phi\partial_y u_1 + \partial_z u_2)^2 + K_{22}[\cos\phi(\partial_y u_2 - \partial_z(u_1\cos\phi)) + \sin\phi(\partial_z(-u_1\sin\phi) - \partial_x u_2)]^2 + K_{33}[(-u_2 p_0\sin\phi + \cos\phi\partial_x(-u_1\sin\phi) + \sin\phi\partial_y(-u_1\sin\phi))^2 + (u_2 p_0\cos\phi + \cos\phi\partial_x(u_1\cos\phi) + \sin\phi\partial_y(u_1\cos\phi))^2 + (\cos\phi\partial_x u_2 + \sin\phi\partial_y u_2)^2]\}, \quad (2.22)$$

where $\partial_l \equiv \partial/\partial l$, $l = x, y, z$.

Since an equilibrium CLC is spatially homogeneous in a plane orthogonal to the z axis, it is convenient to pass to the 2D Fourier spectrum. Henceforth, we will use a continuous 2D Fourier transform in the form

$$f(\mathbf{r}) = \int\frac{d\mathbf{q}}{(2\pi)^2}f(\mathbf{q}, z)e^{i\mathbf{q}\cdot\mathbf{r}_\perp},\quad f(\mathbf{q}, z) = \int d\mathbf{r}_\perp f(\mathbf{r})e^{-i\mathbf{q}\cdot\mathbf{r}_\perp}. \quad (2.23)$$

In this case, the distortion energy (2.22) assumes the form

$$\delta F = \int\frac{d^2\mathbf{q}}{(2\pi)^2}\delta F_\mathbf{q}, \quad (2.24)$$

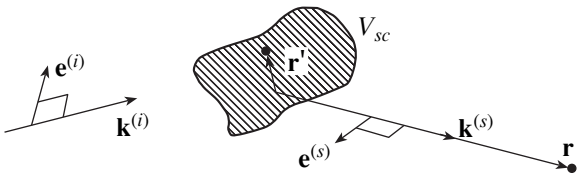

Fig. 2. Geometry of a conventional experiment on light scattering: $\mathbf{k}^{(i)}$ is the wave vector of the incident wave, V_{sc} is the scattering volume, $\mathbf{k}^{(s)}$ is the wave vector of the scattered wave, and $\mathbf{r}$ is the point of observation.

where

$$\delta F_{\mathbf{q}} = \frac{1}{2}\int dz\{K_{11}|\partial_z u_2 + i(-\sin\phi q_x + \cos\phi q_y)u_1|^2 + K_{22}|-\partial_z u_1 + iu_2(\cos\phi q_y - \sin\phi q_x)|^2 + K_{33}[|u_2 p_0 + i(\cos\phi q_x + \sin\phi q_y)u_1|^2 + |u_2|^2(\cos\phi q_x + \sin\phi q_y)^2]\}. \quad (2.25)$$

Integrating by parts and disregarding the terms outside the integral, we can represent quantity $\delta F_{\mathbf{q}}$ as the quadratic form

$$\delta F_{\mathbf{q}} = \frac{1}{2}\int \mathbf{u}^*(\mathbf{q}, z)\hat{\mathcal{A}}(\mathbf{q}, z)\mathbf{u}(\mathbf{q}, z)dz, \quad (2.26)$$

where

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}.$$

Matrix $\hat{\mathcal{A}}$ is a second-order differential operator. In a coordinate system with the x axis directed along vector $\mathbf{q}$ ($q_x = q$, $q_y = 0$), this matrix has the form

$$\hat{\mathcal{A}} = K_{11}\begin{pmatrix} q^2\sin^2\phi & iq\sin\phi\partial_z \\ iq\partial_z\sin\phi & -\partial_z^2 \end{pmatrix} + K_{22}\begin{pmatrix} -\partial_z^2 & -iq\partial_z\sin\phi \\ -iq\sin\phi\partial_z & q^2\sin^2\phi \end{pmatrix} + K_{33}\begin{pmatrix} q^2\cos^2\phi & -ip_0 q\cos\phi \\ ip_0 q\cos\phi & q^2\cos^2\phi + p_0^2 \end{pmatrix}, \quad (2.27)$$

where $\partial_z^2 \equiv \partial^2/\partial z^2$.

The fluctuation probability is proportional to $\exp[-\delta F_{\mathbf{q}}/k_B T]$, where k_B is the Boltzmann constant and T is temperature. In accordance with the general principles of statistical mechanics [36], the calculation of the correlation function can be reduced to the inversion of matrix $\hat{\mathcal{A}}$, which is equivalent to solving the equation

$$\hat{\mathcal{A}}(\mathbf{q}, z)\hat{G}(\mathbf{q}; z, z_1) = k_B T\delta(z - z_1)\hat{I}. \quad (2.28)$$

To solve this equation unambiguously, we must supplement it with the boundary conditions. In an unbounded system, we can use for such conditions the principle of attenuation of correlations, i.e., the condition $\hat{G}(\mathbf{q}; z, z_1) \longrightarrow \hat{0}$ for $z \longrightarrow \pm\infty$.

3. GENERAL THEORY OF SINGLE LIGHT SCATTERING IN A LAYERED MEDIUM

From the optical point of view, cholesterics have the form of a spatially inhomogeneous medium whose properties vary along the cholesteric axis. Normal waves and the field of a point source in such a medium have a complex structure. In addition, the correlation function $\hat{\mathcal{G}}(\mathbf{r}_1, \mathbf{r}_2)$ of permittivity fluctuations depends in this case not only on the difference $\mathbf{r}_1 - \mathbf{r}_2$ of the spatial coordinates, but also on their absolute values. For this reason, the problem of light scattering in CLCs has specific features. In order to illustrate them, we will briefly describe the conventional approach to solving the problem of light scattering in a homogeneous isotropic medium with a permittivity of $\varepsilon_{lk}^0 = \varepsilon_0\delta_{lk}$.

3.1. Homogeneous Medium

In the usual formulation of the problem of light scattering, it is assumed that a plane wave

$$\mathbf{E}^{(i)}(\mathbf{r}) = E_0\mathbf{e}^{(i)}\exp(i\mathbf{k}^{(i)}\cdot\mathbf{r})$$

is incident on the sample, where E_0 is the field amplitude, $\mathbf{e}^{(i)}$ is the polarization vector, $\mathbf{k}^{(i)}$ is the wave vector of the incident wave ($\mathbf{e}^{(i)} \perp \mathbf{k}^{(i)}$), and the scattered field with polarization vector $\mathbf{e}^{(s)}$ is detected at large distances from scattering volume V_{sc} (Fig. 2). In this case, we can assume that the scattered field is a quasi-plane wave with wave vector $\mathbf{k}^{(s)}$. To disregard refraction at the boundary, we assume that the sample is surrounded by a homogeneous medium with permittivity ε_0. In this case, $|\mathbf{k}^{(i)}| = |\mathbf{k}^{(s)}| = k$, where $k = k_0\sqrt{\varepsilon_0}$ is the wave number in the medium.

In an isotropic medium, we have $\hat{T}^0(\mathbf{r}, \mathbf{r}') \equiv \hat{T}^0(\mathbf{r} - \mathbf{r}')$, where

$$T_{0\alpha\beta}(\mathbf{R}) \approx P_{\alpha\beta}(\mathbf{R})\frac{e^{ikR}}{4\pi R}, \tag{3.1}$$

in the approximation of a far-field (wave) zone ($kR \gg 1$) and

$$P_{\alpha\beta}(\mathbf{R}) = \delta_{\alpha\beta} - \frac{R_\alpha R_\beta}{R^2} \tag{3.2}$$

is the transverse projector onto a plane perpendicular to $\mathbf{R}$ [37]. The presence of tensor $P_{\alpha\beta}$ in Eq. (3.1) ensures the transverse form of the field of a point source in the far-field zone.

At large distances from the sample, when $R = |\mathbf{r} - \mathbf{r}'| \gg V_{sc}^{1/3} \sim r'$, we can carry out the substitution $|\mathbf{r} - \mathbf{r}'| \approx r$ in the nonexponential factors of formula (3.1) and use the "Fraunhoffer" approximation $|\mathbf{r} - \mathbf{r}'| \approx r - \mathbf{r}' \cdot \mathbf{r}/r$ in the exponent (the latter approximation also presumes the fulfillment of the condition $kV_{sc}^{2/3} \ll r$). As a result, we obtain from expression (3.1) the "plane-wave" approximation for the Green function,

$$T_{0\alpha\beta}(\mathbf{r} - \mathbf{r}') \approx P_{\alpha\beta}(\mathbf{r})\frac{e^{ikr}}{4\pi r}e^{-i\mathbf{k}^{(s)} \cdot \mathbf{r}'}, \tag{3.3}$$

where $\mathbf{k}^{(s)} = k\mathbf{r}/r$ is the wave vector of the scattered wave.

Setting $\mathbf{E}^0(\mathbf{r}) = \mathbf{E}^{(i)}(\mathbf{r})$ in Eq. (2.9) and using formula (3.3), we obtain the following expression for the scattered field:

$$E_\alpha^{(s)}(\mathbf{r}) = E_0\frac{k_0^2 e^{ikr}}{4\pi r}P_{\alpha\beta}(\mathbf{r})e_\gamma^{(i)} \times \int\limits_{V_{sc}} d\mathbf{r}'\delta\varepsilon_{\beta\gamma}(\mathbf{r}')e^{i(\mathbf{k}^{(i)} - \mathbf{k}^{(s)}) \cdot \mathbf{r}'}. \tag{3.4}$$

Thus, we find that the scattered field is determined by the 3D Fourier component of permittivity fluctuations $\delta\hat{\varepsilon}(\mathbf{Q})$ in volume V_{sc}, where $\mathbf{Q} = \mathbf{k}^{(s)} - \mathbf{k}^{(i)}$ is the scattering vector. Considering that $e_\alpha^{(s)}P_{\alpha\beta} = e_\beta^{(s)}$, we obtain, in particular, the following expression for the scattered field component with polarization $\mathbf{e}^{(s)}$:

$$E^{(s)}(\mathbf{r}) = E_0\frac{k_0^2 e^{ikr}}{4\pi r}\mathbf{e}^{(s)}\delta\hat{\varepsilon}(\mathbf{Q})\mathbf{e}^{(i)}. \tag{3.5}$$

Consequently, the corresponding intensity (Poynting vector modulus) of scattered light,

$$I_{(i)}^{(s)} = \frac{c}{8\pi}\sqrt{\varepsilon_0}\left\langle |E^{(s)}|^2 \right\rangle, \tag{3.6}$$

has the form

$$I_{(i)}^{(s)} = \frac{V_{sc}I_0^{(i)}k_0^4}{(4\pi)^2 r^2}e_\alpha^{(s)}e_\beta^{(s)}\mathcal{G}_{\alpha\nu\beta\mu}(\mathbf{Q})e_\mu^{(i)}e_\nu^{(i)}, \tag{3.7}$$

where $I_0^{(i)}$ is the intensity of incident light and $\hat{\mathcal{G}}(\mathbf{Q})$ is the 3D Fourier transform of the correlation function of permittivity fluctuations; in the case of a homogeneous medium, this function depends only on the difference of the coordinates, $\hat{\mathcal{G}}(\mathbf{r}, \mathbf{r}') = \hat{\mathcal{G}}(\mathbf{r} - \mathbf{r}')$. Here, we take into account the relation $\langle \delta\hat{\varepsilon}(\mathbf{Q}) \otimes \delta\hat{\varepsilon}^*(\mathbf{Q}) \rangle = V_{sc}\hat{\mathcal{G}}(\mathbf{Q})$, where $\otimes$ is the symbol of the tensor product.

It should be emphasized that, from the symmetry point of view, the fact that the scattered field in formula (3.5) is expressed in terms of a single 3D Fourier harmonic of fluctuations $\delta\hat{\varepsilon}$ is a consequence of spatial homogeneity of the system relative to its optical properties.

3.2. A Medium with Periodic Inhomogeneities

Let us now consider the situation of interest, when the properties of a medium change periodically.

In a simple approach that enables us to take into account periodic inhomogeneities, the so-called kinematic approximation in the theory of diffraction is used [8, 9]. We write permittivity tensor $\hat{\varepsilon}(\mathbf{r})$ in the form

$$\varepsilon_{\alpha\beta}(\mathbf{r}) = \varepsilon_0\delta_{\alpha\beta} + \Delta\varepsilon_{\alpha\beta}^0(\mathbf{r}) + \delta\varepsilon_{\alpha\beta}(\mathbf{r}),$$

where the terms $\Delta\hat{\varepsilon}^0(\mathbf{r})$ and $\delta\hat{\varepsilon}(\mathbf{r})$ take into account periodic inhomogeneities of the structure and random fluctuations, respectively. Treating $\Delta\hat{\varepsilon}^0(\mathbf{r}) + \delta\hat{\varepsilon}(\mathbf{r})$ as a perturbation, we can assume that the incident and scattered fields propagate in a homogeneous medium; analogously to expression (3.4), we then obtain the corresponding scattered field with polarization $\mathbf{e}^{(s)}$ in the form

$$E^{(s)}(\mathbf{r}) = E_0\frac{k_0^2 e^{ikr}}{4\pi r} \times [\mathbf{e}^{(s)} \cdot \Delta\hat{\varepsilon}^0(\mathbf{Q})\mathbf{e}^{(i)} + \mathbf{e}^{(s)} \cdot \delta\hat{\varepsilon}(\mathbf{Q})\mathbf{e}^{(i)}]. \tag{3.8}$$

The first term in the sum on the right-hand side of this equation corresponds to scattering from the periodic structure (diffraction), while the second term describes conventional Rayleigh scattering (3.4).

Formulas (3.4) and (3.8) correspond to the single scattering approximation. However, while this approximation in expression (3.4) is substantiated by the smallness of thermal fluctuations $\delta\hat{\varepsilon}(\mathbf{r})$, periodic part

$\Delta\hat{\varepsilon}_0(\mathbf{r})$ of inhomogeneities in relation (3.8) is usually not small and the single scattering approximation is generally invalid in this case. The application of a formula of type (3.8) for describing light scattering can be justified only in the case of very small periodic inhomogeneities or for very thin samples.

The problem of including the effect of small inhomogeneities on light scattering can be solved correctly if we pass from the description of scattering in terms of normal waves $\mathbf{E}_0(\mathbf{r}) \propto \exp(i\mathbf{k}\cdot\mathbf{r})$ and Green function $\hat{T}_0(\mathbf{r}-\mathbf{r}')$ of a homogeneous medium with permittivity $\varepsilon_0\delta_{\alpha\beta}$ to normal waves $\mathbf{E}^0(\mathbf{r})$ and Green function $\hat{T}^0(\mathbf{r},\mathbf{r}')$ of a periodically inhomogeneous medium with permittivity $\varepsilon^0_{\alpha\beta} = \varepsilon_0\delta_{\alpha\beta} + \Delta\varepsilon^0_{\alpha\beta}(\mathbf{r})$.

In this case, the field singly scattered from random fluctuations $\delta\hat{\varepsilon}$ is described by formula (2.9). This formula takes into account single scattering from fluctuations $\delta\hat{\varepsilon}(\mathbf{r})$ and all orders of scattering from periodic structure $\Delta\hat{\varepsilon}_0(\mathbf{r})$ (diffraction). The latter statement follows from the fact that Eqs. (2.7) and (2.8) defining $\mathbf{E}^0(\mathbf{r})$ and $\hat{T}^0(\mathbf{r},\mathbf{r}')$ in integral form can be written as

$$\mathbf{E}^0(\mathbf{r}) = \mathbf{E}_0(\mathbf{r}) + k_0^2\int \hat{T}_0(\mathbf{r}-\mathbf{r}')\Delta\hat{\varepsilon}_0(\mathbf{r}')\mathbf{E}^0(\mathbf{r}')d\mathbf{r}',$$

$$\hat{T}^0(\mathbf{r},\mathbf{r}') = \hat{T}_0(\mathbf{r}-\mathbf{r}') + k_0^2\int \hat{T}_0(\mathbf{r}-\mathbf{r}'')\Delta\hat{\varepsilon}_0(\mathbf{r}'')\hat{T}^0(\mathbf{r}'',\mathbf{r}')d\mathbf{r}''.$$

Iterating these equations, we indeed obtain all orders in $\Delta\hat{\varepsilon}_0$ on the right-hand side.

It should be noted that, in contrast to the case of homogeneous medium (3.1), Green function $\hat{T}^0(\mathbf{r},\mathbf{r}')$ is no longer a function of the difference $\mathbf{r}-\mathbf{r}'$ and normal waves $\mathbf{E}^0(\mathbf{r})$ do not have the simple form of a plane wave $\propto\exp(i\mathbf{k}\cdot\mathbf{r})$. Accordingly, plane-wave approximation (3.3) for the Green function in the coordinate representation is not valid either. Consequently, the intensity of single scattering in such a medium is not proportional to the 3D Fourier transform of permittivity fluctuations $\delta\varepsilon$ on wave vector $\mathbf{Q} = \mathbf{k}^{(s)} - \mathbf{k}^{(i)}$.

Thus, conventional algorithm (3.1)–(3.7) for calculating the intensity of scattering is inapplicable in this case and another approach is required.

In addition to the difficulty associated with the inclusion of periodic inhomogeneities, which can be overcome by using formula (2.9) instead of (3.4), there exists one more difficulty in the scattering problem for inhomogeneous systems. In contrast to formula (3.4), formula (2.9) describes the scattered field only inside the medium. However, in experiment, the intensities of scattering outside the medium are measured. For scattering media that are homogeneous on the average, this problem is usually solved as follows. In the simplest case, we assume that the scattering volume is placed in a homogeneous medium with permittivity ε_0, which enables us to disregard refraction at the sample boundary. In a more consistent approach, the refraction at the sample boundaries is taken into account. Since the incident wave in the homogeneous medium is plane, and the scattered wave in the far-field zone in the sample can also be regarded as quasi-plane, the problem of refraction can be solved using the conventional Fresnel formulas. However, in this case, there is a subtle point associated with a nontrivial correction of solid angled in the case of refraction even in an isotropic system. This problem for anisotropic scattering media is considered in [38].

Optical properties of the scattering system with periodic inhomogeneities in the scattering medium and of the surrounding homogeneous medium differ drastically. Normal waves and the Green function inside and outside the scattering volume are substantially different (in particular, the incident and scattered waves can be treated as plane only outside the sample); this does not allow us to disregard the presence of the boundary.

In order to overcome these difficulties, we can use the following algorithm for computing the intensity of single scattering in media with one-dimensional periodic inhomogeneities. Suppose that the scattering volume has the form of a flat layer $0 \le z \le L$ with large transverse dimensions $L_\perp \gg L$; a plane wave is incident on this volume from the side $z = -\infty$, while the scattered field is registered in the region $z > L$, i.e., in the front hemisphere. The latter is not of principal importance since scattering in the rear hemisphere ($z < 0$) can also be considered in the same way.

We first determine incident field $\mathbf{E}^{(i)}_{\text{in}}(\mathbf{r})$ in the medium, which is generated by incident plane wave $\mathbf{E}^{(i)}_{\text{out}}(\mathbf{r})$ with wave vector $\mathbf{k}^{(i)}$ outside the medium. Here and below, subscripts "in" and "out" correspond to the quantities calculated inside and outside a inhomogeneous medium, respectively. The relation between the field components inside and outside the sample at its boundary can easily be found on the basis of the general boundary conditions in electrodynamics. For flat-layered media with boundaries parallel to the layers, the wave inside the medium has the form

$$\mathbf{E}^{(i)}_{\text{in}}(\mathbf{r}) = \mathscr{E}^{(i)}(\mathbf{k}^{(i)}_\perp, z)e^{i\mathbf{k}^{(i)}_\perp\cdot\mathbf{r}_\perp}, \qquad (3.9)$$

where function $\mathscr{E}^{(i)}(\mathbf{k}^{(i)}_\perp, z)$ is determined by the properties of the flat-layered medium, the polarization of incident wave $\mathbf{E}^{(i)}_{\text{out}}(\mathbf{r})$, and its amplitude. Consequently, we require only the corresponding relation for the Fou-

rier components of fields $\mathbf{E}_{\text{out}}^{(i)}(\mathbf{k}_\perp^{(i)}, z)$ and $\mathbf{E}_{\text{in}}^{(i)}(\mathbf{k}_\perp^{(i)}, z)$ in coordinates x and y.

It should be noted that, in view of the identity $k_z^2 + k_\perp^2 = k_0^2\varepsilon_0$, which is valid outside the flat-layered medium, it is sufficient to specify vector $\mathbf{k}_\perp$ and the sign of component k_z in order to define the total wave vector $\mathbf{k}$. For this reason, it is sufficient to specify vector $\mathbf{k}_\perp^{(i)}$ if we know the direction of incidence of the wave on the sample (positive or negative relative to z). The same is also true of wave vector $\mathbf{k}_\perp^{(s)}$ of the scattered wave:

$$\mathbf{E}_{\text{in}}^{(s)}(\mathbf{r}) = \mathscr{E}^{(s)}(\mathbf{k}_\perp^{(s)}, z)e^{i\mathbf{k}_\perp^{(i)}\cdot\mathbf{r}_\perp}. \tag{3.10}$$

Scattered field $\mathbf{E}_{\text{in}}^{(s)}(\mathbf{k}_\perp^{(s)}, L)$ at boundary $z = L$ inside the inhomogeneous sample can be determined using formulas (2.9) and (3.9). We have

$$\begin{aligned}\mathbf{E}_{\text{in}}^{(s)}(\mathbf{k}_\perp^{(s)}, L) &= k_0^2\int_0^L dz'\hat{T}^0(\mathbf{k}_\perp^{(s)}; L, z')\\ &\times\delta\hat{\varepsilon}(\mathbf{k}_\perp^{(s)} - \mathbf{k}_\perp^{(i)}, z')\mathscr{E}^{(i)}(\mathbf{k}_\perp^{(i)}, z').\end{aligned} \tag{3.11}$$

It can be seen that, in addition to knowledge of functions $\mathscr{E}^{(i)}(\mathbf{k}_\perp^{(i)}, z)$, we will require in this case the expression for the Fourier components of Green function $\hat{T}^0(\mathbf{k}_\perp^{(s)}; z, z')$.

Using the boundary conditions, we can determine the relation between field $\mathbf{E}_{\text{out}}^{(s)}(\mathbf{r})$ outside the sample and field $\mathbf{E}_{\text{in}}^{(s)}(\mathbf{r})$ inside the sample at its boundary. In fact, we require the corresponding relation only for Fourier components $\mathbf{E}_{\text{out}}^{(s)}(\mathbf{k}_\perp^{(s)}, L)$ and $\mathbf{E}_{\text{in}}^{(s)}(\mathbf{k}_\perp^{(s)}, L)$.

Field $\mathbf{E}_{\text{out}}^{(s)}(\mathbf{r})$ at observation point $\mathbf{r}$ far away from the sample can be determined from the values of field $\mathbf{E}_{\text{out}}^{(s)}(\mathbf{r}_\perp, L)$ at the boundary of the scattering volume outside the sample with the help of the Kirchhoff method [39].

3.3. Kirchhoff Method

We will first elucidate the application of the Kirchhoff method in the case of a scalar wave field. We consider an arbitrary domain Γ bounded by a closed surface Σ that is completely outside an inhomogeneous sample (i.e., in a homogeneous medium). Inside domain Γ, field $E(\mathbf{r}) = E_{\text{out}}(\mathbf{r})$ satisfies the Helmholtz equation

$$(\Delta + k_0^2\varepsilon_0)E(\mathbf{r}) = 0. \tag{3.12}$$

Let function $T(\mathbf{r}, \mathbf{r}') = T_{\text{out}}(\mathbf{r}, \mathbf{r}')$ satisfy the equation

$$(\Delta + k_0^2\varepsilon_0)T(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}') \tag{3.13}$$

for all $\mathbf{r}, \mathbf{r}' \in \Gamma$. Then Eqs. (3.12) and (3.13) lead to the Kirchhoff–Helmholtz integral theorem [39]

$$\begin{aligned}E(\mathbf{r}) &= \int_\Sigma d^2\mathbf{r}'(T(\mathbf{r}, \mathbf{r}')\nabla_{\mathbf{r}'}E(\mathbf{r}')\\ &- E(\mathbf{r}')\nabla_{\mathbf{r}'}T(\mathbf{r}, \mathbf{r}'))\cdot\mathbf{s}(\mathbf{r}'),\end{aligned} \tag{3.14}$$

where $\mathbf{r} \in \Gamma$ is an arbitrary point, $\mathbf{r}' \in \Sigma$, and $\mathbf{s}(\mathbf{r}')$ is the outward normal to surface Σ at point $\mathbf{r}'$.

Equation (3.13) does not define function $T(\mathbf{r}, \mathbf{r}')$ unambiguously; additional boundary conditions are required. If we take $T|_\Sigma = 0$ as the boundary condition to this equation, field $E(\mathbf{r})$ at the point of observation in this case can be expressed in terms of the values of field $E(\mathbf{r}')$ on surface Σ:

$$E(\mathbf{r}) = -\int_\Sigma d^2\mathbf{r}'E(\mathbf{r}')\nabla_{\mathbf{r}'}T(\mathbf{r}, \mathbf{r}')\cdot\mathbf{s}(\mathbf{r}'). \tag{3.15}$$

The form of the Green function satisfying the condition $T|_\Sigma = 0$ is determined by the shape of the sample. Let us consider the simplest case when surface Σ is a part of plane $z = L$ with a large transverse dimension $L_\perp$, which is enclosed in a large hemisphere. If Green function $T(\mathbf{r}, \mathbf{r}')$ satisfies the radiation conditions at infinity (it is precisely these functions that will be chosen in our subsequent analysis), the contribution to integral (3.15) from this hemisphere tends to zero as its size increases. In this case, the boundary condition $T|_\Sigma = 0$ is reduced to $T|_{z=L} = 0$; using the mirror mapping method, we obtain

$$T(\mathbf{r}, \mathbf{r}') = \frac{1}{4\pi}\left(\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|} - \frac{e^{ik|\mathbf{r}-\mathbf{r}_1'|}}{|\mathbf{r}-\mathbf{r}_1'|}\right), \tag{3.16}$$

where $\mathbf{r}_1'$ is the mirror image of point $\mathbf{r}'$ relative to the flat boundary $z = L$.

We assume that the field is measured at point $\mathbf{r} = (\mathbf{r}_\perp, z)$, $z - L \gg L_\perp$. Then we can apply an approximation of type (3.3) in both terms of formula (3.16). Considering that $\mathbf{s}(\mathbf{r}')\cdot\nabla_{\mathbf{r}'} = -\partial/\partial z'$ in our geometry, we obtain from Eq. (3.15)

$$\begin{aligned}E(\mathbf{r}) &= \frac{-ik_0\sqrt{\varepsilon_0}}{2\pi}\frac{e^{ikr}}{r}\frac{z}{r}e^{-ik_z^{(s)}L}\\ &\times\int_\Sigma d^2\mathbf{r}_\perp' E(\mathbf{r}_\perp', L)e^{-i\mathbf{k}_\perp^{(s)}\cdot\mathbf{r}_\perp'}.\end{aligned} \tag{3.17}$$

Thus, for $L_\perp \gg \lambda$, where λ is the wavelength of light, we find that the field at point $\mathbf{r}$ is proportional to the trans-

verse Fourier component of the field on plane Σ:

$$E(\mathbf{r}) = \frac{-ik_0\sqrt{\varepsilon_0}}{2\pi}\frac{e^{ikr}}{r}\frac{z}{r}e^{-ik_z^{(s)}L}E(\mathbf{k}_\perp^{(s)}, L). \quad (3.18)$$

Then the corresponding intensity has the form

$$I \propto |E(\mathbf{r})|^2 = \frac{k_0^2\varepsilon_0}{4\pi^2}\frac{1}{r^2}\left(\frac{z}{r}\right)^2\left|E(\mathbf{k}_\perp^{(s)}, L)\right|^2. \quad (3.19)$$

Let us now consider the actual vector form $\mathbf{E}(\mathbf{r})$ of the electromagnetic field. Field $\mathbf{E}(\mathbf{r}) = \mathbf{E}_{out}(\mathbf{r})$ outside the inhomogeneous sample satisfies the wave equation

$$(\mathrm{curl}\,\mathrm{curl} - k_0^2\varepsilon_0)\mathbf{E}(\mathbf{r}) = 0. \quad (3.20)$$

It can easily be verified that the system of three coupled equations (3.20) is equivalent to the system

$$\begin{cases}(\Delta + k_0^2\varepsilon_0)\mathbf{E}(\mathbf{r}) = 0,\\ \mathrm{div}\mathbf{E}(\mathbf{r}) = 0.\end{cases} \quad (3.21)$$

The first equation in this system indicates that each of the three vector components of the field satisfies the scalar Helmholtz equation

$$(\Delta + k_0^2\varepsilon_0)E_\alpha(\mathbf{r}) = 0, \quad (3.22)$$

while the second equation shows that all three components together satisfy the additional condition $\mathrm{div}\mathbf{E} = 0$, which corresponds to a transverse electromagnetic field. Consequently, a scalar formula of the type of Kirchhoff equation (3.18) is formally applicable to each of three field components E_α. However, we disregarded the transversality condition $\mathrm{div}\mathbf{E} = 0$ in Eqs. (3.21). In order to take this condition into account, we multiply the set of three scalar formulas (3.18) for each component of the field by projector $\hat{P}(\mathbf{r})$ (3.2), which ensures the transversality of the field in the far-field zone. As a result, we obtain a vector analog of the Kirchhoff formula in the form[1]

$$\mathbf{E}(\mathbf{r}) = \frac{-ik_0\sqrt{\varepsilon_0}}{2\pi}\frac{e^{ikr}}{r}\frac{z}{r}e^{ik_z^{(s)}L}\hat{P}(\mathbf{r})\mathbf{E}(\mathbf{k}_\perp^{(s)}, L). \quad (3.23)$$

[1] It is well known [39] that direct extension of the scalar Kirchhoff formula to the vector case gives rise to the problem of violation of the transversality condition $\mathrm{div}\mathbf{E} = 0$. In order to eliminate this contradiction, we can use the vector Kirchhoff–Kotler formula (see, for example, [40]). However, in the approximation of the far-field zone of the sample, the nontransversality of field $\mathbf{E}$ is on the order of $\lambda/L_\perp \ll 1$ when the conventional Kirchhoff formulas are used. In such a situation, the application of simple vector formula (3.23), in which the field transversality is ensured by projector $P_{\alpha\beta}(\mathbf{r})$, is equivalent to applying the simple Kirchhoff–Kotler method.

In this case, in view of the relation $e_\alpha^{(s)}P_{\alpha\beta} = e_\beta^{(s)}$, we obtain, analogously to Eqs. (3.5)–(3.7), the formula for the intensity of the scattered field component with polarization $\mathbf{e}^{(s)}$:

$$I^{(s)} = \frac{\sqrt{\varepsilon_0}c^2}{8\pi}\frac{k_0^2\varepsilon_0}{4\pi^2}\frac{1}{r^2}\left(\frac{z}{r}\right)^2\left\langle\left|\mathbf{e}^{(s)}\cdot\mathbf{E}_{out}^{(s)}(\mathbf{k}_\perp^{(s)}, L)\right|^2\right\rangle. \quad (3.24)$$

In the case of spatially homogeneous systems, this formula is transformed into expression (3.7). Indeed, carrying out the Fourier transformation with respect to transverse variables x and y in formula (2.9), we obtain

$$\mathbf{E}^{(s)}(\mathbf{k}_\perp^{(s)}, L) = E_0k_0^2\int_0^L dz'\hat{T}^0(\mathbf{k}_\perp^{(s)}, L - z') \times \delta\hat{\varepsilon}(\mathbf{k}_\perp^{(s)} - \mathbf{k}_\perp^{(i)}, z')\mathbf{e}^{(i)}e^{ik_z^{(i)}z'}. \quad (3.25)$$

The Fourier transform of Green function $\hat{T}^0(\mathbf{r})$ with respect to coordinates x and y has the form

$$\hat{T}^0(\mathbf{k}_\perp^{(s)}, L - z') = \frac{i}{2k_z^{(s)}}e^{ik_z^{(s)}|L-z'|}\hat{P}(\mathbf{k}^{(s)}), \quad (3.26)$$

where $k_z^{(s)} = \sqrt{k_0^2\varepsilon_0 - k_\perp^{(s)2}}$. For a homogeneous system, in which the values of ε_0 inside and outside the sample coincide, we have

$$\mathbf{E}_{in}^{(s)}(\mathbf{k}_\perp^{(s)}, L) = \mathbf{E}_{out}^{(s)}(\mathbf{k}_\perp^{(s)}, L).$$

Since $\mathbf{k}^{(s)} = k_0\sqrt{\varepsilon_0}\,\mathbf{r}/r$, we can write

$$k_z^{(s)} = k_0\sqrt{\varepsilon_0}z/r, \quad \hat{P}(\mathbf{k}^{(s)}) = \hat{P}(\mathbf{r}).$$

Substituting Eq. (3.26) into (3.25) and taking into account the condition $z' < L$, we obtain

$$\mathbf{E}_{(out)}^{(s)}(\mathbf{k}_\perp^{(s)}, L) = \frac{iE_0k_0^2}{2k_z^{(s)}}\hat{P}(\mathbf{r})\delta\hat{\varepsilon}(\mathbf{k}_s - \mathbf{k}_i)\mathbf{e}^{(i)}e^{ik_z^{(s)}L}. \quad (3.27)$$

Substituting this expression into formula (3.23) and forming the dot product of the left- and right-hand sides and vector $\mathbf{e}^{(s)}$, we obtain Eq. (3.5).

Applying formula (3.24) to the problem of light scattering in CLCs, we cannot disregard the difference between fields $\mathbf{E}_{in}$ and $\mathbf{E}_{out}$. In order to find the relation between these fields, we take into account the fact that the tangential component of the field does not change

upon the transition through the interface between two media,

$$\begin{aligned} E^{(i)}_{\text{out}\perp}(\mathbf{k}^{(i)}_\perp, 0) &= E^{(i)}_{\text{in}\perp}(\mathbf{k}^{(i)}_\perp, 0), \\ E^{(s)}_{\text{out}\perp}(\mathbf{k}^{(s)}_\perp, L) &= E^{(s)}_{\text{in}\perp}(\mathbf{k}^{(s)}_\perp, L), \end{aligned} \tag{3.28}$$

while components $E_{\text{out}\,z}(\mathbf{k}^{(s)}_\perp, L)$ and $E_{\text{out}\,z}(\mathbf{k}^{(i)}_\perp, 0)$ can be obtained from the condition $\text{div}\mathbf{D} = 0$.

In the case when the pitch is much larger than the wavelength, $d \gg \lambda$, we can use the geometrical optics approximation for the interior of the CLC. Fields $\mathbf{E}_{\text{in}}$ and $\mathbf{E}_{\text{out}}$ at the boundaries inside and outside the medium satisfy linear equations of the form

$$\begin{aligned} \mathbf{E}^{(i)}_{\text{in}}(\mathbf{k}^{(i)}_\perp, 0) &= \hat{M}^{\text{out}\to\text{in}}(\mathbf{k}^{(i)}_\perp, 0)\mathbf{E}^{(i)}_{\text{out}}(\mathbf{k}^{(i)}_\perp, 0), \\ \mathbf{E}^{(s)}_{\text{out}}(\mathbf{k}^{(s)}_\perp, 0) &= \hat{M}^{\text{in}\to\text{out}}(\mathbf{k}^{(s)}_\perp, L)\mathbf{E}^{(s)}_{\text{in}}(\mathbf{k}^{(s)}_\perp, L), \end{aligned} \tag{3.29}$$

where transition matrices $\hat{M}^{\text{out}\to\text{in}}$ and $\hat{M}^{\text{in}\to\text{out}}$ can easily be derived from Eqs. (3.28) and the condition $\text{div}\mathbf{D} = 0$.

Using formulas (3.11), (3.29) and relation $\langle \delta\hat{\varepsilon}(\mathbf{k}_\perp, z) \otimes \delta\hat{\varepsilon}^*(\mathbf{k}_\perp, z')\rangle = S_\perp \hat{\mathcal{G}}(\mathbf{k}_\perp; z, z')$, where $S_\perp$ is the cross-sectional area of the sample, we find the corresponding quantity in formula (3.24), which defines the intensity of single scattering:

$$\begin{aligned} \left\langle \left| \mathbf{e}^{(s)} \cdot \mathbf{E}^{(s)}_{\text{out}}(\mathbf{k}^{(s)}_\perp, L) \right|^2 \right\rangle &= k_0^4 S_\perp e^{(s)}_\alpha e^{(s)}_\gamma \\ \times \hat{M}^{\text{in}\to\text{out}}_{\alpha\beta}(\mathbf{k}^{(s)}_\perp, L) \hat{M}^{\text{in}\to\text{out}}_{\gamma\delta}(\mathbf{k}^{(s)}_\perp, L) &\int_0^L dz_1 \\ \times \int_0^L dz_2 T^0_{\beta\rho}(\mathbf{k}^{(s)}_\perp; L, z_1) T^{0*}_{\delta\varphi}(\mathbf{k}^{(s)}_\perp; L, z_2) & \\ \times \mathcal{G}_{\rho\nu\varphi\mu}(\mathbf{k}^{(s)}_\perp - \mathbf{k}^{(i)}_\perp; z_1, z_2) \mathcal{E}^{(i)}_\nu(\mathbf{k}^{(i)}_\perp, z_1) \mathcal{E}^{(i)*}_\mu(\mathbf{k}^{(i)}_\perp, z_2). & \end{aligned} \tag{3.30}$$

Here, we have omitted for simplicity transition matrices $\hat{M}^{\text{out}\to\text{in}}(\mathbf{k}^{(i)}_\perp, 0)$, which make it possible to express the field inside the medium in terms of the field incident on the sample.

4. OPTICS OF CLCs WITH A LARGE PITCH

Let us consider the problem of propagation of waves in CLCs. It is more convenient for our purposes to consider directly, instead of wave equation (2.7), the system of Maxwell equations (2.4) for an equilibrium CLC. Taking into account the relations $\text{div}\hat{\varepsilon}^0\mathbf{E} = 0$ and $\text{div}\mathbf{H} = 0$ in the $(\mathbf{q}, z)$ representation, we can reduce the problem to a system of equations of the form

$$\frac{\partial}{\partial\xi}\begin{pmatrix} E_x \\ E_y \\ H_x \\ -H_y \end{pmatrix} = -i\Omega \begin{pmatrix} 0 & 0 & 0 & 1 - q^2/k_0^2\varepsilon_\perp \\ 0 & 0 & 1 & 0 \\ \varepsilon^0_{xy}(\xi) & \varepsilon^0_{yy}(\xi) - q^2/k_0^2 & 0 & 0 \\ \varepsilon^{(0)}_{xx}(\xi) & \varepsilon^0_{xy}(\xi) & 0 & 0 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ H_x \\ -H_y \end{pmatrix}, \tag{4.1}$$

where $\xi = p_0 z$ is a dimensionless variable, $\Omega = k_0/p_0 = 2d/\lambda$ is a dimensionless parameter, and λ is the wavelength of light. The x and y coordinate axes are chosen in the same way as in Eq. (2.27); i.e., the direction of the x axis coincides with the direction of vector $\mathbf{q}$. In accordance with relations (2.2) and (2.3), the permittivity tensor components have the form

$$\begin{aligned} \varepsilon^0_{xx} &= \varepsilon_\perp + \varepsilon_a \cos^2\phi, \quad \varepsilon^0_{xy} = \varepsilon_a \sin\phi\cos\phi, \\ \varepsilon^0_{yy} &= \varepsilon_\perp + \varepsilon_a \sin^2\phi. \end{aligned}$$

Since we consider an equilibrium CLC in this section, we have omitted superscript "0" on field component $\mathbf{E}(\mathbf{r})$ (the subscript on the components of point source field $\hat{T}(\mathbf{r}, \mathbf{r}')$ will also be omitted).

A system of type (4.1) was used in [41] for describing the propagation of waves in CLCs. An analogous system of equations was numerically solved in [41] for $\lambda \sim d$. In our case, where $\lambda \ll d$ ($\Omega \gg 1$), direct numerical methods are ineffective in view of the rapidly oscillating solution; for constructing the solution, the vector generalization of the WKB method is more effective here.

4.1. Normal Waves

An asymptotic WKB solution to system (4.1) was obtained in [29]. According to [29], four normal waves exist for a given $\mathbf{q}$ in CLCs with a large pitch. Two of these waves propagate in the direction of positive values of z, while the other two waves propagate in the opposite direction. Reconstructing components E_z from

the equation $\mathrm{div}\hat{\varepsilon}^0\mathbf{E} = 0$, we can write these waves in the form

$$\mathbf{E}_{\pm}^{(j)}(\mathbf{r}) = E_0^{(j)}A^{(j)}(\mathbf{q}; z, z_0)\mathbf{e}^{(j)}(\mathbf{q}, z) \times \exp\left(i\mathbf{q}\cdot\mathbf{r}_\perp \pm i\int_{z_0}^{z} k_z^{(j)}(\mathbf{q}, z')dz'\right), \quad (4.2)$$

where $j = 1, 2$; the plus and minus signs correspond to the direction of wave propagation. Constants $E_0^{(j)}$ characterize the initial field amplitude on the plane z_0. Here, $k_z^{(j)}(\mathbf{q}, z) \sim k_0$; $\mathbf{e}^{(j)}(\mathbf{q}, z)$ are unit vectors, and $A^{(j)}(\mathbf{q}; z, z_0)$ are the amplitude factors. All these quantities are smoothly varying functions over the scale of λ. Consequently, waves (4.2) are locally plane waves with wave vectors $\mathbf{k}^{(j)}(\mathbf{q}, z) = (\mathbf{q}, \pm k_z^{(j)}(\mathbf{q}, z))$ and polarization vectors $\mathbf{e}^{(j)}(\mathbf{q}, z)$. In the main order in large parameter Ω, we have

$$k_z^{(1)}(\mathbf{q}, z) \equiv k_z^{(1)}(q) = \sqrt{\varepsilon_\perp k_0^2 - q^2},$$
$$k_z^{(2)}(\mathbf{q}, z) = \sqrt{\varepsilon_\parallel k_0^2 - q^2 - \frac{\varepsilon_a}{\varepsilon_\perp}(\mathbf{q}\cdot\mathbf{n}^0(z))^2}, \quad (4.3)$$

$$A^{(1)}(\mathbf{q}; z, z_0) = 1, \quad A^{(2)}(\mathbf{q}; z, z_0) = \sqrt{\frac{\varepsilon_\perp^2 k_0^2 + \varepsilon_a(\mathbf{q}\cdot\mathbf{n}^0(z))^2}{\varepsilon_\perp^2 k_0^2 + \varepsilon_a(\mathbf{q}\cdot\mathbf{n}^0(z_0))^2}}\sqrt{\frac{k_z^{(2)}(\mathbf{q}, z_0)}{k_z^{(2)}(\mathbf{q}, z)}}, \quad (4.4)$$

and polarization vectors $\mathbf{e}^{(j)}(\mathbf{q}, z)$ in the coordinate system used above have the form

$$\mathbf{e}^{(1)}(\mathbf{q}, z) = \frac{1}{\sqrt{k_0^2\varepsilon_\perp - q^2\cos^2\phi}}(k_z^{(1)}\sin\phi, -k_z^{(1)}\cos\phi, -q\sin\phi),$$
$$\mathbf{e}^{(2)}(\mathbf{q}, z) = \frac{\sqrt{\varepsilon_\perp}}{\sqrt{(k_0^2\varepsilon_\perp - q^2\cos^2\phi)(k_0^2\varepsilon_\perp^2 + \varepsilon_a q^2\cos^2\phi)}}((k_0^2\varepsilon_\perp - q^2)\cos\phi, k_0^2\varepsilon_\perp\sin\phi, -qk_z^{(2)}\cos\phi). \quad (4.5)$$

Since vectors $\mathbf{e}^{(j)}(\mathbf{q}, z)$ in Eqs. (4.5) are real-valued, waves (4.2) in the main order in Ω locally exhibit a linear polarization.[2]

Quantities $A^{(j)}(\mathbf{q}; z, z_0)$ in Eqs. (4.4) can be written in the form

$$A^{(j)}(\mathbf{q}; z, z_0) = \frac{B^{(j)}(\mathbf{q}, z)}{B^{(j)}(\mathbf{q}, z_0)}, \quad (4.6)$$

where

$$B^{(1)}(\mathbf{q}, z) \equiv B^{(1)}(q) = \sqrt{\frac{k_0}{k_z^{(1)}(q)}},$$
$$B^{(2)}(\mathbf{q}, z) = \frac{\sqrt{\varepsilon_\perp^2 k_0^2 + \varepsilon_a(\mathbf{q}\cdot\mathbf{n}^0(z))^2}}{\varepsilon_\perp\sqrt{k_0 k_z^2(\mathbf{q}, z)}}. \quad (4.7)$$

Formulas (4.2)–(4.7) are quite formal in appearance. The physical meaning of waves $\mathbf{E}^{(j)}(\mathbf{r})$ will become clearer if we pay attention to the fact that polarization vectors $\mathbf{e}^{(j)}(\mathbf{q}, z)$ in Eqs. (4.5) satisfy the conditions

$$\mathbf{e}^{(1)}(\mathbf{q}, z) \perp \mathbf{n}^0(z), \quad \mathbf{e}^{(1)}(\mathbf{q}, z) \perp \mathbf{k}^{(1)}(\mathbf{q}, z), \quad \hat{\varepsilon}^0(z)\mathbf{e}^{(2)}(\mathbf{q}, z) \perp \mathbf{k}^{(2)}(\mathbf{q}, z), \quad (4.8)$$

and that vector $\mathbf{e}^{(2)}(\mathbf{q}, z)$ lies in the plane of vectors $\mathbf{k}^{(2)}(\mathbf{q}, z)$ and $\mathbf{n}^0(z)$. It should also be noted that, in accordance with Eqs. (4.3), the wave numbers

$$k^{(j)}(\mathbf{q}, z) = \sqrt{q^2 + k_z^{(j)2}(\mathbf{q}, z)}$$

of these waves satisfy the relations

$$k^{(1)2}(\mathbf{q}, z) = k_0^2\varepsilon_\perp,$$
$$k^{(2)2}(\mathbf{q}, z) = k_0^2\frac{\varepsilon_\perp\varepsilon_\parallel}{\varepsilon_\perp + \varepsilon_a\cos^2\theta}, \quad (4.9)$$

where θ is the angle between $\mathbf{n}^0(z)$ and $\mathbf{k}^{(2)}(\mathbf{q}, z)$. It should be borne in mind here that the second of these equalities is an equation for $k^{(2)}(\mathbf{q}, z)$ since

$$\cos\theta = \mathbf{n}^0(z)\cdot\mathbf{k}^{(2)}(\mathbf{q}, z)/k^{(2)}(\mathbf{q}, z) = q\cos\phi(z)/k^{(2)}(\mathbf{q}, z).$$

A comparison of formulas (4.8) and (4.9) with the conventional formulas for polarizations and wave vectors of normal waves in a homogeneous uniaxial anisotropic medium [37] shows that two waves (4.2)–(4.5) are locally the ordinary (superscript (1)) and extraordinary (superscript (2)) waves at a given point in the CLC.

Thus, formulas (4.2)–(4.5) correspond to the adiabatic mode of wave propagation. These formulas can be

[2] In the next order in the large parameter, vectors $\mathbf{e}^{(j)}(\mathbf{q}, z)$ acquire an imaginary correction and, hence, waves (4.2) become weakly polarized elliptically [31] (cf. formula (A.12) in Appendix).

treated as a generalization of the well-known Mauguin solution [24] to the case of oblique incidence. The physical meaning of these formulas is as follows. When a normal wave with number j is incident on the plane $z = z_0$, it acquires a phase incursion $\int_{z_0}^{z} k_z^{(j)}(\mathbf{q}, z')dz'$ as a result of propagation in the medium and falls on the z plane as a normal wave with the same number j. Since $\mathbf{e}^{(j)}(\mathbf{q}, z)$ differs from $\mathbf{e}^{(j)}(\mathbf{q}, z_0)$, the polarization vector rotates in this case. The fact that amplitude factors $A^{(j)}(\mathbf{q}, z)$ in Eqs. (4.4) become functions of z is associated with the energy conservation that must be ensured for a wave propagating in an inhomogeneous medium without absorption (see Eqs. (4.11)–(4.14) below).

Wave vector $\mathbf{k}^{(j)}(\mathbf{q}, z)$ at a given point of the CLC is directed along the normal to the wave front. For the ordinary ray, wave vector $\mathbf{k}^{(1)} = \mathbf{k}^{(1)}(q)$ does not depend on point z, while wave vector $\mathbf{k}^{(2)} = \mathbf{k}^{(2)}(\mathbf{q}, z)$ for the extraordinary ray varies in magnitude and direction depending on z. At the same time, the direction of polarization vectors $\mathbf{e}^{(j)}(\mathbf{q}, z)$ vary with z for both types of waves. However, in any case, for a fixed value of $\mathbf{q}$, wave vector $\mathbf{k}^{(j)}(\mathbf{q}, z)$ always lies in the same plane (containing vectors $\mathbf{q}$ and $\mathbf{e}_z$) both for the ordinary and the extraordinary ray.

The trajectory of wave propagation in an anisotropic medium is characterized by the ray vector of the wave, which is directed along the Poynting vector. For a plane wave in an anisotropic medium, the Poynting vector has the form [37]

$$\mathbf{S}(\mathbf{r}) = \frac{c}{8\pi k_0}[\mathbf{k}|E|^2 - \mathbf{E}(\mathbf{E}^* \cdot \mathbf{k})]. \tag{4.10}$$

In particular, for waves (4.2), we obtain

$$\mathbf{S}^{(j)}(\mathbf{r}) = \frac{c|E_0^{(j)}|^2}{8\pi k_0} A^{(j)2}(\mathbf{q}, z) \times [\mathbf{k}^{(j)}(\mathbf{q}, z) - \mathbf{e}^{(j)}(\mathbf{q}, z)(\mathbf{k}^{(j)}(\mathbf{q}, z) \cdot \mathbf{e}^{(j)}(\mathbf{q}, z))]. \tag{4.11}$$

Since $\mathbf{e}^{(1)} \cdot \mathbf{k}^{(1)} = 0$, we have $\mathbf{S}^{(1)}(\mathbf{q}, z) \parallel \mathbf{k}^{(1)}(q)$ and the ordinary ray has a rectilinear trajectory. At the same time, $\mathbf{e}^{(2)} \cdot \mathbf{k}^{(2)} \neq 0$ in the general case; in accordance with expression (4.11), vector $\mathbf{S}^{(2)}(\mathbf{q}, z)$ does not remain in the same plane upon a variation of z. Consequently, the trajectory of the extraordinary ray, a tangent to which at each point must be parallel to $\mathbf{S}^{(j)}(\mathbf{r})$ at this point, does not lie in the same plane. Figure 3 shows a typical trajectory of the extraordinary ray, calculated on the basis of formula (4.11).

Let us analyze the consequences of the energy conservation law $\operatorname{div}\mathbf{S}^{(j)} = 0$ for waves (4.2). In our case, we have $\operatorname{div}\mathbf{S}^{(j)} = \partial_z S_z^{(j)}(\mathbf{q}, z) = 0$. Consequently, component $S_z^{(j)}(\mathbf{q}, z)$ is independent of z. Then, we obtain from Eq. (4.11)

$$A^{(j)2}(\mathbf{q}, z) = C_0^{(j)}(\mathbf{q}) \times k_0[k_z^{(j)}(\mathbf{q}, z) - e_z^{(j)}(\mathbf{q}, z)(\mathbf{k}^{(j)}(\mathbf{q}, z) \cdot \mathbf{e}^{(j)}(\mathbf{q}, z))]^{-1}, \tag{4.12}$$

where $C_0^{(j)}$ are arbitrary dimensionless functions of $\mathbf{q}$. Using Eqs. (4.5), we find

$$\frac{k_z^{(j)} - e_z^{(j)}\mathbf{k}^{(j)} \cdot \mathbf{e}^{(j)}}{k_0} = \begin{cases} k_z^{(1)}(q)/k_0 & \text{for } j = 1 \\ \dfrac{k_z^{(2)}(\mathbf{q}, z)k_0\varepsilon_\perp^2}{k_0^2\varepsilon_\perp^2 + \varepsilon_a q^2 \cos^2\phi(z)} & \text{for } j = 2. \end{cases} \tag{4.13}$$

In accordance with Eqs. (4.7), the right-hand side of this equation coincides with $1/B^{(j)2}(\mathbf{q}, z)$ and we

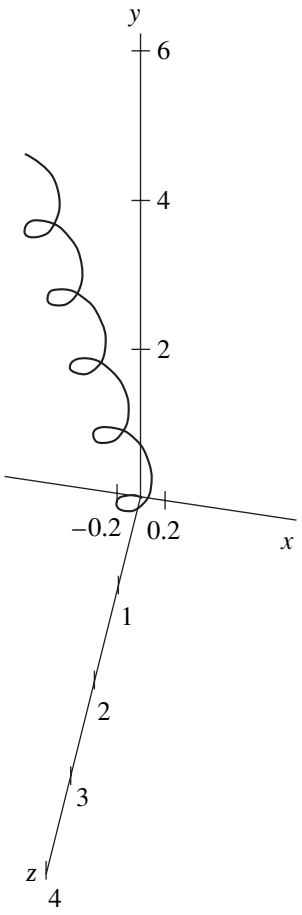

Fig. 3. Trajectory of the extraordinary ray in a CLC. Calculations were made for $\varepsilon_a = 2.0$ and $\varepsilon_\parallel = 2.5$. All distances are given in units of d.

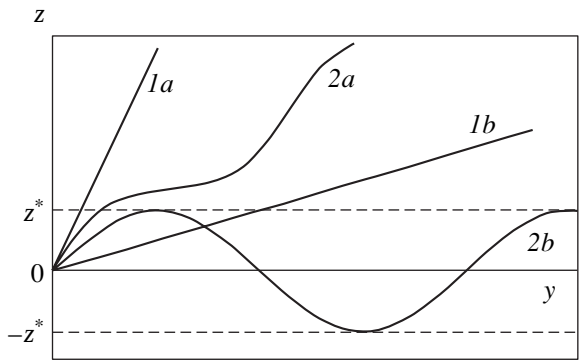

Fig. 4. Types of trajectories of rays in CLC: *1a* and *1b* are ordinary rays, curve *2a* is the extraordinary ray outside the waveguide channel, and curve *2b* is the extraordinary ray trapped in the waveguide channel.

obtain from Eq. (4.12)

$$A^{(j)2}(\mathbf{q}, z) = C_0^{(j)}(\mathbf{q})B^{(j)2}(\mathbf{q}, z). \quad (4.14)$$

This expression transforms into Eq. (4.6) if we set $C_0^{(j)}(\mathbf{q}) = 1/B^{(j)2}(\mathbf{q}, z_0)$. The latter choice is convenient since amplitude $E_0^{(j)}$ in Eq. (4.2) is independent of the choice of initial point z_0 in this case.

Let us consider the conditions under which waves (4.2) can propagate to the bulk of the CLC. Let us first consider the ordinary wave. It should be noted that $k_z^{(1)}$ is real-valued under the condition $q^2 \le k_0^2\varepsilon_\perp$. In this case, the ordinary wave propagates to the region with any values of z. For $q^2 > k_0^2\varepsilon_\perp$, quantity $k_z^{(1)}$ becomes imaginary and the ordinary wave does not propagate in the medium with arbitrary values of z and z_0.

For the extraordinary wave, the conditions of propagation are more complex and depend on the relation between parameters q and z. In this case, the following situations are possible.

(i) If $q^2 > k_0^2\max(\varepsilon_\parallel, \varepsilon_\perp)$, the value of $k_z^{(2)}(\mathbf{q}, z)$ is imaginary for any z and such a wave does not propagate in the CLC.

(ii) If $q^2 \le k_0^2\min(\varepsilon_\parallel, \varepsilon_\perp)$, the value of $k_z^{(2)}(\mathbf{q}, z)$ will be real for all values of z and the wave propagates in the medium into the region with any z.

(iii) If $k_0^2\min(\varepsilon_\parallel, \varepsilon_\perp) < q^2 \le k_0^2\max(\varepsilon_\parallel, \varepsilon_\perp)$, the extraordinary wave can propagate in the medium only for definite values of z. The range of corresponding values of z is determined by the inequality $\cos^2\phi(z) \le \varepsilon_\perp(k_0^2\varepsilon_\parallel - q^2)/q^2\varepsilon_a$ for $\varepsilon_a > 0$ and the inequality $\cos^2\phi(z) \ge \varepsilon_\perp(k_0^2\varepsilon_\parallel - q^2)/q^2\varepsilon_a$ for $\varepsilon_a < 0$ (it should be noted that the condition $0 < \varepsilon_\perp(k_0^2\varepsilon_\parallel - q^2)/q^2\varepsilon_a < 1$ is satisfied in the range of q in question).

Thus, in the latter case, the effect of trapping of the extraordinary ray in the CLC is observed [31, 32]. The physical pattern of the effect is that the extraordinary ray begins to rotate and the quantity $k_z^{(2)}(\mathbf{q}, z)$ vanishes at a certain point $z = z^*(\mathbf{q})$ and then reverses its sign. In a certain sense, this effect is analogous to the total internal reflection from a certain plane in the medium. Since the refractive index is a periodic function, such a ray will be alternately reflected from two planes perpendicular to the z axis. Consequently, a plane wave channel is formed, in which the extraordinary wave can propagate to the region of indefinitely large values of $\mathbf{r}_\perp$, remaining within a period in z. Figure 4 schematically shows the projections of the trajectories of the ordinary and extraordinary rays onto the yz plane, which demonstrate the effect of waveguide propagation.

4.2. The Field of a Point Source

The corresponding problem (2.8) for the field $\hat{T}^0(\mathbf{q}; z, z_1)$ of a point source can be reduced to a system of equations of the form

$$\hat{\mathcal{L}}(z)\hat{T}^0(\mathbf{q}; z, z_1) = \delta(z - z_1)\hat{I}, \quad (4.15)$$

where

$$\hat{\mathcal{L}}(z) = \begin{pmatrix} -\frac{\partial^2}{\partial z^2} - k_0^2\varepsilon_{xx}^0 & -k_0^2\varepsilon_{xy}^0 & iq\frac{\partial}{\partial z} \\ -k_0^2\varepsilon_{xy}^0 & -\frac{\partial^2}{\partial z^2} + q^2 - k_0^2\varepsilon_{yy}^0 & 0 \\ iq\frac{\partial}{\partial z} & 0 & q^2 - k_0^2\varepsilon_\perp \end{pmatrix}$$

is a second-order linear differential operator.

We obtained the solution to this system in [32] with the help of the vector WKB method. The construction of the solution is based on the method developed in [29] for a homogeneous equation. The solution of inhomogeneous equation (4.15) is constructed as a superposition of solutions (4.2) to the corresponding homogeneous equation in two regions $z > z_1$ and $z < z_1$ separately by choosing the superposition coefficients that ensure the required type of the singularity on the right-hand side of Eq. (4.15). The result [32] has the form

$$\hat{T}^0(\mathbf{q}; z, z_1) = \hat{T}^{(1)}(\mathbf{q}; z, z_1) + \hat{T}^{(2)}(\mathbf{q}; z, z_1), \quad (4.16)$$

where

$$T^{(j)}_{\alpha\beta}(\mathbf{q}; z, z_1) = \frac{i}{2k_0} B^{(j)}(\mathbf{q}, z) B^{(j)}(\mathbf{q}, z_1) e^{(j)}_{\alpha}(\mathbf{q}, z) \times e^{(j)}_{\beta}(\mathbf{q}, z_1) \exp\left(i \left| \int_{z_1}^{z} k^{(j)}_z(\mathbf{q}, z') dz' \right| \right). \quad (4.17)$$

Let us analyze the limit $p_0 \longrightarrow 0$ in Eq. (4.17), which corresponds to a homogeneous uniaxial anisotropic medium (nematic liquid crystal in our physical context). In this limit, quantities $\mathbf{n}^0(z)$, $k^{(j)}_z(\mathbf{q}, z)$, $B^{(j)}(\mathbf{q}, z)$, and $\mathbf{e}^{(j)}(\mathbf{q}, z)$ become independent of point z: $\mathbf{n}^0(z) = \mathbf{n}^0$, $k^{(j)}_z(\mathbf{q}, z) = k^{(j)}_z(\mathbf{q})$, $B^{(j)}(\mathbf{q}, z) = B^{(j)}(\mathbf{q})$, and $\mathbf{e}^{(j)}(\mathbf{q}, z) = \mathbf{e}^{(j)}(\mathbf{q})$, while functions $\hat{T}^{(j)}(\mathbf{q}; z, z_1)$ become dependent only on the difference in the spatial coordinates: $\hat{T}^{(j)}(\mathbf{q}; z, z_1) = \hat{T}^{(j)}(\mathbf{q}; z - z_1)$. In this case, formula (4.17) assumes the form

$$T^{(j)}_{\alpha\beta}(\mathbf{q}; z - z_1) = \frac{i}{2k_0} B^{(j)2}(\mathbf{q}) e^{(j)}_{\alpha}(\mathbf{q}) e^{(j)}_{\beta}(\mathbf{q}) e^{ik^{(j)}_z(\mathbf{q})|z - z_1|}. \quad (4.18)$$

The expression for the field of a point source in a homogeneous anisotropic medium in the 3D Fourier representation has the form [42]

$$T^0_{\alpha\beta}(\mathbf{Q}) = \frac{1}{k_0^2} \sum_{j=1,2} \frac{e_{(j)\alpha}(\mathbf{Q}) e_{(j)\beta}(\mathbf{Q})}{\mathbf{e}_{(j)}(\mathbf{Q}) \hat{\varepsilon}^0 \mathbf{e}_{(j)}(\mathbf{Q})} \times \frac{k^{(2)}_{(j)}(\mathbf{Q})}{Q^2 - k^{(2)}_{(j)}(\mathbf{Q}) - i0} - \frac{Q_\alpha Q_\beta}{\mathbf{Q}\hat{\varepsilon}^0\mathbf{Q}}. \quad (4.19)$$

Here, $\mathbf{Q}$ is the 3D wave vector, $\mathbf{e}_{(j)}$ are the polarization vectors, and $k_{(j)}$ are the wave numbers of two (ordinary and extraordinary) plane waves propagating in the homogeneous anisotropic medium. The last term on the right-hand side of this equation is associated with the near field of a dipole and is immaterial for the subsequent analysis. In the case of a uniaxial medium, the corresponding polarization vectors and the wave numbers of the ordinary and extraordinary waves have the form

$$e_{(1)}(\mathbf{Q}) = \frac{\mathbf{Q} \times \mathbf{n}^0}{Q}, \quad k_{(1)}(\mathbf{Q}) = k_0 \sqrt{\varepsilon_\perp},$$
$$e_{(2)}(\mathbf{Q}) = \frac{\mathbf{n}^0(\mathbf{Q}\hat{\varepsilon}^0\mathbf{Q}) - \mathbf{Q}(\mathbf{Q}\hat{\varepsilon}^0\mathbf{n}^0)}{\sqrt{(\mathbf{Q}\hat{\varepsilon}^0\mathbf{Q})^2 - 2(\mathbf{Q}\hat{\varepsilon}^0\mathbf{Q})(\mathbf{Q}\hat{\varepsilon}^0\mathbf{n}^0)(\mathbf{Q}\cdot\mathbf{n}^0) + Q^2(\mathbf{Q}\hat{\varepsilon}^0\mathbf{n}^0)^2}}, \quad (4.20)$$
$$k_{(2)}(\mathbf{Q}) = k_0 Q \sqrt{\frac{\varepsilon_\perp \varepsilon_\parallel}{\varepsilon_\perp Q^2 + \varepsilon_a (\mathbf{Q}\cdot\mathbf{n}^0)^2}}.$$

We write wave vector $\mathbf{Q}$ in the form $\mathbf{Q} = (\mathbf{q}, q_z)$ and carry out the inverse Fourier transformation in variable q_z in Eq. (4.19):

$$T^0_{\alpha\beta}(\mathbf{q}; z) = \int_{-\infty}^{\infty} \frac{dq_z}{2\pi} T^0_{\alpha\beta}(\mathbf{q}; q_z) e^{iq_z z}. \quad (4.21)$$

The main contribution to the asymptotic form of the integral for $z \gg \lambda$ is determined by the residues at the first-order poles on two dispersion surfaces,

$$q^2 + q_z^2 - k^2_{(j)}(\mathbf{q}; q_z) = 0, \quad (4.22)$$

$j = 1, 2$. Denoting the two solutions to each dispersion equation in q_z as $q_z = \pm q^{(j)}_z(\mathbf{q})$, we obtain from relation (4.19)

$$T^0_{\alpha\beta}(\mathbf{q}; z) = \frac{i}{k_0^2} \sum_{j=1,2} k^2_{(j)} \left(2q_z - \frac{\partial k^2_{(j)}}{\partial q_z} \right)^{-1} \times \frac{e_{(j)\alpha} e_{(j)\beta}}{\mathbf{e}_{(j)}\hat{\varepsilon}^0\mathbf{e}_{(j)}} e^{iq^{(j)}_z|z|}, \quad (4.23)$$

where the values of $\mathbf{e}_{(j)}$ and $k_{(j)}$ are calculated on the wave vector $\mathbf{Q}^{(j)}(\mathbf{q}) \equiv (\mathbf{q}, q^{(j)}_z(\mathbf{q}))$.

Substituting expressions (4.20) for $k^2_{(j)}(\mathbf{Q})$ into Eq. (4.22) and solving the obtained equations with $j = 1, 2$ for q_z, we find that, in both cases, $q^{(j)}_z = k^{(j)}_z(\mathbf{q})$, where $k^{(j)}_z(\mathbf{q})$ is defined in relations (4.3) (for $j = 2$, we must take into account the fact that $\mathbf{Q}\cdot\mathbf{n}^0 = \mathbf{q}\cdot\mathbf{n}^0$). Thus, $\mathbf{Q}^{(j)}(\mathbf{q}) = (\mathbf{q}, k^{(j)}_z(\mathbf{q})) = \mathbf{k}^{(j)}(\mathbf{q})$. It can easily be verified using relations (4.20) that the values of $\mathbf{e}_{(j)}(\mathbf{k}^{(j)}(\mathbf{q}))$

and $k_{(j)}(\mathbf{k}^{(j)}(\mathbf{q}))$ in this case coincide with $\mathbf{e}^{(j)}(\mathbf{q})$ and $k^{(j)}(\mathbf{q})$ in Eqs. (4.3) and (4.5) for $p_0 = 0$. The identity

$$\frac{k_z^{(j)}(\mathbf{q})B^{(j)2}(\mathbf{q})}{k_0} = \frac{k^{(j)2}(\mathbf{q})}{k_0^2\mathbf{e}^{(j)}(\mathbf{q})\hat{\varepsilon}^0\mathbf{e}^{(j)}(\mathbf{q})} \times\left(q_z - \frac{1}{2}\frac{\partial k_{(j)}^2(\mathbf{q}, q_z)}{\partial q_z}\bigg|_{q_z = k_z^{(j)}(\mathbf{q})}\right)^{-1} \tag{4.24}$$

can also be verified easily for both cases $j = 1, 2$. As a result, relation (4.23) coincides with Eqs. (4.16) and (4.17).

5. CORRELATION FUNCTION FOR DIRECTOR FLUCTUATIONS IN A CLC WITH A LARGE PITCH

Problem (2.28) for operator (2.27) can be reduced to solving a nonhomogeneous system of two second-order differential equations with periodic coefficients. The solution to the corresponding homogeneous system (which has the form of two matrix functions of z and z_1, i.e., the exponential function with the exponent linear in z and z_1 and a periodic function with period $2d = 2\pi/p_0$) satisfies the Floquet theorem [43]. In order to find the exponents and the Fourier harmonics of a periodic function, a standard algorithm for Hill-type equations is used [43], leading to an infinite system of trinomial recurrence relations. This approach was used in [20, 21] for a detailed analysis of the problem of director fluctuations in CLCs. In these publications, the lower harmonics of the 3D Fourier spectrum of correlation function $\hat{g}$ were determined under the assumption that $q/p_0 \ll 1$, when the contributions from the Fourier harmonics rapidly decrease with increasing numbers of the harmonics. From the physical point of view, the limit $q/p_0 \longrightarrow 0$ corresponds to a smecticlike CLC, which is confirmed by the result obtained in [20]: the main contribution to the correlation function is

$$\hat{g}(\mathbf{q}, k_z) \sim q^2/(k_z^2 + c_0 p_0^{-2} q^4),$$

where c_0 is a dimensionless constant. Director fluctuations in smectics A are characterized precisely by this type of a correlator [34].

The situation for CLCs with a large pitch corresponds to the opposite limiting case of a nematic-like CLC, for which $p_0/q \longrightarrow 0$. From the point of view of the conventional Floquet approach, the problem is that the main contribution to the correlation function comes from a broad spectrum of Fourier harmonics with large numbers for periodic factors. In such a situation, the methods based on the Floquet theory are ineffective and it is more expedient to apply the WKB method in the large parameter $q/p_0 \gg 1$.

5.1. General Algorithm for Calculating the Correlation Function in Media with One-Dimensional Inhomogeneities

Equation (2.28), together with the condition of decreasing for $\hat{G}(z, z_1)$ as $z \longrightarrow \pm\infty$, has the form of an equation for the Green function. It should be noted that Eq. (2.28) becomes homogeneous for $z \neq z_1$. We will first solve the homogeneous equations for the cases when $z > z_1$ and $z < z_1$ separately. Then, using the continuity conditions for function $\hat{G}$ and the jump of its derivative for $z = z_1$, we will construct the Green function.

The system of homogeneous equations has the form

$$\left[-\begin{pmatrix} K_{22} & 0 \\ 0 & K_{11} \end{pmatrix}\frac{d^2}{d\xi^2} + i\tilde{\Omega}(K_{11} - K_{22})\sin\phi\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\frac{d}{d\xi}\right.$$
$$\left. + \begin{pmatrix} \tilde{\Omega}^2(K_{11}\sin^2\phi + K_{33}\cos^2\phi) & -i\tilde{\Omega}\cos\phi(K_{22} + K_{33}) \\ i\tilde{\Omega}\cos\phi(K_{11} + K_{33}) & \tilde{\Omega}_2(K_{22}\sin^2\phi + K_{33}\cos^2\phi) + K_{33} \end{pmatrix}\right]\mathbf{u}(\xi) = 0, \tag{5.1}$$

where $\tilde{\Omega} = q/p_0$. Equation (5.1) forms a system of two second-order differential equations. This system has four linearly independent solutions. Using four linearly independent column vectors of the solutions to Eq. (5.1), we construct two matrices $\hat{u}_1(\xi)$ and $\hat{u}_2(\xi)$, such that $\hat{u}_1(\xi) \longrightarrow \hat{0}$ for $\xi \longrightarrow +\infty$ and $\hat{u}_2(\xi) \longrightarrow \hat{0}$ for $\xi \longrightarrow -\infty$. Such a choice ensures the required behavior of $\hat{G}(\xi, \xi_1)$ at infinity.

We will seek the Green function in the form

$$\hat{G}(\xi, \xi_1) = \begin{cases} \hat{u}_1(\xi)\hat{v}_1(\xi_1) & \text{for} \quad \xi \geq \xi_1, \\ \hat{u}_2(\xi)\hat{v}_2(\xi_1) & \text{for} \quad \xi < \xi_1, \end{cases} \tag{5.2}$$

where $\hat{v}_1$ and $\hat{v}_2$ are 2×2 matrices. To find eight elements of these matrices, we will use the conditions imposed on the Green function in the vicinity of point $\xi = \xi_1$. These conditions include the continuity of the function itself and a jump of its first derivative, such that Eq. (2.28) is satisfied,

$$\hat{G}(\xi_1 + 0, \xi_1) = \hat{G}(\xi_1 - 0, \xi_1),$$
$$\hat{K}\left(\frac{\partial \hat{G}}{\partial \xi}\bigg|_{\xi = \xi_1 - 0} - \frac{\partial \hat{G}}{\partial \xi}\bigg|_{\xi = \xi_1 + 0}\right) = \frac{k_B T}{p_0}\hat{I}, \tag{5.3}$$

where

$$\hat{K} = \begin{pmatrix} K_{22} & 0 \\ 0 & K_{11} \end{pmatrix}.$$

Substituting relations (5.2) into (5.3), we obtain the system of eight equations for the elements of matrices $\hat{v}_{1,2}$:

$$\begin{cases} \hat{u}_1(\xi_1)\hat{v}_1(\xi_1) = \hat{u}_2(\xi_1)\hat{v}_2(\xi_1), \\ \hat{u}_1'(\xi_1)\hat{v}_1(\xi_1) - \hat{u}_2'(\xi_1)\hat{v}_2(\xi_1) = -k_B T p_0^{-1}\hat{K}^{-1}. \end{cases} \tag{5.4}$$

This formula can easily be derived by integrating Eq. (2.28) with respect to z in an infinitely small interval containing point z_1. Solving system (5.4), we obtain

$$\hat{v}_j = \frac{k_B T}{p_0}\hat{u}_j^{-1}\hat{w}\hat{K}^{-1}, \tag{5.5}$$

where $\hat{w}(\xi) = (\hat{u}_2'(\xi)\,\hat{u}_2^{-1}(\xi) - \hat{u}_1'(\xi)\,\hat{u}_1^{-1}(\xi))^{-1}$, $j = 1, 2$.

Substituting relation (5.5) into (5.2), we finally obtain the following expression for the Green function:

$$\hat{G}(\xi, \xi_1) = k_B T p_0^{-1} \times \begin{cases} \hat{u}_1(\xi)\hat{u}_1^{-1}(\xi_1)\hat{w}(\xi_1)\hat{K}^{-1} & \text{for} \quad \xi \geq \xi_1, \\ \hat{u}_2(\xi)\hat{u}_2^{-1}(\xi_1)\hat{w}(\xi_1)\hat{K}^{-1} & \text{for} \quad \xi < \xi_1. \end{cases} \tag{5.6}$$

The choice of matrices $\hat{u}_1$ and $\hat{u}_2$ is ambiguous (for example, their columns can be multiplied by arbitrary constants); however, this ambiguity disappears in formula (5.6) due to factors $\hat{w}$.

Pay attention to the fact that $\hat{G}(\xi, \xi_1) \longrightarrow 0$ for $|\xi - \xi_1| \longrightarrow \infty$ in view of our choice of matrices $\hat{u}_{1,2}$; i.e., the boundary conditions for $\hat{G}$ are satisfied.

This scheme for constructing the Green function with the help of the solutions to the homogeneous equation is universal and does not necessitate the use of large parameter $\tilde{\Omega}$.

5.2. Application of the WKB Method for Determining the Correlation Function

The presence of the large parameter in our problem will be used here for explicitly constructing solutions **u** to homogeneous equation (5.1). We will find these solutions using the vector WKB method. Introducing the vector

$$\mathbf{v} = \frac{1}{i\tilde{\Omega}}\frac{d\mathbf{u}}{d\xi},$$

we reduce Eq. (5.1) to the system of four first-order equations:

$$\frac{d\Psi}{d\xi} = (i\tilde{\Omega}\hat{B} + \hat{C})\Psi. \tag{5.7}$$

Here, we used the notation

$$\Psi = \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}, \quad \hat{B} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ b_1 & 0 & 0 & b_3 \\ 0 & b_2 & b_4 & 0 \end{pmatrix},$$
$$\hat{C} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & c_1 & 0 & 0 \\ c_2 & c_3 & 0 & 0 \end{pmatrix}, \tag{5.8}$$

where

$$b_1 = -\frac{K_{11}\sin^2\phi + K_{33}\cos^2\phi}{K_{22}},$$
$$b_2 = -\frac{K_{22}\sin^2\phi + K_{33}\cos^2\phi}{K_{11}},$$
$$b_3 = \left(\frac{K_{11}}{K_{22}} - 1\right)\sin\phi,$$
$$b_4 = \left(1 - \frac{K_{22}}{K_{11}}\right)\sin\phi,$$
$$c_1 = -\left(1 + \frac{K_{33}}{K_{22}}\right)\cos\phi,$$
$$c_2 = \left(1 + \frac{K_{33}}{K_{11}}\right)\cos\phi,$$
$$c_3 = -i\frac{1}{\tilde{\Omega}}\frac{K_{33}}{K_{11}}.$$

The solution to system (5.7) can be written in the form

$$\mathbf{\Psi}(\xi) = \hat{M}(\xi, \xi_0)\mathbf{\Psi}(\xi_0), \tag{5.9}$$

where $\hat{M}(\xi, \xi_0)$ is the evolution matrix for this system, $\hat{M}(\xi, \xi_0) = \hat{I}$. We will find the evolution matrix in the main order in parameter $\tilde{\Omega}$.

The procedure for constructing the evolution matrix is described in the Appendix. In accordance with Eq. (A.17), we have

$$\hat{M}(\xi, \xi_0) \approx \hat{U}(\xi)\widehat{\mathrm{diag}}\left\{\exp\left[-\int_{\xi_0}^{\xi}\left(\tilde{\Omega}\mu_l(x) + \left(\hat{U}^{-1}(x)\frac{\partial \hat{U}(x)}{\partial x} - \hat{U}^{-1}(x)\hat{C}(x)\hat{U}(x)\right)_{ll}\right)dx\right]\right\}\hat{U}^{-1}(\xi_0), \tag{5.10}$$

where $i\mu_l$ are the eigenvalues of matrix $\hat{B}$ ($l = 1$–4) and the columns of matrix $\hat{U}$ are formed by the eigenvectors of matrix $\hat{B}$. Here, we have used the notation $\widehat{\mathrm{diag}}(x_l)$ for the diagonal matrix with elements $x_1, x_2, \ldots$ on the diagonal. It should be noted that expression (5.10) is inapplicable when the eigenvalues of matrix $\hat{B}$ converge. It can be seen from expression (A.15) that the value of V_{lm} becomes large in this case and the condition $|V_{lm}| \ll \tilde{\Omega}$ of applicability of the WKB method (A.18) may be violated.

Formula (5.10) gives the solution to Eq. (5.7) for any initial conditions $\mathbf{\Psi}(\xi_0)$. This solution can be treated as a linear combination of four linearly independent vectors forming the columns of evolution matrix $\hat{M}(\xi, \xi_0)$ with four coefficients (elements of vector $\mathbf{\Psi}(\xi_0)$).

In order to construct the Green function, we require solutions **u** in the form of the first two components of solution $\mathbf{\Psi}(\xi)$. As the columns of matrices $\hat{u}_{1,2}(\xi)$, we can use the vectors whose components are the first two elements of the columns of matrix $\hat{M}(\xi, \xi_0)\,\hat{U}(\xi_0)$ or linear combinations of these columns.

Let us first determine the evolution matrix $\hat{M}(\xi, \xi_0)$ (5.10). For this purpose, we must know the eigenvalues and eigenvectors of matrix $\hat{B}$. The eigenvalues can be determined from the relation

$$\det(\hat{B} - i\mu\hat{I}) = 0, \tag{5.11}$$

which has the form of a biquadratic equation. Solving this equation, we obtain

$$\mu_l(\xi) = \sqrt{\sin^2\phi(\xi) + \frac{K_{33}}{K_{ll}}\cos^2\phi(\xi)}, \quad l = 1, 2, \quad \mu_3 = -\mu_1, \quad \mu_4 = -\mu_2. \tag{5.12}$$

Eigenvectors $\mathbf{\psi}_l$ satisfy the relation $\hat{B}\mathbf{\psi}_l = i\mu_l\mathbf{\psi}_l$. We find these vectors and use them to construct matrix $\hat{U}(\xi) = (\mathbf{\psi}_l(\xi), \mathbf{\psi}_2(\xi), \mathbf{\psi}_3(\xi), \mathbf{\psi}_4(\xi))$:

$$\hat{U} = \begin{pmatrix} -i\mu_1^{-1}\sin\phi & 1 & i\mu_1^{-1}\sin\phi & -1 \\ -1 & -i\mu_2^{-1}\sin\phi & -1 & -i\mu_2^{-1}\sin\phi \\ \sin\phi & i\mu_2 & \sin\phi & i\mu_2 \\ -i\mu_1 & \sin\phi & i\mu_1 & -\sin\phi \end{pmatrix}. \tag{5.13}$$

The arbitrariness in the choice of vectors $\mathbf{\psi}_l(\xi)$ and, hence, matrix $\hat{U}(\xi)$ is associated with the normalization factors, which may depend on ξ. It can easily be seen, however, that the right-hand side of expression (5.10) is independent of the choice of normalizations of vectors $\mathbf{\psi}_l(\xi)$. Then we determine matrix $\hat{U}^{-1}(\xi)$:

$$\hat{U}^{-1} = \frac{-1}{2K_{33}\cos^2\phi} \times \begin{pmatrix} iK_{11}\mu_1\sin\phi & K_{22}\mu_2^2 & K_{22}\sin\phi & -iK_{11}\mu_1 \\ -K_{11}\mu_1^2 & iK_{22}\mu_2\sin\phi & iK_{22}\mu_2 & K_{11}\sin\phi \\ -iK_{11}\mu_1\sin\phi & K_{22}\mu_2^2 & K_{22}\sin\phi & iK_{11}\mu_1 \\ K_{11}\mu_1^2 & iK_{22}\mu_2\sin\phi & iK_{22}\mu_{22} & -K_{11}\sin\phi \end{pmatrix}. \tag{5.14}$$

Disregarding the term on the order of $1/\tilde{\Omega}$ (i.e., element c_3) in matrix $\hat{C}$ (5.8), since it makes zero contribution

to the evolution matrix in approximation (5.10) used here, we obtain

$$(\hat{U}^{-1}\hat{C}\hat{U})_{ll} = \frac{(K_{11}-K_{22})\sin\phi}{2K_{33}\cos\phi}(-1)^{l+1}, \quad l = 1, \ldots, 4. \tag{5.15}$$

The diagonal elements of matrix $\hat{U}^{-1}\partial\hat{U}/\partial\xi$ can be represented in a form convenient for integration:

$$\left(\hat{U}^{-1}\frac{\partial\hat{U}}{\partial\xi}\right)_{ll} = \frac{1}{2}\frac{(\cos\phi/\mu_l)'}{\cos\phi/\mu_l} + \frac{1}{2}\frac{(\cos\phi)'}{\cos\phi} - \frac{K_{11}-K_{22}}{2K_{33}}\frac{(\cos\phi)'}{\cos\phi}(-1)^{l+1}, \tag{5.16}$$

$l = 1, \ldots, 4$. Subtracting this expression from (5.15) and integrating, we obtain

$$\int_{\xi_0}^{\xi}\left(\hat{U}^{-1}\hat{C}\hat{U} - \hat{U}^{-1}\frac{\partial\hat{U}}{\partial\xi'}\right)_{ll}d\xi' = -\frac{1}{2}\ln\frac{\mu_l(\xi_0)|\cos\phi|}{\mu_l(\xi)|\cos(\xi_0+\phi_0)|} - \frac{1}{2}\ln\frac{|\cos\phi|}{|\cos(\xi_0+\phi_0)|}. \tag{5.17}$$

Substituting this relation into Eq. (5.10), we obtain

$$\hat{M}(\xi, \xi_0) = \hat{U}(\xi)\frac{|\cos(\xi_0+\phi_0)|}{|\cos\phi|} \times \widehat{\mathrm{diag}}\left\{\sqrt{\frac{\mu_l(\xi)}{\mu_l(\xi_0)}}\exp\left(-\tilde{\Omega}\int_{\xi_0}^{\xi}\mu_l(\xi')d\xi'\right)\right\}\hat{U}^{-1}(\xi_0). \tag{5.18}$$

In order to find the correlation function, we must construct matrices $\hat{u}_1(\xi)$ and $\hat{u}_2(\xi)$. Choosing the first two components of the columns of matrix $\hat{M}(\xi, \xi_0)\hat{U}(\xi_0)$, which exhibit the required behavior at infinity, we obtain

$$\hat{u}_1(\xi) = \begin{pmatrix} -i\mu_1^{-1}\sin\phi & 1 \\ -1 & -i\mu_2^{-1}\sin\phi \end{pmatrix}\exp(\hat{\Phi}_-),$$
$$\hat{u}_2(\xi) = \begin{pmatrix} i\mu_1^{-1}\sin\phi & -1 \\ -1 & -i\mu_2^{-1}\sin\phi \end{pmatrix}\exp(\hat{\Phi}_+), \tag{5.19}$$

where

$$\exp(\hat{\Phi}_\pm) = \frac{|\cos(\xi_0+\phi_0)|}{|\cos\phi|} \times \widehat{\mathrm{diag}}\left\{\sqrt{\frac{\mu_l(\xi)}{\mu_l(\xi_0)}}\exp\left(\pm\tilde{\Omega}\int_{\xi_0}^{\xi}\mu_l(\xi')d\xi'\right)\right\}, \quad l = 1, 2. \tag{5.20}$$

Thus, the conditions $\hat{u}_1(\xi) \longrightarrow \hat{0}$ for $\xi \longrightarrow +\infty$ and $\hat{u}_2(\xi) \longrightarrow \hat{0}$ for $\xi \longrightarrow -\infty$ are satisfied for matrices (5.19)

Substituting matrices $\hat{u}_{1,2}(\xi)$ from relations (5.19) into expression (5.6) for the correlation function and neglecting the terms on the order of $1/\tilde{\Omega}$ in the preexponential factors, we obtain

$$\hat{G}(\xi, \xi_1) = \frac{k_B T}{2qK_{33}\cos\phi(\xi_1)\cos\phi(\xi)} \times \begin{pmatrix} \mathrm{sgn}(\xi-\xi_1)\sin\phi(\xi) & i\,\mathrm{sgn}(\xi-\xi_1)\mu_2(\xi) \\ -i\mu_1(\xi) & \sin\phi(\xi) \end{pmatrix} \times \begin{pmatrix} \dfrac{\exp\left(-\tilde{\Omega}\left|\displaystyle\int_{\xi}^{\xi_1}\mu_1 d\xi'\right|\right)}{\sqrt{\mu_1(\xi)\mu_1(\xi_1)}} & 0 \\ 0 & \dfrac{\exp\left(-\tilde{\Omega}\left|\displaystyle\int_{\xi}^{\xi_1}\mu_2 d\xi'\right|\right)}{\sqrt{\mu_2(\xi)\mu_2(\xi_1)}} \end{pmatrix} \times \begin{pmatrix} \mathrm{sgn}(\xi_1-\xi)\sin\phi(\xi_1) & i\mu_1(\xi_1) \\ i\,\mathrm{sgn}(\xi_1-\xi)\mu_2(\xi_1) & -\sin\phi(\xi_1) \end{pmatrix}. \tag{5.21}$$

It is convenient to split function $\hat{G}$ into two parts associated with indices μ_1 and μ_2. Returning from dimensionless variable ξ to variable z, we finally obtain

$$\hat{G}(\mathbf{q}; z_1, z_2) = \hat{G}_1(\mathbf{q}; z_1, z_2) + \hat{G}_2(\mathbf{q}; z_1, z_2), \tag{5.22}$$

where

$$\hat{G}_j(\mathbf{q}; z_1, z_2) = \frac{k_B T}{2qK_{33}\cos\phi(z_1)\cos\phi(z_2)} \times \exp\left(-q\left|\int_{z_1}^{z_2}\mu_j(z)dz\right|\right)\hat{W}^{(j)}(\mathbf{q}; z_1, z_2), \tag{5.23}$$

$$\hat{W}^{(1)}(\mathbf{q}; z_1, z_2) = \begin{pmatrix} -\dfrac{\sin\phi(z_1)\sin\phi(z_2)}{\sqrt{\mu_1(z_1)\mu_1(z_2)}} & i\,\mathrm{sgn}(z_1 - z_2)\sin\phi(z_1)\dfrac{\sqrt{\mu_1(z_2)}}{\sqrt{\mu_1(z_1)}} \\ i\,\mathrm{sgn}(z_1 - z_2)\sin\phi(z_2)\dfrac{\sqrt{\mu_1(z_1)}}{\sqrt{\mu_1(z_2)}} & \sqrt{\mu_1(z_1)\mu_1(z_2)} \end{pmatrix},$$

$$\hat{W}^{(2)}(\mathbf{q}; z_1, z_2) = \begin{pmatrix} \sqrt{\mu_2(z_1)\mu_2(z_2)} & -i\,\mathrm{sgn}(z_1 - z_2)\sin\phi(z_2)\dfrac{\sqrt{\mu_2(z_1)}}{\sqrt{\mu_2(z_2)}} \\ -i\,\mathrm{sgn}(z_1 - z_2)\sin\phi(z_1)\dfrac{\sqrt{\mu_2(z_2)}}{\sqrt{\mu_2(z_1)}} & -\dfrac{\sin\phi(z_1)\sin\phi(z_2)}{\sqrt{\mu_2(z_1)\mu_2(z_2)}} \end{pmatrix}, \tag{5.24}$$

Here, $\cos\phi = \mathbf{q}\cdot\mathbf{n}_0/q$ and $\sin\phi = \sqrt{g^2 - (\mathbf{q}\cdot\mathbf{n}_0)^2}/q$.

Figure 5 shows the correlation function component G_{11} expressed in relative units. The figure was obtained as a result of numerical calculations based on formulas (5.22)–(5.24). Corrections decrease exponentially with increasing distance $|z - z_1|$. It is also worth noting that correlations increase at $z = z_1$ as the value of $\phi(z)$ approaches $\pi/2$. The latter circumstance is determined by the fact that the values of $\cos\phi(z_{1,2})$ in the denominator of the common multiplier in relation (5.23) approach zero. This indicates that we approach the region in which the WKB method is inapplicable and formula (5.27) is not valid. This question will be discussed in greater detail at the end of this section.

Matrices $\hat{W}^{(j)}(\mathbf{q}; z_1, z_2)$ can be written in the form

$$W^{(j)}_{\alpha\beta}(\mathbf{q}; z_1, z_2) = l^{(j)}_{\alpha}(\mathbf{q}; z_1, z_2) l^{(j)*}_{\beta}(\mathbf{q}; z_2, z_1), \tag{5.25}$$

where

$$\mathbf{l}^{(1)}(\mathbf{q}; z, z') = \left(i\,\mathrm{sgn}(z - z')\frac{\sin\phi(z)}{\sqrt{\mu_1(z)}}, \sqrt{\mu_1(z)}\right),$$
$$\mathbf{l}^{(2)}(\mathbf{q}; z, z') = \left(\sqrt{\mu_2(z)}, -i\,\mathrm{sgn}(z - z')\frac{\sin\phi(z)}{\sqrt{\mu_2(z)}}\right). \tag{5.26}$$

Carrying out summation in k and l in formula (2.20), we obtain the final expression for the correlation function of director fluctuations from formulas (2.19), (2.20), (5.22)–(5.26):

$$g_{\alpha\beta}(\mathbf{q}; z_1, z_2) = \frac{k_B T}{2qK_{33}\cos\phi(z_1)\cos\phi(z_2)} \times \sum_{j=1}^{2} \exp\left(-q\left|\int_{z_1}^{z_2} \mu_j(z)\,dz\right|\right) \times f^{(j)}_{\alpha}(\mathbf{q}; z_1, z_2) f^{(j)*}_{\beta}(\mathbf{q}; z_2, z_1), \tag{5.27}$$

where $\mathbf{f}^{(j)}(\mathbf{q}; z, z') = \sum_{k=1,2} l^{(j)}_k(\mathbf{q}; z, z')\mathbf{h}^{(k)}(z)$. In the coordinate system used by us here, vectors $\mathbf{f}^{(j)}$ have the form

$$\mathbf{f}^{(1)}(\mathbf{q}; z, z') = \frac{1}{\sqrt{\mu_1(z)}}\left(-i\,\mathrm{sgn}(z - z')\sin^2\phi(z), \frac{i}{2}\mathrm{sgn}(z - z')\sin 2\phi(z), \mu_1(z)\right),$$
$$\mathbf{f}^{(2)}(\mathbf{q}; z, z') = \sqrt{\mu_2(z)}\left(-\sin\phi(z), \cos\phi(z), -i\,\mathrm{sgn}(z - z')\frac{\sin\phi(z)}{\mu_2(z)}\right). \tag{5.28}$$

It should be noted, in particular, that

$$\mathbf{f}^{(j)}(\mathbf{q}; z, z')\cdot\mathbf{n}^{(0)}(z) = 0,$$
$$\left|\mathbf{f}^{(j)}(\mathbf{q}; z, z')\right|^2 = \mu_j(z) + \frac{\sin^2\phi(z)}{\mu_j(z)}.$$

Relations (5.27) and (2.14) lead to the permittivity correlation function $\mathcal{G}_{\beta\delta\gamma\nu}(\mathbf{q}; z_1, z_2)$ in the form

$$\mathcal{G}_{\beta\delta\gamma\nu}(\mathbf{q}; z_1, z_2) = \frac{k_B T \varepsilon_a^2}{2qK_{33}\cos\phi(z_1)\cos\phi(z_2)} \times \sum_{j=1}^{2} \exp\left(-q\left|\int_{z_1}^{z_2} \mu_j(z)\,dz\right|\right)\mathcal{M}^{(j)}_{\beta\delta\gamma\nu}(\mathbf{q}; z_1, z_2), \tag{5.29}$$

where

$$\begin{aligned}\mathcal{M}^{(j)}_{\beta\delta\gamma\nu}(z_1, z_2) &= n^{(0)}_\beta(z_1)n^0_\gamma(z_2)\\ &\times f^{(j)}_\delta(z_1, z_2) f^{(j)*}_\nu(z_2, z_1)\\ &+ n^0_\delta(z_1)n^0_\gamma(z_2) f^{(j)}_\beta(z_1, z_2) f^{(j)*}_\nu(z_2, z_1)\\ &+ n^0_\beta(z_1)n^0_\nu(z_2) f^{(j)}_\delta(z_1, z_2) f^{(j)*}_\gamma(z_2, z_1)\\ &+ n^0_\delta(z_1)n^0_\nu(z_2) f^{(j)}_\beta(z_1, z_2) f^{(j)*}_\gamma(z_2, z_1).\end{aligned} \tag{5.30}$$

Let us consider the domain of applicability for formulas (5.27) and (5.29). In addition to the inequality $q \gg p_0$, two constraints associated with inequalities (A.19) and (A.20) may appear. Since $\mu_{1,2} \sim 1$, inequality (A.20) gives $|z_1 - z_2| \ll q/p_0^2$.

The constraint (A.19) is most significant for our analysis. This is due to the fact that, in accordance with relations (5.12), eigenvalues μ_1 and μ_2 coincide when $\cos\phi = 0$. Consequently, formula (5.27) becomes meaningless if point z_* (such that $\cos\phi(z_*) = 0$) gets into the domain between points z_1 and z_2.[3]

Let us consider the constraint imposed by this effect on the domain of applicability of formulas (5.27) and (5.29). For this purpose, we introduce a new variable $\zeta = p_0 z + \phi_0 - \pi/2 = \phi - \pi/2$ and expand μ_l into a series in the vicinity of point $\zeta = 0$

$$\begin{aligned}\mu_l &\approx 1 - \frac{1}{2}C_l\zeta^2, \quad l = 1, 2,\\ \mu_3 &= -\mu_1, \quad \mu_4 = -\mu_2,\end{aligned} \tag{5.31}$$

where $C_l = 1 - K_{33}/K_{ll}$. It should be noted that the corresponding lines $\mu_1(\zeta)$ and $\mu_2(\zeta)$ do not intersect at point $\zeta = 0$, but are just contiguous to each other. We can obtain the following estimate for matrix element V_{12} in the vicinity of point $\zeta = 0$:

$$V_{12} \sim \frac{1}{\zeta^3}(1 + O(\zeta^2)).$$

In this case, the first condition in (A.18) has the form

$$\tilde{\Omega}|\zeta|^3 \gg 1. \tag{5.32}$$

This means that expression (5.27) is applicable only in the cases when $\tilde{\Omega}|\zeta|^3 \gg 1$ and $\tilde{\Omega}|\zeta_1|^3 \gg 1$, and there are no points between ζ and ζ_1 at which the values of $\mu_{1,2}$ coincides (i.e., points at which $\cos\phi = 0$).

[3] In addition, since the equality $\mu_1 = \mu_2$ is satisfied identically for $K_{11} = K_{22}$, the fulfillment of inequality $|K_{11} - K_{22}| \gg 2(K_{11}K_{22}/K_{33})(p_0/q)$ is required; in particular, our formulas (5.27) and (5.29) do not permit the application of the one-constant approximation for the Frank energy (2.1).

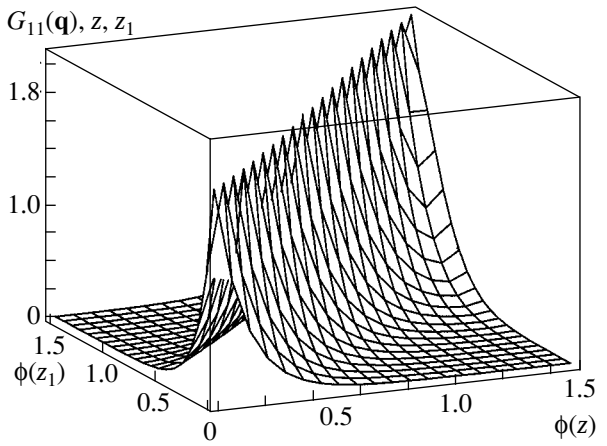

Fig. 5. Component of the CLC correlation function $G_{11}(\mathbf{q}; z, z_1)$ expressed in relative units as a function of coordinates z and z_1 of two points.

Analysis of the behavior of the correlation function in the vicinity of the points for which $\cos\phi = 0$ requires the application of the approaches used for analyzing turning points in the WKB method. This question was considered in detail in our previous publication [44], where it was shown that the correlation function of fluctuations is actually bounded in the vicinity of regions $\cos\phi(z_1) = 0$ and $\cos\phi(z_2) = 0$. It is important, however, that the main contribution to integral (3.30) in the problem of light scattering considered here comes from the domain $|z_1 - z_2| \ll d$ (see Section 6.1 below). In addition, in the case when both points $z_{1,2}$ are simultaneously in the region $\cos\phi(z_{1,2}) = 0$, the singularity in the correlation function is cancelled out. This can easily be verified since, in the limit $z_1 \longrightarrow z_2$ and $\cos\phi(z_1)$, $\cos\phi(z_2) \longrightarrow 0$, the exponents $\mu_1(z)$, $\mu_2(z) \longrightarrow 1$ in formula (5.23) and the condition $\sum_{j=1,2}\hat{W}^{(j)}(\mathbf{q}; z_1, z_2) \longrightarrow 0$ is satisfied. The absence of a singularity in this case can be explained by the fact that a CLC is practically indistinguishable from a nematic in the region of close values of z_1 and z_2, and it was proved above that correlation function (5.21) for an NLC is finite for $\cos\phi_0 = 0$.

Thus, in the scattering problem, we can confine our analysis to expression (5.27) for the correlation function, provided that inequalities $q \gg p_0$ and

$$|z_1 - z_2| \ll q/p_0^2 \tag{5.33}$$

hold.

5.3. Limiting Cases

Let us analyze the behavior of expression (5.27) for $p_0 \longrightarrow 0$. This limiting transition corresponds to a transition to an NLC. If we set $p_0 = 0$, formula (5.22) assumes the form

$$\hat{G}(\mathbf{q}; z-z_1) = \frac{k_B T e^{-q\mu_1|z-z_1|}}{2qK_{33}\cos^2\phi_0} \times \begin{pmatrix} -\mu_1^{-1}\sin^2\phi_0 & i\,\mathrm{sgn}(z-z_1)\sin\phi_0 \\ i\,\mathrm{sgn}(z-z_1)\sin^2\phi_0 & \mu_1 \end{pmatrix}$$

$$+ \frac{k_B T e^{-q\mu_2|z-z_1|}}{2qK_{33}\cos^2\phi_0} \times \begin{pmatrix} \mu_2 & -i\,\mathrm{sgn}(z-z_1)\sin\phi_0 \\ -i\,\mathrm{sgn}(z-z_1)\sin\phi_0 & -\mu_2^{-1}\sin^2\phi_0 \end{pmatrix}. \tag{5.34}$$

Carrying out Fourier transformation (4.21) with respect to variable $z - z_1$, we obtain

$$\hat{G}(\mathbf{q}, k_z) = \frac{k_B T}{(K_{33}q^2\cos^2\phi_0 + K_{11}(q^2\sin^2\phi_0 + k_z^2))(K_{33}q^2\cos^2\phi_0 + K_{22}(q^2\sin^2\phi_0 + k_z^2))}$$
$$\times \begin{pmatrix} (K_{33}\cos^2\phi_0 + K_{22}\sin^2\phi_0)q^2 + K_{11}k_z^2 & qk_z\sin\phi_0(K_{11}-K_{22}) \\ qk_z\sin\phi_0(K_{11}-K_{22}) & (K_{33}\cos^2\phi_0 + K_{11}\sin^2\phi_0)q^2 + K_{22}k_z^2 \end{pmatrix}. \tag{5.35}$$

It is worth noting that, for $\phi_0 \longrightarrow \pi/2$, formula (5.34) acquires an indeterminacy that can easily be expanded and leads to a finite expression for $\phi_0 = \pi/2$. Consequently, correlation function (5.35) has no singularities in the vicinity of this point.

If we pass to a coordinate system in which $\mathbf{n}_0 = \mathbf{e}_z$ and $\mathbf{q} \parallel \mathbf{e}_x$, matrix $\hat{G}$ becomes diagonal and formula (5.35) is transformed to the well-known formula for the correlation function of NLC [34]:

$$G_{\alpha\beta}(\mathbf{k}) = \frac{k_B T}{K_{33}k_z^2 + K_{\alpha\alpha}k_\perp^2}\delta_{\alpha\beta}.$$

It would be interesting to compare our result (5.22) for correlation matrix $\hat{G}(\mathbf{q}; z_1, z_2)$, which is valid in the limit $q \gg p_0$ (nematic-like CLCs), with the well-known result for $\hat{G}(\mathbf{q}; z_1, z_2)$, which was obtained in the opposite limiting case $q \ll p_0$ (smecticlike CLCs) [20, 21]. For simplicity, we consider the one-constant approximation, in which the Frank moduli are equal ($K_{11} = K_{22} = K_{33} = K$). We will be interested in the behavior of $\hat{G}(\mathbf{q}; z_1, z_2)$ as a function of modulus q of the wave vector and of quantity $z_1 - z_2$.

For $q \gg p_0$, the components of correlation matrix $\hat{G}$ exhibit, in accordance with relation (5.22), a characteristic dependence on q and $z_1 - z_2$ of the form

$$G_{\beta\gamma}(\mathbf{q}; z_1, z_2) \sim \frac{1}{q}\exp(-q|z_1 - z_2|). \tag{5.36}$$

On the other hand, the three-dimensional Fourier transforms of the elements of the correlation matrix for $q \ll p_0$ can be written as [21]

$$G_{11}(\mathbf{q}, q_z) = \frac{2p_0^2 k_B T}{K(2p_0^2 q_z^2 + q^2 q_z^2 + q^4)},$$
$$G_{22}(\mathbf{q}, q_z) = \frac{k_B T}{K(p_0^2 + q^2 + q_z^2)}, \tag{5.37}$$
$$G_{12} = G_{21} = 0.$$

These expressions were obtained by averaging over many pitches of a helix. For this reason, modes u_1 and u_2 do not correlate (the nondiagonal elements of the correlation matrix are equal to zero) in contrast to our results, in which nearly local fluctuations are taken into account. It is noteworthy that director fluctuations in a plane perpendicular to the helix axis (G_{11}) in this limiting case are of a type analogous to fluctuations of displacements of layers in smectic-A. At the same time, fluctuations along the helix axis (G_{22}) are of the same type as fluctuations of the director in a nematic, but are limited by the pitch of a cholesteric helix.

Passing to the $(\mathbf{q}, z)$ representation in formula (5.37), we obtain

$$G_{11}(\mathbf{q}; z_1 - z_2) \sim \frac{p_0^2}{q^2\sqrt{2p_0^2 + q^2}}\exp\left(-\frac{q^2|z_1 - z_2|}{\sqrt{2p_0^2 + q^2}}\right),$$
$$G_{22}(\mathbf{q}; z_1 - z_2) \sim \frac{1}{\sqrt{p_0^2 + q^2}} \tag{5.38}$$
$$\times \exp(-\sqrt{p_0^2 + q^2}|z_1 - z_2|).$$

Although these expressions are applicable only for $q \ll p_0$, we formally extrapolate them to region $q \gg p_0$ and compare with result (5.36) for $\hat{G}$. This enables us to estimate the behavior of fluctuation modes in a wide range of q. For $q \gg p_0$, we obtain, instead of relations (5.38),

$$G_{11}(\mathbf{q}; z_1 - z_2) \sim \frac{p_0^2}{q^3}\exp(-q|z_1 - z_2|),$$
$$G_{22}(\mathbf{q}; z_1 - z_2) \sim \frac{1}{q}\exp(-q|z_1 - z_2|). \quad (5.39)$$

The exponential factors in relations (5.39) and (5.36) coincide. However, the coefficients of the exponent for mode u_1, which corresponds to fluctuations in a plane perpendicular to the cholesteric axis, are different in these formulas. On the other hand, the coefficients for mode u_2 corresponding to fluctuations along the helix axis coincide. The latter circumstance suggests that the expression for G_{22} derived by us for region $q \gg p_0$ is applicable in a wider range of q values. The first mode turns out to be more sensitive to the pitch and its behavior differs strongly in the cases when $q \ll p_0$ and $q \gg p_0$.

6. LIGHT SCATTERING IN CLCs

Let us consider a CLC sample in the form of a flat layer of thickness L with large transverse dimensions (Fig. 6). A plane wave with wave vector $\mathbf{k}^{(i)}$ is incident on the sample. We analyze the scattered wave with wave vector $\mathbf{k}^{(s)}$ in the far-field region of the sample.

For simplicity, we confine our analysis to the case when the polarization of incident light outside the sample is chosen so that incident wave $E_{\text{in}}^{(i)}(\mathbf{r})$ of only one of two admissible types (4.2) is formed in the sample; otherwise, we must carry out summation with respect to (i) inside the medium. Analogously, we choose the polarization of scattered line outside the sample in such a way that it corresponds to scattered wave $E_{\text{in}}^{(s)}(\mathbf{r})$ of only one of admissible types (4.2) inside the sample. This allows us to avoid summation with respect to (s) inside the medium in the subsequent analysis. Thus, indices (i) and (s) can be identified here with numbers 1 or 2 depending on the type of the incident and scattered waves.

In this geometry, in accordance with relations (3.24) and (3.30), the intensity of singly scattered light in

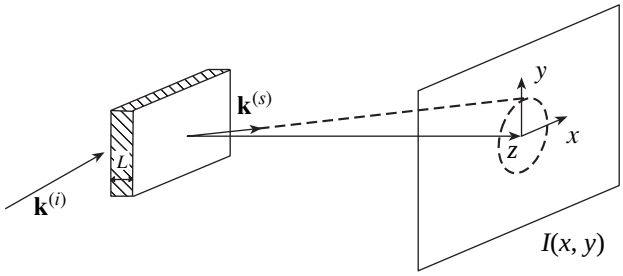

Fig. 6. Geometry of light scattering in CLC.

CLCs is defined by the expression

$$I^{(s)} = \frac{\sqrt{\varepsilon_0}c^2 k_0^6 \varepsilon_0}{8\pi}\frac{S_\perp}{4\pi^2 r^2}\left(\frac{z}{r}\right)^2 e_\alpha^{(s)} e_\gamma^{(s)}$$
$$\times M_{\alpha\beta}^{\text{in}\to\text{out}}(\mathbf{k}_\perp^{(s)}, L) M_{\gamma\delta}^{\text{in}\to\text{out}}(\mathbf{k}_\perp^{(s)}, L)\int_0^L dz_1 \quad (6.1)$$
$$\times \int_0^L dz_2 T_{\beta\rho}^0(\mathbf{k}_\perp^{(s)}; L, z_1) T_{\delta\varphi}^{0*}(\mathbf{k}_\perp^{(s)}; L, z_2)$$
$$\times \mathcal{G}_{\rho\nu\varphi\mu}(\mathbf{k}_\perp^{(s)} - \mathbf{k}_\perp^{(i)}; z_1, z_2)\mathcal{E}_\nu^{(i)}(\mathbf{k}_\perp^{(i)}, z_1)\mathcal{E}_\mu^{(i)*}(\mathbf{k}_\perp^{(i)}, z_2).$$

This expression for the scattering intensity contains conjugate pairs of incident fields and the Green functions. Using relations (4.2) and (4.17), we obtain

$$\mathcal{E}_\nu^{(i)}(\mathbf{k}_\perp^{(i)}, z_1)\mathcal{E}_\mu^{(i)*}(\mathbf{k}_\perp^{(i)}, z_2) = E_0^{(i)2} A^{(i)}(\mathbf{k}_\perp^{(i)}; z_1, 0)$$
$$\times A^{(i)}(\mathbf{k}_\perp^{(i)}; z_2, 0)\exp\left[-i\int_{z_1}^{z_2} k_z^{(i)}(\mathbf{k}_\perp^{(i)}, z')dz'\right] \quad (6.2)$$
$$\times e_\nu^{(i)}(\mathbf{k}_\perp^{(i)}; z_1) e_\mu^{(i)}(\mathbf{k}_\perp^{(i)}; z_2),$$

$$T_{\beta\rho}^{(s)}(\mathbf{k}_\perp^{(s)}; L, z_1) T_{\delta\varphi}^{(s)*}(\mathbf{k}_\perp^{(s)}; L, z_2) = \frac{1}{4k_0^2}$$
$$\times \exp\left[i\int_{z_1}^{z_2} k_z^{(s)}(\mathbf{k}_\perp^{(s)}, z')dz'\right] B^{(s)2}(\mathbf{k}_\perp^{(s)}, L) B^{(s)}(\mathbf{k}_\perp^{(s)}, z_1) \quad (6.3)$$
$$\times B^{(s)}(\mathbf{k}_\perp^{(s)}, z_2) e_\beta^{(s)}(\mathbf{k}_\perp^{(s)}, L) e_\delta^{(s)}(\mathbf{k}_\perp^{(s)}, L)$$
$$\times e_\rho^{(s)}(\mathbf{k}_\perp^{(s)}, z_1) e_\varphi^{(s)}(\mathbf{k}_\perp^{(s)}, z_2).$$

Substituting expressions (6.3) for the Green functions, (5.29) for the correlation function, and (6.2) for the incident field into formula (6.1), we find that the

scattering intensity assumes the form of the sum of two double integrals corresponding to two fluctuation modes of the director,

$$I^{(s)} = \sum_{j=1}^{2} I_j = J_0 \sum_{l=1}^{2} \int_0^L dz_1 \int_0^L dz_2$$

$$\times \exp\left[i \int_{z_1}^{z_2} (k_z^{(s)}(\mathbf{k}_\perp^{(s)}, z') - k_z^{(i)}(\mathbf{k}_\perp^{(i)}, z')) dz' \right]$$

$$\times \exp\left(-q \left| \int_{z_1}^{z_2} \mu_j(z) dz \right| \right) A^{(i)}(\mathbf{k}_\perp^{(i)}; z_1, 0) \tag{6.4}$$

$$\times A^{(i)}(\mathbf{k}_\perp^{(i)}; z_2, 0) B^{(s)}(\mathbf{k}_\perp^{(s)}, z_1) B^{(s)}(\mathbf{k}_\perp^{(s)}, z_2)$$

$$\times e_\rho^{(s)}(\mathbf{k}_\perp^{(s)}, z_1) e_\varphi^{(s)}(\mathbf{k}_\perp^{(s)}, z_2)$$

$$\times \mathcal{M}_{\rho\nu\varphi\mu}^{(j)}(\mathbf{q}; z_1, z_2) e_\nu^{(i)}(\mathbf{k}_\perp^{(i)}, z_1) e_\mu^{(i)}(\mathbf{k}_\perp^{(i)}, z_2),$$

where $\mathbf{q} = \mathbf{k}_\perp^{(s)} - \mathbf{k}_\perp^{(i)}$, and

$$J_0 = E_0^2 \frac{\sqrt{\varepsilon_0} c^2}{8\pi} \frac{k_0^4 \varepsilon_0}{16\pi^2} \frac{S_\perp}{r^2} \left(\frac{z}{r} \right)^2 B^{(s)2}(\mathbf{k}_\perp^{(s)}, L)$$

$$\times e_\alpha^{(s)} e_\gamma^{(s)} M_{\alpha\beta}^{\text{in}\to\text{out}}(\mathbf{k}_\perp^{(s)}, L)$$

$$\times M_{\gamma\delta}^{\text{in}\to\text{out}}(\mathbf{k}_\perp^{(s)}, L) e_\beta^{(s)}(\mathbf{k}_\perp^{(s)}, L) e_\delta^{(s)}(\mathbf{k}_\perp^{(s)}, L).$$

6.1. Large Parameters in the Expression for Intensity

The presence of large parameters $\Omega = k_0/p_0$ and $\tilde{\Omega} = q/p_0$ in our system makes it possible to considerably simplify general expression (6.4) for the scattering intensity. For this purpose, it is convenient to return to the dimensionless variable $\xi = p_0 z$. Then the expression for I_j assumes the form

$$I_j = \int_0^{Lp_0} \int_0^{Lp_0} d\xi_1 d\xi_2 F_j(\xi_1, \xi_2) \tag{6.5}$$

$$\times \exp(i\Phi_{\text{opt}}(\xi_1, \xi_2) + \Phi_{\text{cor}}^{(j)}(\xi_1, \xi_2)),$$

where Φ_{opt} is the phase associated with the incident field and the Green function,

$$\Phi_{\text{opt}}(\xi_1, \xi_2) = \frac{1}{p_0}$$

$$\times \left[-\int_{\xi_1}^{\xi_2} k_z^{(i)}(\mathbf{k}_\perp^{(i)}, \xi') d\xi' + \int_{\xi_1}^{\xi_2} k_z^{(s)}(\mathbf{k}_\perp^{(s)}, \xi') d\xi' \right] \tag{6.6}$$

$$= \frac{1}{p_0} \int_{\xi_1}^{\xi_2} \Delta k_z(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi') d\xi'.$$

Here, $\Delta k_z(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi) = k_z^{(s)}(\mathbf{k}_\perp^{(s)}, \xi) - k_z^{(i)}(\mathbf{k}_\perp^{(i)}, \xi)$, and $k_z^{(s)}$ and $k_z^{(i)}$ are equal to $k_z^{(1)}$ or $k_z^{(2)}$ depending on the types of the incident and scattered waves. "Phase" $\Phi_{\text{cor}}^{(j)}(\xi_1, \xi_2)$ of the correlation function has the form

$$\Phi_{\text{cor}}^{(j)}(\xi_1, \xi_2) = -\frac{q}{p_0} \left| \int_{\xi_1}^{\xi_2} \mu_j(\mathbf{q}, \xi') d\xi' \right|. \tag{6.7}$$

In accordance with formulas (5.27) and (5.28), function $F_j(\xi_1, \xi_2)$ associated with the preexponential factors in the expressions for the incident field, Green function, and correlation function can be written in the form

$$F_j(\xi_1, \xi_2) = F_{1j}(\xi_1, \xi_2) + i \operatorname{sgn}(\xi_2 - \xi_1) F_{2j}(\xi_1, \xi_2),$$

where $F_{1j}(\xi_1, \xi_2)$ and $F_{2j}(\xi_1, \xi_2)$ are smooth functions.

Integral (6.5) has the form of a dependence on large parameters $k_0/p_0 \sim q/p_0$, which is typical of problems that can be solved by the steepest descent method. In our case, the problem is complicated by the singularity in the phase function on the line $\xi_1 = \xi_2$, which is associated with the sign of the modulus in the exponent in Eq. (6.7) as well as the singularity of the type $\operatorname{sgn}(\xi_1 - \xi_2)$ of function $F_j(\xi_1, \xi_2)$ on the same line.

Real-valued exponent (6.7) attains is maximum on line $\xi_1 = \xi_2$. On the same line, this factor has a singularity associated with the presence of the sign in the modulus. As a result of the latter circumstance, oscillations associated with phase Φ_{opt} are not eliminated completely. Consequently, line $\xi_1 = \xi_2$ makes the main contribution to the asymptotic form of the entire integral (6.5).[4]

Passing to new variables $\xi_+ = (\xi_1 + \xi_2)/2$ and $\xi_- = (\xi_2 - \xi_1)/2$, we expand the phase functions in the vicin-

[4] The contributions from the stationary points of oscillating factor with phase (6.6), which lie outside line $\xi_1 = \xi_2$, are multiplied by exponentially small factors with phase (6.7) and can be discarded.

ity of line $\xi_- = 0$ to within first-order terms. This gives

$$\Phi_j(\xi_+, \xi_-) = i\Phi_{\text{opt}}(\xi_1, \xi_2) + \Phi_{\text{cor}}^{(j)}(\xi_1, \xi_2) \approx \frac{2i}{p_0}\Delta k_z(\xi_+)\xi_- - \frac{2q}{p_0}\mu_j(\xi_+)|\xi_-|. \quad (6.8)$$

Thus, the integrals determining the intensity of scattered light have a structure of the form

$$I_j = 2\int_0^{Lp_0} d\xi_+ \int_{-\infty}^{+\infty} d\xi_- \times [F_{1j}(\xi_+, \xi_-) + i\,\text{sgn}(\xi_-)F_{2j}(\xi_+, \xi_-)]e^{\Phi_j(\xi_+, \xi_-)}. \quad (6.9)$$

We have extended the integration with respect to variable ξ_- from $-\infty$ to $+\infty$ since the contribution in ξ_- comes only from a narrow neighborhood of $\xi_- = 0$.

Splitting the integration domain into two regions $(-\infty, 0)$ and $(0, +\infty)$ and replacing the preexponential smooth functions $F_{1j, 2j}$ by their values for $\xi_- = 0$, we obtain the following expression for the internal integral in formula (6.9):

$$\int_{-\infty}^{+\infty} F_j(\xi_+, \xi_-)e^{\Phi_j(\xi_+, \xi_-)}d\xi_- = \int_{-\infty}^{0} d\xi_- \exp\left(\frac{2i}{p_0}\Delta k_z(\xi_+)\xi_- + \frac{2q}{p_0}\mu_j(\xi_+)\xi_-\right) \times (F_{1j}(\xi_+, 0) - iF_{2j}(\xi_+, 0)) + \int_0^{+\infty} d\xi_- \exp\left[\frac{2i}{p_0}\Delta k_z(\xi_+)\xi_- - \frac{2q}{p_0}\mu_j(\xi_+)\xi_-\right] \times (F_{1j}(\xi_+, 0) + iF_{2j}(\xi_+, 0)) = \frac{qp_0\mu_j(\xi_+)F_{1j}(\xi_+, 0)}{q^2\mu_j^2(\xi_+) + \Delta k_z^2(\xi_+)} - \frac{k_0p_0\Delta k_z(\xi_+)F_{2j}(\xi_+, 0)}{q^2\mu_j^2(\xi_+) + \Delta k_z^2(\xi_+)}. \quad (6.10)$$

Thus, we obtain the following expression for the scattering intensity in the form of a simple integral in the main order in the large parameter:

$$I_j = 2p_0\int_0^{Lp_0} d\xi_+ \times \frac{q\mu_j(\xi_+)F_{1j}(\xi_+, 0) - \Delta k_z(\xi_+)F_{2j}(\xi_+, 0)}{q^2\mu_j^2(\xi_+) + \Delta k_z^2(\xi_+)}. \quad (6.11)$$

6.2. Main Geometries of Scattering

Let us consider the scattering of light with various polarizations.

Scattering of the (o)–(o) type does not exist. This can easily be verified if we take into account the fact that the expression for the scattering intensity contains scalar products of the polarization vectors of incident and scattered waves and the director vector appearing in expression (5.30) for the correlation function, while the polarization vector of the ordinary wave is orthogonal to the director. In this case, the situation is similar to that in an NLC.

Let us consider scattering of the (o)–(e) type. Carrying out summation over recurring indices and integration with respect to the difference variable in expression (6.4), we obtain

$$I(\mathbf{e}^{(1)}, \mathbf{e}^{(2)}) = \frac{J_0}{p_0}\frac{\varepsilon_a^2 k_B T}{qK_{33}}\frac{k_0^2\varepsilon_\perp^2 + \varepsilon_a k_\perp^{(s)2}\cos^2(\phi(L) - \gamma_s)}{\varepsilon_\perp^2 k_0 k_z^{(2)}(\mathbf{k}_\perp^{(s)}, L)}\int_0^{Lp_0} d\xi_+ \frac{V_0(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi_+)}{\cos^2(\xi_+ + \phi_0)} \times \sum_{j=1}^{2}\frac{q\mu_j(\xi_+)V_j(\mathbf{k}_\perp^{(i)}, \xi_+) - [k_z^{(2)}(\mathbf{k}_\perp^{(s)}, \xi_+) - k_z^{(1)}(k_\perp^{(i)})]W_j(\mathbf{k}_\perp^{(i)}, \xi_+)}{q^2\mu_j^2(\xi_+) + [k_z^{(2)}(\mathbf{k}_\perp^{(s)}, \xi_+) - k_z^{(1)}(k_\perp^{(i)})]^2}, \quad (6.12)$$

where γ_s is the angle between vectors $\mathbf{q}$ and $\mathbf{k}_\perp^{(s)}$ and γ_i is the angle between vectors $\mathbf{q}$ and $\mathbf{k}_\perp^{(i)}$, and

$$V_0(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi) = \frac{k_0}{k_z^{(2)}(\mathbf{k}_\perp^{(s)}, \xi)}\frac{k_0^2\varepsilon_\perp - k_\perp^{(s)2}\cos^2(\phi(\xi) - \gamma_s)}{k_0^2\varepsilon_\perp - k_\perp^{(i)2}\cos^2(\phi(\xi) - \gamma_i)}, \quad (6.13)$$

$$V_1(\mathbf{k}_\perp^{(i)}, \xi) = -\frac{k_z^{(1)2}(k_\perp^{(i)})}{k_0^2\varepsilon_\perp}\frac{\sin^2\phi(\xi)}{\mu_1(\xi)} + \frac{k_\perp^{(i)2}}{k_0^2\varepsilon_\perp}\sin^2(\phi(\xi)-\gamma_i)\mu_1(\xi),$$

$$V_2(\mathbf{k}_\perp^{(i)}, \xi) = \frac{k_z^{(1)2}(k_\perp^{(i)})}{k_0^2\varepsilon_\perp}\mu_2(\xi) - \frac{k_\perp^{(i)2}}{k_0^2\varepsilon_\perp}\sin^2(\phi(\xi)-\gamma_i)\frac{\sin^2\phi(\xi)}{\mu_2(\xi)}, \quad (6.14)$$

$$W_1(\mathbf{k}_\perp^{(i)}, \xi) = W_2(\mathbf{k}_\perp^{(i)}, \xi) = \frac{2k_\perp^{(i)}k_z^{(1)}(k_\perp^{(i)})}{k_0^2\varepsilon_\perp}\sin(\phi(\xi)-\gamma_i)\sin\phi(\xi). \quad (6.15)$$

The intensity of the (e)–(o) scattering can easily be derived from the intensity of the (o)–(e) scattering by carrying out the substitutions $\mathbf{e}^{(1)} \rightleftarrows \mathbf{e}^{(2)}$ and $\mathbf{k}^{(s)} \rightleftarrows \mathbf{k}^{(i)}$.

The calculation of the intensity of the (e)–(e) scattering is different since, after convolutions, the contribution to scattering comes from all four terms of correlation functions (5.30). Summing over recurring indices and integrating with respect to the difference variable in expression (6.4), we obtain the intensity of the (e)–(e) scattering in the form

$$I(\mathbf{e}^{(2)}, \mathbf{e}^{(2)}) = \frac{J_0}{p_0}\frac{\varepsilon_a^2 k_B T}{qK_{33}}\frac{k_z^{(2)}(\mathbf{k}_\perp^{(i)}, 0)[k_0^2\varepsilon_\perp^2 + \varepsilon_a k_\perp^{(s)2}\cos^2(\phi(L)-\gamma_s)]}{k_z^{(2)}(\mathbf{k}_\perp^{(s)}, L)[k_0^2\varepsilon_\perp^2 + \varepsilon_a k_\perp^{(i)2}\cos^2(\phi-\gamma_i)]\varepsilon_\perp^2}\int_0^{Lp_0}\frac{d\xi_+}{\cos^2(\xi_+ + \phi_0)}$$
$$\times\sum_{j=1}^{2}\sum_{k=1}^{4}V_0^{(k)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi_+)\frac{q\mu_j(\xi_+)V_j^{(k)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi_+) - [k_z^{(2)}(\mathbf{k}_\perp^{(s)}, \xi_+) - k_z^{(2)}(\mathbf{k}_\perp^{(i)}, \xi_+)]W_j^{(k)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi_+)}{q^2\mu_j^2(\xi_+) + [k_z^{(2)}(\mathbf{k}_\perp^{(s)}, \xi_+) - k_z^{(2)}(\mathbf{k}_\perp^{(i)}, \xi_+)]^2}. \quad (6.16)$$

Here, the following notation has been introduced:

$$V_0^{(1)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi) = \frac{k_0^2\varepsilon_\perp - k_\perp^{(s)2}\cos^2(\phi(\xi)-\gamma_s)}{k_0^2\varepsilon_\perp - k_\perp^{(i)2}\cos^2(\phi(\xi)-\gamma_i)} \times\frac{k_\perp^{(i)2}\cos^2(\phi(\xi)-\gamma_i)}{k_z^{(2)}(\mathbf{k}_\perp^{(s)}, \xi)k_z^{(2)}(\mathbf{k}_\perp^{(i)}, \xi)},$$

$$V_0^{(2)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi) = V_0^{(3)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi) = \frac{k_\perp^{(s)}k_\perp^{(i)}\cos(\phi(\xi)-\gamma_s)\cos(\phi(\xi)-\gamma_i)}{k_z^{(2)}(\mathbf{k}_\perp^{(i)}, \xi)k_z^{(2)}(\mathbf{k}_\perp^{(s)}, \xi)},$$

$$V_0^{(4)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi) = \frac{k_0^2\varepsilon_\perp - k_\perp^{(i)2}\cos^2(\phi(\xi)-\gamma_i)}{k_0^2\varepsilon_\perp - k_\perp^{(s)2}\cos^2(\phi(\xi)-\gamma_s)} \times\frac{k_\perp^{(s)2}\cos^2(\phi(\xi)-\gamma_s)}{k_z^{(2)}(\mathbf{k}_\perp^{(i)}, \xi)k_z^{(2)}(\mathbf{k}_\perp^{(s)}, \xi)}, \quad (6.17)$$

$$V_j^{(1)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi) = \tilde{V}_j(\mathbf{k}_\perp^{(i)}, \mathbf{k}_\perp^{(i)}, \xi),$$

$$V_j^{(2)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi) = V_j^{(3)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi) = \tilde{V}_j(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi),$$

$$V_j^{(4)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi) = \tilde{V}_j(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(s)}, \xi),$$

where $j = 1, 2$, and

$$\tilde{V}_1(\mathbf{k}_1, \mathbf{k}_2, \xi) = -\frac{k_1k_2}{k_0^2}\frac{\sin^2\phi(\xi)}{\mu_1(\xi)}\times\sin(\phi(\xi)-\gamma_1)\sin(\phi(\xi)-\gamma_2) + \frac{k_z^{(2)}(\mathbf{k}_1, \xi)k_z^{(2)}(\mathbf{k}_2, \xi)}{k_0^2}\mu_1(\xi),$$

$$\tilde{V}_2(\mathbf{k}_1, \mathbf{k}_2, \xi) = \frac{k_1k_2}{k_0^2}\mu_2(\xi)\sin(\phi(\xi)-\gamma_1)\sin(\phi(\xi)-\gamma_2) - \frac{k_z^{(2)}(\mathbf{k}_1, \xi)k_z^{(2)}(\mathbf{k}_2, \xi)}{k_0^2}\frac{\sin^2\phi(\xi)}{\mu_2(\xi)}, \quad (6.18)$$

γ_j being the angle between vectors $\mathbf{q}$ and $\mathbf{k}_j$,

$$W_j^{(1)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi) = \tilde{W}_j(\mathbf{k}_\perp^{(i)}, \mathbf{k}_\perp^{(i)}, \xi),$$

$$W_j^{(2)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi) = W_j^{(3)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi) = \tilde{W}_j(\mathbf{k}_\perp^{(i)}, \mathbf{k}_\perp^{(s)}, \xi),$$

$$W_j^{(4)}(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(i)}, \xi) = \tilde{W}_j(\mathbf{k}_\perp^{(s)}, \mathbf{k}_\perp^{(s)}, \xi),$$

where $j = 1, 2$, and

$$\tilde{W}(\mathbf{k}_1, \mathbf{k}_2, \xi) = -\frac{\sin\phi(\xi)}{k_0^2} \times [k_1 k_z^{(2)}(\mathbf{k}_2, \xi)\sin(\phi(\xi) - \gamma_1) + k_2 k_z^{(2)}(\mathbf{k}_1, \xi)\sin(\phi(\xi) - \gamma_2)], \quad \tilde{W}_2(\mathbf{k}_1, \mathbf{k}_2, \xi) = -\tilde{W}_1(\mathbf{k}_1, \mathbf{k}_2, \xi). \tag{6.19}$$

Expressions (6.12) and (6.16) make it possible to calculate the intensity of single scattering of light by a CLC cell to the front hemisphere for an arbitrary orientation of the director in the boundary planes. The intensity of scattering to the rear hemisphere can be calculated with the help of the same methods.

The use of large parameters $\Omega = k_0/p_0$ and $\tilde{\Omega} = q/p_0$ imposes certain limitations on the geometrical conditions of scattering, under which formulas (6.12) and (6.16) are applicable. First, angle γ between vectors $\mathbf{k}_\perp^{(i)}$ and $\mathbf{k}_\perp^{(s)}$ cannot be very small ($\gamma \gg p_0/k_0 \sim \lambda/d$) since expressions for the correlation function in Section 5.2 were derived for $q \gg p_0$. Second, the angles formed by wave vectors $\mathbf{k}^{(i)}$ and $\mathbf{k}^{(s)}$ of the incident and scattered waves of the extraordinary type with the z axis should not be too close to 90°. This is due to the effect of trapping of the extraordinary ray in the flat wave channel, which was described in Section 4.1. Finally, the condition for applicability of the WKB approximation (5.27) for the correlation function, which is associated with inequality (5.33), limits the thickness L of an admissible CLC: $L \ll k_0/p_0^2 \sim \pi d^2/\lambda$. The latter inequality implies that the formulas derived by us make it possible, in particular, to analyze the CLC containing many periods of a helix.

We calculated the intensity of scattered light in the above geometries. In these calculations, we introduced angle ϕ_i between director vector $\mathbf{n}^0$ and vector $\mathbf{k}^{(i)}$ on the plane $z = 0$, as well as angle γ between the transverse components of the wave vectors of incident and scattered light. These angles are shown in Fig. 7.

We calculated scattering intensities $I(\mathbf{e}^{(1)}, \mathbf{e}^{(2)})$ and $I(\mathbf{e}^{(2)}, \mathbf{e}^{(2)})$. Figure 8 shows the constant-intensity lines for light scattered in a cholesteric with total twisting angle $Lp_0 = \pi/2$, when the angle of incidence relative to axis z is equal to $\pi/8$ and $\phi_i = \pi/4$. The maximal intensity for both types of scattering is attained in region $\mathbf{k}^{(s)} \approx \mathbf{k}^{(i)}$. It can be seen that the "spot on the screen" for the (o)–(e) scattering is much broader than for the (e)–(e) scattering. The intensity of the (e)–(e) scattering at the center is formally unlimited, while the intensity for the (o)–(e) scattering is finite. A more complex

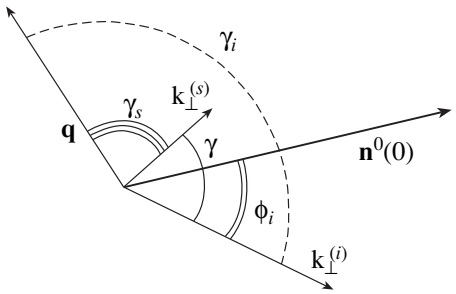

Fig. 7. Projections of the wave vectors onto the *xy* plane.

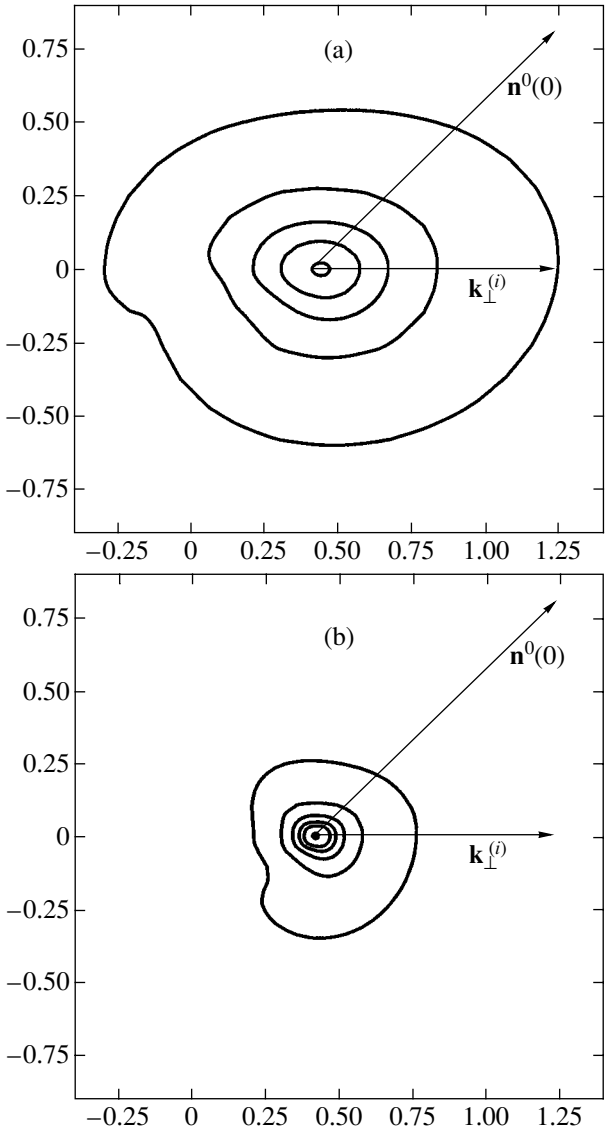

Fig. 8. Isolines of intensity of scattered light for the (o)–(e) (a) and (e)–(e) (b) types of scattering. The distances on the axes are in relative units which are the same for both types of scattering. The intensities are calculated for $\varepsilon_a = 0.5$, $\varepsilon_\perp = 2.0$, $K_{11} = 3.0 \times 10^{-6}$ dyne, $K_{22} = 2.0 \times 10^{-6}$ dyne, and $K_{33} = 5.0 \times 10^{-6}$ dyne; the angle of incidence is $\pi/8$, $\phi_i = \pi/4$, and $Lp_0 = \pi/2$. The peak height for the (o)–(e) scattering is 2.1 relative units. The intensity isolines are plotted at five levels: 2.0, 1.0, 0.5, 0.2, and 0.05.

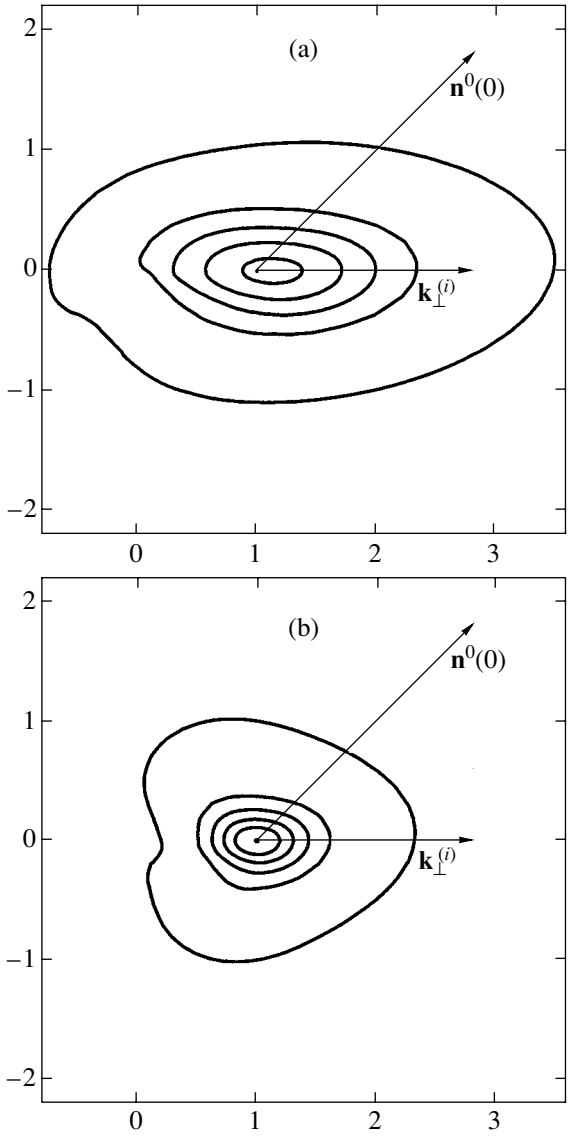

Fig. 9. Isolines of intensity of scattered light for the (o)–(e) (a) and (e)–(e) (b) types of scattering. The distances on the axes are in relative units which are the same for both types of scattering. The values of permittivities and Frank moduli are the same as in Fig. 8. The angle of incidence is $\pi/4$, $\phi_i = \pi/4$, and $Lp_0 = \pi/2$. The peak height for the (o)–(e) scattering is 0.68 relative unit. The intensity isolines are plotted at five levels: 0.5, 0.2, 0.1, 0.05, and 0.01.

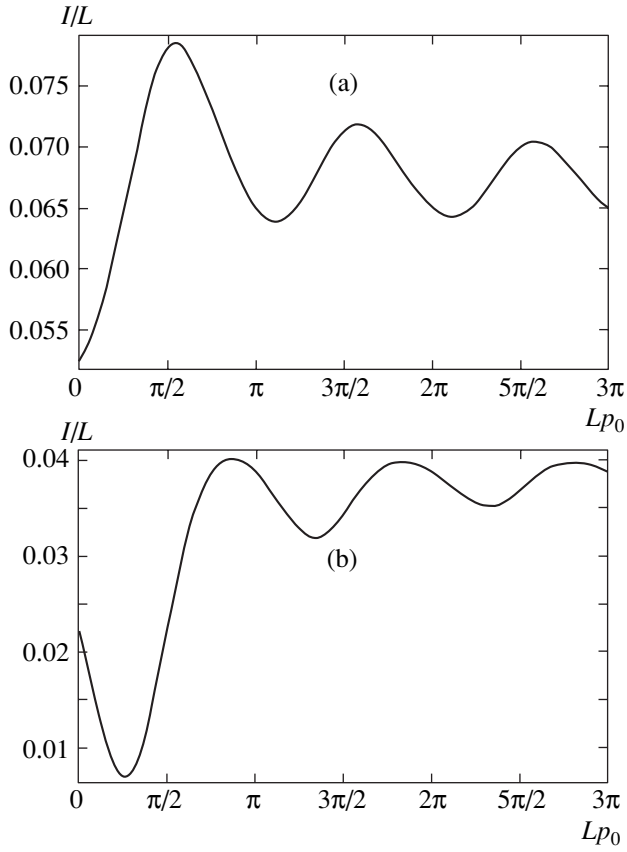

Fig. 10. Dependence of quantity I/L on parameter Lp_0 for scattering of the type (o)–(e) (a) and (e)–(e) (b), expressed in relative units. The values of permittivities and Frank moduli are the same as in Fig. 8. The angle of incidence is $\pi/8$, the angle of scattering is $\pi/4$, $\gamma = \pi/6$, and $\phi_i = \pi/4$.

shape of the peak for the (e)–(e) scattering is also worth noting. Figure 9 shows the same intensities, but for an angle of incidence of $\pi/4$.

A specific feature of the given system is the nonlinear dependence of the intensity of first-order light scattering on the sample volume $V = S_\perp L$. Namely, upon a change in the sample thickness, quantity I/L as a function of parameter Lp_0 begins to oscillate. The dependence of I on $S_\perp$ remains linear in this case. This feature is represented in Fig. 10 for both types of scattering.

We also calculated the degree of polarization P for scattering of the extraordinary wave [39],

$$P = \frac{\left|I(\mathbf{e}^{(2)}, \mathbf{e}^{(2)}) - I(\mathbf{e}^{(2)}, \mathbf{e}^{(1)})\right|}{I(\mathbf{e}^{(2)}, \mathbf{e}^{(2)}) + I(\mathbf{e}^{(2)}, \mathbf{e}^{(1)})}.$$

Intensity $I(\mathbf{e}^{(2)}, \mathbf{e}^{(1)})$ of scattering of the (e)–(o) type can be derived analogously to the intensity (6.12) of the (o)–(e) scattering. The constant-polarization curves are shown in Fig. 11. These curves are rather intricate since the (e)–(e) and (e)–(o) peaks differ significantly. The (e)–(e) scattering peak is much higher and decreases sharply with the angle, while the (e)–(o) scattering peak is more gently sloping and has a considerably smaller height. As a result, in the vicinity of zero scattering angle, the degree of polarization is close to unity; then it rapidly decreases to zero, after which it starts to increase again. In the region where the (e)–(o) scattering intensity increases, the isoline $P = 0.5$ in Fig. 11

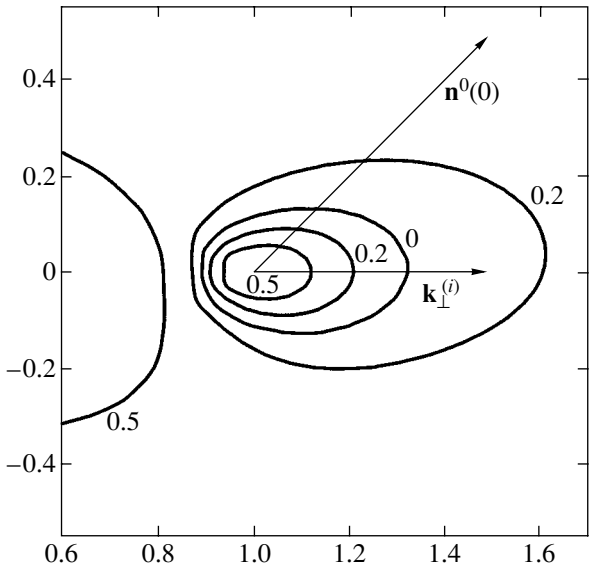

Fig. 11. Isolines of the degree of polarization for the (e)–(o) and (e)–(e) types of scattering. The distances on the axes are in relative units. The values of permittivities and Frank moduli are the same as in Fig. 8. The angle of incidence is $\pi/4$, $\phi_i = \pi/4$, and $Lp_0 = \pi/2$. The isolines of the degree of polarization are plotted at three levels: 0.5, 0.2, and 0.0.

becomes larger and closed for large scattering angles; for this reason, we depict only a part of this line.

7. CONCLUSIONS

We have studied light scattering in cholesterics with a large pitch. In solving this problem, it was found that spatial inhomogeneity of the medium is significant for describing normal waves and the Green function as well as for calculating the spatial correlation functions of permittivity fluctuations and the thermal noise spectra.

The computational method used here may turn out to be helpful in studying light scattering in various layered media and the media with one-dimensional periodicity, when the wavelength of light is smaller than the spatial inhomogeneity of a medium. In particular, the approach developed by us can be used to study the problem of light propagation and scattering in a waveguide channel.

Our calculations show that the proposed theory makes it possible to obtain information on correlation functions (in media whose properties change in space) by measuring the angular and polarization characteristics of scattered light intensity.

ACKNOWLEDGMENTS

The authors are grateful to V.A. Belyakov for fruitful discussions.

This study was supported financially by the Russian Foundation for Basic Research (project nos. 02-02-16577 and 03-02-16173) and partly by the Ministry of Education of the Russian Federation and the St. Petersburg administration (grant nos. PD02-1.2-297 and PD03-1.2-36) as well as NATO (grant no. PST.CLG.979652) and the Royal Society (grant no. 15298).

APPENDIX

Vector WKB Method

Let us consider an equation of the type

$$\frac{\partial}{\partial\xi}\hat{M}(\xi, \xi_0) = (i\tilde{\Omega}\hat{B} + \hat{C})\hat{M}(\xi, \xi_0) \qquad (A.1)$$

with the initial condition $\hat{M}(\xi_0, \xi_0) = \hat{I}$, where $\tilde{\Omega} \gg 1$, $\hat{B} = \hat{B}(\xi)$, and $\hat{C} = \hat{C}(\xi)$. It is convenient to carry out the substitution of the unknown, after which the system becomes diagonal in the main order in large parameter $\tilde{\Omega}$. For this purpose, we represent $\hat{M}(\xi, \xi_0)$ in the form

$$\hat{M}(\xi, \xi_0) = \hat{U}(\xi)\hat{H}(\xi, \xi_0), \qquad (A.2)$$

where $\hat{U}(\xi)$ is a certain nondegenerate matrix that will be chosen later and $\hat{H}$ is a new unknown satisfying the initial condition $\hat{H}(\xi_0, \xi_0) = \hat{U}^{-1}(\xi_0)$.

In this case, Eq. (A.1) assumes the form

$$\frac{\partial\hat{H}}{\partial\xi} = \left(i\tilde{\Omega}\hat{U}^{-1}\hat{B}\hat{U} + \hat{U}^{-1}\hat{C}\hat{U} - \hat{U}^{-1}\frac{\partial\hat{U}}{\partial\xi}\right)\hat{H}. \qquad (A.3)$$

We now choose $\hat{U}(\xi)$ in such a way that matrix $\hat{U}^{-1}\hat{B}\hat{U}$ becomes diagonal, i.e.,

$$\hat{U}^{-1}\hat{B}\hat{U} = \hat{\Lambda}, \qquad (A.4)$$

where $\hat{\Lambda}$ is a diagonal matrix composed of the eigenvalues of matrix $\hat{B}$.[5] In this case, the columns of matrix $\hat{U}$ are eigenvectors of matrix $\hat{B}$. Then Eq. (A.3) assumes the form

$$\frac{\partial\hat{H}}{\partial\xi} = i\tilde{\Omega}\left[\hat{\Lambda} + \frac{i}{\tilde{\Omega}}\left(\hat{U}^{-1}\frac{\partial\hat{U}}{\partial\xi} - \hat{U}^{-1}\hat{C}\hat{U}\right)\right]\hat{H}. \qquad (A.5)$$

If we disregard the term of the relative order of $1/\tilde{\Omega}$ in the brackets on the right-hand side of this formula, the

[5] The situation when matrix $\hat{B}$ cannot be reduced to the diagonal form at some points is considered in [45] as applied to problem of wave propagation and the theory of oscillations and in [44] as applied to fluctuations in CLCs.

system splits, in view of the diagonal form of matrix $\hat{\Lambda}$, into the independent equations

$$\frac{\partial \hat{H}_{(0)}}{\partial \xi} = i\tilde{\Omega}\hat{\Lambda}\hat{H}_{(0)}. \quad \text{(A.6)}$$

The solution to these equations has the form

$$\hat{H}_{(0)}(\xi, \xi_0) = \exp\left[i\tilde{\Omega}\int_{\xi_0}^{\xi}\hat{\Lambda}(\xi')d\xi'\right]\hat{U}^{-1}(\xi_0). \quad \text{(A.7)}$$

The main drawback of this formula of the "zero-order approximation" is that it gives, together with Eq. (A.2), an expression for $\hat{M}(\xi, \xi_0)$, which is not invariant relative to an arbitrary (but depending on point ξ) choice of normalization of eigenvectors of matrix $\hat{B}$. In order to overcome this difficulty, we must obtain a solution in the next order in $\tilde{\Omega}$. For this purpose, we represent $\hat{H}(\xi, \xi_0)$ in the form

$$\hat{H}(\xi, \xi_0) = \hat{U}_{(1)}(\xi)\hat{H}_{(1)}(\xi, \xi_0), \quad \text{(A.8)}$$

where $\hat{U}_{(1)}(\xi)$ is a nondegenerate matrix that has not yet been determined. Substituting relation (A.8) into (A.5), we obtain the equation for $\hat{H}_{(1)}(\xi, \xi_0)$,

$$\frac{\partial \hat{H}_{(1)}}{\partial \xi} = i\tilde{\Omega}\left\{\hat{U}_{(1)}^{-1}\left[\hat{\Lambda} + \frac{i}{\tilde{\Omega}}\left(\hat{U}^{-1}\frac{\partial \hat{U}}{\partial \xi} - \hat{U}^{-1}\hat{C}\hat{U}\right)\right]\hat{U}_{(1)} + \frac{i}{\tilde{\Omega}}\hat{U}_{(1)}^{-1}\frac{\partial \hat{U}_{(1)}}{\partial \xi}\right\}\hat{H}_{(1)} \quad \text{(A.9)}$$

with the initial condition $\hat{H}_{(1)}(\xi_0, \xi_0) = \hat{U}_{(1)}^{-1}(\xi_0)\hat{U}^{-1}(\xi_0)$.

We now choose matrix $\hat{U}_{(1)}(\xi)$ so that

$$\hat{U}_{(1)}^{-1}\left[\hat{\Lambda} + \frac{i}{\tilde{\Omega}}\left(\hat{U}^{-1}\frac{\partial \hat{U}}{\partial \xi} - \hat{U}^{-1}\hat{C}\hat{U}\right)\right]\hat{U}_{(1)} = \hat{\Lambda}_{(1)}, \quad \text{(A.10)}$$

where $\hat{\Lambda}_{(1)}$ is a diagonal matrix composed of the eigenvalues of the matrix appearing in the brackets in Eq. (A.10). Then Eq. (A.9) assumes the form

$$\frac{\partial \hat{H}_{(1)}}{\partial \xi} = \left(i\tilde{\Omega}\hat{\Lambda}_{(1)} - \hat{U}_{(1)}^{-1}\frac{\partial \hat{U}_{(1)}}{\partial \xi}\right)\hat{H}_{(1)}. \quad \text{(A.11)}$$

It should be noted that, for $\tilde{\Omega} \gg 1$, matrix $\hat{\Lambda} + i\tilde{\Omega}^{-1}(\hat{U}^{-1}\partial\hat{U}/\partial\xi - \hat{U}^{-1}\hat{C}\hat{U})$ is close to matrix $\hat{\Lambda}$; consequently, matrices $\hat{\Lambda}$ and $\hat{\Lambda}_{(1)}$ are also close and, hence, matrix $\hat{U}_{(1)}$ in Eq. (A.10) is close to the identity matrix; i.e.,

$$\hat{U}_{(1)}(\xi) \approx \hat{I} + \frac{i}{\tilde{\Omega}}\hat{V}(\xi), \qquad \hat{U}_{(1)}^{-1}(\xi) \approx \hat{I} - \frac{i}{\tilde{\Omega}}\hat{V}(\xi), \quad \text{(A.12)}$$

where $\hat{V} = O(1)$.

It follows from relations (A.12) that the second term in the parentheses in Eq. (A.11) is on the order of $1/\tilde{\Omega}^2$ relative to the first term. Disregarding this term, we obtain

$$\hat{H}_{(1)}(\xi, \xi_0) \approx \exp\left[i\tilde{\Omega}\int_{\xi_0}^{\xi}\hat{\Lambda}_{(1)}(\xi')d\xi'\right]\hat{U}_{(1)}^{-1}(\xi_0)\hat{U}^{-1}(\xi_0). \quad \text{(A.13)}$$

Substituting formulas (A.12) into Eq. (A.10), we obtain the following expressions for new matrices $\hat{\Lambda}_{(1)}$ and $\hat{V}$ in the main orders in $\tilde{\Omega}$:

$$(\hat{\Lambda}_{(1)})_{ll} \approx i\mu_l + \frac{i}{\tilde{\Omega}}\left(\hat{U}^{-1}\frac{\partial \hat{U}}{\partial \xi} - \hat{U}^{-1}\hat{C}\hat{U}\right)_{ll}, \quad \text{(A.14)}$$

$$V_{lm} \approx \frac{1}{i(\mu_m - \mu_l)}\left(\hat{U}^{-1}\frac{\partial \hat{U}}{\partial \xi} - \hat{U}^{-1}\hat{C}\hat{U}\right)_{lm}; \quad l \neq m, \quad \text{(A.15)}$$

here, $\mu_l = -i\Lambda_{ll}$. The determination of diagonal terms V_{ll} of matrix $\hat{V}$ requires the next iteration in our method of successive diagonalization, i.e., a substitution of the form

$$\hat{H}_{(1)}(\xi, \xi_0) = \hat{U}_{(2)}(\xi)\hat{H}_{(2)}(\xi, \xi_0)$$

into (A.11) and the computation of the values of $\hat{U}_{(2)}$ and $\hat{H}_{(2)}$ taking into account corrections on the order of $1/\tilde{\Omega}^2$. However, the contributions from the diagonal part of $\hat{V}$ in the expression

$$\hat{M}(\xi, \xi_0) = \hat{U}(\xi)\hat{U}_{(1)}(\xi)\hat{H}_{(1)}(\xi, \xi_0) \quad \text{(A.16)}$$

for the evolution operator in the "first approximation", which are connected with external factors $\hat{U}_{(1)}(\xi)$ in Eq. (A.16) and $\hat{U}_{(1)}^{-1}(\xi_0)$ in Eq. (A.13), are cancelled with the contribution from the diagonal part of $\hat{V}$, which is due to term $\hat{U}_{(1)}^{-1}\partial\hat{U}_{(1)}/\partial\xi$ in Eq. (A.11). For this reason, we can disregard the diagonal terms of matrix $\hat{V}$ in the order in $\tilde{\Omega}$ we are interested in.

Disregarding corrections on the order of $1/\tilde{\Omega}$ in expressions (A.12) for matrices $\hat{U}_{(1)}$ and $\hat{U}_{(1)}^{-1}$ (i.e., in fact replacing these matrices by identity matrices), we obtain from Eq. (A.16) matrix $\hat{M}(\xi, \xi_0)$ in the WKB approximation:

$$\hat{M}(\xi, \xi_0) = \hat{U}(\xi)\widehat{\mathrm{diag}}\left\{\exp\left[-\int_{\xi_0}^{\xi}\left(\tilde{\Omega}\mu_l(x) + \left(\hat{U}^{-1}(x)\frac{\partial \hat{U}(x)}{\partial x} - \hat{U}^{-1}(x)\hat{C}(x)\hat{U}(x)\right)_{ll}\right)dx\right]\right\}\hat{U}^{-1}(\xi_0). \quad \text{(A.17)}$$

This formula for $\hat{C} = \hat{0}$ is a vector analog of the classical WKB approximation.

The domain of applicability of this formula is defined by the inequalities

$$\tilde{\Omega} \gg 1, \quad |V_{lm}(\xi')| \ll \tilde{\Omega},$$
$$\tilde{\Omega}\left|\int_{\xi_0}^{\xi}(\hat{\Lambda}_{(2)}(\xi') - \hat{\Lambda}_{(1)}(\xi'))_{ll}d\xi'\right| \ll 1, \quad \text{(A.18)}$$

where $\hat{\Lambda}_{(2)}$ is the corresponding diagonal matrix in the second approximation.

The first inequality simply indicates that $\tilde{\Omega}$ is a large parameter. The second inequality is associated with disregard of terms $\pm i\hat{V}(\xi)/\tilde{\Omega}$ in expressions (A.12) in the derivation of Eq. (A.17); in accordance with relation (A.15), this inequality imposes the following constraint on the closeness of eigenvalues μ_l and μ_m in the entire interval from ξ_0 to ξ:

$$\min_{\xi_0 \le \xi' \le \xi}|\mu_l(\xi') - \mu_m(\xi')| \gg \tilde{\Omega}^{-1}. \quad \text{(A.19)}$$

Finally, the third inequality indicates the smallness of the next correction to the exponential term in Eq. (A.17) for any ξ_0, ξ and sets a limit on the admissible width of region $\xi - \xi_0$, in which the WKB formula (A.17) can be used. In order of magnitude, $\Lambda_{(1)ll}(\xi') - \Lambda_{(2)ll}(\xi') \sim \mu_l(\xi')\tilde{\Omega}^{-2}$; consequently, we have

$$|\xi - \xi'| \ll \bar{\mu}_l\tilde{\Omega}, \quad \text{(A.20)}$$

where $\bar{\mu}_l$ is the mean value of μ_l on interval $[\xi_0; \xi]$.

REFERENCES

1. C. W. Oseen, Trans. Faraday Soc. **29**, 883 (1933).
2. H. De Vries, Acta Crystallogr. **4**, 219 (1951).
3. E. I. Kats, Zh. Éksp. Teor. Fiz. **59**, 1854 (1970) [Sov. Phys. JETP **32**, 1004 (1970)].
4. D. W. Berreman and T. J. Scheffer, Phys. Rev. A **5**, 1397 (1972).
5. D. W. Berreman, J. Opt. Soc. Am. **62**, 502 (1972); **63**, 1374 (1973).
6. Sah Yuvaraj and K. A. Suresh, J. Opt. Soc. Am. A **11**, 740 (1994).
7. S. Chandrasekhar, *Liquid Crystals* (Cambridge Univ. Press, Cambridge, 1977; Mir, Moscow, 1980).
8. V. A. Belyakov, *Diffraction Optics of Complex-Structured Periodic Media* (Nauka, Moscow, 1988; Springer, New York, 1992).
9. V. A. Belyakov and S. A. Sonin, *Optics of Cholesteric Liquid Crystals* (Nauka, Moscow, 1982).
10. C. Oldano, Phys. Rev. A **40**, 6014 (1989).
11. S. P. Palto, Zh. Éksp. Teor. Fiz. **119**, 638 (2001) [JETP **92**, 552 (2001)].
12. P. Galatola, Phys. Rev. E **55**, 4338 (1997).
13. P. Hubert, P. Jagemalm, C. Oldano, and M. Rajteri, Phys. Rev. E **58**, 3264 (1998).
14. C. Oldano and S. Ponti, Phys. Rev. E **63**, 011703 (2000).
15. S. Ponti, C. Oldano, and M. Becchi, Phys. Rev. E **64**, 021704 (2001).
16. M. A. Peterson, Phys. Rev. A **27**, 520 (1983).
17. N. A. Nicorovici, R. C. McPhedran, and R. Petit, Phys. Rev. E **49**, 4563 (1994).
18. P. Galatola, Phys. Rev. E **49**, 4552 (1994).
19. A. Yu. Val'kov, V. P. Romanov, and A. N. Shalaginov, Akust. Zh. **37**, 636 (1991) [Sov. Phys. Acoust. **37**, 329 (1991)].
20. T. C. Lubensky, Phys. Rev. A **6**, 452 (1972).
21. M. J. Stephen and J. P. Straley, Rev. Mod. Phys. **46**, 617 (1974).
22. R. L. Stratanovich, Zh. Éksp. Teor. Fiz. **70**, 1290 (1976) [Sov. Phys. JETP **43**, 672 (1976)].
23. M. S. Veshchunov, Zh. Éksp. Teor. Fiz. **76**, 1515 (1979) [Sov. Phys. JETP **49**, 769 (1979)].
24. M. C. Mauguin, Bull. Soc. Fr. Mineral. Cristallogr. **34**, 71 (1911).
25. J. W. Shelton and Y. R. Shen, Phys. Rev. Lett. **25**, 23 (1970); **26**, 538 (1971); Phys. Rev. A **5**, 1867 (1972).
26. V. V. Zheleznyakov, V. V. Kocharovskiĭ, and Vl. V. Kocharovskiĭ, Usp. Fiz. Nauk **141**, 257 (1983) [Sov. Phys. Usp. **26**, 877 (1983)].
27. V. S. Liberman and B. Ya. Zel'dovich, Phys. Rev. E **49**, 2389 (1994).
28. A. Yu. Savchenko and B. Ya. Zel'dovich, Phys. Rev. E **50**, 2287 (1994).
29. A. Yu. Val'kov, R. V. Grinin, and V. P. Romanov, Opt. Spektrosk. **83**, 239 (1997) [Opt. Spectrosc. **83**, 221 (1997)].
30. E. V. Aksenova, V. P. Romanov, and A. Yu. Val'kov, Phys. Rev. E **59**, 1184 (1999).
31. E. V. Aksenova, A. Yu. Val'kov, and V. P. Romanov, Mol. Cryst. Liq. Cryst. **359**, 351 (2001).

32. E. V. Aksenova, A. Yu. Val'kov, and V. P. Romanov, Opt. Spektrosk. **91**, 1030 (2001) [Opt. Spectrosc. **91**, 969 (2001)].

33. M. A. Anisimov, *Critical Phenomena in Liquids and Liquid Crystals* (Nauka, Moscow, 1987).

34. P. de Gennes, *The Physics of Liquid Crystals* (Clarendon Press, Oxford, 1974; Mir, Moscow, 1977).

35. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 7th ed. (Nauka, Moscow, 1988; Pergamon Press, Oxford, 1975).

36. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Pergamon Press, Oxford, 1980), Part 1.

37. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 2nd ed. (Nauka, Moscow, 1982; Pergamon, New York, 1984).

38. M. Lax and D. F. Nelson, in *Proceedings of III Rochester Conference on Coherent and Quantum Optics* (Plenum, New York, 1973), p. 415; in *Proceedings of I Soviet-American Conference on Theory of Light Scattering in Condensed Media, Moscow, 1975* (Nauka, Moscow, 1976), Vol. 2, p. 452.

39. M. Born and E. Wolf, *Principles of Optics*, 3rd ed. (Pergamon Press, Oxford, 1965; Nauka, Moscow, 1970).

40. M. B. Vinogradova, O. V. Rudenko, and A. P. Sukhorukov, *The Theory of Waves*, 1st ed. (Nauka, Moscow, 1979).

41. D. W. Berreman and T. J. Scheffer, Phys. Rev. Lett. **25**, 577 (1970).

42. M. Lax and D. F. Nelson, Phys. Rev. B **4**, 3694 (1971).

43. V. A. Yakubovich and V. M. Strazhinskiĭ, *Linear Differential Equations with Periodic Coefficients and Their Applications* (Nauka, Moscow, 1987).

44. E. V. Aksenova, V. P. Romanov, and A. Yu. Val'kov, in *Proceedings of the International Seminar on Day on Diffraction-2001* (St. Petersburg, 2001), p. 7.

45. M. V. Perel', *Problems of Wave Diffraction and Propagation* (Leningr. Gos. Univ., Leningrad, 1990), No. 23, p. 58; Izv. Vyssh. Uchebn. Zaved., Radiofiz. **33**, 1208 (1990).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 98, No. 1, 2004, pp. 93–101.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 125, No. 1, 2004, pp. 103–113.

FLUIDS

Equilibrium Adsorption of Ligands on DNA (for Chitosan)

Yu. D. Nechipurenko[a,b,d,*], A. M. Wolf[c], V. I. Salyanov[a,d], and Yu. M. Yevdokimov[a,d]

[a]*Engelgardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia*
[b]*Department of Physics, Moscow State University, Moscow, 119992 Russia*
[c]*Humboldt University, 10099 Berlin, Germany*
[d]*Bioanalytical Technologies Company, Moscow, 117049 Russia*
**e-mail: nech@smtp.ru*

Received July 18, 2003

Abstract—The binding of ligands to DNA molecules in solution is characterized by a distribution function determining the probability that a certain number of ligands per DNA molecule are adsorbed. As a result of the binding to polycations, rigid linear double-stranded DNA molecules (with negative charges of the phosphate groups neutralized by the positive charges of ligands) exhibit a phase transition. Experimental data on the binding of chitosan to DNA are interpreted assuming that chitosan exhibits equilibrium adsorption on DNA. When the number of chitosan molecules adsorbed on DNA exceeds a certain critical value, the DNA molecule covered by chitosan becomes capable of interacting with like DNA molecules. This interaction (attraction) results in the formation of particles constituting a liquid-crystalline dispersion. It is shown that the proposed model with certain parameters provides a description of certain experimental data characterizing the formation of cholesteric liquid-crystalline dispersions. An analysis of the experimental data makes it possible to determine both the size of a site occupied by an amino sugar unit of chitosan adsorbed on DNA and the energy of interaction of this unit with DNA. © *2004 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

The equilibrium binding of ligand molecules to linear polymers in solutions is considered as the Langmuir adsorption process, whereby the linear polymer macromolecule represents a one-dimensional matrix—a lattice of reactive centers—and the ligand is adsorbed on these centers of the matrix. The principles of this approach were formulated by Hill [1, 2] and by Magee, Gibbs, and Zimm [3]. This phenomenon attracts much attention and there are several dozens of papers developing various adsorption models and establishing relationships describing the binding of biologically active ligands to DNA matrices (see, e.g., [4–13]). A ligand bound to DNA usually covers several base pairs on a DNA molecule, rendering these sites inaccessible for the other ligand molecules. From the standpoint of physics, the behavior of adsorbed ligands is analogous to the behavior of a lattice gas with allowance for the excluded volume.

The binding of ligands to DNA is most adequately described in terms of a distribution function that determines the probability of finding a DNA molecule with a certain number of adsorbed ligands in solution. This function was originally introduced by Reiter and Epstein [14, 15] for description of the binding of extended ligands to DNA. In [16], this function was used for analyzing the arrangement of ligands on DNA molecules. Poland [17, 18] thoroughly considered the application of such a distribution function to description of the binding of various ligands to macromolecules.

In this study, the probability distribution function is used to describe a process in which the noncooperative "critical neutralization" of DNA phosphates by the ligands bearing positively charged groups leads to DNA condensation. The condensed form of double-stranded (duplex) DNA attracts the attention of researchers by offering a model describing certain peculiarities of the state of DNA in biological objects such as viruses, protozoan chromosomes, etc. [19].

There are two possible modes of condensation for the double-stranded molecules of nucleic acids (NAs):

(i) "Entropy condensation" is a process in which the driving force is a change in the system entropy. This type of condensation takes place when NA molecules (or segments of the same high-molecular-weight NA molecule) are separated from an aqueous polymer or an aqueous salt solution [20, 21]. It should be noted that, when NA is separated from an aqueous polymer solution, the polymer molecules do not enter into the composition of a new NA phase. The process of NA phase separation can be described to a sufficiently good accuracy within the framework of the Flory theory (see, e.g., [20]).

(ii) "Enthalpy condensation" is a process in which the driving force is attraction between neighboring NA molecules (or segments of the same NA molecule). The intermolecular attraction forces are mostly of the electrostatic nature, including the London dispersion interaction and interactions of the dipole–induced dipole type. These forces are rather weak at large intermolecular distances R but grow rapidly when NA molecules approach one another, the interaction energy varying approximately in proportion to $1/R^5$ [21]. Apparently, when a surface charge density on the NA molecules is sufficiently low (as a result of neutralization of the negative charges of the phosphate groups of NAs by the positive charges of polycation groups), the dispersion forces equilibrate and then exceed the electrostatic repulsion between adjacent NA molecules, inducing the condensation process.

It should be noted that, in the case of high-molecular-weight NA molecules, condensation is an intramolecular process (manifested as compactization or globulization), whereas for low-molecular-weight NAs, this is an intermolecular process.

In the case of enthalpy condensation, the process can develop according to one of two possible scenarios: (a) neutralization of the negative charges of the phosphate groups of NAs by the positive charges of polycations and attraction of the adjacent molecules in NA–polycation complexes; (b) the formation of polycation crosslinks between the neighboring NA molecules, determining the effective proximity of these molecules. Condensation of the latter type was observed in experiments with polycations representing spermine, spermidine, etc. (see, e.g., [13, 22] and references therein).

Irrespective of the particular enthalpy condensation scenario, the added polycations enter into the composition of a new phase. Experimental data show that the condensation of NAs is always switched on when the polycation concentration in solution reaches a certain "critical" level. This process is described by a characteristic S-shaped curve of the degree of NA condensation versus polycation concentration [23–27].

Recently [28], it was demonstrated that the use of chitosan as a polycation leads to the formation of a liquid-crystalline phase of NAs and that the character of spatial packing of the adjacent NA molecules in this phase can be controlled by changing the spacing of amino groups in chitosan molecules at a constant molecular weight of chitosan (i.e., at a constant length of the molecule). The interest in chitosan—a biodegradable polymer called a "polymer of the 21st century"—has stimulated investigations into the properties of liquid crystals formed as a result of chitosan–DNA interaction [28].

In this study, we have attempted to develop a theory describing the interaction between DNA molecules (of several hundred base pairs in size) and chitosan—an extended ligand occupying several base pairs upon binding to DNA. The results will be compared to experimental data characterizing a phase transition in DNA caused by neutralization of the negative charges of the phosphate groups of DNA by the positive charges of the amino groups of chitosan. Our description is based on the following assumptions:

(i) Chitosan binds to DNA in a noncooperative manner, whereby the solution contains both DNA molecules free of chitosan and those on which one, two, etc., chitosan molecules are adsorbed—up to the maximum DNA coverage by chitosan (it is assumed that the bound chitosan molecules neither overlap nor dangle from DNA ends).

(ii) An essential part in chitosan binding to DNA is the interaction of the charged amino groups of chitosan with DNA phosphates, whereby the charge of the latter groups is gradually neutralized when bound chitosan covers a DNA molecule.

(iii) When the chitosan coverage reaches a certain critical level, DNA molecules pass to a modified state required for the formation of a cholesteric liquid-crystalline phase called liquid-crystalline dispersion.

(iv) The abnormal optical activity of a liquid-crystalline form of the DNA–chitosan complex is directly proportional to the concentration of DNA molecules in the modified state (the same is valid for an "apparent" optical density characterizing the light scattering from liquid-crystalline dispersion particles consisting of DNA–chitosan complexes).

It should be noted that we will neither consider nor take into account any interactions between DNA molecules covered by chitosan (nor shall we consider the possible crosslinking, whereby one chitosan molecule binds to two DNA molecules). Thus, we only estimate the number of DNA molecules capable of entering into such interactions.

2. A MODEL OF LIGAND ADSORPTION ON DNA MOLECULES

Let us consider a solution at constant temperature and pressure, containing DNA molecules with adsorbed ligands as well as free DNA and ligand molecules. A DNA molecule represents a matrix—a linear polymer containing N reactive centers (the function of such centers is performed by base pairs with atoms of the sugar-phosphate core). A bound ligand occupies L successive centers, rendering them inaccessible for binding other ligand molecules. It should be noted that we do not consider particular sterical details and peculiarities of the ligand–DNA interactions. These L centers will be referred to as the binding site. If the reactive centers of a given matrix are numbered from one end to the other, the first binding site comprises a sequence of the first L centers. Let K be the chemical equilibrium constant for the reaction of ligand binding to one binding site of the matrix. We will consider the case when

ligand binding to any site is characterized by the same equilibrium binding constant (the so-called homopolymer binding). The solution under consideration contains DNA molecules with various numbers q of adsorbed ligands. In the saturated state, a DNA molecule containing N centers is capable of adsorbing $q_{\max} = N/L$ ligand molecules. The equilibrium properties of the system representing a DNA matrix with adsorbed ligands are determined when the free energy $\Delta f(q, N)$ of the system is calculated. This function can be written as follows (see [29]):

$$-\Delta f(q, N)/RT = \ln[(N - qL + q)!/q!(N - qL)!] + q\ln Km, \quad (1)$$

where m is the free ligand concentration in solution, R is the universal gas constant, and T is the temperature.

The first term in the right-hand part of formula (1) corresponds to the statistical entropy, representing the number of permutations of the ligands and individual vacant centers. If a DNA molecule has bound q ligands, there are $N - qL$ vacant centers on this molecule. The statistical entropy is equal to the logarithm of the degree of degeneracy of a macroscopic state of the matrix with adsorbed ligands, or to the logarithm of the number of permutations in the sequence of elements of two types—ligands and free centers:

$$\Delta S(q, N) = \ln\frac{(N - qL + q)!}{q!(N - qL)!} = \ln C^q_{N - qL + q}. \quad (2)$$

It should be recalled that we are considering one matrix bearing adsorbed ligands. The properties of this subsystem involve an uncertainty related to the statistical entropy. Indeed, if it is known that there are q ligands adsorbed on the matrix, no certain answer can be given to the question as to which particular reactive centers are occupied and which are vacant. The statistical entropy of the matrix with adsorbed ligands is a measure of this uncertainty.

Let us denote by $M(q)$ the concentration of matrices with q adsorbed ligands ($M(0)$ being the concentration of matrices free of adsorbed ligands). In accordance with the Boltzmann distribution, we have

$$\ln[M(q)/M(0)] = -\Delta f(q, N)/RT, \quad (3)$$

where quantity $\Delta f(q, N)$ is defined by Eq. (1). Once the free ligand concentration m in solution is known, the concentration of matrices bearing q ligands can be calculated. Indeed, if the concentration of ligand-free matrices is $M(0)$, the concentration of matrices bearing q ligands according to Eq. (3) is

$$M(q) = M(0)\exp(-\Delta f(q, N)/RT). \quad (4)$$

Consider a system including a single matrix. The matrix can occur in one of the $q_{\max} + 1$ states: it can be either free of ligands or bear 1, 2, …, $q_{\max}$ ligand molecules. The probability that the matrix occurs in a certain state exponentially depends on the free energy corresponding to this state. This probability determines the number of matrices bearing q ligands in solution. Denoting by C_M the total concentration of matrices in solution, we can write

$$\sum_{q=0}^{q_{\max}} M(q) = \sum_{q=0}^{q_{\max}} M(0)\exp(-\Delta f(q, N)/RT) = C_M. \quad (5)$$

Now let us arbitrarily chose one matrix among all contained in solution. The probability of finding a matrix bearing q ligands is given by the ratio of the concentration of such matrices $M(q)$ to the total concentration of matrices in solution C_M:

$$\Phi(q) = M(q)/C_M. \quad (6)$$

There is a certain distribution of matrices between states $\Phi(q)$ for each concentration m of free ligands in solution. This distribution function, presenting an exhaustive characteristic of the system under consideration, can be written as

$$\Phi(q) = \frac{C^q_{N - qL + q}(Km)^q}{\sum_{q=1}^{q_{\max}} C^q_{N - qL + q}(Km)^q}, \quad (7)$$

with the obvious condition that

$$\sum_{q=0}^{q_{\max}} \Phi(q) = 1.$$

The distribution function is independent of the concentration of matrices in solution, but depends on the lengths of both the ligand and the matrix. As for the dependence on the concentration m of free ligands in solution, the range of these vales featuring significant changes in the distribution function is determined by the equilibrium binding constant K. Indeed, note that $\Phi(q)$ depends only on the product Km. The average number of ligands $\langle q\rangle$ adsorbed on one matrix is also determined by the Km value. For this average, Eq. (7) yields

$$\langle q\rangle = \sum_{q=0}^{q_{\max}} q\Phi(q) = \frac{\sum_{q=1}^{q_{\max}} qC^q_{N - qL + q}(Km)^q}{\sum_{q=0}^{q_{\max}} C^q_{N - qL + q}(Km)^q}. \quad (8)$$

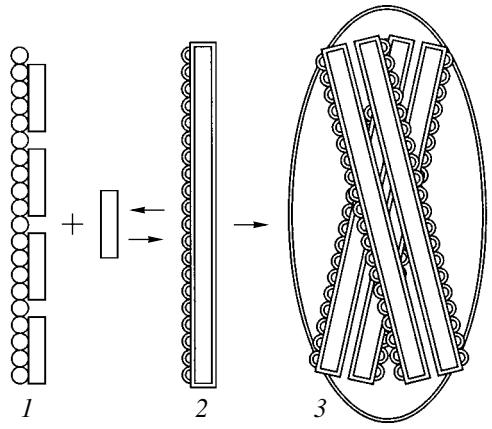

Fig. 1. Schematic diagrams showing DNA in states of different coverage by adsorbed ligands and the condensation of DNA–ligand complexes upon reaching critical coverage. DNA may exist in three different states in solution: (*1*) normal state, whereby the ligand coverage is below critical; (*2*) modified state, when the ligand coverage is above the critical level; (*3*) liquid-crystalline dispersion of DNA–ligand complexes.

In the experiments with polycations bound to DNA, it is convenient to use a relative quantity called the coverage of a matrix by ligands. If a given matrix binds q ligands, each covering L reactive centers of the matrix, the coverage of the matrix is defined as $\alpha = qL/N$.

3. THE BINDING OF POLYCATIONS TO DNA AND THE EQUILIBRIUM OF TWO STATES

We assume that a rigid double-stranded DNA molecule (with a molecular weight not exceeding 10^6 daltons (D)) having bound a certain critical number q_{crit} of ligands becomes capable of interacting with other like DNA molecules, this interaction eventually providing for the DNA transition to a condensed state. In particular, DNA molecules coated by chitosan become capable of interacting with each other, provided that the negative charges of the phosphate groups of DNA are compensated to a sufficiently large degree by the amino groups of chitosan. This interaction makes possible the formation of a liquid-crystalline dispersion [28].

Below, a DNA matrix upon binding $q \geq q_{crit}$ ligands (i.e., with DNA charges compensated to a sufficiently large degree) will be referred to as occurring in the modified state (or state 2), in contrast to the normal state of the DNA matrix (state 1) observed for $q < q_{crit}$ (Fig. 1). If chitosan is considered as a polycation, we may suggest that, as the chitosan concentration in solution is increased, a growing proportion of DNA molecules in solution passes to state 2, thus making liquid-crystalline dispersion formation possible. In this paper, we are not interested in the process of dispersion formation: consideration is restricted to estimating, using the above distribution function, the fraction of DNA molecules capable of participating in the process of liquid-crystalline dispersion formation.

State 2 is distinguished by the fact that the DNA matrix bearing a critical number of chitosan molecules behaves as a molecule acquiring new properties manifested by the ability to form liquid-crystalline dispersion. If an overwhelming majority of DNA molecules that have passed into state 2 enter into liquid-crystalline dispersion particles, we can assume that the DNA fraction in liquid-crystalline dispersion is a parameter characterizing the ratio of DNA molecules in states 1 and 2. Then, state 2 represents the liquid-crystalline phase that can be readily detected by measuring the anomalous amplitude of the optical response signal in the circular dichroism spectrum or the signal of light scattering from liquid-crystalline dispersion particles (apparent optical density) [28].

In order to calculate the proportion of DNA molecules occurring in state 2, we will find the relations describing the probability of finding a DNA molecule in solution with a number of bound ligands q greater than q_{crit}. Denoting this probability by R, we can write

$$R = \sum_{q = q_{crit}}^{q_{max}} \Phi(q) = \frac{\sum_{q = q_{crit}}^{q_{max}} C^q_{N - qL + q}(Km)^q}{\sum_{q = 0}^{q_{max}} C^q_{N - qL + q}(Km)^q}. \quad (9)$$

Let C_{DNA} denote the total concentration of DNA (in base pairs), so that $C_{DNA} = NC_M$. Denoting by C_{lig} the total concentration of ligands in solution, we have

$$C_{lig} = m + \sum_{q_{crit}}^{q_{max}} q\Phi(q)C_M$$
$$= m + \frac{\sum_{q = q_{crit}}^{q_{max}} qC^q_{N - qL + q}(Km)^q}{N\sum_{q = 0}^{q_{max}} C^q_{N - qL + q}(Km)^q} C_{DNA}. \quad (10)$$

Using Eqs. (9) and (10), it is possible to calculate the fraction of molecules in state 2 as a function of the ligand concentration ion solution (see [29]).

Indeed, once the K and L values are known, Eqs. (9) and (10) can be solved by numerical methods and the function $R(C_{lig})$ can be constructed for the given values of the matrix length N and the DNA solution concentration C_{DNA}. This is most conveniently done by using a set of the free ligand concentrations m in solution, ranging from $m = 0.01/K$ to $1/K$ (as can be seen, this very interval features significant binding of ligands to

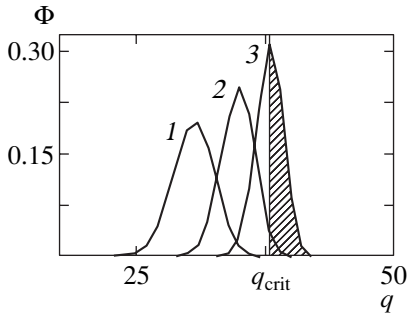

Fig. 2. The distribution functions $\Phi(q)$ characterizing ligand binding to DNA in solution for various concentrations m of a ligand covering 17 base pairs on a DNA molecule having a length of $N = 800$ base pairs. Curves *1–3* were calculated by formula (7) for $Km = 0.001$, 0.163, and 0.872, respectively. The cross-hatched region corresponds to DNA molecules binding 38 or more ligands and occurring in state 2 according to the model of Fig. 1.

DNA). As a result, we obtain an $R(C_{lig})$ curve describing the DNA transition from a state in which less than q_{crit} ligands are adsorbed, on average, on one DNA molecule to the state in which the average number of adsorbed ligands $\langle q \rangle$ is greater than q_{crit}.

Figure 2 illustrates variation of the distribution function $\Phi(q)$ for a DNA solution with increasing free ligand concentration m. As concentration m grows, the number of DNA molecules with $q > q_{crit}$ and, hence, the $R(C_{lig})$ value increase.

In practice, however, the problem is formulated differently: given the experimental distribution of $R(C_{lig})$, it is required to evaluate the size L of a binding site occupied by the ligand bound to DNA and to estimate the constant K of interaction between a ligand and the binding site. Below we will demonstrate how this problem can be solved for the binding of chitosan to DNA in cases when this interaction results in the formation of an liquid-crystalline dispersion of chitosan-covered DNA.

4. THEORETICAL CURVES OF THE CHITOSAN BINDING TO DNA COMPARED WITH EXPERIMENTAL DATA ON LIQUID-CRYSTALLINE DISPERSION FORMATION: RESULTS AND DISCUSSION

Previously [28], experimental data on the binding of chitosan to DNA and the formation of a liquid-crystalline DNA dispersion were obtained by methods of circular dichroism and apparent optical density. Figure 3 shows the corresponding families of characteristic S-shaped curves describing this binding process. The experimental techniques are described elsewhere [28]. The size of DNA molecules was about 800 base pairs. The chitosan preparation employed was deacetylated to 85% (i.e., it contained 85% active amino groups). The molecular weight of chitosan used in various experiments varied from 2 to 32 kD, which corresponds to the number of amino sugar residues changing approximately from 10 to 160 (one such residue had a molecular weight of about 200 D).

We calculated the theoretical curves of chitosan binding to DNA for each experiment using the results of independent measurements of circular dichroism and the apparent optical density. Based on the published data, it was assumed that a DNA molecule passes to the mod-

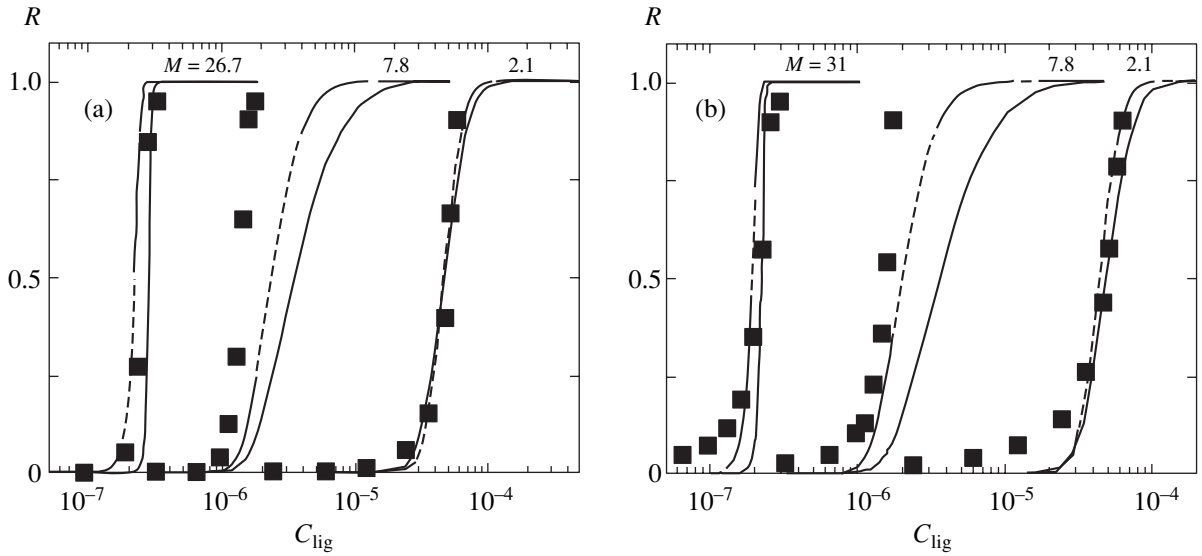

Fig. 3. A comparison of the experimental data (black squares) and theoretical results (curves) describing the binding of chitosan with various molecular weights to DNA: R is the response signal amplitude normalized to maximum for the measurements of (a) circular dichroism and (b) optical absorption; C_{lig} is the ligand (chitosan) concentration in solution; M *i*s the molecular weight of chitosan. Solid curves show the results of theoretical calculations performed for DNA molecules occurring in the modified state upon chitosan absorption to a coverage of $\alpha = 0.8$; dashed curves correspond to $\alpha = 0.7$.

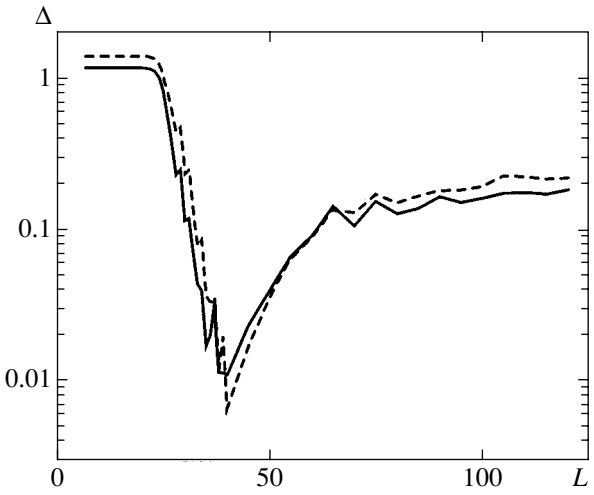

Fig. 4. A plot of the sum $\Delta(L)$ of square deviations of theoretical curves of from experimental data for chitosan with a molecular weight of 13.6 kD. Each point is constructed by selecting K so as to ensure the best fit of theory to experiment for a given L (in base pairs). Solid and dashed curves refer to the calculations using data on the optical absorption and circular dichroism, respectively.

ified state at a ligand coverage of $\alpha = 0.8$ (the calculations were also performed for $\alpha = 0.7$, since there is some evidence that the α value can be lower than 0.8 [30]). Taking certain L values, we determined the binding constants K corresponding to the best fit of theory to experiment. Then, once the critical coverage α is known, we can determine the values of $q_{crit} = \alpha N/L$ for the given L and N and calculate the corresponding theoretical curves using Eqs. (9) and (10).

A criterion for selecting the optimum binding constant K was the sum of square deviations of the theoretical values of the fraction of DNA molecules in the liquid-crystalline state, from the experimental values. For each experimental curve, it was found that this fraction for the signal maximum corresponds approximately to unity. Further increase in the chitosan concentration in solution gives rise to irreversible processes related to precipitation of liquid-crystalline dispersion particles.

Our approach can be illustrated by an analysis of two experimental curves obtained from the measurements of circular dichroism and the apparent optical density for chitosan with a molecular weight of 13.6 D (corresponding to about 68 amino sugar residues per molecule). Taking into account that one sugar residue is about 5 Å in length and that two adjacent DNA base pairs are spaced by 3.4 Å, we can suggest that the maximum possible size of a binding site occupied on DNA by a ligand consisting of 68 residues (assuming it to be parallel to the long DNA axis) amounts to $L = 68 \times 5/3.4 = 100$ base pairs (reactive centers of the matrix). It should be noted that this estimate is definitely overstated, since a chitosan molecule can wind itself around a DNA molecule as a helix (isogeometric with the phosphate core shape), in which case the binding site will be smaller by half (covering $L \sim 50$ base pairs). In this purely sterical model, we consider chitosan as a physical object, ignoring the fact that only every second group bears a charge (i.e., neglecting details and the nature of the chitosan–DNA interaction): we are interested only in estimating the size of the DNA segment covered by chitosan (or, in terms of our model approach, the number of reactive centers of the matrix rendered inaccessible for binding other ligand molecules).

Figure 4 presents the plots of the sum Δ of square deviations calculated for L ranging from 10 to 120 base pairs (i.e., for the interval definitely containing the size of the binding site for one chitosan molecule). As can

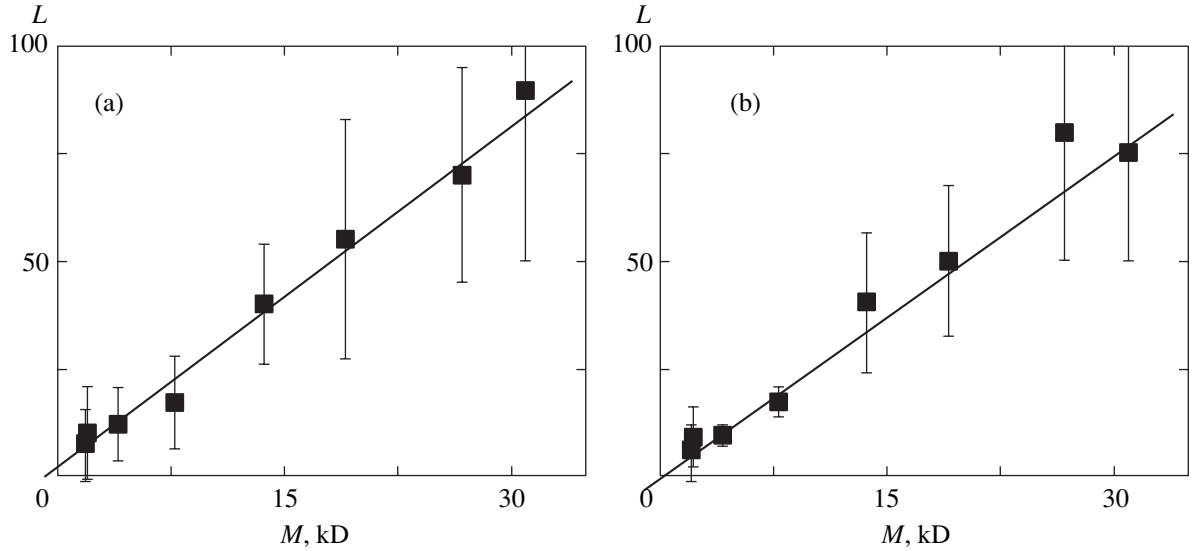

Fig. 5. Plots of the size L of the DNA binding site (in base pairs) for chitosan of various molecular weights M, providing for the best fit of theory to experiment for (a) circular dichroism and (b) apparent optical density.

Table 1. Estimates of the size L of the binding sites covered by chitosan molecules of various molecular weights on DNA and estimates of the equilibrium interaction constants K for chitosan binding to these sites. Upper and lower panels show the L and K values determined from an analysis of the circular dichroism and the optical absorption data, respectively

Molecular weight, kD	L	L_{min}	L_{max}	$\alpha = 0.7$		$\alpha = 0.8$	
				K_{min}, M^{-1}	K_{max}, M^{-1}	K_{min}, M^{-1}	K_{max}, M^{-1}
2	7	3	20	2×10^4	6×10^4	2×10^5	2×10^5
2.1	10	3	25	2×10^4	5×10^4	2×10^5	2×10^5
4.1	12	7	24	3×10^5	1×10^6	1×10^6	6×10^6
7.8	17	11	33	4×10^5	1×10^7	3×10^6	2×10^{10}
13.6	40	32	60	2×10^6	5×10^7	1×10^7	4×10^{10}
19	55	39	95	6×10^5	8×10^9	4×10^6	4×10^{10}
26.7	70	60	110	2×10^6	8×10^7	8×10^6	3×10^9
31	90	70	150	2×10^6	6×10^8	5×10^6	9×10^{10}
Molecular weight, kD	L	L_{min}	L_{max}	$\alpha = 0.7$		$\alpha = 0.8$	
				K_{min}, M^{-1}	K_{max}, M^{-1}	K_{min}, M^{-1}	K_{max}, M^{-1}
2	5	3	16	2×10^4	6×10^4	2×10^4	1×10^5
2.1	9	6	20	2×10^4	4×10^4	9×10^4	2×10^5
4.1	9	8	12	2×10^5	1×10^6	4×10^6	6×10^6
7.8	17	13	20	1×10^6	9×10^6	8×10^6	9×10^9
13.6	40	32	65	2×10^6	4×10^{10}	9×10^6	4×10^{10}
19	50	40	75	1×10^6	7×10^7	2×10^7	4×10^{10}
26.7	80	60	120	1×10^6	3×10^9	1×10^7	3×10^{10}
31	75	60	110	1×10^6	9×10^{10}	6×10^6	9×10^{10}

be seen, the best fit of theory to experiment is obtained for $L = 40$ base pairs. For each L value, we determined the constant K minimizing the sum $\Delta(L)$ (these minimum values are plotted in Fig. 4). The minimization procedure was performed using MATLAB software.

Table 1 summarizes the results of calculations for all experimental curves. The upper and lower panel show the K and L values determined from an analysis of the circular dichroism and the optical absorption data, respectively; L_{min} and L_{max} correspond to the minimum and maximum binding sites, respectively; K_{min} and K_{max} are the minimum and maximum binding constants, respectively, as determined by least squares.

Figure 5 shows the results of analysis of the experimental data obtained by methods of circular dichroism and optical absorption (apparent optical density). As can be seen, the data obtained by the two techniques are close and give approximately the same estimates: the size of the binding site of chitosan on DNA is proportional to the molecular weight of chitosan. This fact indicates that long chitosan molecules bind to DNA in the same manner as short ones. Thus, the relation between the size of the binding site and the molecular weight of chitosan can be expressed as

$$L = cM, \tag{11}$$

where c is the proportionality factor. From an analysis of the circular dichroism spectra, this coefficient is estimated as $c = 2.64 \pm 0.25$ kD^{-1}, while the optical absorption date refine this estimate as $c = 2.79 \pm 0.15$ kD^{-1}. Adopting the latter estimate, it is possible to calculate

Table 2. The energy parameters of chitosan binding to DNA determined by an analysis of the results of circular dichroism and optical absorption measurements

α	Experimental method	a	b, 10^4 M^{-1}
0.7	Circular dichroism	0.13 ± 0.08	1.9 ± 1.2
	Optical absorption	0.09 ± 0.05	2.8 ± 1.9
0.8	Circular dichroism	0.14 ± 0.08	8.7 ± 5.6
	Optical absorption	0.10 ± 0.07	11.8 ± 4

Table 3. The equilibrium interaction constants for various polycations bound to double-stranded DNA molecules

Compound	Conditions	K, M^{-1}	Refs.
Poly(L-lysine)	Mol. weight 36.600	8.3×10^9	[31]
Thiolated poly(L-lysine)	Mol. weight 36.600	1.3×10^{10}	[31]
Histonelike (sperm-specific protein from *Spisula solidissima*)	(300 amino acid residues; protaminelike protein)	$(1–9) \times 10^8$	[32]
Spermidine	17 mM	1.4×10^6	[23]
Spermine	1 mM	1.87×10^5	[33]
Pentalysine	0.1 M	4.7×10^3 (binding to poly(A)-poly(T))	[34]
Cetyltrimethylammonium bromide (CTAB)		1.2×10^3	[35]

that five amino sugar residues (weighing about 1 kD) occupy approximately 2.8 base pairs of DNA. The length of a sugar-phosphate DNA core per base pair is about 7.2 Å, so that five amino sugar residues cover a 20-Å-long DNA segment. These five residues possess a somewhat greater intrinsic length of ~25 Å (the difference probably reflects the fact that about 1/5 of the amino sugar chain of chitosan can dangle from DNA, forming loops and/or other structures).

Assuming that the mode of chitosan binding to DNA is independent of the size of this ligand, the binding energy should vary in an additive manner depending on the size L of the binding site occupied by chitosan and, hence, the binding constant should exponentially depend on this size. Thus, we may suggest that the binding constant obeys the relation

$$K = b\exp(aL). \tag{12}$$

Table 2 presents the estimates of parameters a and b obtained from our analysis of the experimental data for two values of the critical DNA coverage by chitosan, $\alpha = 0.7$ and 0.8. Note that the values of constants obtained from this analysis are, on the whole, consistent with the data characterizing the binding of polycations to DNA published by various researchers (Table 3). According to Table 2, variations in the free energy of binding per base pair are relatively small: the binding constant increases only by a factor of about 2.8 when the size of the chitosan binding site increases by ten base pairs.

It should be also noted that we have attempted to describe the formation of a liquid-crystalline DNA dispersion as a result of the chitosan binding to DNA based on some other models, in particular, the Porshke model [13]. According to this model, NAs exhibit compactization via the formation of complexes in which one ligand molecule binds to two DNA molecules. However, we failed to obtain a satisfactory description of the available experimental data on this basis. This is not surprising, since we are dealing with a liquid-crystalline phase in which DNA molecules retain their mobility, which is evidence for the absence of chitosan crosslinks.

5. CONCLUSIONS

We have proposed a model according to which reaching a certain coverage of DNA by chitosan in solution induces the transition of DNA into a modified state providing for an effective interaction between chitosan–DNA complexes and the formation of a liquid-crystalline DNA dispersion.

We have considered a model of noncooperative binding of ligands to DNA and shown, for the example of chitosan binding to DNA, how to evaluate the fraction of modified DNA fragments "prepared" to the phase transitions. A description of the phase transitions in single DNA molecules as a result of cooperative interactions between adsorbed ligands was recently proposed by Lando, Teif, *et al.* [36, 37].

ACKNOWLEDGMENTS

The authors are grateful to G.V. Gursky for his interest in this study and to S.V. Semenov for useful critical remarks.

This study was supported in part by the Russian Foundation for Basic Research, project no. 03-04-48212.

REFERENCES

1. T. L. Hill, J. Polym. Sci. **23**, 549 (1957).
2. T. L. Hill, *Statistical Mechanics: Principles and Selected Applications* (McGraw-Hill, New York, 1956; Inostrannaya Literatura, Moscow, 1960).
3. W. S. Magee, J. H. Gibbs, and B. H. Zimm, Biopolymers **1**, 133 (1963).
4. D. Crothers, Biopolymers **6**, 575 (1968).
5. A. S. Zasedatelev, G. V. Gursky, and M. V. Vol'kenshteĭn, Mol. Biol. (Moscow) **5**, 245 (1971).

6. G. V. Gursky, A. S. Zasedatelev, and M. V. Vol'kenshteĭn, Mol. Biol. (Moscow) **6**, 479 (1972).
7. J. A. Schellman, Biopolymers **14**, 999 (1975).
8. I. R. Epstein, Biophys. Chem. **8**, 327 (1978).
9. Yu. D. Nechipurenko, Biofizika **27**, 391 (1982).
10. G. V. Gursky and A. S. Zasedatelev, Sov. Sci. Rev. D. Physicochem. Biol. **5**, 53 (1984).
11. Yu. D. Nechipurenko and G. V. Gursky, Dokl. Akad. Nauk SSSR **281**, 213 (1985).
12. Yu. D. Nechipurenko and G. V. Gursky, Biophys. Chem. **24**, 195 (1986).
13. D. Porschke, J. Mol. Biol. **222**, 423 (1991).
14. J. Reiter and I. R. Epstein, J. Phys. Chem. **91**, 4813 (1987).
15. J. Reiter and I. R. Epstein, Biophys. Chem. **33**, 1 (1989).
16. B. Jovanovic and Yu. D. Nechipurenko, Mol. Biol. **24**, 478 (1990).
17. D. Poland, J. Chem. Phys. **113**, 4774 (2000).
18. D. Poland, Biopolymers **58**, 477 (2001).
19. F. Livolant and A. Leforestier, Progr. Polym. Sci. **21**, 1115 (1996).
20. R. L. Rill, T. E. Strzelecka, M. W. Davidson, and D. H. van Winkle, Physica A (Amsterdam) **176**, 87 (1991).
21. V. A. Bloomfield, R. W. Wilson, and D. C. Rau, Biophys. Chem. **11**, 339 (1980).
22. D. Y. Lando and V. B. Teif, J. Biomol. Struct. Dyn. **20**, 215 (2002).
23. D. Porschke, Biochemistry **23**, 4821 (1984).
24. A. Osland and K. Kleppe, Nucleic Acids Res. **4**, 685 (1977).
25. E. Raspaud, M. Olivera de la Cruz, J.-L. Sikorav, and F. Livolant, Biophys. J. **74**, 381 (1998).
26. S. G. Skuridin, A. T. Dembo, V. S. Efimov, and Yu. M. Yevdokimov, Dokl. Akad. Nauk **365**, 400 (1999).
27. M. Saminathan, T. Antony, A. Shirahata, *et al.*, Biochemistry **38**, 3821 (1999).
28. Yu. M. Yevdokimov, V. I. Salyanov, S. V. Semenov, *et al.*, Mol. Biol. **36**, 532 (2002).
29. Yu. D. Nechipurenko and G. V. Gursky, Biofizika **48**, 773 (2003).
30. Yu. D. Nechipurenko, A. M. Wolf, and Yu. M. Yevdokimov, Biofizika **48**, 802 (2003).
31. T. Wink, J. de Beer, W. E. Hennink, *et al.*, Anal. Chem. **71**, 801 (1999).
32. L. J. Libertini, J. Ausio, and K. E. van Holde, Biopolymers **27**, 1459 (1988).
33. R. Marquet, C. Houssier, and E. Fredericq, Biochim. Biophys. Acta **825**, 365 (1985).
34. M. T. Record, T. M. Lohman, and P. de Haseth, J. Mol. Biol. **107**, 145 (1976).
35. S. M. Mel'nikov, V. G. Sergeev, and K. Yoshikawa, J. Am. Chem. Soc. **117**, 9951 (1995).
36. D. Lando and V. Teif, J. Biomol. Struct. Dyn. **17**, 903 (2000).
37. V. B. Teif, S. G. Haroutiunian, V. I. Vorob'ev, *et al.*, J. Biomol. Struct. Dyn. **19**, 1093 (2002).

Translated by P. Pozdeev